Uniqueness of invariant distributions for split-merge transformations and the Poisson-Dirichlet law

Persi Diaconis,1 Eddy Mayer-Wolf,2 Ofer Zeitouni,3 and Martin P. W. Zerner 4

July 2, 2002.

Abstract

We consider a Markov chain on the space of (countable) partitions of the interval [0, 1], obtained first by size biased sampling twice (allowing repetitions) and then merging the parts (if the sampled parts are distinct) or splitting the part uniformly (if the same part was sampled twice). We prove a conjecture of Vershik stating that the Poisson-Dirichlet law is the unique invariant measure for this Markov chain. Our proof uses a combination of probabilistic, combinatoric, and representation-theoretic arguments.

Key Words: Partitions, coagulation, fragmentation, invariant measures, Poisson-Dirichlet. AMS 2000 Mathematics Subject Classification. Primary 60K35; secondary 60J27, 60G55.
1 Introduction

Let $\Omega_1$ denote the space of (ordered) partitions of $[0, 1]$, that is
$$\Omega_1 := \left\{ p \in [0, 1]^\mathbb{N} : p_1 \geq p_2 \geq \ldots \geq 0, \ |p|_1 = 1 \right\},$$
where $|x| = \sum_i |x_i|$ for any finite or countable sequence $(x_i)$. By size-biased sampling according to a point $p \in \Omega_1$ we mean picking the $j$-th part $p_j$ with probability $p_j$. Our interest in this paper is in the following Markov chain on $\Omega_1$, which we call a continuous coagulation-fragmentation process (CCF): size-bias sample (with replacement) two parts from $p$. If the same part was picked twice, split it (uniformly), and reorder the partition. If different parts were picked, merge them, and reorder the partition.

We denote by DCF$^{(n)}$ (discrete coagulation-fragmentation) the Markov chain describing the evolution of the cycle lengths of permutations of $\{1, \ldots, n\}$ under random transpositions. The CCF process appears in a variety of contexts, but of particular relevance to us is its occurrence as a natural limit of DCF$^{(n)}$, when $n$ increases, see [12] for a discussion of this and its link with the space of “virtual permutations”.

For any $n \in \mathbb{N}$ denote
$$\mathcal{P}_n := \{ \ell = (\ell_i)_{i \geq 1} \in \{0, 1, \ldots, n\}^\mathbb{N} : \ell_1 \geq \ell_2 \geq \ldots \geq 0, \ |\ell|_1 = n \} \subset n\Omega_1.$$ (Elements in $\mathcal{P}_n$ may be thought of as being of length $n$; the remaining entries are necessarily zero).

A sequence $\ell \in \mathcal{P}_n$ is uniquely determined by its type $(N_\ell(k) = \#\{i : \ell_i = k\})_{k=1}^n$, with $N_\ell = \sum_{k=1}^n N_\ell(k)$ denoting $\ell$’s total number of parts.

The long-time behaviour of the DCF$^{(n)}$, viewed as an evolution in $\mathcal{P}_n$, is well understood. In particular, see e.g. [2], it possesses a unique stationary distribution given by the Ewens formula:

$$\pi_S^{(n)}(\ell) = \left( \prod_{k=1}^n k^{N_\ell(k)} N_\ell(k)! \right)^{-1} = \left( \prod_{i=1}^n \ell_i \prod_{k=1}^n N_\ell(k)! \right)^{-1}, \quad \ell \in \mathcal{P}_n. \quad (1.1)$$

It is well known, at least since [7, 9, 14], that the measures $\pi_S^{(n)}(\cdot)$ on $\Omega_1$ converge weakly to the Poisson-Dirichlet distribution $\hat{\mu}_1$ (a precise definition of $\hat{\mu}_1$ is given in Section 2.1). It has been shown in more than one way (cf. [5, 11, 12]) that $\hat{\mu}_1$ is invariant for the CCF transition. This fact, and hints coming from the theory of virtual permutations, led Vershik (see [12]) to the

**Conjecture 1.1 (Vershik)** $\hat{\mu}_1$ is the unique invariant distribution for the CCF.

Our goal in this article is to prove Vershik’s conjecture. A naive approach toward the proof would be to use the link with the DCF$^{(n)}$ and the fact that the latter converges to the distribution $\pi_S^{(n)}$ exponentially fast. However, the rate of that convergence deteriorates with $n$. To overcome this difficulty, our strategy consists of the following steps:
1. We provide a-priori estimates (Proposition 2.1) showing that every invariant distribution for the CCF leads to a good control on the number of “small parts”.

2. We couple the DCF(\(\kappa\)) and the CCF in such a way that whenever they start from initial distributions with such control on the tails, the decoupling time is roughly \(\sqrt{n}\) (Theorem 3.1).

3. For initial conditions as above, and for an appropriate class of test functions, we show by using some harmonic analysis on the symmetric group that the DCF(\(\kappa\)) achieves near equilibrium before the decoupling time (Theorem 4.1).

These steps are then combined in Theorem 5.1 to yield the proof of Vershik’s conjecture.

We next review some of the literature on this question. Tsilevich, in [12], proves that \(\hat{\mu}_1\) is the only CCF-invariant measure that is also invariant under additional symmetry conditions. Pitman, in [11], proves that \(\hat{\mu}_1\) is the only CCF-invariant measure which is also invariant under size-biased sampling. Related results appear in [5]. In another direction, it is shown in [10] that \(\hat{\mu}_1\) is the only CCF-invariant measure that is analytic in the sense that for any \(k\), the law of an independently size-biased sample (with replacement) possesses an analytic density. Finally, Tsilevich in [13] shows that the law of the CCF, initialized at \(p = (1, 0, \ldots)\), and stopped at a Binomial(\(n, 1/2\)) random time, converges to \(\hat{\mu}_1\).

We conclude this introduction by noting that in [10], we have introduced a slightly more general model of split-merge transformations, by allowing either the split or the merge operations to be rejected with a certain probability. An invariant measure for these generalizations is the Poisson-Dirichlet law of parameter \(\theta > 0\). The discrete counterpart of this chain has been analyzed in [3, Section 4]. While it is plausible that the techniques of the current paper can be adapted to that setup using the results of [3], we do not pursue this generalization here.

## 2 Continuous and Discrete Coagulation-Fragmentation

### 2.1 Preliminaries and CCF

Given a topological space \(W\), its Borel \(\sigma\)-algebra will be denoted by \(B_W\), and the space of probability measures on \((W, B_W)\) by \(\mathcal{M}_1(W)\). By a slight abuse of notations, \(\mathcal{M}_1(V)\) will also be \(\mathcal{M}_1(W)\)'s subspace of probability measures whose support is contained in a given closed subset \(V\) of \(W\). The total variation of a measure \(\nu\) is denoted by \(\|\nu\|_{\text{var}}\).

We equip \(\Omega_1\) with its relative \(\cdot, 1\)-topology which, on \(\Omega_1\), coincides with the weak (coordinatewise convergence) topology.

On \(\Omega_1\) we consider the Markov chain CCF in which two segments \(p_i\) and \(p_j\) of a given partition \(p\) are size-bias sampled with replacement and then, if \(i \neq j\) they merge into one of length \(p_i + p_j\) (coagulation), while if \(i = j\), \(p_i\) splits into two new parts \(up_i, (1 - u)p_i\) with \(u \sim U[0, 1]\) independent of all the rest (fragmentation). In either case the new partition is then rearranged nonincreasingly.
Recall that the Poisson-Dirichlet law $\hat{\mu}_1$ is invariant for the CCF transition. Indeed, $\hat{\mu}_1$ itself has been defined in a variety of manners ([1, 7]) which are well known to be equivalent. Perhaps the simplest is the GEM description in which segments are successively and uniformly removed from whatever remains of $[0, 1]$, and then rearranged nonincreasingly. Namely, let $Y_1=1$ and for $n \in \mathbb{N}$ define $X_n=U_n Y_n$, (the removed part at stage $n$) and $Y_{n+1}=Y_n-X_n$ (the remaining segment from which the $(n+1)$-th part is to be removed), where the $U_n$'s are independent $U[0, 1]$ variables. Since $Y_{n+1}=(1-U_n)Y_n$ it follows that $1-Y_{n+1}=\sum_{i=1}^{n} X_i$ increases almost surely to $1$ as $n \to \infty$. The distribution on $\Omega_1$ of the nonincreasing rearrangement $(p_i)_i$ of $(X_n)_n$ is called the Poisson–Dirichlet law (with parameter $\theta = 1$) and denoted $\hat{\mu}_1$.

As has been mentioned in the Introduction, it is the ultimate goal of this work to show that the Poisson-Dirichlet law is the only CCF-invariant probability distribution. It will be crucial for the main argument to establish in advance that any such invariant distribution does not put too much weight on very small parts:

Proposition 2.1 Let $\mu \in \mathcal{M}_1(\Omega_1)$ be CCF-invariant. Then

$$\int \sum_{i \geq 1} p_i^\alpha \, d\mu < \infty \quad \text{for all } \alpha > 2/5. \quad (2.1)$$

The proof is deferred to the Appendix.

2.2 DCF

In this section we formally introduce the coagulation-fragmentation chain on the discrete version of $\Omega_1$, in which the partition points lie on a finite equidistant grid in $[0, 1]$, or its equivalent state space $\mathcal{P}_n$, the set of integer partitions of a fixed $n \in \mathbb{N}$ defined in the Introduction. It will be helpful to view $\mathcal{P}_n$ as the conjugacy classes of the permutation group $S_n$.

The DCF($n$) Markov chain on $\mathcal{P}_n$ is defined similarly to the CCF chain on $\Omega_1$. Identify each $\ell \in \mathcal{P}_n$ with a partition $\bigcup A_i$ of $\{1, 2, \ldots, n\}$, where for each $i$, $\# A_i = \ell_i$, and sample two independent integers $x, y$ uniformly from $\{1, \ldots, n\}$ and without replacement, say $x \in A_i$ and $y \in A_j$. If $i \neq j$ replace $A_i$ and $A_j$ by $A_i \cup A_j$ while if $i = j$ (in which case $\ell_i \geq 2$ since $x \neq y \in A_i$) replace $A_i$ by two of its subsets, consisting respectively of $A_i$'s $k$ smallest elements and of the $\ell_i-k$ remaining ones, where $k$ is uniformly sampled from $\{1, \ldots, \ell_i-1\}$ independently of $x$ and $y$. In either case relabel and rearrange the new $A_i$'s if necessary.

The transition matrix $K^{(n)}$ of DCF($n$) is described as follows. Let $1 \leq j \neq k \leq n$ and $\ell, \ell' \in \mathcal{P}_n$ be such that $N_{\ell'}(j)=N_{\ell}(j)-1$, $N_{\ell'}(k)=N_{\ell}(k)-1$, $N_{\ell'}(j+k)=N_{\ell}(j+k)+1$ and $N_{\ell'}(q)=N_{\ell}(q)$ for all $q \not\in \{j, k, j+k\}$. Then

$$K^{(n)}(\ell, \ell') = \frac{2jk}{n(n-1)} N_{\ell}(j) N_{\ell}(k) \quad \text{merge}$$

$$K^{(n)}(\ell', \ell) = \frac{2(j+k)}{n(n-1)} N_{\ell}(j+k) \quad \text{split} \quad (2.2)$$

4
and all other entries are zero.

It is customary to think of the representation \( \{1, 2, \ldots, n\} = \bigcup A_i \) above as the notation for the conjugacy class of a permutation \( \sigma \in S_n \). Seen this way, the \( \text{DCF}^{(n)} \) transition is nothing but the action of a random transposition on \( S_n \)'s conjugacy classes. Since the random transposition's unique stationary probability measure is the uniform law on \( S_n \) (being a finite group convolution), one concludes that the \( \text{DCF}^{(n)} \)'s unique stationary probability measure is the one induced on \( S_n \)'s conjugacy classes by the uniform law, namely \((1.1)\) (the Ewens sampling formula). In fact, \( \text{DCF}^{(n)} \) is reversible with respect to \( \pi_S^{(n)} \), which can also be checked directly by using \((2.2)\) and \((1.1)\) to verify the detailed balance equation \( K^{(n)}(\ell, \ell') \pi_S^{(n)}(\ell) = K^{(n)}(\ell', \ell) \pi_S^{(n)}(\ell') \).

### 3 Coupling of CCF and DCF

In order to successfully approximate a CCF chain by \( \text{DCF}^{(n)} \) chains as \( n \to \infty \) it is necessary to couple them on a common probability space.

**Theorem 3.1** For all \( \mu \in \mathcal{M}_1(\Omega_1) \) and \( \alpha < 1/2 \) satisfying

\[
\int \sum_{i \geq 1} p_i \, d\mu < \infty, \tag{3.1}
\]

it is possible to define for all \( n \geq 1 \) a CCF Markov chain \( p(k) (k \geq 0) \) with initial distribution \( \mu \) and a \( \text{DCF}^{(n)} \) Markov chain \( \ell(k) (k \geq 0) \) on the same probability space with probability measure \( Q_\mu^{(n)} \) and expectation \( E_\mu^{(n)} \) in such a way that

\[
\begin{align*}
\lim_{n \to \infty} Q_\mu^{(n)} \left[ N_{\ell(0)} \geq n^\beta \right] &= 0 \quad \text{for all } \alpha < \beta \quad \text{and} \\
\lim_{n \to \infty} E_\mu^{(n)} \left[ p(\lfloor n^\beta \rfloor) - \frac{\ell(\lfloor n^\beta \rfloor)}{n} \right] &= 0 \quad \text{for all } \beta < 1/2.
\end{align*}
\tag{3.2}
\]

**Proof:** Fix \( n \geq 1 \). We shall construct a Markov chain \( (c_k, d_k, e_k) (k \geq 0) \) on the state space

\[
\Omega^{(n)}_{\text{calc}} := \{(c, d, e) \mid c : [0, n) \to \mathbb{Z} \text{ measurable, } \mathbb{Z}\backslash c[[0, n) \text{ infinite}, \\
\quad d : \{1, \ldots, n\} \to \mathbb{Z}, \quad e \in \{0, 1\}\}
\]

Here \( c \) and \( d \) describe a continuous partition of \([0, n)\) and a discrete partition of \(\{1, \ldots, n\}\), respectively. The interpretation of \( c \) and \( d \) in terms of elements of \( \Omega_1 \) and \( \mathcal{P}_n \) is given by the functions \( \pi_c : \Omega^{(n)}_{\text{calc}} \to \Omega_1 \) and \( \pi_d : \Omega^{(n)}_{\text{calc}} \to \mathcal{P}_n \), respectively, defined by

\[
\begin{align*}
\pi_c(c, d, e) &:= \text{sort} \left( \left( \frac{\text{Leb}(c^{-1}([i]))}{n} \right)_{i \in \mathbb{Z}} \right) \quad \text{and} \\
\pi_d(c, d, e) &:= \text{sort} \left( \left( \frac{\text{Leb}(d^{-1}([i]))}{n} \right)_{i \in \mathbb{Z}} \right).
\end{align*}
\tag{3.4}
\]

(Continued on next page...)
Here Leb denotes the Lebesgue measure and sort \((x_i)_i\) is the sequence obtained by arranging the \(x_i\)'s in decreasing order, ignoring the 0's if there are infinitely many positive \(x_i\)'s. Thus two points \(x, y \in [0, n)\) belong to the same set in the partition of \([0, n)\) which is described by \(c\) iff \(c(x) = c(y)\). Analogously, \(x, y \in \{1, \ldots, n\}\) belong to the same set in the partition of \(\{1, \ldots, n\}\) described by \(d\) iff \(d(x) = d(y)\). The CCF Markov chain \(p(k)\) and the DCF\(^{(n)}\) Markov chain \(\ell(k)\) will be realized as

\[
p(k) := \pi_c(c_k, d_k, e_k) \quad \text{and} \quad \ell(k) := \pi_d(c_k, d_k, e_k).
\] (3.6)

The flag \(e_k\) indicates whether the coupling between the two processes \(p(k)\) and \(\ell(k)\) is considered to be still in force \((e = 0)\) or to have already broken down \((e = 1)\).

The distribution of \((c_0, d_0, e_0)\), that is the initial distribution of the Markov chain, is defined as the image of \(\mu\) under the function \(\Phi^{(n)} = (\Phi_1^{(n)}, \Phi_2^{(n)}, 0) : \Omega_1 \rightarrow \Omega_{c_d e}^{(n)}\) which assigns to each element of \(\Omega_1\) an equivalent function \(c\) and an approximating function \(d\) as follows (see Figure 3):

\[
\Phi_1^{(n)}(p)(x) := \sum_{j \geq 1} j \mathbf{1} \left\{ x \in [0, np_j) + n \sum_{i=1}^{j-1} p_i \right\} \quad (x \in [0, n))
\]

\[
\Phi_2^{(n)}(p)(m) := \Phi_1^{(n)}(p)(m - 1) \quad (m \in \{1, \ldots, n\}).
\]

Thus \(p(0) = p\) and \(\ell(0) = \pi_d(\Phi^{(n)}(p))\) are the initial continuous and discrete partitions generated by \(p \in \Omega_1\).

**Proof of (3.2):** To bound \(N_{\ell(0)}\), the number of parts in \(\ell(0)\), observe that all the pieces in \(p\) of size less than \(1/n\) can give rise to at most \(\sum_i np_i \mathbf{1}_{np_i < 1}\) parts (singletons) in \(\ell(0)\). Therefore,

\[
N_{\ell(0)} = \# \Phi_2^{(n)}(p)[\{1, \ldots, n\}] = \sum_i \mathbf{1}_{np_i \geq 1} + np_i \mathbf{1}_{np_i < 1}
\]

\[
\leq \sum_i (np_i)^\alpha \mathbf{1}_{np_i \geq 1} + (np_i)^\alpha \mathbf{1}_{np_i < 1} = n^\alpha \sum_i p_i^\alpha. \quad (3.7)
\]

Consequently, due to assumption (3.1),

\[
E_{\mu}^{(n)}[N_{\ell(0)}] = O(n^\alpha) \quad (3.8)
\]
and hence, for all $\beta > \alpha$,

$$Q^{(n)}_\mu \left[ N_{\ell(0)} > n^\beta \right] \leq n^{-\beta} E^{(n)}_\mu [ N_{\ell(0)} ] = O(n^{\alpha-\beta}),$$

thus proving (3.2).

We now define informally the kernel of the Markov chain $(c_k, d_k, e_k)$ with state space $\Omega^{(n)}_{cde}$. Assume that the current state of the Markov chain is $(c, d, e)$. To compute the state $(c, d, e)$ the Markov chain is going to jump to in the next step we generate four random variables $\xi_1, \xi_2, \zeta_1$ and $\zeta_2$ such that $\xi_1$ and $\xi_2$ and $(\zeta_1, \zeta_2)$ are independent of each other and of everything else and such that the $\xi_i$ are uniformly distributed on $[0, n)$ and $(\zeta_1, \zeta_2)$ is uniformly distributed on $[0, n)^2 \setminus \bigcup_{j=1}^{n} [j-1, j)^2$. The $\xi_i$ will serve to sample uniformly with replacement from $[0, n)$ whereas the $\zeta_i$ will be used to sample uniformly without replacement from $\{1, \ldots, n\}$ in case the $\xi_i$ have chosen the same atom in $d$ twice. The new continuous partition $\tilde{e}$ is then defined as follows:

If $c(\xi_1) \neq c(\xi_2)$:

$$\tilde{e}(x) = \begin{cases} c(\xi_1) & \text{if } c(x) = c(\xi_2) \\ c(x) & \text{else.} \end{cases} \quad (3.9)$$

If $c(\xi_1) = c(\xi_2)$:

$$\tilde{e}(x) = \begin{cases} \text{new } (c, d) & \text{if } c(x) = c(\xi_1) \text{ and } x > \xi_1 \\ c(x) & \text{else.} \end{cases} \quad (3.10)$$

In (3.9) two different sets have been selected and are merged by assigning the set hit by $\xi_2$ the number of the set selected by $\xi_1$. In (3.10) one set is chosen twice. Since $\xi_1$ is conditionally uniformly distributed on this set we can reuse it as splitting point for that set: The part to the left of $\xi_1$ retains its old number whereas the part to its right gets a new number new $(c, d)$ which is not in the range of $c$ or $d$. Note that it is always possible to find such a new number since $\mathbb{Z}\setminus c([0, n])$ is assumed to be infinite. It is obvious that $p(k)$ defined in (3.6) is a CCF Markov chain.

The rule for merges in the discrete partition is analogous to (3.9):

If $d([\xi_1]) \neq d([\xi_2])$:

$$\bar{d}(m) = \begin{cases} d([\xi_1]) & \text{if } d(m) = d([\xi_2]) \\ d(m) & \text{else.} \end{cases} \quad (3.11)$$

The rule for splitting is slightly more complicated. If the same part (but not the same atom) is sampled twice by the $\xi_i$ then again, $\xi_1$ determines the point at which the set $d^{-1}([\xi_1])$ is going to be split: The points to the left of $[\xi_1]$ and the points to the right of $[\xi_1]$ will constitute the two new fragments. The point $[\xi_1]$ itself will be attached to the left or the right part in such a way that the splitting rule for DCF$^{(n)}$ is imitated. This is done as follows:

If $d([\xi_1]) = d([\xi_2])$ and $[\xi_1] \neq [\xi_2]$:

$$\bar{d}(m) = \begin{cases} \text{new } (c, d) & \text{if } d(m) = d([\xi_1]) \text{ and } m > [\xi_1] \\ \text{or } m = [\xi_1] \text{ and } \xi_1 < [\xi_1] + \frac{#d^{-1}([\xi_1]) \cap [0, \xi_1]}{#d^{-1}([\xi_1])} - 1 & \text{else.} \end{cases} \quad (3.12)$$

7
If however $[\xi_1] = [\xi_2]$, which means that the same atom in $d$ has been sampled twice, then $\xi_1$ and $\xi_2$ are disregarded and $\bar{d}(n)$ is defined as in (3.11) and (3.12) but with $(\xi_1, \xi_2)$ replaced by $(\zeta_1, \zeta_2)$ in order to sample without replacement. The process $\ell(k)$ defined in (3.6) is a DCF$(n)$ Markov chain.

It remains to define $\bar{e}$:

$$\bar{e} = \begin{cases} 
1 & \text{if } [\xi_1] = [\xi_2] \text{ or } c(\xi_1) \neq d(\xi_1) \text{ or } c(\xi_2) \neq d(\xi_2) \\
e & \text{else}.
\end{cases}$$

In the case $\bar{e} = 1$ the coupling has broken down: Either the same atom in the discrete partition has been sampled twice by the $\xi_i$'s or at least one of the $\xi_i$'s belongs to non-corresponding sets in the continuous and the discrete partition. The time $\tau := \inf\{k \geq 1 : e_k = 1\}$ is regarded as the decoupling time of the chains $p(k)$ and $\ell(k)$.

The definition of the transition kernel for the Markov chain on $\Omega^{(n)}_{cde}$ is now complete. It is summarized in Figures 3 to 3.

**Proof of (3.3):** We denote by

$$g_k := \text{Leb}\left(\{x \in [0,n) : c_k(x) \neq d_k([x])\}\right).$$

the discrepancy between $c_k$ and $d_k$. For $k = 0$, this is the roundoff error caused by the approximation of $c_0$ by $d_0$; its size is the length of the shaded area in Figure 3. Note that

$$g_0 \leq N_{(0)}$$

because any part in $d_0$ might disagree with $c_0$ at most in its rightmost atom. Moreover, $g_k$ can increase in each step by at most 1 as long as $k < \tau$: Indeed, if two parts are merged, $g_k$ does not increase at all (it might even decrease) whereas it might increase by at most $\text{Leb}\left(\{[\xi_1], [\xi_1]\}\right) = 1$ in case of splitting. Hence, $g_{k+1} \leq g_k + 1$ if $k < \tau$ and therefore,

$$g_k \leq g_0 + k \quad \text{on the event } \{k < \tau\}. \quad (3.14)$$
Since the $| \cdot |_1$-diameter of $\Omega_i$ is at most 2 we have

$$
E^{(n)}_{\nu} \left[ p([n^\beta]) - \frac{\ell([n^\beta])}{n} \right]_{1} 
\leq \quad E^{(n)}_{\nu} \left[ p([n^\beta]) - \frac{\ell([n^\beta])}{n} \right]_{1} , \quad [n^\beta] \leq \tau + 2Q^{(n)}_{\nu}[\tau < [n^\beta]].
$$

We are going to bound the first term in (3.15) first. It is easy to see that $|p-q|_1 \geq |\text{sort}(p) - \text{sort}(q)|_1$ for any two summable sequences $p = (p_i)_i$ and $q = (q_i)_i$ of non-negative numbers. Indeed, if $p_i > p_j$ and $q_i < q_j$, then swapping $q_i$ and $q_j$ would not increase $|p-q|_1$. Therefore, by definitions (3.4), (3.5) and (3.6) on the event $\{[n^\beta] < \tau\}$,

$$
\left| p([n^\beta]) - \frac{\ell([n^\beta])}{n} \right|_{1} \leq \quad \frac{1}{n} \sum_{i \geq 1} \left| \text{Leb} \left( c_{[n^\beta]}^{-1} \{ \{i\} \} \right) - \#d_{[n^\beta]}^{-1} \{ \{i\} \} \right|
\leq \quad \frac{1}{n} \sum_{i \geq 1} \text{Leb} \left\{ x : i \in \{ c_{[n^\beta]}(x), d_{[n^\beta]}(x) \} , c_{[n^\beta]}(x) \neq d_{[n^\beta]}(x) \right\}
\leq \quad \frac{2}{n} \sigma_{[n^\beta]} \leq \quad \frac{2}{n}(\sigma_0 + [n^\beta]) \leq \quad \frac{2}{n}(N_{\ell_0} + [n^\beta])
$$

by (3.14) and (3.13). Consequently, due to (3.8), the first term in (3.15) is of order $O(n^{\alpha-1} + n^{\beta-1})$, thus going to 0 as $n \to \infty$.

To show that the second term in (3.15) goes to 0 as well we assume without loss of generality that $\alpha < \beta < 1/2$. Consider the probability that a chain which has not decoupled until the $k$th step will decouple in the $(k+1)$th step. Given $\varrho_0, \ldots, \varrho_k$, the event that $\xi_1$ samples two different
parts in $c_k$ and $d_k$ has probability $g_k/n$. The same holds for $\xi_2$. Moreover, the event that one atom in $d_k$ is sampled twice, i.e. that $[\xi_1] = [\xi_2]$ has probability $1/n$. Therefore, the probability that either of these events occurs and the chain decouples is at most $(2g_k + 1)/n$. On the event \( \{\tau > k, g_0 < n^\beta\} \) this can be bounded from above due to (3.14) by $(2(n^\beta + k) + 1)/n$ which is less than $5n^{\beta-1}$ if $k \leq n^\beta$. Thus we get by induction over $k$,

$$Q^{(n)}_{\mu}[\tau > k, g_0 < n^\beta] \geq (1 - 5n^{\beta-1})^k Q^{(n)}_{\mu}[g_0 < n^\beta]$$

for all $k \leq n^\beta$ and hence

$$Q^{(n)}_{\mu}[\tau \geq [n^\beta]] \geq \left(1 - 5n^{\beta-1}\right)^n Q^{(n)}_{\mu}[g_0 < n^\beta]. \quad (3.16)$$

Due to $2\beta - 1 < 0$, the first factor in (3.16) converges to one as $n \to \infty$. The same holds for the second factor due to (3.13) and (3.2). Consequently, also the second term in (3.15) goes to 0, which completes the proof of (3.3).

\section{DCF$^{(n)}$ convergence}

It was mentioned in the Introduction, that the uniform rate of convergence to $\pi^{(n)}_S$ is too weak to combine properly with $n \to \infty$. However, according to the following theorem (to be proved in subsection 4.2), the situation is better when starting off from partitions with relatively few parts and restricting our attention to a certain family $C$ of $\Omega_1$-neighborhoods to be defined below. For every $n \in \mathbb{N}$ and $\beta \in (0, 1]$, thus, denote accordingly

$$\mathcal{P}_{n, \beta} = \left\{ \ell \in \mathcal{P}_n : N_\ell < n^\beta \right\} = \left\{ \ell \in \mathcal{P}_n : \ell_{[n^\beta]} = 0 \right\}.$$  

As for the definition of $C$, for each $k \in \mathbb{N}$ let

$$I_k = \left\{ (a, b) = (a_i, b_i)_{i=1}^k : 0 < a_i < b_i < 1 \quad \sum_{i=1}^k b_i < 1, \quad a_k > 1 - \sum_{i=1}^k a_i \right\} \quad (4.1) \quad$$
and denote $\delta_{a,b} = \min \left\{ 1 - \sum_{i=1}^{k} b_i, \ a_k - (1 - \sum_{i=1}^{k} a_i) \right\}$. Then, for each $(a, b) \in I_k$, define

$$C_{a,b} = \left\{ x \in \Omega_1 : x_i = (a_i, b_i) \text{ for } i = 1, \ldots, k \right\}$$

which is nonempty if and only if $0 < a_i < \min b_j$ for $i = 1, \ldots, k$, in which case the conditions on $(a, b) \in I_k$ guarantee that

$$C_{a,b} = \left\{ x = (x', x'') : x' \in G_{a,b}, x'' \in (1 - |x'|_1) \Omega_1 \right\}$$

where $(\cdot, \cdot)$ denotes concatenation and $G_{a,b}$ is the (nonempty) subset of points in $\prod_{i=1}^{k} (a_i, b_i)$ whose coordinates are nonincreasing, and moreover that

$$\delta_{a,b} < |x''|_1 < x_k' - \delta_{a,b}, \quad \forall (x', x'') \in C_{a,b}.$$  \hspace{1cm} (4.4)

Finally,

$$C = \{ C_{a,b} : (a, b) \in I_k, k \geq 1 \}.$$ \hspace{1cm} (4.5)

The family $C$ of $\Omega_1$-neighborhoods will be shown in Section 5 to be sufficiently rich to characterize $\hat{\mu}_1$ uniquely. At the same time, and as a result of their special features (4.3) and (4.4), the convergence of the $\text{DCF}^{(n)}$ to its equilibrium is fast on the sets in $C$:

**Theorem 4.1** Fix $\beta \in (0, \frac{1}{2})$. For each $n \in \mathbb{N}$ let $(X^{(n)}(k))_{k \geq 0}$ be a $\text{DCF}^{(n)}$ Markov chain with underlying probability measure $P^{(n)}$ and initial distribution $\mu_0^{(n)} \in M_1(P, \beta)$. Then for any $C \in C$, $\beta' > \beta$ and integer sequence $k = k_n \geq n^{3\beta'}$

$$\Delta_C^{(n)}(k) := P^{(n)} \left( X^{(n)}(k) \in nC \right) - \pi_S^{(n)}(nC) \xrightarrow[n \to \infty]{} 0.$$ \hspace{1cm} (4.6)

### 4.1 Characters in $S_n$ - Background

Recall that the partition space $P_n$ can be viewed as the quotient of the permutation group $S_n$ under conjugacy. Thus the natural inner product on $F_n := \{ f : P_n \rightarrow \mathbb{R} \}$ is

$$\langle f, g \rangle = \langle f, g \rangle_n = \sum_{\gamma \in P_n} f(\gamma) g(\gamma) \pi_S^{(n)}(\gamma).$$

The fact mentioned earlier that $\pi_S^{(n)}$ is a reversing measure for the $\text{DCF}^{(n)}$ means precisely that $K^{(n)}$ is selfadjoint with respect to this inner product.

The following basic facts regarding the character theory of $S_n$, as well as the full theory, can be found, for example, in [6], and their relevance to random group actions (such as transpositions in our case) in [2] and [4]. The characters $\{ \chi \}$ of $S_n$ (traces of the irreducible representations)
are functions on $S_n$, constant on conjugacy classes, and as such can be seen to be functions on $P_n$. They are orthonormal under $\langle \cdot , \cdot \rangle$ and since there are $\#P_n$ of them, they are indexed by the partitions $(\lambda \lambda_{\lambda \in P_n})$ and form an orthonormal base of $F_n$.

Since $K^{(n)}$ represents a random transposition, its dual $K^{(n)^*}$ acts on $M_1(S_n)$ as a convolution

$$K^{(n)^*} \mu = K^{(n)} \ast \mu$$

$$K^{(n)}(\text{transposition}) = \frac{2}{n(n-1)} \text{ and } 0 \text{ otherwise}$$

as a result of which, and of a corollary ([2, Ch. 2, Prop. 6]) of Schur’s lemma,

a) $K^{(n)}$’s eigenfunctions are the characters $(\chi_{\lambda})_{\lambda \in P_n}$

b) the eigenvalue $\theta^{(n)}_{\lambda}$ corresponding to $\chi_{\lambda}$ is given by $\frac{\chi_{\lambda}(\text{transposition})}{\chi_{\lambda}(\text{identity})}$.

A result of Frobenius in principle provides formulae for all characters. Although in general they can be intractable, this is not so at transpositions and at the identity, thus yielding ([2, D-2,p.40])

$$\theta^{(n)}_{\lambda} = \frac{1}{n(n-1)} \sum_{j} \chi_{j}(\lambda_j - 2j + 1) = \frac{1}{n(n-1)} \left( \sum_{i=1}^{n} \chi_{i}^2 - \sum_{j=1}^{\lambda_1} \chi_{j}^2 \right). \quad (4.6)$$

($\lambda$’s adjoint partition $\lambda'$ is defined below). In particular $\theta^{(n)}_{(n,n,...)} = 1$ and $\chi_{(n,n,...)} \equiv 1$.

For many purposes, a partition $\lambda \in P_n$ can be best described by its Young diagram $\Upsilon_{\lambda}$ (Fig. 5), consisting of $N_{\lambda}$ rows of $\lambda_1, \ldots, \lambda_{N_{\lambda}}$ cells respectively, in terms of which some useful features of $\lambda$ can be defined. The $j$-th cell in row $i$ is denoted $(i, j)$.

- $\lambda' \in P_n$ is the partition whose Young diagram is obtained from $\lambda$’s by transposition; $\Upsilon_{\lambda'} = \Upsilon_{\lambda}^T$
- $B_{\lambda} = \max\{i : (i, i) \in \Upsilon_{\lambda}\} = \max\{i : \lambda_i \geq i\}$ ($\lambda$’s diagonal length)
- $R_{\lambda}(i, j) = \{(u, v) : i \leq u \leq j, j \geq \lambda_{u+1} \leq v \leq \lambda_u\}$ ($\Upsilon_{\lambda}$’s rim segment straddled by $(i, j)$)
- $\Upsilon_{\lambda}^{(i,j)} = \Upsilon_{\lambda} \setminus R_{\lambda}(i, j)$ defines $\lambda^{(i,j)}$ (a diagram obtained from $\lambda$’s by removing a rim segment is a Young diagram; this defines the partition $\lambda^{(i,j)}$)

In addition, for any $\gamma \in P_n$, define $\gamma^r = (\gamma_1, \ldots, \gamma_r, \ldots) \in P_{n-r}$, the partition obtained from $\gamma$ by removing its $r$-th part. The following Murnaghan–Nakayama rule (see [4, Theorem 3.4]) provides a way of recursively evaluating characters: for all $\lambda, \gamma \in P_n$ and $1 \leq r \leq N_{\gamma}$

$$\chi_{\lambda}(\gamma) = \sum_{(i,j) : \#R_{\lambda}(i,j) = r} (-1)^{\lambda_j-i} \chi_{\lambda^{(i,j)}}(\gamma^r) \quad (4.7)$$
\[ \lambda = (8, 8, 7, 4, 4, 1, 1, 0, \ldots) \]
\[ \lambda' = (7, 5, 5, 5, 3, 3, 3, 2, 0, \ldots) \]
\[ \gamma = (10, 9, 7, 3, 2, 2, 0, \ldots) \]
\[ \gamma' = (6, 6, 4, 3, 3, 3, 2, 2, 1, 0, \ldots) \]

Murnaghan-Nakayama rule: \[ \chi_\lambda(\gamma) = \chi_{\lambda'}(\gamma')(\gamma^2) - \chi_{\lambda'}(\gamma^3) \]

Figure 5: Young diagrams of \( \lambda, \gamma \in \mathcal{P}_{33} \). Two \( \lambda \)-cells, \((1, 4)\) and \((2, 3)\), generate rim segments of size 9, the latter shown explicitly, which the Murnaghan-Nakayama rule “peels off” together with the deletion of \( \gamma_2 \).

in the sense that the sum is zero if its index set is empty, and \( \chi_\emptyset(0) = 1 \). Thus, for a fixed order in which \( \gamma \)'s parts are chosen, \( \chi_\lambda(\gamma) \) can be calculated by covering all possible ways of successively stripping off \( \gamma \)-sized rim segments from \( \lambda \)'s diagram, and \( \chi_\lambda(\gamma) = 0 \) if it is impossible to exhaust \( \Upsilon_\lambda \) entirely in this way. In particular

\[ N_\gamma < B_\lambda \implies \chi_\lambda(\gamma) = 0 \quad (4.8) \]

since any rim segment of \( \lambda \) contains at most one diagonal cell \((i, i)\).

### 4.2 Proof of Theorem 4.1

Before proceeding with the proof itself, it will be helpful to characterize the \( \gamma \in \mathcal{P}_n \) which belong to \( nC = nC_{a,b} \) for given \( k \in \mathbb{N} \) and \((a, b) \in I_k\) (assuming \( C \neq \emptyset \)). It follows from \( C_{a,b} \)'s description (4.3) that any such \( \gamma \) can be expressed as a concatenation \((\gamma', \gamma'')\) where \( \gamma' \in G_{a,b}^{(n)} \) and \( \gamma'' \in \mathcal{P}_{n-|\gamma'|} \), and where \( G_{a,b}^{(n)} \) consists of nonincreasing integer valued \( k \)-sequences \( \gamma' \) which by virtue of (4.4) satisfy

\[ i) \ |\gamma'|_1 < n \quad \text{ii) } \exists \delta = \delta(C) > 0 \text{ such that } \gamma_k > (n-|\gamma'|_1) + \delta n . \quad (4.9) \]

This state of affairs is illustrated in Figure 6.
Figure 6: A partition $\gamma$ in $nC_{a,b}$ splits into its first $k$ rows $\gamma'$ and the remainder $\gamma''$ which is nonempty but smaller in size than $\gamma''$'s last row.

**Proof of Theorem 4.1** Fix $C \in \mathcal{C}$ and define $f_n = 1_{nC}$. Then, in terms of $\mu_n^{(n)}$’s density $g_0^{(n)}(\gamma) = \frac{\mu_n^{(n)}(\gamma)}{\pi_n^{(n)}(\gamma)}$:

$$P^{(n)}(X^{(n)}(k) \in nC) = \sum_{\gamma} \mu_0^{(n)}(\gamma) K^{(n)} f_n(\gamma) = \langle g_0^{(n)} , K^{(n)} f_n \rangle = \sum_{\lambda \in \mathcal{P}_n} \theta_\lambda^{(n)} \langle g_0^{(n)} , \chi_\lambda \rangle \langle f_n, \chi_\lambda \rangle$$

and, since $\theta_\lambda^{(n)} = 1$ and $\chi_{(n,0,\ldots)} = 1$,

$$\pi_n^{(n)}(nC) = \langle f_n, 1 \rangle = \theta_\lambda^{(n)} \langle g_0^{(n)} , \chi_\lambda \rangle \langle f_n, \chi_\lambda \rangle$$

so that

$$\Delta_C^{(n)}(k) = \sum_{(n,0,\ldots) \neq \lambda \in \mathcal{P}_n} \theta_\lambda^{(n)} \langle g_0^{(n)} , \chi_\lambda \rangle \langle f_n, \chi_\lambda \rangle . \tag{4.10}$$

By assumption, $g_0^{(n)}(\gamma) = 0$ whenever $N_\gamma > n^\beta$. On the other hand, $\chi_\lambda(\gamma) = 0$ whenever $B_\lambda > n^\beta$ and $N_\gamma \leq n^\beta$ by the consequence (4.8) of Murnaghan–Nakayama’s rule. Thus (4.10) becomes

$$\Delta_C^{(n)}(k) = \sum_{(n,0,\ldots) \neq \lambda \in \mathcal{P}_n, B_\lambda \leq n^\beta} \theta_\lambda^{(n)} \langle g_0^{(n)} , \chi_\lambda \rangle \langle f_n, \chi_\lambda \rangle .$$

Now choose an $\eta$ such that $1 - (\beta' - \beta) < \eta < 1$ and let $n_0 = 5^{\frac{1}{1-\eta}}$. Then, for all $n \geq n_0$,

$$\Delta_C^{(n)}(k) = \left( \sum_{\lambda \in \mathcal{P}_n'} + \sum_{\lambda \in \mathcal{P}_n''} + \sum_{\lambda \in \mathcal{P}_n'''} \right) \theta_\lambda^{(n)} \langle g_0^{(n)} , \chi_\lambda \rangle \langle f_n, \chi_\lambda \rangle , \tag{4.11}$$

where

$$\mathcal{P}_n' = \mathcal{P}_n'(\eta, \beta) = \left\{ \lambda \in \mathcal{P}_n : B_\lambda \leq n^\beta, \lambda_1 \leq n - 2n^\eta, N_\lambda \leq n - 2n^\eta \right\}$$

$$\mathcal{P}_n'' = \mathcal{P}_n''(\eta, \beta) = \left\{ \lambda \in \mathcal{P}_n : B_\lambda \leq n^\beta, n - 2n^\eta < \lambda_1 < n \right\}$$

$$\mathcal{P}_n''' = \mathcal{P}_n'''(\eta, \beta) = \left\{ \lambda \in \mathcal{P}_n : B_\lambda \leq n^\beta, n - 2n^\eta < N_\lambda \right\}$$
(Our choice of \( n_0 \) ensures that \( P_{n_0}' \) and \( P_{n_0}'' \) are disjoint and that \( (n, 0, \ldots) \notin P_{n_0}' \). It turns out that for \( n \) large enough, the terms in (4.11) vanish for all \( \lambda \in P_n' \cup P_n'' \), whereas when \( \lambda \in P_n' \) the factor \( |\theta_\lambda^{(n)}| \) is sufficiently separated from 1:

**Lemma 4.2** \( \exists n_1 = n_1(\beta, C) \) such that \( \langle f_n, \chi_\lambda \rangle = 0, \forall n \geq n_1, \forall \lambda \in P_n' \cup P_n'' \).

**Lemma 4.3** For all \( \lambda \in P_n' \), \( |\theta_\lambda^{(n)}| \leq \frac{\Lambda_{n1}}{n} \) and thus \( \exists n_2 = n_2(\eta) \) such that

\[
|\theta_\lambda^{(n)}| \leq e^{-n^{n-1}} \quad \forall n \geq n_2, \forall \lambda \in P_n'.
\]  

(4.12)

**Proof of Lemma 4.2** Consider first \( \lambda \in P_n'' \). Now, \( C = C_{a,b} \) for some \( k \in \mathbb{N} \) and \( (a, b) \in I_k \), so that, as discussed at the beginning of the section and illustrated in Figure 6, \( \gamma \) can be split into \((\gamma', \gamma'')\) and

\[
\langle f_n, \chi_\lambda \rangle = \sum_{\gamma' \in C_{a,b}^{(n)}} \sum_{\gamma'' \in P_{n-|\gamma'|_1}} \pi_S^{(n)}(\gamma', \gamma'') \chi_\lambda(\gamma', \gamma'').
\]  

(4.13)

(Note that property (4.9i) guarantees that the inner sum is not vacuous, i.e., \(|\gamma''|_1 > 0\).)

We shall show that for every fixed \( \gamma' \in C_{a,b}^{(n)} \) the inner sum in (4.13) equals zero. First apply Murnaghan–Nakayama’s rule (4.7) \( k \) times to \( \chi_\lambda(\gamma', \gamma'') \) by successively stripping rim segments from \( \lambda \), of lengths \( \gamma'_i \) at each stage \( i, i = 1, \ldots, k \). On the one hand \( \lambda_1 > (1-\delta)n \) for \( n \geq n'' = n''_1(\beta, \eta, C) \) (since \( \lambda \in P_n'' \)), and on the other \( \gamma'_i > \delta n, \ i = 1, \ldots, k \) (by (4.9ii)). This implies that at each of these \( k \) reduction stages precisely one rim segment can be stripped off, namely the last \( \gamma'_i \) cells of whatever remains of \( \lambda_i, i = 1, \ldots, k \). Summing up

\[
\chi_\lambda(\gamma', \gamma'') = \chi_{\lambda^*}(\gamma'')
\]  

(4.14)

where \( \lambda^* \in P_{n-|\gamma'|_1} \) is defined by \( \lambda^* = \lambda_1 - n \) and \( \lambda^*_j = \lambda_j, \ j \geq 2 \). As for the first factor of the summand in (4.13), note that (4.9ii) implies \( \gamma'_i > \gamma''_i \) (see Figure 6) and thus

\[
\frac{\pi_S^{(n)}(\gamma', \gamma'')}{\pi_S^{(n-|\gamma'|_1)}(\gamma'')} = \frac{1}{\prod_{i=1}^k \gamma'_i \prod_{j} N_{\gamma'}(j)!} =: R(\gamma').
\]  

(4.15)

Inserting (4.14) and (4.15) in the inner sum of (4.13) we obtain

\[
\sum_{\gamma'' \in P_{n-|\gamma'|_1}} \pi_S^{(n)}(\gamma', \gamma'') \chi_\lambda(\gamma', \gamma'') = R(\gamma')(\chi_{\lambda^*}, 1)|_{n-|\gamma'|_1} = 0
\]

since \( \lambda^* \) is not the trivial partition, that is \( \lambda^* \neq (n-|\gamma'|_1, 0, \ldots) \), (because \( \lambda \neq (n, 0, \ldots) \), and thus \( \chi_{\lambda^*} \) is orthogonal to \( \chi_{\lambda(n-|\gamma'|_1, 0, \ldots)} = 1 \).

The proof for \( \lambda \in P_n''' \) is similar, with \( n \geq n'''_1 = n'''_1(\beta, \eta, C) \), where now the only rim segments which can be stripped off from \( \lambda \) are from its first column. It remains to define \( n_1 = n'''_1 \cap n'''_2 \). \( \square \)
Proof of Lemma 4.3 Using the formula for $\theta^{(n)}_\lambda$ given in (4.6),

$$\theta^{(n)}_\lambda = \frac{1}{n(n-1)} \left( \sum_{i=1}^{n} \lambda_i^2 - \sum_{j=1}^{\lambda_1} \lambda_j^2 \right) \leq \frac{1}{n(n-1)} (\lambda_1 n - \lambda_1) = \frac{\lambda_1}{n},$$

whereas, by duality, $-\theta^{(n)}_\lambda = \theta^{(n)}_\lambda \leq \frac{\lambda_1}{n}$. Moreover, for $\lambda \in \mathcal{P}_n'$,

$$|\theta^{(n)}_\lambda| \leq \left(1 - \frac{2}{n^{1-\eta}}\right) = \left(1 - \frac{2}{n^{1-\eta}}\right)^{n^{1-\eta}} \leq e^{-n^{-\eta}} \quad \text{as soon as} \quad \left(1 - \frac{2}{n^{1-\eta}}\right)^{n^{1-\eta}} \leq 1.$$

We now continue with the estimation of (4.11). As a result of Lemma 4.2 and Lemma 4.3, and recalling that $k \geq n^{\beta'}$, it holds for any $n \geq n_1 \vee n_2$ that

$$\left| \Delta^{(n)}_C(k) \right| \leq \sum_{\lambda \in \mathcal{P}_n'} e^{-n^{\beta'+\eta}} |\langle \varphi^{(n)}_0, \chi_\lambda \rangle| |\langle f_n, \chi_\lambda \rangle|. \quad (4.16)$$

To estimate the number of terms in (4.16), note that the Young diagram $\Upsilon_\lambda$ of any $\lambda \in \mathcal{P}_n$ with $B_\lambda = s$ consists of an $s \times s$ square of cells, with (certainly no more than $n-1$) cells added to each one of the square’s $s$ rows and $s$ columns. Ignoring the various additional constraints, there are $n^{2s}$ ways of making such additions, and thus for any $t > 0$, $\#\{\lambda \in \mathcal{P}_n : B_\lambda \leq t\} \leq tn^{2t}$, so that

$$\#\mathcal{P}_n' \leq \#\{\lambda \in \mathcal{P}_n : B_\lambda \leq n^\beta\} \leq n^\beta n^{2n^\beta} \leq e^{3n^\beta \log n}.$$

As for the terms in (4.16), $|\langle f_n, \chi_\lambda \rangle| \leq 1$ by the Cauchy–Schwartz inequality, and

$$\sup_{\lambda \in \mathcal{P}_n'} |\langle \varphi^{(n)}_0, \chi_\lambda \rangle| \leq \sup_{\lambda \in \mathcal{P}_n'} |\chi_\lambda(\gamma)| \leq n^{n^\beta} = e^{n^\beta \log n}$$

where the second inequality follows from applying Murnaghan–Nakayama’s rule at most $n^\beta$ times, each time with not more that $n$ terms in the sum (4.7).

The above and (4.16) imply that for all $n \geq \max\{n_0, n_1, n_2\}$

$$\left| \Delta^{(n)}_C(k) \right| \leq \exp \left\{ -n^\beta \left( n^{(\beta'-\beta) - (1-\eta)} - 4 \log n \right) \right\}. $$

Eventually, thus, $\left| \Delta^{(n)}_C(k) \right| \leq e^{-n^{\beta'}},$ which concludes the proof of Theorem 4.1. \qed
5 Proof of Vershik’s conjecture

This section is devoted to the proof of Conjecture 1.1, which we restate as

**Theorem 5.1** If \( \mu \in \mathcal{M}_1(\Omega_1) \) is CCF-invariant then \( \mu \) is the Poisson–Dirichlet measure \( \hat{\mu}_1 \).

The main ingredients in its proof have been established in Sections 2, 3 and 4 and are, respectively, the a priori finite moment estimate Proposition 2.1, the couplings with approximating DCF\((n)\)'s of Theorem 3.1, and the fast convergence to equilibrium of the DCF\((n)\) chains in the sense of Theorem 4.1.

**Proof of Theorem 5.1:**
Let \( \Omega_1' = \{ p \in \Omega_1 : \exists \text{ infinitely many } n \in \mathbb{N} \text{ such that } p_n > \sum_{j > n} p_j \} \). We shall show that

\[
\hat{\mu}_1(\Omega_1') = 1 \tag{5.1}
\]

\[
\{ C \cap \Omega_1' : C \subset \mathcal{B}_{\Omega_1'} \} \text{ is measure determining on } (\Omega_1', \mathcal{B}_{\Omega_1'}) \tag{5.2}
\]

\[
\mu(C) = \hat{\mu}_1(C) \quad \forall C \in \mathcal{C} \tag{5.3}
\]

which together imply in particular that \( \mu(\Omega_1') = 1 \), and indeed the theorem’s statement as well.

**Proof of (5.1):** Recall \( \hat{\mu}_1 \)'s description as the law of the nonincreasing rearrangement of the uniform stickbreaking process \( X_n \) (with \( Y_n = 1 - \sum_{j < n} X_j \) the remaining stick length prior to the \( n \)-th break and \( X_n = U_n Y_n \)) and define \( \tau_1 = 1 \), \( \tau_{k+1} = \min\{n > \tau_k : X_n \wedge (Y_n - X_n) < X_j, \forall j \leq \tau_k \} \) for \( k \geq 1 \). Since a.s. \( X_n \uparrow 0 \), each \( \tau_k \) is finite. We claim that

\[
A_k := \{ \tau_k > \frac{1}{2} \} = \{ X_{\tau_k} > \sum_{j > \tau_k} X_j \} \text{ are independent, } \quad P(A_k) = \frac{1}{2} \quad \forall k . \tag{5.4}
\]

This implies that a.s. \( U_{\tau_k} > \frac{1}{2} \) infinitely often, and these \( n = \tau_k \) will be the ones alluded to in \( \Omega_1' \)'s definition. Indeed, on \( A_k \) \( \sum_{j > \tau_k} X_j < X_i \quad \forall i \leq \tau_k \), so that the nondecreasing permutation of the \( X_i \)'s decouples on \( [1, \tau_k] \) and \( (\tau_k, \infty) \) and thus \( p_{\tau_k} = \min_{i \leq \tau_k} X_i > \sum_{j > \tau_k} X_j = \sum_{j > \tau_k} p_j \).

To prove (5.4), represent the splitting variables as \( U_n = \begin{cases} V_n & \text{if } \eta_n = 1 \\ 1 - V_n & \text{if } \eta_n = 0 \end{cases} \), where \( V_n \sim U[0, 1] \) and \( \eta_n \sim \text{Bernoulli}(0.5) \) are independent of each other, and write \( A_k = (B_k \setminus C_k) \cup (B_k^C \setminus C_k^C) \), with \( B_k = \{ V_k > \frac{1}{2} \} \) and \( C_k = \{ \eta_k = 1 \} \). The \( \tau_k \) are \( \mathcal{F} \)-stopping times, where \( \mathcal{F}_n = \sigma(V_1, \ldots, V_n, \eta_0, \ldots, \eta_{n-1}) \) (arbitrarily set \( \eta_0 = 1 \)) so that \( B_k \in \mathcal{F}_{\tau_k} \) and \( C_k \) is independent of \( \mathcal{F}_{\tau_k} \) (in particular \( P(C_k) = 0.5 \)) for all \( k \). For any \( D \in \mathcal{F}_{\tau_k} \)

\[
P(D \cap A_k) = P((D \cap B_k) \cap C_k) + P((D \cap B_k^C) \cap C_k^C) = P(D \cap B_k)P(C_k) + P(D \cap B_k^C)P(C_k^C) = \frac{1}{2} P(D) .
\]

17
Choosing first $D = \Omega_1$ and then $D = \cap \{ \Omega_j \}$ with $J \subset \{1, \ldots, k - 1\}$ (indeed, $A_j \in \mathcal{F}_{\tau_j+1} \subset \mathcal{F}_{\tau_k}$ for $j < k$), we respectively obtain $P(A_k) = 0.5$ and the independence of the $A_n$’s. We have thus proved (5.4) and thus (5.1).

**Proof of (5.2):** Fix $\varepsilon > 0$, $p \in \Omega_1'$, and choose $k$ large enough so that $0 < q := \sum_{j > k} p_j < \min(p_k, \frac{\varepsilon}{k})$. Then let $\delta = \frac{\sqrt{n(p_k - q)}}{k+2}$ and $a_i = p_i - \delta$, $b_i = p_i + \delta$ for $i = 1, \ldots, k$. We claim that $(a, b) \in I_k$. Indeed,

\[
\sum_{i \leq k} b_i = \sum_{i \leq k} (p_i + \delta) \leq 1 - q + \frac{k}{k+2} < 1,
\]

whereas

\[
\sum_{i \leq k} a_i = (p_k - \delta) + \sum_{i \leq k} (p_i - \delta) = (p_k + 1 - q) - (k + 1) \delta > 1.
\]

By definition $p \in C_{a,b}$. Moreover, for any $x \in C_{a,b}$

\[
|x - p|_1 \leq 2 \sum_{j=1}^k |x_j - p_j| + 2 \sum_{j=k+1}^\infty p_j \leq 2k\delta + 2q \leq 2(1 + \frac{k}{k+1})q < \varepsilon,
\]

which shows that for any open $l_1$-ball $B_\varepsilon(p)$ in $\Omega_1'$ there is some $C \subset C$ such that $p \in C \cap \Omega_1' \subset B_\varepsilon(p)$. In other words, $\{C \cap \Omega_1', C \in C\}$ generates $\Omega_1$’s topology.

To conclude the proof of (5.2) we need to check that $C$ is closed under intersections. For any $j \leq k$ then, let $(a_1, b_1) \in I_j$ and $(a_2, b_2) \in I_k$, and if $j < k$ denote $a_{1j} = 0$ and $b_{1j} = 1$ for $i = j + 1, \ldots, k$. It follows immediately that $(a, b)$ defined by $a_i = a_{1i} \lor a_{2i}$ and $b_i = b_{1i} \land b_{2i}$ for $i = 1, \ldots, k$ belongs to $I_k$, and $C_{a_1, b_1} \cap C_{a_2, b_2} = C_{a,b}$.

**Proof of (5.3):** First note that if $(a, b) \in I_k$ then $(1+\varepsilon)a, (1-\varepsilon)b) \in I_k$ for all $\varepsilon$ in some neighborhood of 0, and if $C := C_{a,b} \neq \emptyset$, then so is $C^{(\varepsilon)} := C_{(1+\varepsilon)a, (1-\varepsilon)b}$ for all $\varepsilon$ in a neighborhood of 0.

Once we show that for all $\varepsilon > 0$ small enough,

\[
\hat{\mu}_1(C^{(\varepsilon)}) \geq \mu(C) \geq \hat{\mu}_1(C^{(\varepsilon)}),
\]

let $\varepsilon \searrow 0$ and use $\hat{\mu}_1(\partial C) = 0$ to obtain $\mu(C) = \hat{\mu}_1(C)$ for every $C \in C$, thus proving (5.3) and with it the theorem.

Let $\frac{2}{3} < \alpha < \beta < \gamma < \frac{1}{2}$ be three otherwise arbitrary numbers. Since by Proposition 2.1 $\mu$ satisfies (3.1), we consider for every $n \in \mathbb{N}$ the probability measure $Q^{(n)} = Q^{(n)}_\mu$ introduced in Proposition 3.1 which is defined on a space which supports both a CCF Markov chain $\mathcal{P}(\cdot)$ with $\mu$ as its stationary marginal and a DCF Markov chain $\mathcal{P}(\cdot)$ which “emulates” $\mathcal{P}(\cdot)$ in terms of its initial law (cf. (3.2)) and in the sense that they remain close after $n\gamma$ units of time (cf. (3.3)). For any $n \in \mathbb{N}$,

\[
\mu(C) - \hat{\mu}_1(C^{(\varepsilon)}) = Q^{(n)}(\mathcal{P}[\{n^\gamma\}] \in C) - \hat{\mu}_1(C^{(\varepsilon)}) \geq Q^{(n)}(\mathcal{P}[\{n^\gamma\}] \in C) - Q^{(n)}(\frac{1}{n} \mathcal{P}(\{n^\gamma\}) \in C^{(\varepsilon)}) \geq \frac{Q^{(n)}(\mathcal{P}(\{n^\gamma\}) \in C^{(\varepsilon)})}{Q^{(n)}} = \frac{Q^{(n)}(\mathcal{P}(\{n^\gamma\}) \in C^{(\varepsilon)})}{Q^{(n)}} + \left(\pi^{(n)}(nC^{(\varepsilon)}) - \hat{\mu}_1(C^{(\varepsilon)})\right) = D_1^{(\varepsilon)} - D_2^{(\varepsilon)} + D_3^{(\varepsilon)}.
\]
The first term is estimated using a simple union bound with \( \varepsilon' := \varepsilon \min \{ a_i : 1 \leq i \leq k \} \), and (3.3):
\[
D_1(\varepsilon) \geq -Q^{(n)} \left( \left| p([n^\gamma]) - \frac{1}{n} \ell^{(n)}([n^\gamma]) \right| > \varepsilon' \right) \geq \frac{1}{\varepsilon'} E_{Q^{(n)}} \left[ p([n^\gamma]) - \frac{1}{n} \ell^{(n)}([n^\gamma]) \right] \to 0.
\]

To estimate \( D_2(\varepsilon) \) we would like to apply Theorem 4.1 to the sequence of discrete processes \( \ell^{(n)}(\cdot) \). Their initial laws, however, are guaranteed by (3.2) to be only nearly supported on \( \mathcal{P}_{n,\beta} \), respectively, but not totally as required by Theorem 4.1. Define thus \( \tilde{Q}^{(n)}(\cdot) := Q^{(n)} \left( \cdot \mid \ell^{(n)}(0) \in \mathcal{P}_{n,\beta} \right) \); obviously \( \tilde{Q}^{(n)}(\ell^{(n)}(0) \in \mathcal{P}_{n,\beta}) = 1 \), and under \( \tilde{Q}^{(n)} \), \( \ell^{(n)}(\cdot) \) remains a DCF\(^{(n)}\) chain. Then,
\[
D_2(\varepsilon) \leq \left| \tilde{Q}^{(n)} \left( \frac{\ell^{(n)}([n^\gamma])}{n} \right) \in nC^{(\varepsilon)} \right| - \pi^{(n)}(nC^{(\varepsilon)}) \right| + \| \tilde{Q}^{(n)} - Q^{(n)} \|_{\text{var}} \to 0.
\]

Here we applied Theorem 4.1 for the first term, while \( \| \tilde{Q}^{(n)} - Q^{(n)} \|_{\text{var}} \leq \frac{Q^{(n)}(\ell^{(n)}(0) \notin \mathcal{P}_{n,\beta})}{Q^{(n)}(\ell^{(n)}(0) \in \mathcal{P}_{n,\beta})} \to 0 \) by (3.2).

Finally, recall that \( \pi_S^{(n)}(n \cdot) \to \mu_1 \) weakly ([9, 14]), and since \( C^{(\varepsilon)} \) satisfies \( \mu_1(\partial C^{(\varepsilon)}) = 0 \), it follows that \( \lim_{n \to \infty} \pi_S^{(n)}(nC^{(\varepsilon)}) = \mu_1(C^{(\varepsilon)}) \). Thus \( \lim_{n \to \infty} D_3(\varepsilon) = 0 \). Consequently
\[
\mu(C) - \tilde{\mu}_1(C^{(\varepsilon)}) \geq \lim_{n \to \infty} D_1(\varepsilon) - \lim_{n \to \infty} D_2(\varepsilon) + \lim_{n \to \infty} D_3(\varepsilon) \geq 0.
\]

The reverse inequality \( \tilde{\mu}_1(C^{(-\varepsilon)}) \geq \mu(C) \) is obtained similarly from
\[
\tilde{\mu}_1(C^{(-\varepsilon)}) - \mu(C) \geq -D_1(-\varepsilon) - D_2(-\varepsilon) - D_3(-\varepsilon).
\]

\( \square \)

6 Appendix

Proof of Proposition 2.1 Consider the partition of \((0, 1]\) by \( J_n := (2^{-n-1}, 2^{-n}] \) \((n \geq 0)\) and define on \( \Omega_1 \) the random variables
\[
W_n := \sum_{i \geq 1} p_i 1_{p_i > 2^{-n}} \quad (n \geq 1).
\]

Fix \( n \geq 1 \). If two intervals are merged then \( W_n \) can only increase and if some interval is split then \( W_n \) can only decrease. We call the increment in the case of merging \( \Delta_+ \geq 0 \) and the loss in the case of splitting \( \Delta_- \leq 0 \). Given \( p \), we can bound \( \Delta_+ \) by
\[
\Delta_+ \geq \sum_{i \neq j} p_i^2 p_j 1_{p_i \in J_n} 1_{p_j > 2^{-n-1}}
\]
\[
= \left( \sum_i p_i^2 1_{p_i \in J_n} \right) \left( \sum_j p_j 1_{p_j > 2^{-n-1}} \right) - \sum_i p_i^3 1_{p_i \in J_n}
\]
\[
\geq 2^{-2n-2} \left( \sum_j 1_{p_j \in J_n} \right) \left( \sum_i p_i 1_{p_i > 2^{-n-1}} \right) - 2^{-2n} \sum_i p_i 1_{p_i \in J_n}.
\]
and compute $\Delta_-$ as

$$
\Delta_- = \sum_i p_i^2 \int_0^1 x p_i 1_{x p_i \leq 2^{-n} < p_i} + (1-x)\left[ x p_i 1_{(1-x) p_i \leq 2^{-n} < p_i} \right] dx
$$

$$
= 2 \sum_i p_i^2 1_{2^{-n} < p_i} \int_0^1 1_{x \leq 2^{-n}/p_i} = 2 \sum_i p_i^2 1_{2^{-n} < p_i} \left[ \frac{2^n}{2} \right]_0^{2^{-n}/p_i} = 2^{-2n} \sum_i p_i 1_{2^{-n} < p_i}.
$$

Therefore,

$$
\Delta_+ - \Delta_- \geq 2^{-2n} \left( \frac{1}{4} \left( \sum_i 1_{p_i \in J_n} \right) \left( \sum_j p_j 1_{p_j > 2^{-n} - 1} \right) - \sum_i p_i 1_{2^{-n} - 1 < p_i} \right)
$$

$$
\geq 2^{-2n} \left( \frac{1}{4} \left( \sum_i 1_{p_i \in J_n} \right) \left( \sum_j p_j 1_{p_j > 2^{-n} - 1} \right) - 1 \right).
$$

Since $\int \Delta_+ - \Delta_- \ d\mu = 0$ due to stationarity this implies

$$
4 \geq \int \left( \sum_i 1_{p_i \in J_n} \right) \left( \sum_j p_j 1_{p_j > 2^{-n} - 1} \right) \ d\mu \quad (6.1)
$$

$$
\geq 2^{-n-1} \left( \sum_i 1_{p_i \in J_n} \right)^2 \ d\mu \geq 2^{-n-1} \left( \int \sum_i 1_{p_i \in J_n} \ d\mu \right)^2.
$$

Since this holds for any $n \geq 1$ we get

$$
\int \sum_i 1_{p_i \in J_n} \ d\mu = O(2^{\beta n}) \quad (n \to \infty) \quad (6.2)
$$

with $\beta = 1/2$. Therefore, for any $\alpha > \beta = 1/2$,

$$
\int \sum_i p_i^\alpha \ d\mu \leq \sum_{n \geq 0} 2^{-\alpha n} \int \sum_i 1_{p_i \in J_n} \ d\mu \leq c \sum_{n \geq 0} 2^{n(\beta - \alpha)} < \infty, \quad (6.3)
$$

for some constant $c > 0$, thus proving (2.1) for $\alpha > 1/2$. We shall now use this result to extend it to all $2/5 < \alpha < 1$, as required. To this end, observe that we have due to (6.1) for arbitrary
$0 < \beta < 1,$

$$4 \geq \int \left( \sum_i 1_{p_i \in J_n} \right) \left( \sum_j p_j 1_{p_j, p_i > 2^{n-1}} \right) 1 \left\{ \sum_j p_j 1_{p_j > 2^{n-1}} > 2^{-n\beta} \right\} d\mu$$

$$\geq 2^{-n\beta} \int \sum_i 1_{p_i \in J_n} 1 \left\{ \sum_j p_j 1_{p_j > 2^{n-1}} > 2^{-n\beta} \right\} d\mu$$

$$= 2^{-n\beta} \left( \int \sum_i 1_{p_i \in J_n} d\mu - \int \sum_i 1_{p_i \in J_n} 1 \left\{ \sum_j p_j 1_{p_j > 2^{n-1}} \leq 2^{-n\beta} \right\} d\mu \right)$$

$$\geq 2^{-n\beta} \left( \int \sum_i 1_{p_i \in J_n} d\mu - \frac{2^{-n\beta}}{2^{-n-1}} \mu \left[ \sum_j p_j 1_{p_j > 2^{n-1}} \leq 2^{-n\beta} \right] \right)$$

$$\geq 2^{-n\beta} \left( \int \sum_i 1_{p_i \in J_n} d\mu - 2^{-n\beta} + 1 \mu \left[ \forall i : p_i \leq 2^{-n\beta} \right] \right). \quad (6.4)$$

To bound the $\mu$-probability in the last expression we recall from (6.3) that for $1/2 < \alpha < 1$,

$$\infty > \int \sum_i p_i^\alpha d\mu \geq \int \left( \sum_i p_i^\alpha \right) 1 \left\{ \forall i : p_i \leq 2^{-n\beta} \right\} d\mu \quad (6.5)$$

for any $n \geq 0$. On the event $\{ \forall i : p_i \leq 2^{-n\beta} \}$, by Jensen’s inequality,

$$\sum_i p_i^\alpha = \sum_i p_i \cdot p_i^{\alpha-1} \geq \left( \sum_i p_i \cdot p_i \right)^{\alpha-1} \geq 2^{n\beta(1-\alpha)}.$$ 

Therefore, for all $1/2 < \alpha < 1$ due to (6.5), $\mu \left[ \forall i : p_i \leq 2^{-n\beta} \right] = O(2^{-n\beta(1-\alpha)})$ as $n \to \infty$. Substituting this into (6.4) we get that for any $1/2 < \alpha < 1$

$$\int \sum_i 1_{p_i \in J_n} d\mu = O(2^n \beta + 2^{-n\beta + n(1-\alpha)}) = O(2^{n \max\left( \beta, 1-\beta(2-\alpha) \right)}).$$

The choice $\beta = (3-\alpha)^{-1}$ minimizes $\max\left( \beta, 1-\beta(2-\alpha) \right)$ and therefore yields that (6.2) and consequently also (6.3) and (2.1) hold for any $\alpha, \beta > (3-1/2)^{-1} = 2/5$. \hfill \Box

**Remark:** A posteriori, once it has been established that $\mu$ must be the Poisson-Dirichlet law, Proposition 2.1 holds for all $\alpha > 0$ since by [7, (20)]

$$\int \sum_{i \geq 1} p_i^\alpha d\mu = \int_0^1 z^{\alpha-1} dz = \frac{1}{\alpha}.$$ 

21
References


Persi Diaconis
Dept. of Mathematics and Dept. of Statistics
Stanford University
Stanford, CA 94305, U.S.A.

Eddy Mayer-Wolf
Dept. of Mathematics
Technion
Haifa 32000, Israel
emw@tx.technion.ac.il

Ofer Zeitouni
Dept. of Electrical Engineering and Dept. of Mathematics
Technion
Haifa 32000, Israel
zeitouni@ee.technion.ac.il
www-ee.technion.ac.il/~zeitouni

Martin Zerner
Dept. of Mathematics
Stanford University
Stanford, CA 94305, U.S.A.
zerner@math.stanford.edu
math.stanford.edu/~zerner