Combinatorics, Probability and Computing (2006) 15, 31–62. © 2006 Cambridge University Press
doi:10.1017/S0963548305007054 Printed in the United Kingdom

A Tale of Three Couplings: Poisson–Dirichlet and GEM Approximations for Random Permutations

RICHARD ARRATIA,[1] A. D. BARBOUR[2†] and SIMON TAVARÉ[3‡]

[1]Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA
(e-mail: rarratia@math.usc.edu)

[2]Angewandte Mathematik, Universität Zürich, Winterthurerstrasse 190,
CH-8057, Zürich, Switzerland
(e-mail: a.d.barbour@math.unizh.ch)

[3]Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA
and
Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Cambridge, UK
(e-mail: stavare@usc.edu)

Received 31 January 2005; revised 7 June 2005

For Béla Bollobás on his 60th birthday

For a random permutation of n objects, as $n \to \infty$, the process giving the proportion of elements in the longest cycle, the second-longest cycle, and so on, converges in distribution to the Poisson–Dirichlet process with parameter 1. This was proved in 1977 by Kingman and by Vershik and Schmidt. For soft reasons, this is equivalent to the statement that the random permutations and the Poisson–Dirichlet process can be coupled so that zero is the limit of the expected ℓ_1 distance between the process of cycle length proportions and the Poisson–Dirichlet process. We investigate how rapid this metric convergence can be, and in doing so, give two new proofs of the distributional convergence.

One of the couplings we consider has an analogue for the prime factorizations of a uniformly distributed random integer, and these couplings rely on the 'scale-invariant spacing lemma' for the scale-invariant Poisson processes, proved in this paper.

1. Introduction

Pick a permutation of n objects, with all $n!$ possibilities equally likely. Write $L_1^{(n)}$ for the length of the longest cycle, $L_2^{(n)}$ for the length of the second-longest cycle, and so on, with $L_j^{(n)} = 0$ if the permutation has fewer than j cycles. Kingman [18] and Vershik and

† Supported in part by Schweizerischer Nationalfondsprojekte Nos. 20-61753.00 and 20-67909.02.
‡ Holder of a Royal Society–Wolfson Research Merit Award.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Shmidt [22] proved that these lengths, taken as proportions of n, converge in distribution, to a process $(L_1, L_2, \dots)$ now known as the Poisson–Dirichlet process (with parameter 1):

$$n^{-1}\big(L_1^{(n)}, L_2^{(n)}, \dots\big) \Rightarrow (L_1, L_2, \dots). \tag{1.1}$$

A condition equivalent to (1.1) is that for every fixed k, the distribution of the random vector $n^{-1}(L_1^{(n)}, L_2^{(n)}, \dots, L_k^{(n)})$ converges to that of $(L_1, L_2, \dots, L_k)$.

The convergence in (1.1) is, by definition, based on the product topology for $\mathbb{R}^\infty$, and states that for every bounded continuous $f : \mathbb{R}^\infty \to \mathbb{R}$, we have

$$\lim_{n\to\infty} \mathbb{E} f\big(n^{-1}\big(L_1^{(n)}, L_2^{(n)}, \dots\big)\big) = \mathbb{E} f((L_1, L_2, \dots).$$

Both processes $n^{-1}(L_1^{(n)}, L_2^{(n)}, \dots)$ and $n^{-1}(L_1^{(n)}, L_2^{(n)}, \dots)$ take values in the simplex $\Delta := \{(x_1, x_2, \dots) : \ x_1 + x_2 + \cdots = 1 \text{ and } \forall i, x_i \geqslant 0\}$, and the ℓ_1 distance is a bounded metric on Δ which induces the same topology as the restriction of the product topology to $\Delta \subset \mathbb{R}^\infty$. It then follows from Strassen's theorem, together with the the Kantorovich–Rubinstein theorem (see Chapter 11 of Dudley [10],) that the distributional convergence claimed in (1.1) is *equivalent* to the following statement about couplings: 'The processes $(L_1^{(n)}, L_2^{(n)}, \dots)$ and $(L_1, L_2, \dots)$ can be coupled, that is, simultaneously constructed on a single probability space, so that

$$\mathbb{E} \sum_{i\geqslant 1} |L_i^{(n)}/n - L_i| \to 0 \tag{1.2}$$

as $n \to \infty$.'

In this paper, we investigate how rapidly the left side of (1.2) can converge to zero. Rather than divide the cycle lengths by n, we prefer to multiply the Poisson–Dirichlet components by n, so the quantity of interest is n times the expected ℓ_1 distance involved in (1.2). Thus, for any given coupling we consider

$$\begin{aligned} d_{\mathrm{PD}}(n) &:= \mathbb{E} \sum_{i\geqslant 1} |L_i^{(n)} - nL_i| \\ &= n\,\mathbb{E}\,\big\|n^{-1}\big(\, L_1^{(n)}, L_2^{(n)}, \dots\,\big) - (L_1, L_2, \dots)\big\|_1 . \end{aligned} \tag{1.3}$$

Any coupling in which $d_{\mathrm{PD}}(n) = o(n)$ as $n \to \infty$ serves as a proof of the distributional convergence in (1.1). We give three different couplings that do so, and hence three different proofs of (1.1). The first coupling is the 'natural' coupling involved in the usual proof of (1.1), and what is new is our analysis of the expected ℓ_1 distance for this coupling. Our second and third couplings are novel, and thus provide two new proofs of (1.1). We consider the convergence in (1.1) to be so fundamental that *every* explicit coupling that achieves (1.2) would be interesting, and this, rather than a quest to minimize the distance $d_{\mathrm{PD}}(n)$, is our primary motivation. In particular, fourth and further couplings would be interesting, even if they offer no improvement on the metric bound.

We give a lower bound in Theorem 2.2 stating that for *any* coupling,

$$\liminf_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) \geqslant 1/4.$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

In Section 5, we analyse the obvious natural coupling, based on the canonical cycle notation for a permutation, to show that it achieves

$$\limsup_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) \leqslant \mathbb{E}|Q|,$$

where Q is a random variable defined in (5.12). Simulation shows that $\mathbb{E}|Q| \doteq 0.39357 \pm 0.00001$. In Section 6 we consider a second coupling, which achieves

$$\limsup_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) \leqslant 1/3.$$

(We believe, but do not attempt to prove, that this coupling has $\lim_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) = 1/3$.) In Section 8 we consider a third coupling, which achieves

$$\limsup_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) \leqslant 1/4,$$

and hence by comparison with the lower bound, this third coupling satisfies

$$\lim_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) = 1/4. \tag{1.4}$$

In Section 9 we give the analogous result for θ-biased permutations.

1.1. Age-ordered cycles

The age-ordering of the cycles of a permutation is described in detail and from first principles in Section 4.1. For an introductory overview, it suffices to say that starting from the list of cycle lengths $(L_1^{(n)}, L_2^{(n)}, \dots)$, one can apply auxiliary randomization (a size-biased permutation) to re-order the list to get an 'age-ordered' list $(A_1^{(n)}, A_2^{(n)}, \dots)$, with $L_i^{(n)} = i$th-largest of $A_1^{(n)}, A_2^{(n)}, \dots$. Furthermore, this age-ordered list, rescaled by n, has a distributional limit $(A_1, A_2, \dots)$, called the GEM process, which has a remarkably simple structure.

In each of our three couplings, we simultaneously construct the age-ordered cycle lengths $(A_1^{(n)}, A_2^{(n)}, \dots)$ and the scaled GEM $(nA_1, nA_2, \dots)$, aiming to make them close to one another. Then we construct the Poisson–Dirichlet $(L_1, L_2, \dots)$ from the GEM $(A_1, A_2, \dots)$ by applying RANK, that is, we take L_i to be the ith-largest of $A_1, A_2, \dots$. Lemma 3.2 below shows that RANK can only reduce the l_1-distance, so that

$$\sum_{i\geqslant 1} |L_i^{(n)} - nL_i| \leqslant \sum_{i\geqslant 1} |A_i^{(n)} - nA_i|, \tag{1.5}$$

and taking expectations then shows that

$$d_{\mathrm{PD}}(n) \leqslant d_{\mathrm{GEM}}(n) := \mathbb{E}\left\{ \sum_{i\leqslant 1} |A_i^{(n)} - nA_i| \right\}. \tag{1.6}$$

The second and third couplings are rather different from the first. For coupling 1, we show that

$$\lim_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{GEM}}(n) = \mathbb{E}|Q|,$$

so that, by Lemma 3.2, $\limsup_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) \leqslant \mathbb{E}|Q|$. In contrast, for coupling 2, we show that $\liminf_{n\to\infty} (\log n)^{-1}\, d_{\mathrm{PD}}(n) \leqslant 1/3$, but do not get an estimate for $d_{\mathrm{GEM}}(n)$;

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

rather, we show that, for some permutation σ of $\mathbb{N}$, $\mathbb{E}\sum_{i\geqslant 1}|A_i^{(n)} - nA_{\sigma(i)}|$ is small, and then apply Lemma 3.2. We use a similar strategy with coupling 3, matching a rearrangement of the GEM random variables, and then applying Lemma 3.2 to deduce an upper bound on $d_{\mathrm{PD}}(n)$. This gives $\limsup_{n\to\infty}(\log n)^{-1} d_{\mathrm{PD}}(n) \leqslant 1/4$, and comparison with the lower bound from Theorem 2.1 then yields (1.4).

We believe that there is an essential difference between $d_{\mathrm{PD}}(n)$ and $d_{\mathrm{GEM}}(n)$. The idea is that if one starts with a close coupling of $n^{-1}(L_1^{(n)}, L_2^{(n)}, \dots)$ with $(L_1, L_2, \dots)$, and applies a size-biased permutation to each process, then, since the sizes are slightly different, it is *not* possible to use exactly the same permutation for size-biasing. The following conjecture uses a constant strictly greater than $1/4$, to assert that there is an essential difference between $d_{\mathrm{PD}}(n)$ and $d_{\mathrm{GEM}}(n)$; it uses the particular constant $\mathbb{E}|Q|$ for the simple reason that our first coupling achieves this constant.

Conjecture ($100). Any coupling of the age-ordered cycle lengths with the GEM has $\liminf_{n\to\infty}(\log n)^{-1} d_{\mathrm{GEM}}(n) \geqslant \mathbb{E}|Q|$, where Q is specified by (5.12).

1.2. Comparison with prime factorization of an integer

Consider a random positive integer chosen uniformly from 1 to n, and let $P_i(n)$ be the size of its ith-largest prime factor, with $P_i(n) = 1$ if i is greater than the number of factors of the random integer. Billingsley [7] proved in 1972 that

$$(\log n)^{-1}(\log P_1(n), \log P_2(n), \dots) \Rightarrow (L_1, L_2, \dots),$$

with the same Poisson–Dirichlet limit as in (1.1).

As in (1.3), one can consider the ℓ_1 distance to the limit, but this time scaling up by $\log n$ rather than by n. Arratia [2] gave a coupling, similar to our third coupling for permutations, which has

$$\mathbb{E}\sum |\log P_i(n) - (\log n)L_i| = O(\log\log n). \tag{1.7}$$

A conjecture from [2], that $O(1)$ can be achieved in place of $O(\log\log n)$, remains open, with a $100 prize offered for its resolution. Our third coupling for permutations, as well as the coupling in [2], relies on the 'scale-invariant spacing lemma' for the scale-invariant Poisson processes, proved in Section 7 in this paper.

2. A lower bound for all couplings

A lower bound asymptotic to $(1/4)\log n$ for $\mathbb{E}\sum_{i\geqslant 1}|L_i^{(n)} - nL_i|$ can be derived from the fact that the intensity $\mu(a,b)$, defined by

$$\mu(a,b) = \mathbb{E}\sum_{i\geqslant 1} \mathbb{1}(L_i \in (a,b)) \quad \text{for } 0 < a < b \leqslant 1,$$

is given by the explicit expression $\mu(a,b) = \int_a^b x^{-1}\,dx$. An alternate statement of this property of the Poisson–Dirichlet process, from [12] and [17], is

$$\mathbb{E}\sum_{j\geqslant 1}\phi(L_j) = \int_0^1 \phi(x)\frac{dx}{x}, \tag{2.1}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

for any function ϕ that makes the integral absolutely convergent. One can also deduce (2.1) directly from the distributional convergence in (1.1), together with the fact that $\mathbb{E}C_i^{(n)} = 1/i$ for $1 \leqslant i \leqslant n$, where $C_i^{(n)}$ is the number of cycles of length i.

Theorem 2.1. *Let $L_i^{(n)}$ denote the size of the ith-largest cycle of a random permutation, and let L_i be the ith coordinate of the Poisson–Dirichlet process with parameter* 1. Uniformly *over all couplings of these two processes,*

$$\liminf_{n\to\infty} (\log n)^{-1}\mathbb{E}\sum_{i\geqslant 1} |L_i^{(n)} - nL_i| \geqslant \frac{1}{4}. \tag{2.2}$$

Proof. The factor $1/4$ arises as the long term average of the sawtooth function $d(x,\mathbb{Z}) = |x - \lfloor x + .5\rfloor|$. Since each $L_i^{(n)} \in \mathbb{Z}$, any coupling has

$$\mathbb{E}\sum_{i\geqslant 1} |L_i^{(n)} - nL_i| \geqslant \mathbb{E}\sum_{i\geqslant 1} d(nL_i,\mathbb{Z}).$$

Using (2.1), we see that

$$\mathbb{E}\sum_{i\geqslant 1} d(nL_i,\mathbb{Z}) = \int_0^1 d(nx,\mathbb{Z})x^{-1}\,dx = \int_0^n d(x,\mathbb{Z})x^{-1}\,dx \sim \frac{1}{4}\log n.$$

The asymptotic in the last step comes from the fact that $d(x,\mathbb{Z}) = x$ if $0 < x \leqslant 1/2$, so that $\int_0^{1/2} d(x,\mathbb{Z})x^{-1}\,dx = 1/2$, while for $k = 1, 2, \ldots,$

$$d(x,\mathbb{Z}) = \begin{cases} k - x, & \text{if } x \in [k - 1/2, k], \\ x - k, & \text{if } x \in [k, k + 1/2], \end{cases}$$

so that the contribution from 'one sawtooth' is

$$\begin{aligned} \int_{k-.5}^{k+.5} \frac{d(x,\mathbb{Z})}{x}\,dx &= \int_{k-.5}^{k} \frac{k-x}{x}\,dx + \int_{k}^{k+.5} \frac{x-k}{x}\,dx \\ &= k\left(\int_{k-.5}^{k} \frac{dx}{x} - \int_{k}^{k+.5} \frac{dx}{x}\right) \\ &= k\log(k^2/(k^2 - 1/4)) = -k\log(1 - 1/(4k^2)). \end{aligned}$$

Hence

$$\int_0^1 d(nx,\mathbb{Z})x^{-1}\,dx = \frac{1}{2} + \sum_{k=1}^{n-1} -k\log\left(1 - \frac{1}{4k^2}\right) + \int_{n-.5}^{n} \frac{n-x}{x}\,dx, \tag{2.3}$$

and, since $-k\log(1 - \frac{1}{4k^2}) \sim 1/(4k)$, the sum, as well as the entire right side of (2.3), is asymptotic to $(1/4)\log n$. □

3. Lemmas for sorting and ℓ_1 distances

There is a well-known 'rearrangement inequality' stating that, for any nonincreasing sequences $(l_i,\ 1 \leqslant i \leqslant k)$ and $(m_i,\ 1 \leqslant i \leqslant k)$ and permutations ρ and σ, $\sum_1^k l_i m_i \geqslant \sum_1^k l_{\rho(i)} m_{\sigma(i)}$. Lemma 3.1 treats differences rather than products, and we provide a simple

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

direct proof, very similar to the proof of the rearrangement inequality. Lemma 3.2, an extension of Lemma 3.1, implies that the RANK function is a contraction for the ℓ_1 distance.

Lemma 3.1 may be viewed as a special case of the result (1.1.14) in [20], or Exercises 11.8.1–2 in [10]. Recall that for any random variables X, Y, with cumulative distribution functions F, G, the Wasserstein distance, $\inf_{\text{couplings}} \mathbb{E}|X - Y|$, equals $\int_{-\infty}^{\infty} |F(t) - G(t)|\, dt$, and this infimum is achieved by constructing both X and Y using their quantiles applied to a single uniform [0,1] random variable. Dudley [10, p. 342] traces the history back to Gini in 1914 and Dall'Aglio [8]. The special case of Lemma 3.1 is that of empirical distributions: X is chosen uniformly from the multiset $\{l_1, \ldots, l_k\}$ and Y is chosen uniformly from the multiset $\{m_1, \ldots, m_k\}$. But in a sense this is circular, since proofs of the result on $\inf_{\text{couplings}} \mathbb{E}|X - Y|$ start with the discrete case, and then take limits to get the general case.

Lemma 3.1. *If $l_1 \geqslant l_2 \geqslant \cdots \geqslant l_k$ and $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_k$, then for any permutations ρ, σ on $\{1, 2, \ldots, k\}$,*

$$\sum_1^k |\ell_i - m_i| \leqslant \sum_1^k |\ell_{\rho(i)} - m_{\sigma(i)}|. \tag{3.1}$$

Proof. Without loss of generality, the permutation σ can be taken to be the identity with $\sigma(i) = i$ for $i = 1$ to k. Thus our goal is to show that the 'score' $s(\rho) := \sum_1^k |l_{\rho(i)} - m_i|$ is minimized by taking ρ to be the identity permutation.

Any time there is a reversal, *i.e.*, $i < j$ with $l_{\rho(i)} > l_{\rho(j)}$, we can apply the transposition $(i\ j)$ without increasing the score. To check this, write $a = m_i, A = m_j, b = l_{\rho(j)}, B = l_{\rho(i)}$, so that $a \leqslant A$ and $b \leqslant B$. We have $s(\rho) - s(\rho \circ (i\ j)) = |a - B| + |A - b| - (|a - b| + |A - B|) \geqslant 0$, which can be verified by considering cases such as $a < A < b < B$, which yields zero, $a < b < A < B$, which yields $2|A - b|$, or $a < b < B < A$, which yields $2(B - b)$. An arbitrary permutation ρ can be transformed into the identity in a finite number of such steps. □

Lemma 3.2. *If $l_1 \geqslant l_2 \geqslant \cdots$ and $m_1 \geqslant m_2 \geqslant \ldots$ then, for any permutations ρ, σ on $\mathbb{N}$,*

$$\sum_{i \geqslant 1} |l_i - m_i| \leqslant \sum_{i \geqslant 1} |l_{\rho(i)} - m_{\sigma(i)}|.$$

Hence the map RANK is a contraction on ℓ_1.

Proof. As in the previous lemma, without loss of generality σ can be taken to be the identity permutation, so our goal is to show, for an arbitrary permutation of $\mathbb{N}$, that $\sum_{i \geqslant 1} |l_i - m_i| \leqslant \sum_{i \geqslant 1} |l_{\rho(i)} - m_i|$. This in turn is equivalent to showing that, for arbitrary $k \geqslant 1$,

$$\sum_{i=1}^k |l_i - m_i| \leqslant \sum_{i \geqslant 1} |l_{\rho(i)} - m_i|. \tag{3.2}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Fix k. The set of ordered pairs $\{(\rho(1),1),\dots,(\rho(k),k),(1,\rho^{-1}(1)),\dots,(k,\rho^{-1}(k))\}$ is a matching between two sets, say A and B, with

$$k \leqslant n = |A| = |B| \leqslant 2k.$$

Note that $\{1,2,\dots,k\} \subset A \cap B$. Label the elements of these sets so that $A = \{i_1, i_2, \dots, i_n\}$ and $B = \{j_1, j_2, \dots, j_n\}$ with $i_1 < i_2 < \cdots < i_n$ and $j_1 < j_2 < \cdots < j_n$ (and hence $1 = i_1 = j_1, \dots, k = i_k = j_k$). The matching of A and B can be expressed as $\{(1,\tau(1)),\dots,(n,\tau(n))\}$ for a permutation τ of $\{1,\dots,n\}$. We have, writing for example $i(a) \equiv i_a$,

$$\begin{aligned}\sum_{a=1}^{k} |l_a - m_a| = \sum_{a=1}^{k} |l_{i(a)} - m_{j(a)}| &\leqslant \sum_{a=1}^{n} |l_{i(a)} - m_{j(a)}| \\ &\leqslant \sum_{a=1}^{n} |l_a - m_{\tau(a)}| \leqslant \sum_{i \geqslant 1} |l_{\rho(i)} - m_i|.\end{aligned}$$

In the above, the middle inequality is justified by Lemma 3.1 applied to the permutation τ, while the first and third inequalities are simply by inserting additional nonnegative terms. □

4. Canonical cycle notation and the GEM

The easy path to understanding the structure of $(L_1^{(n)}, L_2^{(n)}, \dots)$, which are the cycle lengths taken in decreasing order, is to consider first a more elaborate construction: the cycle lengths in the order produced by writing out the canonical cycle notation for a permutation on $[n] := \{1,2,\dots,n\}$.

4.1. Age order of the cycles

A permutation $\rho \in \mathscr{S}_n$ can be written as an (ordered) product of cycles in the following way: start the first cycle with the integer 1, followed by its image $\rho(1)$, the image of $\rho(1)$ and so on. Once this cycle is completed, the second cycle starts with the smallest unused integer followed by its images, and so on. For example, the permutation $\rho \in \mathscr{S}_{10}$ given by

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 2 & 7 & 3 & 6 & 4 & 5 & 8 & 10 & 1 \end{pmatrix} \tag{4.1}$$

is decomposed as

$$\rho = (1\,9\,10)(2)(3\,7\,5\,6\,4)(8), \tag{4.2}$$

a permutation with two singleton cycles (or fixed points), one cycle of length 3, and one of length 5. This example with $n = 10$ has cycle lengths in *decreasing* order given by $(L_1^{(n)}, L_2^{(n)}, \dots) = (5,3,1,1,0,0,\dots)$, while the cycle lengths in canonical, or *age* order are $(A_1^{(n)}, A_2^{(n)}, \dots) = (3,1,5,1,0,0,\dots)$.

The distribution now known as the GEM arose in population biology, in [14, 11, 19]. Just as the Poisson–Dirichlet distribution (with parameter 1) may be characterized as the limit occurring in (1.1), the GEM distribution with parameter 1 may be characterized as

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

the distributional limit for the proportions of elements in the cycles taken in age order:

$$n^{-1}\big(A_1^{(n)}, A_2^{(n)}, \dots\big) \Rightarrow (A_1, A_2, \dots). \tag{4.3}$$

Conditional on the values $(L_i^{(n)}, i \geqslant 1)$, the joint distribution of $(A_i^{(n)}, i \geqslant 1)$ is obtained by size-biasing. The size-biased order derived from any finite collection $\{l_1, \dots, l_m\} \subset \mathbb{R}_+$ with sum L is the random permutation σ of $\{1, 2, \dots, m\}$ obtained from the relations

$$\mathbb{P}[\sigma(1) = j] = l_j/L;$$

$$\mathbb{P}[\sigma(r+1) = j \mid \sigma(1) = j_1, \dots, \sigma(r) = j_r] = l_j \left\{ L - \sum_{i=1}^{r} l_{j_i} \right\}^{-1},$$

$0 \leqslant r \leqslant m-1$. For an infinite collection $\{l_1, l_2, \dots\}$ with sum $L < \infty$, the same construction can be used to define the size-biased order σ of $\mathbb{N}$, but if $L = \infty$, the algorithm fails. However, the size-biased order can be more generally defined as a random ordering of $\{l_j, j \geqslant 1\}$ with the property that any *finite* collection of them are in their size-biased order. Such a random order can be realized by the following simple construction. Assign to each l_j the value $\lambda_j = l_j^{-1} S_j$, where $(S_j, j \geqslant 1)$ are independent standard exponential random variables, and arrange the l_j in increasing order of λ_j. If the sets $\{l_j, j \geqslant 1\} \cap (M^{-1}, M)$ are finite for each $M \in \mathbb{N}$, the l_j can be indexed by $i \in \mathbb{Z}$ in increasing order of λ_j. The size-biased ordering of a random sequence $\{L_j, j \geqslant 1\}$ is obtained in the same way, by first sampling the multiset consisting of its values, and then using the independent S_js to determine their order.

4.2. Coupling to independent discrete random variables

The probabilistic structure found in writing the canonical cycle notation for a randomly chosen permutation involves independent Bernoulli $(1/i)$ random variables, for $i = 1, 2, \dots$. We call this structure the *Feller coupling*, and trace it back to Feller (1945) and Rényi (1962); see also Arratia, Barbour and Tavaré (1992). In writing the canonical cycle notation for a random $\rho \in \mathscr{S}_n$, say with $n = 10$, one always starts with '(1', and then makes a ten-way choice, between '(1)(2', '(1 2', ..., and '(1 10'. One continues with a nine-way choice, an eight-way choice, ..., a two-way choice, and finally a one-way choice.

Say that D_i, chosen from 1 to i, is used to make the i-way choice – take item D_i from the i available choices listed in increasing order; the example (4.2) has $(D_{10}, D_9, \dots, D_2, D_1) = (9, 9, 1, 1, 5, 3, 3, 2, 1, 1)$. Clearly, the map constructing canonical cycle notation,

$$(D_1, D_2, \dots, D_n) \mapsto \rho,$$

from $[1] \times [2] \times \cdots \times [n]$ to $\mathscr{S}_n$, is one-to-one, and hence a bijection. This is the natural coupling; starting with $D_1, D_2, D_3, \dots$ independent, and with no further randomization, we produce random permutations, simultaneously for $n = 1, 2, \dots$, distributed uniformly over $\mathscr{S}_n$.

Define ξ_i to be the indicator function

$$\xi_i = \mathbb{1}\{D_i = 1\} = \mathbb{1}\{\text{end cycle when there is an } i\text{-way choice}\}. \tag{4.4}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Thus

$$\mathbb{P}[\xi_i = 1] = \frac{1}{i}, \qquad \mathbb{P}[\xi_i = 0] = 1 - \frac{1}{i}, \qquad i \geqslant 1,$$

and

$$\xi_1, \xi_2, \xi_3, \ldots, \quad \text{are independent.}$$

The sequence $\xi_1\xi_2\xi_3 \ldots \xi_n$ read from ξ_n down to ξ_1 determines the list of cycle lengths in age order. The length $A_1^{(n)}$ of the first cycle is the waiting time to the first of $\xi_n, \xi_{n-1}, \ldots$ to take the value 1, the length $A_2^{(n)}$ of the next cycle is the waiting time to the next 1, and so on.

We may view this in terms of 'an artificial 1 in position $n+1$', and consider the sequence $\xi_1\xi_2\xi_3 \ldots \xi_n 1$ of length $n+1$ that begins and ends with 1. Every i-spacing in $1\xi_2\xi_3 \ldots \xi_n 1$, that is, every pattern $10^{i-1}1$ of two ones separated by $i-1$ zeros, corresponds to a cycle of length i. The size of the rightmost spacing in $1\xi_2\xi_3 \ldots \xi_n 1$ gives the size of the first cycle in canonical cycle notation. The example (4.2) with $n = 10$ has $(D_1, \ldots, D_n) = (1, 1, 2, 3, 3, 5, 1, 1, 9, 9)$, so $\xi_1\xi_2\xi_3 \ldots \xi_n 1 = 11000011001$. We read the spacings from right to left to see that the cycle lengths in age order are $(A_1^{(n)}, A_2^{(n)}, \ldots, A_4^{(n)}) = (3, 1, 5, 1)$, with $(A_1^{(n)}, A_2^{(n)}, \ldots) = (3, 1, 5, 1, 0, 0, \ldots)$.

The single set of realizations $(D_i, i \geqslant 1)$ enables one to generate a coupled set of realizations of uniform random permutations $\sigma(n)$ for each $n \geqslant 1$, for which, for each fixed j, the random variables

$$C_j^{(n)} := \sum_{r=1}^{n-j-1} \mathbb{1}\{\xi_r = 1, \xi_{r+1} = \cdots = \xi_{r+j-1} = 0, \xi_{r+j} = 1\} + \mathbb{1}\{\xi_{n-j} = 1, \xi_{n-j+1} = \cdots = \xi_n = 0\},$$

the numbers of cycles of lengths j in $\sigma(n)$, converge a.s. as $n \to \infty$ to random variables $C_j^{(\infty)}$. The joint distribution of $\{C_j^{(\infty)}, j \geqslant 1\}$ is that of independent Poisson random variables with means $1/j$: see Arratia and Tavaré [6] and Diaconis and Pitman [9].

Now let $M_0 = 1$ and $M_i = \min\{j > m_{i-1}, \xi_j = 1\}$, $i \geqslant 1$; set $M_i^{(n)} = M_i \wedge n$. Then the differences $B_i^{(n)} = M_i^{(n)} - M_{i-1}^{(n)}$, $i \geqslant 1$, form the cycle lengths of $\sigma(n)$ in reverse age (size-biased) order, followed by an infinite string of zeros, and we also have $B_i^{(n)} \to B_i := M_i - M_{i-1}$ a.s. for each i. A random vector with the same distribution as $(B_i^{(n)}, i \geqslant 1)$ can equally be realized by sampling $(C_j^{(n)}, 1 \leqslant j \leqslant n)$, and then arranging the $C_j^{(n)}$ elements of size j, $1 \leqslant j \leqslant n$, together in reversed size-biased order, again filling out with zeros. Now define the events

$$E_{1m}^{(n)} := \{C_j^{(n^2)} = C_j^{(\infty)}, 1 \leqslant j \leqslant m\}, \qquad E_{2m}^{(n)} := \{B_j^{(n^2)} = B_j, 1 \leqslant j \leqslant n\}$$

and

$$E_{3m}^{(n)} := \left\{\min_{j>n} B_j^{(n^2)} \leqslant m\right\}.$$

Then simple calculation shows that $\lim_{n\to\infty} \mathbb{P}[E_{lm}^{(n)}] = 0$ for each fixed m, $l = 1, 2, 3$. However, all elements of the sequence $(B_i, i \geqslant 1)$ of size at most m are realized in

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

the correct order on the set $E_{1m}^{(n)} \cap E_{2m}^{(n)} \cap E_{3m}^{(n)}$, by taking $(C_j^{(n^2)}, 1 \leqslant j \leqslant n^2)$, performing the size-biased ordering to get $(B_j^{(n^2)}, 1 \leqslant j \leqslant n)$, and then taking the elements of size at most m in $(B_j^{(n^2)}, 1 \leqslant j \leqslant n)$; since, on this set, we could equally well have begun with $(C_j^{(\infty)}, 1 \leqslant j \leqslant m)$ in place of $(C_j^{(n^2)}, 1 \leqslant j \leqslant m)$, it follows that the infinite sequence $(B_i, i \geqslant 1)$ is itself in reversed size-biased order. Hence, to construct the age-ordered distributions, we just need to sample independent Poisson random variables with means $1/j$, $j \geqslant 1$, and reverse size-bias the outcome, to obtain the sequence $(B_i, i \geqslant 1)$, and from it to deduce the $(\xi_j, j \geqslant 1)$.

5. The first coupling: using canonical cycle notation

A simple calculation shows that $A_1^{(n)}$ is uniformly distributed over $\{1, 2, \ldots, n\}$ – it might be constructed from a uniform [0,1] random variable U using the ceiling function, $A_1^{(n)} = \lceil nU \rceil$. Similarly, $A_2^{(n)}$ is uniform on 1 to $n - A_1^{(n)}$, meaning that conditional on $n - A_1^{(n)} = m$, if $m > 0$ then $A_2^{(n)}$ is uniformly distributed over $\{1, 2, \ldots, m\}$, and otherwise $A_2^{(n)}$ is zero. We could take $A_2^{(n)} = \lceil (n - A_1^{(n)})U' \rceil$ with U, U' i.i.d. Likewise $A_3^{(n)}$ is uniform on 1 to $n - A_1^{(n)} - A_2^{(n)}$, and so on.

The easy way to view this is in terms of the number of elements *not* yet used after k cycles have been taken:

$$N_0 = n; \qquad N_k \equiv N_k^{(n)} := n - A_1^{(n)} - \cdots - A_k^{(n)} \quad \text{for } k \geqslant 1. \tag{5.1}$$

Then, conditional on $N_0 = n, N_1 = n_1, \ldots, N_k = n_k$, if $n_k > 0$ then N_{k+1} is uniformly distributed over $\{0, 1, \ldots, n_k - 1\}$, while if $n_k = 0$ then $N_{k+1} = 0$. There is an obvious coupling which achieves this: take $U_1, U_2, \ldots$ to be i.i.d. uniform [0,1] random variables and from them construct the process $A_1^{(n)}, A_2^{(n)}, \ldots$ via

$$N_0^{(n)} = n, \qquad N_k^{(n)} := \lfloor N_{k-1}^{(n)} U_k \rfloor \quad \text{for } k \geqslant 1, \tag{5.2}$$

$$A_k^{(n)} := N_{k-1}^{(n)} - N_k^{(n)} \quad \text{for } k \geqslant 1. \tag{5.3}$$

This yields the process $(A_1^{(n)}, A_2^{(n)}, \ldots)$ distributed exactly as the process of cycle lengths in age order, for a random permutation.

Define

$$S_0 = 1, \qquad S_k := U_1 U_2 \cdots U_k, \qquad A_k := S_{k-1} - S_k, \quad \text{for } k \geqslant 1. \tag{5.4}$$

Since $\lfloor nx \rfloor / n \to x$ for all x, by induction on $k \geqslant 0$ we see that $N_k^{(n)}/n \to S_k$ as $n \to \infty$, and by differencing, that $A_k^{(n)}/n \to A_k$. This proves, using only elementary analysis, that

$$n^{-1}\left(A_1^{(n)}, A_2^{(n)}, \ldots\right) \to (A_1, A_2, \ldots) \tag{5.5}$$

under the coupling (5.3), for *all* realizations of the uniform random variables U_i; the weak convergence in (4.3) naturally follows.

With or without coupling, (5.4) is a simple description of the GEM(1) limit $(A_1, A_2, \ldots)$; it is due to Ignatov [15, 16] and others; see [5, p. 119] for further history.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

5.1. Analysis of the expected ℓ_1 distance

The goal of this section is to explain and prove the following theorem, stating that the expected ℓ_1 distance for the coupling above is asymptotic to a certain constant times $\log n$ as $n \to \infty$.

Theorem 5.1. *For the process of cycle lengths in age order, of a permutation chosen uniformly at random from $\mathcal{S}_n$, in comparison with the GEM process* (5.3)*, scaled up by a factor of n, the expected ℓ_1 distance* (1.6)*, for the coupling given by* (5.2)–(5.4)*, has the following asymptotic value as $n \to \infty$:*

$$d_{\mathrm{GEM}}(n) \sim \mathbb{E}|Q| \log n, \tag{5.6}$$

where the distribution of Q is given by (5.12).

To start the proof, let us write $t_k(n) = \mathbb{E}|T_k(n)|$ for the kth term of (1.6), the scaled-up expected ℓ_1 distance, for the coupling given by (5.2)–(5.4):

$$t_k(n) := \mathbb{E}|T_k(n)| \quad \text{where } T_k(n) := nA_k - A_k^{(n)}. \tag{5.7}$$

We use curly braces to denote the fractional part function, $\{x\} := x - \lfloor x \rfloor$, so that $\lfloor x \rfloor = x - \{x\}$.

Lemma 5.2. *For $k \geqslant 0$,*

$$N_k^{(n)} - nS_k = -\sum_{j=1}^{k} \{N_{j-1}^{(n)} U_j\} U_{j+1} \cdots U_k. \tag{5.8}$$

Proof. For $k = 0$, both sides are zero. For $k \geqslant 1$,

$$\begin{aligned} N_k^{(n)} - nS_k &= \lfloor N_{k-1}^{(n)} U_k \rfloor - nS_{k-1} U_k \\ &= N_{k-1}^{(n)} U_k - \{N_{k-1}^{(n)} U_k\} - nS_{k-1} U_k \\ &= -\{N_{k-1}^{(n)} U_k\} + U_k \big(N_{k-1}^{(n)} - nS_{k-1}\big). \end{aligned} \tag{5.9}$$

Using induction to expand $(N_{k-1}^{(n)} - nS_{k-1})$, we get (5.8). □

We use the notation $\overline{U} := 1 - U$ for the complement of a uniformly distributed variable. Starting from (5.7), and using (5.9) to get the third line below,

$$\begin{aligned} T_k(n) &= (nS_{k-1} - nS_k) - \big(N_{k-1}^{(n)} - N_k^{(n)}\big) \\ &= \big(N_k^{(n)} - nS_k\big) - \big(N_{k-1}^{(n)} - nS_{k-1}\big) \\ &= -\{N_{k-1}^{(n)} U_k\} - \big(N_{k-1}^{(n)} - nS_{k-1}\big)\overline{U}_k \\ &= -\{N_{k-1}^{(n)} U_k\} + \overline{U}_k \sum_{j=1}^{k-1} \{N_{k-j-1}^{(n)} U_{k-j}\} \prod_{l=k-j+1}^{k-1} U_l, \end{aligned} \tag{5.10}$$

an empty product being taken to be 1.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Now let $W_1, W_2, \ldots$ and $V_1, V_2, \ldots$ be mutually independent uniform random variables, and consider the random variables

$$Q_k = -V_0 + \sum_{j=1}^{k-1} \left\{ V_j \prod_{l=1}^{j} W_l \right\} \tag{5.11}$$

and

$$Q = -V_0 + \sum_{j \geqslant 1} \left\{ V_j \prod_{l=1}^{j} W_l \right\}. \tag{5.12}$$

Since $\mathbb{E}|Q - Q_k| = 2^{-k}$ is summable, $Q_k \to Q$ almost surely and in L_1. Clearly $\mathbb{E}Q = 0$, but for $\mathbb{E}|Q|$ we have no closed form expression.

(**Heuristic.** Imagine k large, but n much larger, say larger even than e^k, so that nS_k is large. We claim that T_k is close in distribution to Q_k, and hence close to Q. Superficially, this does not appear to be so – in the fourth line of the display (5.10), there are k independent uniform variables, while in (5.11), there are $2k-1$ independent uniforms. However, when N is large, $\{NU\}$ is approximately independent of U, in the sense that the low order bits of U give the approximate value of U, while higher order bits of U control the value of $\{NU\}$.)

To carry out this approximation, start with the i.i.d. uniform [0,1] variables $W_1, W_2, \ldots$ and $V_1, V_2, \ldots$. Note that Q_k from (5.11) is exactly equal in distribution to

$$Q'_k = -V_k + \overline{W}_k \sum_{j=1}^{k-1} \left\{ V_{k-j} \prod_{l=k-j+1}^{k-1} W_l \right\}. \tag{5.13}$$

Given $n \geqslant 1$, we *define*

$$N_0 := n; \qquad N_k := \lfloor N_{k-1} W_k \rfloor, \quad k \geqslant 1,$$

and

$$U_k := (N_k + V_k)/N_{k-1} \quad \text{if } N_{k-1} \geqslant 1; \qquad U_k := V_k \quad \text{if } N_{k-1} = 0.$$

It is then an elementary calculation to *check* that we have satisfied (5.2); *i.e.*, we check that $N_k = \lfloor N_{k-1} U_k \rfloor$, that U_k is uniform, and that $U_1, U_2, \ldots$ are independent. It is also immediate that

$$\{N_{k-1} U_k\} = V_k \quad \text{if } N_{k-1} \geqslant 1, \tag{5.14}$$

and that

$$|U_k - W_k| \leqslant 1/\max(1, N_{k-1}). \tag{5.15}$$

Combining this coupling with (5.10), we have

$$T_k = -V_k + \overline{U}_k \sum_{j=1}^{k-1} \left\{ V_{k-j} \prod_{l=k-j+1}^{k-1} U_l \right\}, \tag{5.16}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

which matches formula (5.13) for Q'_k, except that Q'_k has the variables W which are independent of the Vs, while T_k has the variables U which are dependent on the Vs. We use (5.15) in Lemma 5.3 to show that $\mathbb{E}|T_k - Q'_k|$ is $O(n^{-\delta})$ small, uniformly in $k \leqslant (1-\varepsilon)\log n$. This, together with $\mathbb{E}|Q'_k| = \mathbb{E}|Q_k| \to \mathbb{E}|Q|$, implies that as $n \to \infty$,

$$\sum_{k \leqslant (1-\varepsilon)\log n} \mathbb{E}|T_k| \sim \sum_{k \leqslant (1-\varepsilon)\log n} \mathbb{E}|Q_k| \sim (1-\varepsilon)\mathbb{E}|Q| \log n.$$

Combining this with estimates for the remaining ranges, and arguing with fixed but arbitrarily small $\varepsilon > 0$, the final outcome is that the expected ℓ_1 distance given by (1.6), for this particular coupling, is asymptotic to $\mathbb{E}|Q| \log n$, as was stated formally in Theorem 5.1.

For the range $(1-\varepsilon)\log n < k \leqslant (1+\varepsilon)\log n$, it suffices to use (5.10) to see that for all k, n, $\mathbb{E}|T_k| < 2$; we have $2\varepsilon \log n$ terms in this range, so the net contribution, at most $4\varepsilon \log n$, is made small relative to $\log n$ by taking ε small.

For $k \geqslant k_0 := (1+\varepsilon)\log n$, large deviation theory easily shows that $\sum_{k \geqslant k_0} \mathbb{E}T_k \to 0$. Here are the details. Consider a 'good event' G_k, defined by

$$G_k = \left\{ \frac{\log S_k}{k} < (-1+\varepsilon/2) \right\}.$$

As in the proof of Lemma 5.3, standard large deviation theory gives the upper bound $\mathbb{P}(G_k^c) \leqslant e^{-\delta k}$, where $\delta = \delta(\varepsilon) > 0$. Using (5.2) and (5.4), we have $0 \leqslant N_k^{(n)} \leqslant nS_k$ so that $|T_k| \leqslant nS_k$, and on the good event this gives $|T_k| \leqslant \exp(\log n + (-1+\varepsilon/2)k)$. For $k = k_0$, this upper bound is no more than $n^{-\varepsilon/2+\varepsilon^2/2}$, and goes down exponentially fast with increasing k, so the sum over all $k \geqslant k_0$ is comparable to the bound at k_0. For the contribution from the bad event, observe first that from (5.10), we have $|T_k| \leqslant k$ always, and $\sum_{k \geqslant k_0} k e^{-\delta k} = O(e^{-\delta k_0/2}) = O(n^{-\delta/2})$.

The proof of Theorem 5.1 is completed by the following lemma.

Lemma 5.3. *Fix $\varepsilon > 0$, and let $k_0 \equiv k_0(n) := \lfloor (1-\varepsilon)\log n \rfloor$. There exists a $\delta > 0$, so that uniformly in $1 \leqslant k \leqslant k_0$,*

$$\mathbb{E}|T_k - Q'_k| = O(n^{-\delta}). \tag{5.17}$$

Proof. Consider a 'good event' G_n, defined by

$$G_n = \left\{ \frac{\log S_{k_0}}{k_0} > -(1+\varepsilon/3) \right\}.$$

Since $-\log S_k$ is the sum of k i.i.d. exponentially distributed random variables with mean 1, large deviation theory gives an upper bound of the form $\mathbb{P}(G_n^c) \leqslant e^{-I(\varepsilon/3)k_0} \leqslant n^{-\delta_1}$, where $\delta_1 = (1-\varepsilon)I(\varepsilon/3) > 0$.

For the contribution from the bad event, observe first that from (5.10) and (5.13) we have $|T_k - Q'_k| \leqslant 1 + k$ always, so with respect to this contribution (5.17) is satisfied.

Assume that the good event occurs. For $\varepsilon \in (0,1)$, we have the bound $(1+\varepsilon/3)(1-\varepsilon) < 1 - \varepsilon/2$, and hence $nS_{k_0} > n^{\varepsilon/2}$. Since N_k is obtained from nS_k using k applications of the floor function, we have $N_k > n^{\varepsilon/2} - k_0$ for every $k \leqslant k_0$. Finally, consider the difference

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

of Q'_k from (5.13) and the coupling expression (5.16) for T_k. One converts Q'_k to T_k by replacing factors W_i by U_i, in expressions that are products of factors in $(0,1)$, so that such a replacement causes a change not exceeding $|U_i - W_i|$ in absolute value, and by (5.15) this is at most $1/N_{k_0} < 1/(n^{\varepsilon/2} - k_0)$. There are $1 + 2 + \cdots + k - 1$ places to make such a replacement, and hence for every $k \leqslant k_0$, on the good event

$$|T_k - Q'_k| \leqslant k_0^2/N_{k_0} = O\big(n^{-\varepsilon/4}\big). \qquad \square$$

6. The second coupling: to exploit $\mathbb{E}|U - U'| = 1/3$

Our second coupling achieves

$$d_{\mathrm{PD}}(n) \sim \frac{1}{3}\log n,$$

with the constant $1/3$ arising as the expected value of the difference of two independent uniform (0,1) random variables. To do this, we consider the 'scale-invariant' Poisson process $\mathscr{X}$ on $(0,\infty)$ with intensity dx/x, restricted to $(0,n)$, and for each $i = 2,3,\ldots$, we *throw away* any points representing a second, third, or further arrival of $\mathscr{X}$ in an interval $(i-1,i]$. The 'first arrival' in $(i-1,i]$, if there is one, corresponds to $\xi_i = 1$ in the Feller coupling.

For large i, the 'first arrival' in $(i-1,i]$ is approximately at $i - U$, where U is uniform, and if the Feller process has adjacent arrivals at $i < j$, corresponding to a cycle of length $j - i$, there is a matching spacing in $\mathscr{X}$ of length approximately $(j - U') - (i - U) = (j - i) + (U - U')$; with $\mathbb{E}|U - U'| = 1/3$, this accounts for a $(1/3)\log n$ contribution to $d_{\mathrm{PD}}(n)$. The act of restoring 'extra' arrivals in an interval $(i-1,i]$ can be viewed as splitting a small piece, of length at most 1, from a spacing already accounted for as $(j - U') - (i - U)$, and matching these additional small pieces with zeros in the age-ordered cycle process – since $A_j^{(n)} = 0$ for j greater than the number of cycles of the random permutation. Then Lemma 3.2 is applied.

6.1. Details for this coupling

In the point process $\mathscr{X}$, for any $0 < a < b < n$, the number $\mathscr{X}(a,b]$ of points in the interval $(a,b]$ is Poisson with mean $\mathbb{E}\mathscr{X}(a,b] = \int_a^b dx/x = \log(b/a)$, and disjoint intervals have independent counts. If we label the points of $\mathscr{X}$ in decreasing order, so that

$$n > X_1 > X_2 > \cdots,$$

then the sequence $X_1, X_2, X_3 \ldots$ is distributed exactly as $nU_1, nU_1U_2, nU_1U_2U_3, \ldots$, which are the points in (5.4), scaled up by n. Thus the GEM, scaled up by n, is

$$(nA_1, nA_2, nA_3, \ldots) = (n - X_1, X_1 - X_2, X_2 - X_3, \ldots). \tag{6.1}$$

We extend of course by taking $L_i :=$ ith-largest of the $A_1, A_2, \ldots$.

Let

$$N_i = \mathscr{X}(i-1,i], \quad i = 2,3,\ldots \tag{6.2}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

so that N_i is Poisson with mean

$$\lambda_i := \mathbb{E}N_i = \log(i/(i-1)) = 1/i + 1/(2i^2) + 1/(3i^3) + \cdots,$$

and $\mathbb{P}(N_i = 0) = 1 - (1/i)$. Taking $N_1 = \infty$, we define

$$\xi_i := \min(1, N_i), \quad i = 1, 2, 3, \ldots, \tag{6.3}$$

thereby coupling the scale-invariant Poisson process $\mathscr{X}$ with the Bernoulli process $\xi_1, \xi_2, \ldots$ of the Feller coupling (4.4). The paragraph following (4.4) gives the age-ordered cycle lengths $A_1^{(n)}, A_2^{(n)}, \ldots$ in terms of spacings between successive ones in $1\xi_n\xi_{n-1}\cdots\xi_3\xi_2 1$. In detail, write

$$K := \xi_1 + \xi_2 + \cdots + \xi_n$$

so that K represents the number of cycles in our random cycle structure. Define $M_0 \equiv n+1$ and for $j = 1$ to K, let M_j be the index of the jth 1 in $\xi_n \cdots \xi_2\xi_1$, so that $M_K = 1$ always,

$$A_j^{(n)} = M_{j-1} - M_j, \qquad j = 1 \text{ to } K, \tag{6.4}$$

and $A_j^{(n)} = 0$ for $j > K$. Of course we take $L_i^{(n)} :=$ ith-largest of $A_1^{(n)}, A_2^{(n)}, \ldots$. At this stage, we have defined our second coupling; it only remains to estimate $d_{\mathrm{PD}}(n)$.

The Poisson process $\mathscr{X}$, restricted to $(i-1, i]$, can be constructed from the count N_i, together with an i.i.d. sequence $X_{i,1}, X_{i,2}, \ldots$ of locations in $(i-1, i]$ distributed with density $f_i(x) := 1/(x\lambda_i)$. This density can be expressed as a mixture of the uniform density on $(i-1, i]$, taken with weight $1/(i\lambda_i) = 1/(1 + 1/(2i) + 1/(3i^2) + \cdots$, and some other density on $(i-1, i]$, taken with weight $d(i) := 1 - 1/(i\lambda_i) = 1/(2i) + O(i^{-2})$. Thus we can take V_i uniform on (0,1), so that

$$\mathbb{E}|(i - V_i) - X_{i,1}| < d(i), \tag{6.5}$$

for every $i \geqslant 2$. For the interval (0,1], there is an infinite number of arrivals (but not in i.i.d. locations); we may label these in decreasing order as $X_{1,1} > X_{1,2} > \cdots > 0$. Note that $X_{1,1}$ is uniformly distributed in (0,1), so we may define $d(1) = 0$ and consider (6.5) to also hold for $i = 1$. Finally, note that the $\xi_1, \xi_2, \ldots$ from (6.3) and the $V_1, V_2, \ldots$ from (6.5) can be taken to be mutually independent.

The Poisson process $\mathscr{X}$ restricted to $(0, n]$ is realized as the random set of points

$$\{X_{i,j} : 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant N_i\}. \tag{6.6}$$

We get a subset $\hat{\mathscr{X}}$ of $\mathscr{X}$ by taking only those $X_{i,j}$ with $N_i > 0$ and $j = 1$, so that the cardinality of $\hat{S}$ is K. Label these points as $\hat{X}_j$ with $n \geqslant \hat{X}_1 > \hat{X}_2 > \cdots > \hat{X}_K$, with $\hat{X}_K \in (0, 1]$. For $j = 1$ to K, let M_j be the index of the jth 1 in $\xi_n \cdots \xi_2\xi_1$, so that $\hat{X}_j = X_{M_j,1} \in (M_j - 1, M_j]$ and $M_K = 1$ always. *If the spacings construction* (6.1) were applied to $\hat{\mathscr{X}}$ in place of $\mathscr{X}$, the ℓ_1 distance to the age-ordered list of cycle lengths, that is, the ℓ_1 distance between $(n - \hat{X}_1, \hat{X}_1 - \hat{X}_2, \ldots, \hat{X}_{K-1} - \hat{X}_K, 0, 0, \ldots)$ and $(A_1^{(n)}, A_2^{(n)}, \ldots, A_K^{(n)}, 0, 0, \ldots)$, would be

$$D = |(n - \hat{X}_1) - (n + 1 - M_1)| + \sum_{j=2}^{K} |(\hat{X}_{j-1} - \hat{X}_j) - (M_{j-1} - M_j)|. \tag{6.7}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Looking back at (6.5), we write $U_j := V_{M_j}$ so that

$$|\hat{X}_j - (M_j - U_j)| \leqslant d(M_j) \tag{6.8}$$

and hence

$$\left| D - \sum_{j=2}^{K} |U_{j-1} - U_j| \right| \leqslant 2 + 2\sum_{j=2}^{K} d(M_j). \tag{6.9}$$

Since $\mathbb{E}K = 1 + 1/2 + \cdots + 1/n \sim \log n$, and $\lim_{i\to\infty} d(i) = 0$, and always $M_j \geqslant j$, it follows first that the expectation of the left side of (6.9) is $o(\log n)$; then since $\mathbb{E}\sum_2^K |U_{j-1} - U_j| \sim \mathbb{E}|U - U'| \log n = (1/3)\log n$, it follows that $\mathbb{E}D \sim (1/3)\log n$.

What is the effect of replacing $\mathscr{X}$ with $\hat{\mathscr{X}}$ in the spacings construction? There is an easy bound, provided that we consider the spacings arranged from large to small. Namely, the ℓ_1 distance between $\mathrm{RANK}(n - \hat{X}_1, \hat{X}_1 - \hat{X}_2, \dots)$ and $n(L_1, L_2, \dots) = \mathrm{RANK}(n - X_1, X_1 - X_2, \dots)$ is at most

$$2\sum_{j=2}^{n} (N_i - 1)^+ + 2. \tag{6.10}$$

(For the first sum above: any second, third, or further arrival in $(i-1, i]$ serves to split off a piece of a spacing involving the first arrival, but the split off piece has length at most 1, so the effect on the ℓ_1 distance of the ranked lists of deleting one of these extra arrivals is at most $1 + 1$, using Lemma 3.2. The second term in (6.10), 2, comes from splitting the interval $(0, \hat{X}_K)$ into an infinite number of subpieces.) Since $\mathbb{E}N_i \sim 1/i$, the expectation of the positive part of $(N_i - 1)$ is $O(1/i^2)$, and the expected ℓ_1 distance in (6.10) is $O(1)$ as $n \to \infty$.

Now comes a subtle point: the rearrangement inequality in Lemma 3.2 implies that the ℓ_1 distance between $\mathrm{RANK}(n - \hat{X}_1, \hat{X}_1 - \hat{X}_2, \dots)$ and $(L_1^{(n)}, L_2^{(n)}, \dots) = \mathrm{RANK}(A_1^{(n)}, A_2^{(n)}, \dots)$ is *at most* D, and hence has mean asymptotically *at most* $(1/3)\log n$. We believe that there exists a matching lower bound, asymptotic to $(1/3)\log n$, but in view of the third coupling, it does not seem worth pursuing this.

The net result of these arguments is that we have proved

Theorem 6.1. *The second coupling, given by* (6.1)–(6.4)*, achieves*

$$\limsup_{n\to\infty} (\log n)^{-1} d_{\mathrm{PD}}(n) \leqslant 1/3.$$

7. The scale-invariant spacings lemma

The following 'scale-invariant spacings lemma' was first presented in [1], an unpublished manuscript, and was used in a coupling for prime factors with the Poisson–Dirichlet process in [2]; see also [4]. It will be used in our third coupling, in Section 8.

Start with the Poisson process $\mathscr{P}$ with intensity $\theta/x\,dx$. Consider the process $\mathscr{Y}$ with a point for each spacing in $\mathscr{P}$. To be precise, the points of $\mathscr{P}$ can with probability one be labelled $X_i \in (0, \infty)$ for $i \in \mathbb{Z}$ so that

$$\cdots < X_2 < X_1 < 1 < X_0 < X_{-1} < \cdots, \tag{7.1}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

with $\lim_{i\to\infty} X_i = 0, \lim_{i\to-\infty} X_i = \infty$. The spacings are the points $Y_i := X_i - X_{i+1}$ for $-\infty < i < \infty$, and

$$\mathscr{Y} := \sum_{i\in\mathbb{Z}} \delta_{Y_i} \tag{7.2}$$

is a random counting measure on $(0,\infty)$; here $\delta_{Y_i}(A) = 1$ if $Y_i \in A$ and $= 0$ otherwise.

Lemma 7.1 (The scale-invariant spacing lemma). *For any $\theta > 0$, the random measures $\mathscr{P}$ and $\mathscr{Y}$ have the same distribution.*

Proof. Start with a Poisson process $\mathscr{R}$ on $(0,\infty)^2$ having points (W, Y) with intensity $\theta \exp(-wy)\, dy\, dw$. The intensity for the projection π_1 on the w coordinate is $\theta/w\, dw$, and similarly for the y-projection π_2; *i.e.*, each projection is a copy of the process $\mathscr{P}$. Label the points of $\mathscr{R}$ as (W_i, Y_i) in decreasing order of their w-coordinates, say with $W_{-1} > 1 > W_0$, so that $\cdots > W_{-1} > W_0 > W_1 > W_2 > \cdots$. Define, for $-\infty < j < \infty$,

$$X_j := \sum_{-\infty<i\leqslant j} Y_i, \tag{7.3}$$

and let $\mathscr{X}$ be the process with these points. Since the spacings of $\mathscr{X}$ are by construction the points Y_i of a process which has the same distribution as $\mathscr{P}$, the goal is to show that $\mathscr{X}$ also has the same distribution as $\mathscr{P}$. We do this by calculating, for $k = 0, 1, 2, \ldots$, for $0 < x_0 < x_1 < \cdots < x_k$, the intensity for the process $\mathscr{X}$ to have points at x_i for $i = 0$ to k and no points in (x_i, x_{i+1}) for $i = 0$ to $k-1$.

Consider, for $c > 0$, the restriction of $\mathscr{R}$ to $(c,\infty)\times(0,\infty)$. The intensity function of the y-projection of this process is

$$f_c(y) = \int_{w>c} \theta \exp(-wy)\, dw = \theta \exp(-cy)/y.$$

For the case $c = 1$, this intensity function arises in the study of the Poisson–Dirichlet process [17], and the sum of the y-coordinates has the Gamma distribution with parameter θ and density $g(x;\theta) = x^{\theta-1}e^{-x}/\Gamma(\theta)$. For general $c > 0$, since $f_c(y) = cf_1(cy)$, the process with parameter c is the same as the process with parameter 1, rescaled by dividing each y-coordinate by c. In particular, the sum S_c of the y-coordinates of the points in this process has density function $g(x;\theta,c) = c(xc)^{\theta-1}e^{-xc}/\Gamma(\theta)$.

For $\mathscr{X}$ to have a point at x, there must be some value c such that $S_c = x$. Taking w_0 to be the infimum of such c, the process $\mathscr{R}$ must have a point on the line $w = w_0$ (not necessarily the point labelled (W_0, Y_0)), and $\mathscr{R}$ must have x for the sum of the y-coordinates of the points in $(w_0,\infty)\times(0,\infty)$. Thus the intensity function for the pair (x, w_0) is $(\theta/w_0)g(x;\theta,w_0)$. Integrating out w_0 yields

$$\int_{w>0} \frac{\theta}{w} \frac{w(xw)^{\theta-1}e^{-xw}}{\Gamma(\theta)}\, dw = \frac{\theta}{x}$$

which shows that $\mathscr{X}$ has the same intensity as the Poisson process $\mathscr{P}$.

For a Poisson process on $(0,\infty)$ with intensity $f(x)\, dx$, for $0 < a < b$, the intensity function to have two consecutive points at a, b is $f(a)f(b)\exp(-\int_a^b f(x)\, dx)$; and for the

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

process $\mathscr{P}$ this simplifies to $(\theta/a)(\theta/b)(a/b)^{\theta} = \theta^2 a^{\theta-1} b^{-\theta-1}$. Similarly, the intensity function for $\mathscr{P}$ to have three consecutive points at $a < b < c$ simplifies to $\theta^3 a^{\theta-1} b^{-1} c^{-\theta-1}$, and for four consecutive points at $a < b_1 < b_2 < c$ the expression simplifies to $\theta^4 a^{\theta-1} b_1^{-1} b_2^{-1} c^{-\theta-1}$. Since the case with two points does not illustrate the full pattern, we show below the details of computing the intensity for $\mathscr{X}$ to have three consecutive points at $a = x_0$, $b = x_0 + y_0$, $c = x_0 + y_0 + y_1$, to check that it agrees with the intensity for $\mathscr{P}$.

The configuration of three consecutive points specified for $\mathscr{X}$ above requires that for some $0 < w_1 < w_0$, $\mathscr{R}$ has points at (w_0, y_0) and (w_1, y_1), with no points in the strip $(w_1, w_0) \times (0, \infty)$, and with the sum S_{w_0} of y-coordinates of points in $(w_0, \infty) \times (0, \infty)$ equal to x_0. The intensity function for this, with respect to $dx_0\, dy_0\, dy_1\, dw_0\, dw_1$, is a product with four factors:

$$\theta \exp(-w_0 y_0) \cdot \theta \exp(-w_1 y_1) \cdot (w_1/w_0)^{\theta} \cdot w_0 (x_0 w_0)^{\theta-1} \exp(-x_0 w_0)/\Gamma(\theta),$$

which reduces to

$$\theta^2 x_0^{\theta-1} \Gamma(\theta)^{-1} w_1^{\theta} \exp(-w_1 y_1) \exp(-w_0(x_0 + y_0)).$$

Integrating over $0 < w_1 < w_0$ yields

$$\begin{aligned}
&\frac{\theta^2 x_0^{\theta-1}}{\Gamma(\theta)} \int_0^{\infty} dw_1\, w_1^{\theta} e^{-w_1 y_1} \int_{w_1}^{\infty} dw_0\, e^{-w_0(x_0+y_0)} \\
&\quad = \frac{\theta^2 x_0^{\theta-1}}{\Gamma(\theta)} \int_0^{\infty} dw_1\, w_1^{\theta} e^{-w_1 y_1} \frac{e^{-w_1(x_0+y_0)}}{x_0 + y_0} \\
&\quad = \frac{\theta^2 x_0^{\theta-1}}{(x_0+y_0)\Gamma(\theta)} \int_0^{\infty} dw_1 \left(\frac{w_1(x_0 + y_0 + y_1)}{x_0 + y_0 + y_1} \right)^{\theta} e^{-w_1(x_0+y_0+y_1)} \\
&\quad = \frac{\theta^2 \Gamma(\theta+1)}{\Gamma(\theta)} x_0^{\theta-1} \frac{1}{x_0 + y_0} (x_0 + y_0 + y_1)^{-\theta-1} \\
&\quad = \theta^3 a^{\theta-1} b^{-1} c^{-\theta-1}
\end{aligned}$$

as desired.

For the general case, to calculate the intensity for $k+1$ given points to be consecutive points of $\mathscr{X}$, take $b_0 < b_1 < \cdots < b_k$ and set

$$y_i = b_{i+1} - b_i, 0 \leqslant i \leqslant k-1; \qquad a = b_0; \qquad c = b_k.$$

The intensity for $\mathscr{R}$ to have points at (w_j, y_j) for $j = 0$ to $k-1$, with $0 < w_{k-1} < \cdots < w_1 < w_0$, to have no points in $\cup_1^{k-1}(w_{j-1}, w_j) \times (0, \infty)$, and to satisfy $S_{w_0} = a$, is

$$\prod_{j=0}^{k-1} \theta e^{-w_j y_j} (w_{k-1}/w_0)^{\theta} w_0 (a w_0)^{\theta-1} e^{-x_0 w_0}/\Gamma(\theta).$$

Integrating over $w_{k-1} < w_{k-2} < \cdots < w_0$, the innermost integral is still

$$\int_{w_1}^{\infty} dw_0 \exp(-w_0(x_0 + y_0)) = \int_{w_1}^{\infty} dw_0 \exp(-w_0 b_1),$$

which produces the 'b_1^{-1}' factor along with the function $\exp(-w_1 b_1)$. This combines with the already present factor $\exp(-w_1 y_1)$, so the next integration is $\int_{w_2}^{\infty} dw_1 \exp(-w_1 b_2)$, which

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

produces the 'b_2^{-1}' factor. Continuing to integrate, the final result is

$$\theta^k \Gamma(\theta+1)\Gamma(\theta)^{-1} a^{\theta-1} \left\{ \prod_{j=1}^{k-1} b_j^{-1} \right\} c^{-\theta-1} = \theta^{k+1} a^{\theta-1} \left\{ \prod_{j=1}^{k-1} b_j^{-1} \right\} c^{-\theta-1},$$

as required. □

8. The third coupling: to exploit $\mathbb{E}|U - 1/2| = 1/4$

8.1. Motivating and defining the coupling

As in the proof of Lemma 7.1, start with the Poisson process $\mathscr{R}$ on $(0,\infty)^2$ with intensity $\theta \exp(-wy)\, dy\, dw$, now for $\theta = 1$. The joint intensity $e^{-wy}\, dw\, dy$ may be factored as the marginal intensity $(1/y)\, dy$ for Y, times the conditional intensity $ye^{-wy}\, dw$ for W given $Y = y$. Hence the Y values form the Poisson process on $(0,\infty)$ with intensity dy/y, and given a Y arrival at y, its 'label' W is exponentially distributed with mean $1/y$. So we construct our realization of $\mathscr{R}$ by first sampling the Y-process. If its points are labelled in any way as $(Y_j^*,\ j \geqslant 1)$, attach to each Y_j^* the associated $W_j^* = S_j^*/Y_j^*$, where $(S_j^*,\ j \geqslant 1)$ is an independent sequence of independent standard exponential random variables.

We now re-label, dropping the $*$s. Instead of labelling as at (7.3), we now label so that

$$\cdots > W_2 > W_1 > W_0 > W_{-1} > W_{-2} > \cdots, \tag{8.1}$$

and we define $X_i = \sum_{j \geqslant i} Y_j$, so that, by Lemma 7.1, both of the sets $\{X_i, i \in \mathbb{Z}\}$ and $\{Y_i, i \in \mathbb{Z}\}$ are distributed as the scale-invariant Poisson process on $(0,\infty)$ with intensity dx/x.

The event that there is a point of X at any integer n, or that $Y_i = Y_j$ for some $i \neq j$, has probability 0, and we remove all such outcomes from our probability space! Translate the indexing so that X_1 is the largest X_i before n. Thus

$$0 < \cdots < X_2 < X_1 < n < X_0 < X_{-1} < X_{-2} < \cdots < \infty, \tag{8.2}$$

the spacings are indexed with

$$Y_i := X_i - X_{i+1} \in (0,\infty), \quad \text{for } i \in \mathbb{Z},$$

and

$$X_1 = \sum_{j \geqslant 1} Y_j < n \quad \text{and} \quad X_0 = \sum_{j \geqslant 0} Y_j > n. \tag{8.3}$$

The GEM, scaled up by a factor of n, is constructed from the subintervals of $(0,n)$ with boundaries $X_1, X_2, \ldots$; that is,

$$(nA_1, nA_2, nA_3 \ldots) = \left(\left(n - \sum_{i \geqslant 1} Y_i \right), Y_1, Y_2, \ldots \right).$$

The Poisson–Dirichlet is formed by applying the function RANK, so that L_i is the ith-largest of $A_1, A_2, \ldots$, for $i \geqslant 1$.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

We next define a deterministic step function $f : (0, \infty) \to \mathbb{Z}_+$ to be applied to the points $Y_i, i \in \mathbb{Z}$. The function f is defined, using the Euler constant γ, by

$$\begin{aligned} (0, \exp(-\gamma)) &\mapsto 0, \\ [\exp(-\gamma), \exp(-\gamma + 1)) &\mapsto 1, \\ [\exp(-\gamma + 1), \exp(-\gamma + 1 + 1/2)) &\mapsto 2, \end{aligned} \tag{8.4}$$

and so on. For $k = 1, 2, \ldots$, the map f takes an interval of $x^{-1}\,dx$ measure $1/k$ onto k, so that the number Z_k of times that k appears in the multiset $\{f(Y_i), i \in \mathbb{Z}\}$ is Poisson with $\mathbb{E}Z_k = 1/k$, with $Z_1, Z_2, \ldots$ independent. Also, the interval $(0, e^{-\gamma})$, which is mapped to zero, has infinite dx/x measure, so the multiset $\{f(Y_i), i \in \mathbb{Z}\}$ has an infinite number of copies of zero.

A small table of approximate values shows that starting the step boundaries with offset $\exp(-\gamma)$ gives us a function which is very close to 'round to the nearest integer':

$$\begin{aligned} f^{-1}(0) &\doteq (0, 0.561459483566885), \\ f^{-1}(1) &\doteq [0.561459483566885, 1.52620511159586), \\ f^{-1}(2) &\doteq [1.52620511159586, 2.51628683093936), \\ f^{-1}(3) &\doteq [2.51628683093936, 3.51176116633948), \\ f^{-1}(10) &\doteq [9.50437851808436, 10.5039627325698), \\ f^{-1}(100) &\doteq [99.5004187539487, 100.500414587371), \\ f^{-1}(1000) &\doteq [999.500041687504, 1000.50004164584), \\ f^{-1}(10000) &\doteq [9999.50000416701, 10000.5000041666). \end{aligned}$$

The sequence Y_i is exactly in a size-biased permutation: this is clear from our construction. The sequence $f(Y_i)$ is not exactly in a size-biased permutation, although it comes close. *If* the sequence $f(Y_i)$ *were* in a size-biased permutation, it *would have* all the zeros coming first, followed by positive integers tending from small to large, and the indicator function of the set

$$\left\{ 1 + \sum_{j>i} f(Y_j) : \ i \in \mathbb{Z} \right\} \tag{8.5}$$

would be distributed exactly as the variables $\xi_1, \xi_2, \ldots$ for the Feller coupling in Section 4.2, because

$$(Z_1, Z_2, \ldots) =_d \left(C_1^{(\infty)}, C_2^{(\infty)}, \ldots \right),$$

and the values of each are then arranged in (reversed) size-biased order. (A history of this result, tracing it back to Rényi and Ignatov [15, 16], is given in Section 2.1 of [2].)

In order to have a size-biased permutation of the multiset $\{f(Y_i), i \in \mathbb{Z}\}$ which is close to the identity permutation on $\mathbb{Z}$, we simply re-use the exponential random variables S_j^*,

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Table 1.

i	8	7	6	5	4	3	2	1	0	−1
Y_i	0.125	0.25	0.5	2.8	6.6	0.1	1.6	7.8	20.2	2.3
$f(Y_i)$	0	0	0	3	7	0	2	8	20	2
$X_i = \sum_{j:j\geqslant i} Y_j$	0.25	0.5	1.0	3.8	10.4	10.5	12.1	19.9	40.1	42.4
$\sum_{j:j\geqslant i} f(Y_j)$	0	0	0	3	10	10	12	20	40	42

defining new W-coordinates $\widetilde{W}_j^* = S_j^*/f(Y_j^*)$, and hence, in the labelling (8.2),

$$\widetilde{W}_i = \frac{S_i}{f(Y_i)} = W_i \frac{Y_i}{f(Y_i)}. \tag{8.6}$$

Since the S_j^* are i.i.d. standard exponential, the points $f(Y_i)$, taken in order of decreasing tilde labels $\widetilde{W}_i$, are in a size-biased permutation, tending from small to large.

Use the $f(Y_i)$, taken in order of their tilde labels, to form the sequence $\xi_1, \xi_2, \ldots$ for the Feller coupling, and construct the age-ordered cycle lengths $A_1^{(n)}, A_2^{(n)}, \ldots$. Finally, we take $L_i^{(n)} :=$ ith-largest of the $A_1^{(n)}, A_2^{(n)}, \ldots$.

To summarize: we have defined a coupling, with the $L_i, L_i^{(n)}, A_i, A_i^{(n)}$ for $i \geqslant 1$ all realized together. It only remains to estimate the ℓ_1 distances $d_{\mathrm{PD}}(n)$ and $d_{\mathrm{GEM}}(n)$!

A technicality. To be careful, whenever $Y_i < e^{-\gamma}$, so that $f(Y_i) = 0$, we take $\widetilde{W}_i = \infty$. With probability 1, the non-infinite values among the $\widetilde{W}_i$ are all distinct, and on this good event, there is a permutation σ of the integers, with the property that for distinct i, j, $\sigma(i) > \sigma(j)$ if and only $\widetilde{W}_i > \widetilde{W}_j$ or ($\widetilde{W}_i = \widetilde{W}_j = \infty$ and $Y_i < Y_j < e^{-\gamma}$). Modulo translation, this permutation is unique. The Feller coupling variables $\xi_1, \xi_2, \ldots$ are defined by

$$\xi_i = 1 \text{ if and only if } i \in \left\{ 1 + \sum_{\sigma(j)>\sigma(k)} f(Y_j) : \ k \in \mathbb{Z} \right\}. \tag{8.7}$$

8.2. An example

Suppose that the sequence (Y_i), taking the index i *decreasing* through $\mathbb{Z}$, has negative powers of two, until the partial sum is exactly 1, followed by the values 2.8, 6.6, 0.1, 1.6, 7.8, 20.2, 2.3 and 5.4. We take $20 \leqslant n \leqslant 40$.

Note that in (8.3), the placement of the origin, $i = 0$, depends on n via $X_1 < n < X_0$. Thus, in case $20 \leqslant n \leqslant 40$ the last three columns, starting with 7.8, 20.2 and 2.3, are labelled with $i = 1, 0, -1$; if $n = 41$ they are labelled with $i = 2, 1, 0$; and if $n = 19$ they are labelled with $i = 0, -1, -2$.

The example, with $n = 35$. The scaled-up GEM has first component $nA_1 = n - X_1 = 35 - 19.9 = 15.1$, $nA_2 = Y_1 = 7.8$, $nA_3 = Y_2 = 1.6$, $nA_4 = Y_3 = 0.1$, $nA_5 = Y_4 = 6.6$, $nA_6 = 2.8$, $nA_7 = 0.5, \ldots$

The example, with minimal differences in the size-biased permutations. This is the most common situation, in which the labels $\widetilde{W}$ defined in (8.6), excluding cases where $f(Y_i) = 0$,

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

have the same relative order as in (8.1). In this example, the Feller coupling, as at (8.5), is based on 1 plus the partial sums of the sequence

$$\ldots, 0, 0, 0, 3, 7, 0, 2, 8, 20, 2, \ldots,$$

so that $\xi_1\xi_2\ldots$ has ones in positions indexed by

$$1, 4, 11, 13, 21, 41, 43, \ldots. \tag{8.8}$$

We have $A_1^{(n)} = 15$ (the spacing from 35+1 down to 21), $A_2^{(n)} = 8 = f(Y_1)$, $A_3^{(n)} = 2 = f(Y_2)$, $A_4^{(n)} = 7 = f(Y_4)$ (notice that the value $f(Y_3) = 0$ does not serve as the amount of spacing between two successive ones in $\xi_1\xi_2\cdots\xi_n 1$) and then $A_5^{(n)} = 3 = f(Y_5)$ and $A_j^{(n)} = 0$ for any $j > 5$. Using the rearrangement Lemma 3.2 to give an upper bound on $d_{\mathrm{PD}}(n)$, we will match $nA_j = Y_{j-1}$ with $A_j^{(n)} = f(Y_{j-1})$ for $j = 2, 3$ and we will match $nA_j = Y_{j-1}$ with $A_{j-1}^{(n)} = f(Y_{j-1})$ for $j \geqslant 5$, so the 'gap' at $j = 4$ contributes nothing to $d_{\mathrm{PD}}(n)$. This example shows why our third coupling may be good for $d_{\mathrm{PD}}(n)$, but not good enough for $d_{\mathrm{GEM}}(n)$ – because when $X_i = x$ and $X_i - X_{i+1} < e^{-\gamma}$ we will then usually have to match $A_j^{(n)} = f(Y_j)$ with $nA_j = Y_{j-1}$ for all $j > i$, for a contribution to $d_{\mathrm{GEM}}(n)$ that is order of x, and $\int_1^n x\, e^{-\gamma}/x\, dx/x$ is order of $\log n$.

The example, with dislocation in the size-biased permutations. Note that since $Y_2 = 1.6$ is left of $Y_1 = 7.8$, we have of necessity that $W_2 > W_1$. But in (8.6), $\widetilde{W}_2$ is obtained from W_2 by multiplying by the factor $Y_2/f(Y_2) = 1.6/2$, while $\widetilde{W}_1$ is obtained from W_1 by multiplying by the somewhat larger factor $Y_1/f(Y_1) = 7.8/8$, so if W_1 was only slightly larger than W_1, we will have $\widetilde{W}_2 < \widetilde{W}_1$, and $8 = f(Y_1)$ will come left of $2 = f(Y_2)$. Suppose that this indeed happens. The Feller coupling, as at (8.5), is based on 1 plus the partial sums of the sequence $\ldots, 0, 0, 0, 3, 7, 0, 8, 2, 20, 2, \ldots$, so that $\xi_1\xi_2\ldots$ has ones in positions indexed by $1, 4, 11, 19, 21, 41, 43, \ldots$. We have $A_1^{(n)} = 15$, just as in (8.8), but $A_2^{(n)} = 2 = f(Y_2)$ and $A_3^{(n)} = 8 = f(Y_1)$, in contrast with (8.8), and this contributes a large amount to $d_{\mathrm{GEM}}(n)$. Using the rearrangement Lemma 3.2 to give an upper bound on $d_{\mathrm{PD}}(n)$, we will match $nA_3 = Y_2$ with $A_2^{(n)} = f(Y_2)$ and $nA_2 = Y_1$ with $A_3^{(n)} = f(Y_1)$, so the extra difference in the size-biased permutation does not cause an increase in the upper bound on $d_{\mathrm{PD}}(n)$.

8.3. The principal differences

Our scaled-up GEM is viewed as a list of interval lengths subdividing $(0, n)$, and the Feller process as a list of interval lengths subdividing $[1, n+1]$; in both cases, the sum of the lengths is exactly n. The scaled-up GEM has as its components $nA_1 = n - \sum_{i\geqslant 1} Y_i$ and $Y_1, Y_2, \ldots$. A natural fit to the age-ordered cycle lengths would be that using $(n+1) - (1 + \sum_{i\geqslant 1} f(Y_i))$ and $f(Y_1), f(Y_2), \ldots$, giving a distance of at most $T_1 + T_2$, where

$$T_1 := \sum_{i\geqslant 1} |Y_i - f(Y_i)| \quad \text{and} \quad T_2 = \left| \sum_{i\geqslant 1} (Y_i - f(Y_i)) \right|.$$

Although this match may not quite work, because the $\widetilde{W}$-ordering is not exactly the same as the original, the essence of the proof is nonetheless to show that the error term T_1 is

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

indeed the dominant contribution to $d_{\mathrm{PD}}(n)$. Our first two lemmas concern the asymptotics of T_1 and T_2.

Lemma 8.1.

$$T_1 := \sum_{i \geqslant 1} |Y_i - f(Y_i)| \quad \textit{has} \quad \mathbb{E}T_1 \sim \frac{1}{4} \log n. \tag{8.9}$$

Proof. In the proof of Theorem 2.1, we calculated $\mathbb{E}\sum_{i\geqslant 1} |nL_i - g(nL_i)| \sim (1/4)\log n$, where g is the function 'round to the nearest integer', that is, $g(x) = \lfloor x + .5 \rfloor$. This sum differs from that in (8.9) in two ways. First, since our coupling takes $nL_i = i$th-largest of $(n - \sum_{i\geqslant 1} Y_i), Y_1, Y_2, \ldots$, the first of those differences is no longer present. Secondly, the function g is replaced by f, and the table above makes it clear that $f(x) - g(x) \in \{-1, 0, 1\}$ for all $x > 0$. Write h_i for the harmonic sum $h_i = 1 + 1/2 + \cdots + 1/i$. From the expansion $h_i = \gamma + \log i + 1/(2i) - 1/(12i^2) + O(i^{-4})$ we have

$$\exp(-\gamma + h_i) = i \exp\left(\frac{1}{2i} - \frac{1}{12i^2} + \cdots \right) = i + .5 + \frac{1}{24i} + O(i^{-2}) \tag{8.10}$$

so the set difference $f^{-1}(i) \setminus g^{-1}(i)$ is an interval of length $\sim 1/(24i)$. Hence

$$\mathbb{E} \sum_{i : Y_i < n} |f(Y_i) - g(Y_i)| = \int_0^n |f(x) - g(x)|\, dx/x = O(1); \tag{8.11}$$

note that if we were integrating dx rather than dx/x, this error would be order of $\log n$ and unacceptable. □

Lemma 8.2.

$$T_2 := \left| \sum_{i \geqslant 1} (Y_i - f(Y_i)) \right| \quad \textit{has } \mathbb{E}T_2 = O(\sqrt{\log n}). \tag{8.12}$$

Proof. We will use Lemma 8.1 to approximate f by g, the function which rounds to the nearest integer. Start by considering

$$H_n := \sum_{i : 1/2 \leqslant Y_i < n} (Y_i - g(Y_i)).$$

This is compound Poisson, with values in $[-.5, .5]$, and total Poisson intensity $\int_{1/2}^n dx/x = \log(2n)$. The calculation

$$\begin{aligned} \int_{i-1/2}^{i+1/2} (y - g(y))dy/y &= \int_{i-1/2}^{i+1/2} (y - i)dy/y \\ &= 1 - i\log((i + 1/2)/(i - 1/2)) = O(i^{-2}) \end{aligned}$$

shows that $\mathbb{E}H_n = O(1)$ as $n \to \infty$, and the bound $|y - g(y)| \leqslant 1/2$ shows that $\mathrm{Var}(H_n) = \int_{1/2}^n (y - g(y))^2 dy/y \leqslant (1/4)\log(2n)$; combined, these give $\mathbb{E}H_n^2 \leqslant c \log n$, for some constant

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

$c < \infty$ and for all n. This yields $\mathbb{E}|H_n| \leqslant \sqrt{\mathbb{E}H_n^2} = O(\sqrt{\log n})$. If we change H_n to

$$H_n' := \sum_{i:0<Y_i<n} (Y_i - g(Y_i)),$$

then $H_n - H_n' = \sum_{i:0<Y_i<1/2} Y_i$. The sum is positive a.s., with expectation $1/2$, so $\mathbb{E}|H_n - H_n'| = 1/2$, and hence $\mathbb{E}|H_n'| = O(\sqrt{\log n})$.

Next, from (8.11) in the proof of Lemma 8.1, there is an $O(1)$ error in replacing g by the function f, so now we have

$$\mathbb{E}\left|\sum_{i:Y_i<n} (Y_i - f(Y_i))\right| = O(\sqrt{\log n}).$$

Finally, the function $f(x) - x$ is bounded, and

$$\begin{aligned}\mathbb{E}\left\{\sum_{i\leqslant 0} \mathbb{1}\{Y_i < n\}\right\} &\leqslant 1 + \sum_{i<0} \mathbb{1}\{X_{i-1} - X_i < n\} \\ &\leqslant 1 + \int_n^\infty \frac{n\,dx}{x(n+x)} \leqslant 2,\end{aligned}$$

so that thus

$$\mathbb{E}\left|\sum_{i\geqslant 1}(Y_i - f(Y_i))\right| = O(1) + \mathbb{E}\left|\sum_{i:Y_i<n} (Y_i - f(Y_i))\right| = O(\sqrt{\log n}). \qquad \square$$

8.4. Analysis of the coupling

The remainder of the argument concerns the effect of switching the order from the W-ordering to the $\widetilde{W}$-ordering. To do so, a number of preliminaries are needed. To start with, define

$$f_\infty(w) := \mathbb{P}\left[\sum_{i:W_i>w} Y_i > 1\right].$$

Then we have the following lemma.

Lemma 8.3. *For all $w > 0$,*

$$f_\infty(w) \leqslant \min\{1, 2(1+w)e^{-w}\}.$$

Proof. For any bounded function $g : \mathbb{R}_+ \to \mathbb{R}_+$, let

$$(Tg)(w) := w\int_w^\infty x^{-2}e^{-x}\,dx + w\int_w^\infty x^{-1}\int_0^\infty e^{-xz} g(x(1-z))\,dz\,dx,$$

and, for all $w \geqslant 0$, set $f_0(w) = 0$ and

$$f_n(w) := \mathbb{P}\left[\sum_{i=J(w)}^{J(w)+n} Y_i > 1\right], \quad n \geqslant 1,$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

where $J(w) := \min\{i : W_i > w\}$. Then it is clear that $f_\infty(w) \uparrow f_n(w)$ as $n \to \infty$, and, because

$$\mathbb{P}\left[\sum_{i=J(x)}^{J(x)+n} Y_i > 1 - z\right] = f_n(x(1-z)), \quad x > 0, \quad 0 \leqslant z \leqslant 1,$$

by scaling, it also follows that $f_{n+1} = Tf_n$ for all $n \geqslant 0$. Clearly, $f_0(w) \leqslant K(1+w)e^{-w}$ for all $K, w > 0$, and $f_n(w) \leqslant 1$ for all n and w. Then, if $f_n(w) \leqslant K(1+w)e^{-w}$ for all $w > 0$, it follows from the definition of T that

$$\begin{aligned} f_{n+1}(w) &\leqslant w\int_w^\infty x^{-2}e^{-x}\,dx + w\int_w^\infty x^{-1}\int_0^\infty e^{-xz}K(1+x(1-z))e^{-x(1-z)}\,dz\,dx \\ &= w\int_w^\infty e^{-x}\left\{\frac{1}{x^2} + \frac{K}{x} + \frac{K}{2}\right\} \\ &\leqslant e^{-w} + Ke^{-w} + \tfrac{1}{2}Kwe^{-w}. \end{aligned}$$

Thus it follows that $f_{n+1}(w) \leqslant K(1+w)e^{-w}$ for all w such that $Kw/2 \geqslant 1$. However, $f_{n+1}(w) \leqslant 1 \leqslant K(1+w)e^{-w}$ for all $w < 2/K$ provided that $(K+2)e^{-2/K} \geqslant 1$, true for $K = 2$. Hence, for $K = 2$, it follows by induction that $f_n(w) \leqslant \min\{1, 2(1+w)e^{-w}\}$ for all n, proving the lemma. □

Now, if $W_0^1 := \sup\{w : \sum_{i:W_i>w} Y_i > 1\}$, it follows that

$$\{W_0^1 > w\} = \left\{\sum_{i:W_i>w} Y_i > 1\right\},$$

and hence that

$$\mathbb{E}W_0^1 = \int_0^\infty \mathbb{P}[W_0^1 > w]\,dw \leqslant 2\int_0^\infty (1+w)e^{-w}\,dw = 4.$$

Hence we have proved the following corollary.

Corollary 8.4.

$$\mathbb{E}W_0 \leqslant 4n^{-1}.$$

We now turn to consideration of the differences between the values $\{f(Y_i),\, i \geqslant 1\} = \{f(Y_i),\, W_i > W_0\}$ and $\{f(Y_i),\, \widetilde{W}_i > W_0\}$. The latter set is a 'left-hand segment' of the infinite reverse size-biased multiset $\{f(Y_i),\, i \in \mathbb{Z}\}$, and is close to the coupled reverse age-ordered cycles. The two sets differ because, for some $i < 0$, $\widetilde{W}_i > W_0$, and, for some $i > 0$, $\widetilde{W}_i < W_0$; the effects of these two exchanges are treated in Lemmas 8.5 and 8.6, respectively. We begin by defining

$$c_1 := \min_{y:f(y)\geqslant 1}\{y^{-1}f(y)\}, \qquad c_2 := \max_{y>0}\{y^{-1}f(y)\}.$$

Then, for $y = (y_1, y_2, \dots)$ an increasing sequence of positive reals and $u, v > 0$, we let $A(u, v; y)$ denote the event

$$\{(W_0, Y_0) = (u, v)\} \cap \{\pi_2\mathscr{R}\{(u,\infty) \times \mathbb{R}_+\} = y\}.$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Lemma 8.5. *Defining*

$$\eta_1 := \sum_{i\geqslant 1}\{f(Y_{-i})\mathbb{1}\{\widetilde{W}_{-i} > W_0\}\},$$

we have $\mathbb{E}\eta_1 \leqslant c_2/c_1$.

Proof. We use the fact that, conditional on $A(w_0, y_0; y)$, we have

$$\mathscr{L}(W_{-i},\, i \geqslant 1) = \mathscr{L}\left(w_0 \prod_{j=1}^{i} U_j,\, i \geqslant 1\right),$$

where $U_1, U_2, \ldots$ are independent and uniformly distributed on $(0,1)$. Hence

$$\begin{aligned}
&\mathbb{E}\{f(Y_{-i})\mathbb{1}\{\widetilde{W}_{-i} > W_0\}\} \\
&\quad = \mathbb{E}\left(f(Y_{-i})\mathbb{1}\left\{W_0 \prod_{j=1}^{i} U_j(Y_{-i}/f(Y_{-i})) > W_0\right\}\right) \\
&\quad \leqslant \mathbb{E}\left\{f(Y_{-i})\mathbb{1}\{U_i(Y_{-i}/f(Y_{-i})) > 1\} \prod_{j=1}^{i-1} \mathbb{1}\{U_j > c_1\}\right\} \\
&\quad \leqslant \mathbb{E}\{f(Y_{-i})(Y_{-i} - f(Y_{-i}))_+ / Y_{-i}\}(1-c_1)^{i-1} \leqslant c_2(1-c_1)^{i-1},
\end{aligned}$$

and the lemma follows. □

Lemma 8.6. *Defining*

$$\eta_2 := \sum_{i\geqslant 1}\{f(Y_i\mathbb{1}\{\widetilde{W}_i < W_0\}\},$$

we have $\mathbb{E}\eta_2 \leqslant 4c_2$.

Proof. Conditional on $A(w_0, y_0; y)$, the point set $\{W_i,\, i \geqslant 1\}$ has the same distribution as the set $\{\widehat{W}_j,\, j \geqslant 1\}$, where $\widehat{W}_j := w_0 + y_j^{-1}E_j$ and $E_1, E_2, \ldots$ are independent standard exponential random variables. Hence

$$\begin{aligned}
&\sum_{i\geqslant 1} \mathbb{E}\{f(Y_i)\mathbb{1}\{\widetilde{W}_i < W_0\} \mid A(w_0, y_0; y)\} \\
&\qquad = \sum_{j\geqslant 1} f(y_j)\mathbb{P}\big[(y_j/f(y_j))\big(w_0 + y_j^{-1}E_j\big) < w_0\big].
\end{aligned}$$

But it is easy to see that

$$\begin{aligned}
&f(y_j)\mathbb{P}\big[(y_j/f(y_j))\big(w_0 + y_j^{-1}E_j\big) < w_0\big] \\
&\qquad = f(y_j)\mathbb{P}[E_j < (f(y_j) - y_j)_+ w_0] \leqslant f(y_j)w_0,
\end{aligned}$$

so that

$$\sum_{j\geqslant 1} f(y_j)\mathbb{P}\big[(y_j/f(y_j))\big(w_0 + y_j^{-1}E_j\big) < w_0\big] \leqslant w_0 \sum_{j\geqslant 1} f(y_j) \leqslant w_0 c_2 \sum_{j\geqslant 1} y_j \leqslant w_0 c_2 n.$$

The lemma now follows from Corollary 8.4. □

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

Corollary 8.7. *Enumerate the set* $\{f(Y_i) : i \in \mathbb{Z} \setminus \{0\},\ \widetilde{W}_i > W_0\}$ *as* $\{F_j,\ j \geqslant 1\}$. *Then there is a matching* σ *of the set* $\{F_j,\ j \geqslant 1\}$ *and the set* $\{Y_i,\ i \geqslant 1\}$ *such that*

$$\mathbb{E}\left|\sum_{i\geqslant 1}(Y_i - F_{\sigma(i)})\right| = \mathbb{E}\left|\left(n - \sum_{i\geqslant 1} Y_i\right) - \left(n - \sum_{j\geqslant 1} F_j\right)\right| = O(\sqrt{\log n})$$

and

$$\limsup_{n\to\infty} \{\log n\}^{-1}\, \mathbb{E}\left\{\sum_{i\geqslant 1} |Y_i - F_{\sigma(i)}|\right\} \leqslant \frac{1}{4}.$$

Proof. Match Y_i with $f(Y_i)$ for all $i \geqslant 1$ such that also $\widetilde{W}_i > W_0$, and use Lemmas 8.5 and 8.6 to control the remainder, which is therefore of order $O(1)$. This implies that the first expectation is bounded by $T_2 + O(1)$ and the second by $T_1 + O(1)$, and Lemmas 8.1 and 8.2 complete the bound. □

Since the $f(Y_i)$ in the $\widetilde{W}_i$-ordering are in size-biased order, the set $\{F_j,\ j \geqslant 1\}$ consists of a 'left-hand segment' from the reverse size-biased order, and is close to being that for which the sum is closest to n from below. If $f(Y_0)$ were added to the collection, and the difference between Y_i and $f(Y_i)$ were temporarily neglected, the sum would be the first to exceed n. Thus, from Corollary 8.7, we would be more or less finished if we always had $f(Y_0)$ as next element after the $\{F_j,\ j \geqslant 1\}$ in the $\widetilde{W}_i$-ordering. This need not quite be the case, and the next two lemmas control the possible error made when completing the approximation.

We first consider the case in which $f(Y_0) > Y_0$, and so $\widetilde{W}_0 < W_0$. Here, the main concern is that there may be indices $i < 0$ such that $\widetilde{W}_0 < \widetilde{W}_i < W_0$, so that the corresponding $f(Y_i)$ would be taken before $f(Y_0)$. The possible contribution from indices $i > 0$ satisfying $\widetilde{W}_0 < \widetilde{W}_i < W_0$ is already more than covered by Lemma 8.6.

Lemma 8.8. *If* $f(y_0) > y_0$, *then*

$$\sum_{i\geqslant 1} \mathbb{E}\{(f(Y_{-i}) \wedge f(Y_0))\mathbb{1}\{\widetilde{W}_0 < \widetilde{W}_{-i} \leqslant W_0\} \mid A(w_0, y_0; y)\} \leqslant c_2^2/c_1.$$

Proof. We argue much as for Lemma 8.5, obtaining

$$\begin{aligned}
&\mathbb{E}\{(f(Y_{-i}) \wedge f(y_0))\mathbb{1}\{\widetilde{W}_0 < \widetilde{W}_{-i} \leqslant W_0\} \mid A(w_0, y_0; y)\} \\
&\quad\leqslant \mathbb{E}\left\{(f(Y_{-i}) \wedge f(y_0))\mathbb{1}\{(y_0/f(y_0)) < U_i(Y_{-i}/f(Y_{-i}) \leqslant 1\}\right. \\
&\qquad\qquad \left.\prod_{j=1}^{i-1} \mathbb{1}\{U_j > c_1/c_2\} \,\middle|\, A(w_0, y_0; y)\right\} \\
&\quad\leqslant f(y_0)\{c_2(f(y_0) - y_0)_+ / f(y_0)\}(1 - c_1/c_2)^{i-1} \\
&\quad\leqslant c_2(1 - c_1/c_2)^{i-1}.
\end{aligned}$$

The lemma now follows immediately. □

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

For the complementary case, in which $f(Y_0) < Y_0$ and so $\widetilde{W}_0 > W_0$, we are principally concerned about indices $i > 0$ such that $W_0 < \widetilde{W}_i < \widetilde{W}_0$, so that $f(Y_0)$ would be taken before the corresponding $f(Y_i)$. The possible contribution from indices $i < 0$ satisfying $W_0 < \widetilde{W}_i < \widetilde{W}_0$ is taken care of by Lemma 8.5.

Lemma 8.9. *If $f(y_0) < y_0$, then*

$$\sum_{i \geqslant 1} \mathbb{E}\{f(Y_0)\mathbb{1}\{W_0 < \widetilde{W}_i < \widetilde{W}_0\} \mid A(w_0, y_0; y)\} \leqslant 4c_2.$$

Proof. The argument here is like that of Lemma 8.6. We start by computing

$$\begin{aligned}
&\mathbb{E}\{f(Y_0)\mathbb{1}\{W_0 < \widehat{W}_j(y_j/f(y_j)) < \widetilde{W}_0\} \mid A(w_0, y_0; y)\} \\
&\quad = \mathbb{E}\{f(y_0)\mathbb{1}\{w_0 < \big(w_0 + y_j^{-1}E_j\big)(y_j/f(y_j)) < w_0(y_0/f(y_0))\}\} \\
&\quad \leqslant f(y_0)\mathbb{P}\left[w_0\left\{1 - \frac{y_j}{f(y_j)}\right\} < \frac{E_j}{f(y_j)} < w_0\left\{\frac{y_0}{f(y_0)} - \frac{y_j}{f(y_j)}\right\}\right] \\
&\quad \leqslant f(y_0)f(y_j)w_0\{(y_0/f(y_0)) - 1\} \leqslant w_0 f(y_j).
\end{aligned}$$

Once again, adding over j, this yields

$$\sum_{i \geqslant 1} \mathbb{E}\{f(Y_0)\mathbb{1}\{W_0 < \widetilde{W}_i < \widetilde{W}_0\} \mid A(w_0, y_0; y)\} \leqslant w_0 \sum_{j \geqslant 1} f(y_j),$$

and the lemma follows from Corollary 8.4. □

Using these two lemmas, the main theorem can be proved. Let $\tau(i)$ denote the index of the ith nonzero element of the set $\{f(Y_l), l \in \mathbb{Z}\}$ in decreasing $\widetilde{W}$-ordering, and let I_n be such that

$$\sum_{i=1}^{I_n} f(Y_{\tau(i)}) \leqslant n < \sum_{i=1}^{I_n+1} f(Y_{\tau(i)}).$$

Write $Y_0^{(1)} = n - \sum_{i \geqslant 1} Y_i$ and $Y_i^{(1)} = Y_i$, $i \geqslant 1$, and then $Y_0^{(2)} = n - \sum_{i=1}^{I_n} f(Y_{\tau(i)})$ and $Y_i^{(2)} = f(Y_{\tau(i)})$, $1 \leqslant i \leqslant I_n$, with $Y_i^{(2)} = 0$ for $i > I_n$. These are the scaled GEM and age-ordered cycle lengths to be matched, as realized in our coupling.

Theorem 8.10. *There is a matching ρ of $\{Y_i^{(1)}, i \geqslant 0\}$ and $\{Y_i^{(2)}, i \geqslant 0\}$ such that*

$$\mathbb{E}\left\{\sum_{i \geqslant 0} |Y_i^{(1)} - Y_{\rho(i)}^{(2)}|\right\} \leqslant \frac{1}{4}\log n + o(\log n).$$

Hence, by comparison with the lower bound from Theorem 2.1, the coupling defined in Section 8.1 achieves

$$d_{\mathrm{PD}}(n) \sim \frac{1}{4}\log n.$$

Proof. The proof consists mainly of finding upper bounds for the possible error in particular matchings, in a number of particular cases.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

To start with, consider the case where $f(Y_0) > Y_0$, so that $\widetilde{W}_0 < W_0$, and the set of lengths $\{F_j,\, j \geqslant 1\}$, the values of $f(Y_l)$ for which $\widetilde{W}_l > W_0$, typically needs augmenting in order to have total length n. Let F_{-l}, $1 \leqslant l \leqslant L$, be the values $f(Y_{-i})$ for those $i \geqslant 1$ for which $\widetilde{W}_0 < \widetilde{W}_{-i} < W_0$, taken in $\widetilde{W}$-order, writing L for their total number. Define Q_l, $l \geqslant 0$, to be the undershoot $n - \sum_{j\geqslant 1} F_j - \sum_{s=1}^{l} F_{-s}$ at stage l, when all of the $\{F_j,\, j \geqslant 1\}$ and the first l of the F_{-s} have been taken. Clearly, once the undershoot $Q_l \leqslant 0$, no more elements are taken.

The first sub-case is when the undershoot $Q_0 \leqslant 0$. Here, the F_j already match the Y_i as given by Corollary 8.7, and there is additional mismatch only because of the unmatched element $Y_0^{(1)} = n - \sum_{i\geqslant 1} Y_i$ and the piece of length $\sum_{j\geqslant 1} F_j - n = -Q_0 \geqslant 0$ which has to be removed from the F_js; hence the error can be kept to at most

$$\left(\sum_{j\geqslant 1} F_j - n\right) + \left(n - \sum_{i\geqslant 1} Y_i\right) + \sum_{i\geqslant 1} |Y_i - F_{\sigma(i)}| \leqslant \left|\sum_{j\geqslant 1} F_j - \sum_{i\geqslant 1} Y_i\right| + \sum_{i\geqslant 1} |Y_i - F_{\sigma(i)}|. \tag{8.13}$$

We next consider the sub-case in which $0 < Q_0 \leqslant f(Y_0)$. Here, we begin by taking the successive elements F_{-l}, $1 \leqslant l \leqslant L$, and, when they are exhausted, any remaining interval is then more than covered by the element $f(Y_0)$, Q_L being matched with $Y_0^{(1)}$. Usually, F_{-l} is matched with 0, at a cost of F_{-l}. However, if one of the F_{-l} is big enough to itself cover the whole remaining interval, *i.e.*, $F_{-l} \geqslant Q_{l-1}$, it is used to do so, with Q_{l-1} matched with $Y_0^{(1)}$, and no more are then needed; this happens in particular if $F_{-l} \geqslant f(Y_0)$. (The possible effect of elements arising in the $\widetilde{W}$-interval $(\widetilde{W}_0, W_0)$ from $f(Y_i)$ with $i \geqslant 1$ is controlled by Lemma 8.6, and can introduce an extra error of no more than $2\eta_2$.) We can then bound the error additional to $\sum_{i\geqslant 1} |Y_i - F_{\sigma(i)}|$ by the expression

$$\sum_{l=1}^{L} \mathbb{1}\{Q_{l-1} > 0\}\big[\mathbb{1}\{F_{-l} < f(Y_0)\}(F_{-l}\mathbb{1}\{F_{-l} < Q_{l-1}\} + R_{1,l}\mathbb{1}\{F_{-l} \geqslant Q_{l-1}\}) + \mathbb{1}\{F_{-l} \geqslant f(Y_0)\}R_{1,l}\big] + R_{1,L+1}\mathbb{1}\{Q_L > 0\},$$

where, for $1 \leqslant l \leqslant L+1$, the error $R_{1,l}$ in matching Q_{l-1} and $Y_0^{(1)}$ is

$$R_{1,l} := \left|\left[n - \sum_{j\geqslant 1} F_j - \sum_{s=1}^{l-1} F_{-s}\right] - \left[n - \sum_{i\geqslant 1} Y_i\right]\right| \leqslant \left|\sum_{j\geqslant 1} F_j - \sum_{i\geqslant 1} Y_i\right| + \sum_{s=1}^{l-1} F_{-s}.$$

In this sum, since $Q_0 > 0$, there is exactly one of the $R_{1,l}$, and some or all of those of the F_{-l} that are smaller than $f(Y_0)$. Hence, in this sub-case, the total error is at most

$$\left|\sum_{j\geqslant 1} F_j - \sum_{i\geqslant 1} Y_i\right| + \sum_{i\geqslant 1} |Y_i - F_{\sigma(i)}| + 2\sum_{i\geqslant 1} f(Y_{-i})\mathbb{1}\{f(Y_{-i}) < f(Y_0)\}\mathbb{1}\{\widetilde{W}_0 < \widetilde{W}_{-i} \leqslant W_0\} + 2\eta_2. \tag{8.14}$$

Within the case where $f(Y_0) > Y_0$, there now remains only the possibility that $Q_0 > f(Y_0)$. Here, the previous procedure can be used to match, but using $n' = f(Y_0) + \sum_{j\geqslant 1} F_j$ in place of n throughout. This leaves an interval of length at most $n - n'$ unmatched.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

However, we have

$$n - n' \leqslant \sum_{i \geqslant 0} Y_i - \sum_{j \geqslant 1} F_j - f(Y_0) = Y_0 - f(Y_0) + \sum_{i \geqslant 1} Y_i - \sum_{j \geqslant 1} F_j$$
$$\leqslant 1 + \left| \sum_{j \geqslant 1} F_j - \sum_{i \geqslant 1} Y_i \right|, \tag{8.15}$$

to be added to the error in the previous sub-case.

The case in which $f(Y_0) < Y_0$, so that $W_0 < \widetilde{W}_0$, is argued in rather similar fashion. Here, we shall denote the set $\{f(Y_i),\, W_0 < \widetilde{W}_i < \widetilde{W}_0\} \subset \{F_j,\, j \geqslant 1\}$ by $\{F_j,\, j \in R\}$. The matching of Corollary 8.7 is not quite a matching for our coupling, if R is not empty. To modify the matching to become one, we consider sub-cases. First, if $f(Y_0) \geqslant n - \sum_{j \notin R} F_j$, then $n - \sum_{i \geqslant 1} Y_i$ is matched to $n - \sum_{j \notin R} F_j$ instead of to $n - \sum_{j \geqslant 1} F_j$, and the elements $\{F_j,\, j \in R\}$ are missing in the new matching, so that there is an extra error of at most $2 \sum_{j \in R} F_j$. Note also that, under these circumstances, $\sum_{j \in R} F_j \leqslant f(Y_0)$, so that the extra error is at most

$$2 \sum_{j \in R} (F_j \wedge f(Y_0)). \tag{8.16}$$

The next sub-case has $n - \sum_{j \geqslant 1} F_j \leqslant f(Y_0) < n - \sum_{j \notin R} F_j$, in which case we can match $n - \sum_{i \geqslant 1} Y_i$ with $f(Y_0)$; some of the $\{F_j,\, j \in R\}$ are again missing. The former match differs from the original by at most

$$f(Y_0) - \left(n - \sum_{j \geqslant 1} F_j \right) \leqslant \left(n - \sum_{j \notin R} F_j \right) - \left(n - \sum_{j \geqslant 1} F_j \right) = \sum_{j \in R} F_j,$$

again leading to an upper bound of $2 \sum_{j \in R} F_j$ for the extra error; and an alternative matching with error at most $2f(Y_0)$ could also be achieved by matching $f(Y_0)$ with 0, so that (8.16) is a bound for the extra error in this sub-case, too. In the final sub-case, in which $f(Y_0) < n - \sum_{j \geqslant 1} F_j$, we again match $n - \sum_{i \geqslant 1} Y_i$ to $f(Y_0)$ and the pieces making up the undershoot $n - \sum_{j \geqslant 1} F_j - f(Y_0)$ with 0, leading to an error of at most

$$\left| n - \sum_{i \geqslant 1} Y_i - f(Y_0) \right| + \left(n - \sum_{j \geqslant 1} F_j - f(Y_0) \right)$$
$$\leqslant \left(\sum_{i \geqslant 1} Y_i + Y_0 - n \right) + |f(Y_0) - Y_0| + n - \sum_{j \geqslant 1} F_j - f(Y_0)$$
$$\leqslant 2 + \sum_{i \geqslant 1} Y_i - \sum_{j \geqslant 1} F_j,$$

to replace the original error of $|\sum_{i \geqslant 1} Y_i - \sum_{j \geqslant 1} F_j|$ in matching $n - \sum_{i \geqslant 1} Y_i$ to $n - \sum_{j \geqslant 1} F_j$; thus the increase is here at most 2. Taking expectations, it follows that the overall bound in the case $f(Y_0) < Y_0$ is at most

$$\left| \sum_{j \geqslant 1} F_j - \sum_{i \geqslant 1} Y_i \right| + \sum_{i \geqslant 1} |Y_i - F_{\sigma(i)}| + 2 \sum_{i \geqslant 1} f(Y_0) \mathbb{1}\{W_0 < \widetilde{W}_i < \widetilde{W}_0\} + 2. \tag{8.17}$$

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

The conclusion of the theorem now follows from the bounds (8.13)–(8.17), Lemmas 8.6–8.9 and Corollary 8.7. □

9. θ-biased permutations

To derive a lower bound for $\mathbb{E}\sum_{i\geqslant 1}|L_i^{(n)} - nL_i|$, we use the intensity measure $\mu(dx) = \theta(1-x)^{\theta-1}x^{-1}\,dx$ for $0 < x < 1$, called the 'frequency spectrum' in Ewens [12], corresponding to the Poisson–Dirichlet distribution with parameter θ. We obtain the following result.

Theorem 9.1. *For any $\theta > 0$, let $L_i^{(n)}$ denote the size of the ith-largest component of the Ewens sampling formula, and let L_i be the ith coordinate of the Poisson–Dirichlet process with parameter θ. Uniformly over all couplings of these two processes,*

$$\liminf_{n\to\infty} (\log n)^{-1}\,\mathbb{E}\sum_{i\geqslant 1}|L_i^{(n)} - nL_i| \geqslant \frac{1}{4}\theta. \tag{9.1}$$

Proof. As in Theorem 2.1, any coupling has

$$\begin{aligned}
\mathbb{E}\sum_{i\geqslant 1}|L_i^{(n)} - nL_i| &\geqslant \int_{(0,1]} d(nx,\mathbb{Z})\mu(dx)\\
&= \int_0^1 d(nx,\mathbb{Z})(1-x)^{\theta-1}\frac{\theta}{x}\,dx = \int_0^n d(x,\mathbb{Z})\left(1-\frac{x}{n}\right)^{\theta-1}\frac{\theta}{x}\,dx\\
&= \theta\int_0^n d(x,\mathbb{Z})x^{-1}\,dx + \theta\int_0^n d(x,\mathbb{Z})\left(\left(1-\frac{x}{n}\right)^{\theta-1} - 1\right)\frac{1}{x}\,dx\\
&\sim \frac{\theta}{4}\log n,
\end{aligned}$$

this last following from the monotone convergence theorem and (2.3). □

Theorem 9.2. *For any $\theta > 0$, the coupling of Section 8.1, using exactly the function f given by (8.4), but with the scale-invariant Poisson processes taken to have intensity $\theta/x\,dx$, achieves*

$$d_{\mathrm{PD}}(n) \sim \frac{\theta}{4}\log n.$$

Proof. Every consideration in Section 8, with the factor θ inserted into the intensity for the scale-invariant Poisson process, goes through exactly as it did in the special case $\theta = 1$. □

Acknowledgement

The authors thank Ludger Rüschendorf for the reference to the alternative proof of Lemma 3.1.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press

References

[1] Arratia, R. (1996) Independence of prime factors: total variation and Wasserstein metrics, insertions and deletions, and the Poisson–Dirichlet process. Preprint.

[2] Arratia, R. (2002) On the amount of dependence in the prime factorization of a uniform random integer. In *Contemporary Combinatorics* (B. Bollobás, ed.), Vol. 10 of *Bolyai Society Mathematical Studies*, pp. 29–91.

[3] Arratia, R., Barbour, A. D. and Tavaré, S. (1992) Poisson process approximations for the Ewens Sampling Formula. *Ann. Appl. Probab.* **2** 519–535.

[4] Arratia, R., Barbour, A. D. and Tavaré, S. (1999) The Poisson–Dirichlet distribution and the scale invariant Poisson process. *Combin. Probab. Comput.* **8** 407–416.

[5] Arratia, R., Barbour, A. D. and Tavaré, S. (2003) *Logarithmic Combinatorial Structures: A Probabilistic Approach*, Vol. 1 of *EMS Monographs in Mathematics*, European Mathematical Society Publishing House, Zürich.

[6] Arratia, R. and Tavaré, S. (1992) The cycle structure of random permutations. *Ann. Probab.* **20** 1567–1591.

[7] Billingsley, P. (1972) On the distribution of large prime divisors. *Periodica Mathematica Hungarica* **2** 283–289.

[8] Dall'Aglio, G. (1956) Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuola Norm. Sup. Pisa* **10** 35–74.

[9] Diaconis, P. and Pitman, J. W. (1986) Permutations, record values and random measures. Unpublished lecture notes, Statistics Department, University of California, Berkeley.

[10] Dudley, R. M. (1989) *Real Analysis and Probability*, Wadsworth and Brooks/Cole, Pacific Grove, CA.

[11] Engen, S. (1975) A note on the geometric series as a species frequency model. *Biometrika* **62** 97–699.

[12] Ewens, W. J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3** 87–112.

[13] Feller, W. (1945) The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* **51** 800–832.

[14] Griffiths, R. C. (1979) On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology* **15** 140–158.

[15] Ignatov, T. (1981) Point processes generated by order statistics and their applications. In *Point Processes and Queuing Problems* (P. Bartfái and J. Tomkó, ed.), North-Holland, pp. 109–116.

[16] Ignatov, T. (1982) A constant arising in the asymptotic theory of symmetric groups, and Poisson–Dirichlet measures. *Theory Probab. Appl.* **27** 136–147.

[17] Kingman, J. F. C. (1975) Random discrete distributions. *J. Royal Statist. Soc. Ser. B* **37** 1–22.

[18] Kingman, J. F. C. (1977) The population structure associated with the Ewens sampling formula. *Theoretical Population Biology* **11** 274–283.

[19] McCloskey, J. W. (1965) A model for the distribution of individuals by species in an environment. PhD thesis, Michigan State University.

[20] Rachev, S. T. and Rüschendorf, L. (1998) *Mass Transportation Problems, Part I: Theory*, Springer, New York.

[21] Rényi, A. (1962) On the outliers of a series of observations. *A Magyar Tudományos Akadémia Matematikai és Fizikai Tudományok Osztályának Kőzleményei*, **12** 105–121. Reprinted in *Selected papers of Alfréd Rényi*, Vol. 3, Akadémiai Kiadó (1976), pp. 50–65.

[22] Vershik, A. M. and Shmidt, A. A. (1977) Limit measures arising in the asymptotic theory of symmetric groups I. *Theory Probab. Appl.* **22** 70–85.

https://doi.org/10.1017/S0963548305007054 Published online by Cambridge University Press