Combinatorics, Probability and Computing (1999) 8, 407–416. Printed in the United Kingdom

The Poisson–Dirichlet Distribution and the Scale-Invariant Poisson Process

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

[1] Department of Mathematics, University of Southern California,
Los Angeles, CA 90089-1113, USA
(e-mail: rarratia@math.usc.edu stavare@gnome.usc.edu)

[2] Abteilung für Angewandte Mathematik, Universität Zürich,
Winterthurerstrasse 190, CH-8057, Zürich, Switzerland
(e-mail: adb@amath.unizh.ch)

Received 27 June 1997; revised 16 March 1998

We show that the Poisson–Dirichlet distribution is the distribution of points in a scale-invariant Poisson process, conditioned on the event that the sum T of the locations of the points in (0,1] is 1. This extends to a similar result, rescaling the locations by T, and conditioning on the event that $T \leqslant 1$. Restricting both processes to $(0, \beta]$ for $0 < \beta \leqslant 1$, we give an explicit formula for the total variation distance between their distributions. Connections between various representations of the Poisson–Dirichlet process are discussed.

1. The Poisson–Dirichlet process

This paper gives a new characterization of the Poisson–Dirichlet distribution, showing its relation with the scale-invariant Poisson process. The Poisson–Dirichlet process $(V_1, V_2, \ldots)$ with parameter $\theta > 0$ (Kingman [15, 16], Watterson [25]) plays a fundamental role in combinatorics and number theory: see the exposition in [3]. The coordinates satisfy $V_1 > V_2 > \cdots > 0$ and $V_1 + V_2 + \cdots = 1$ almost surely. The distribution of this process is most directly characterized by the density functions of its finite-dimensional distributions. The joint density of $(V_1, V_2, \cdots, V_k)$ is supported by points $(x_1, \ldots, x_k)$ satisfying $x_1 > x_2 > \cdots > x_k > 0$ and $x_1 + \cdots + x_k < 1$. For the special case $\theta = 1$ the

† Supported in part by NSF grant DMS 96-26412.
‡ Supported in part by Schweizerischer NF Projekt Nr 20-43453.95.

joint density is

$$\rho\left(\frac{1-x_1-x_2-\cdots-x_k}{x_k}\right)\frac{1}{x_1x_2\cdots x_k}, \tag{1.1}$$

where ρ is Dickman's function [9, 21], characterized by $\rho(u)=0$ for $u<0$, $\rho(u)=1$ for $0\leqslant u\leqslant 1$, and $u\rho'(u)+\rho(u-1)=0$ for $u>1$, with ρ continuous for $u>0$ and differentiable for $u>1$. For general $\theta>0$, the expression for the joint density function is (see [25])

$$g_\theta\left(\frac{1-x_1-\cdots-x_k}{x_k}\right)\frac{e^{\gamma\theta}\,\theta^k\,\Gamma(\theta)\,x_k^{\theta-1}}{x_1x_2\cdots x_k}, \tag{1.2}$$

where g_θ is a probability density on $(0,\infty)$ characterized by (2.5).

A well-known construction of the Poisson–Dirichlet process [15, 16, 18] labels the points of the Poisson process $\mathscr{N}$ on $(0,\infty)$ with intensity $\theta e^{-x}/x$ as $\sigma_1,\sigma_2,\ldots$ with $0<\cdots<\sigma_3<\sigma_2<\sigma_1<\infty$. Their sum

$$S=\sigma_1+\sigma_2+\cdots \tag{1.3}$$

has the Gamma distribution with parameter θ and is independent of the renormalized vector $S^{-1}(\sigma_1,\sigma_2,\ldots)$, which has the Poisson–Dirichlet distribution with parameter θ:

$$\mathscr{L}(V_1,V_2,\ldots)=\mathscr{L}(S^{-1}(\sigma_1,\sigma_2,\ldots)). \tag{1.4}$$

A restatement of the independence is that, for any $s>0$,

$$\mathscr{L}(V_1,V_2,\ldots)=\mathscr{L}(s^{-1}(\sigma_1,\sigma_2,\ldots)|S=s). \tag{1.5}$$

2. Scale-invariant Poisson processes on $(0,\infty)$

Let $\mathscr{M}$ be the Poisson process on $(0,\infty)$ with intensity θ/x. The expected number of points in any interval (a,b) with $0<a<b$ is then $\theta\log(b/a)$. Since $\mathscr{M}$ has an intensity measure that is continuous with respect to Lebesgue measure, with probability one $\mathscr{M}$ has no double points. Thus we can identify $\mathscr{M}$ with a random discrete subset of $(0,\infty)$ with almost surely only finitely many points in any interval (a,b) as above. In particular, the points of $\mathscr{M}$ can be labelled X_i for $i\in\mathbb{Z}$ with

$$0<\cdots<X_2<X_1\leqslant 1<X_0<X_{-1}<X_{-2}<\cdots\ . \tag{2.1}$$

The process $\mathscr{M}$ is scale-invariant in that, for any $c>0$, as random sets there is equality in distribution:

$$\{cX_i\ :\ i\in\mathbb{Z}\}\stackrel{\mathrm{d}}{=}\{X_i\ :\ i\in\mathbb{Z}\}, \tag{2.2}$$

or, with the identification of $\mathscr{M}$ as a random set, simply $c\mathscr{M}\stackrel{\mathrm{d}}{=}\mathscr{M}$. Perhaps the simplest way to handle the scale-invariant Poisson process is to start with the translation-invariant Poisson process on $(-\infty,\infty)$ having intensity θ, and apply the exponential map. It is easy to check that, if the points of the translation-invariant Poisson process are labelled T_i for $i\in\mathbb{Z}$ so that $\cdots<T_{-2}<T_{-1}<T_0<0\leqslant T_1<T_2<\cdots$, then setting $X_i=\exp(-T_i)$ gives a realization of the scale-invariant Poisson process labelled to satisfy (2.1). From the familiar property that $W_1=T_1$ and the interpoint distances $W_i:=T_i-T_{i-1}$ for $i=2,3,\ldots$ are independent and exponentially distributed with mean $1/\theta$, so that $\mathbb{P}(\theta W_i\geqslant t)=e^{-t}$

for $t \geqslant 0$, it follows that $U_i := \exp(\theta W_i)$ is uniformly distributed in $(0, 1)$. Hence, for $i = 1, 2, \ldots$ we have $X_i = (U_1 U_2 \cdots U_i)^{1/\theta}$, with independent factors.

With the labelling (2.1), the sum T of locations of all points of the Poisson process $\mathscr{M}$ in (0,1) is

$$T = X_1 + X_2 + \cdots . \tag{2.3}$$

The Laplace transform of the distribution of T is

$$\mathbb{E}\exp(-sT) = \exp\left(-\theta \int_0^1 (1 - \exp(-sx))\frac{dx}{x}\right). \tag{2.4}$$

Computation with this Laplace transform (see Vervaat [24], p. 90, or Watterson [25]) shows that the density g_θ of T, with $g_\theta(x) = 0$ if $x < 0$, satisfies

$$x g_\theta(x) = \theta \int_{x-1}^{x} g_\theta(u) du, \quad x > 0, \tag{2.5}$$

so that

$$x g_\theta'(x) + (1-\theta) g_\theta(x) + \theta g_\theta(x-1) = 0, \quad x > 0. \tag{2.6}$$

Equation (2.6) shows why $\theta = 1$ is special. For the case $\theta = 1$, the density of T is $g_1(t) = e^{-\gamma}\rho(t)$, where γ is Euler's constant and ρ is Dickman's function.

The scale-invariant Poisson processes arise in another connection with the Poisson–Dirichlet process. The size-biased permutation of the Poisson–Dirichlet process has the same distribution as the vector $(1 - X_1, X_1 - X_2, \ldots)$ of spacings of the points of the scale-invariant Poisson process $\mathscr{M}$ in (2.1), starting from 1 and proceeding down: see Ignatov [14] and Donnelly and Joyce [10] for further details. A related property, from [1], is that as random sets with the labelling of (2.1), $\mathscr{M} := \{X_i : i \in \mathbb{Z}\} \stackrel{d}{=} \{X_{i-1} - X_i : i \in \mathbb{Z}\}$.

3. Conditioning the scale-invariant Poisson process

The following characterization of the Poisson–Dirichlet, based on conditioning the Poisson process with intensity θ/x, seems surprisingly to have been overlooked, perhaps because a 'Poisson representation', by rescaling or conditioning the process with intensity $\theta e^{-x}/x$, was already known.

Theorem 3.1. *For any $\theta > 0$, let the scale-invariant Poisson process $\mathscr{M}$ on $(0, \infty)$, with intensity θ/x, have its points falling in $(0, 1]$ labelled so that (2.1) holds. Let $(V_1, V_2, \ldots)$ have the Poisson–Dirichlet distribution with parameter θ. Then*

$$\mathscr{L}((V_1, V_2, \ldots)) = \mathscr{L}(\ (X_1, X_2, \ldots) \mid T = 1). \tag{3.1}$$

Proof. For $x > 0$ let $T(x)$ denote the sum of the locations of the points of $\mathscr{M}$ in $(0, x]$, so that

$$T(x) := \sum_{j \geqslant 1} X_j \mathbb{1}(X_j \leqslant x).$$

Then $T \equiv T(1)$, $T(x)/x$ has the same distribution as T, and $T(x)$ is independent of the Poisson process restricted to (x, ∞). Note that $T(x-)$ is the sum of locations of points in

$(0,x)$, and $T(x-) \stackrel{\mathrm{d}}{=} T(x)$. Let $(x_1,\ldots,x_k)$ satisfy $x_1 > x_2 > \cdots > x_k > 0$. Let $f(\cdot|x_1,\ldots,x_k)$ be the density of T, conditional on $X_i = x_i, 1 \leqslant i \leqslant k$. The joint density of $(X_1,\ldots,X_k,T)$ at $(x_1,\ldots,x_k,y)$ is

$$\exp\left(-\int_{x_1}^{1}\frac{\theta}{u}du\right)\frac{\theta}{x_1}\cdots\exp\left(-\int_{x_k}^{x_{k-1}}\frac{\theta}{u}du\right)\frac{\theta}{x_k}f(y|x_1,\ldots,x_k).$$

Now, for $y > x_1 + \cdots + x_k$,

$$\begin{aligned}\mathbb{P}(T \leqslant y \,|\, X_i = x_i, 1 \leqslant i \leqslant k) &= \mathbb{P}(T(x_k-) \leqslant y - x_1 - \cdots - x_k)\\ &= \mathbb{P}(T \leqslant (y - x_1 - \cdots - x_k)/x_k),\end{aligned}$$

the first equality following from independence, the second from scale invariance. Hence, recalling that g_θ is the density function of T,

$$f(y|x_1,\ldots,x_k) = \frac{1}{x_k}g_\theta\left(\frac{y - x_1 - \cdots - x_k}{x_k}\right).$$

It follows that the conditional density of $(X_1,\ldots,X_k)$, given $T = 1$, is

$$\frac{\theta^k}{x_1\cdots x_k}\,x_k^\theta\,\frac{1}{x_k}\,g_\theta\left(\frac{1 - x_1 - \cdots - x_k}{x_k}\right)/g_\theta(1), \tag{3.2}$$

which simplifies to the expression in (1.2). The equality of the normalizing constants, the fact that $e^{\gamma\theta}\Gamma(\theta) = 1/g_\theta(1)$, is automatic since (1.2) and (3.2) are both probability densities, with all the variable factors in agreement. □

An alternate proof of Theorem 3.1 can be extracted from Perman [17], which gives a general treatment of Poisson processes conditioned on the sum of the locations.

The following corollary about conditioning on $T = t$ for $0 < t \leqslant 1$ extends Theorem 3.1, and Theorem 3.1 is the special case $t = 1$ of Corollary 3.1.

Corollary 3.1. *For any $t \in (0,1]$, the distribution of $t^{-1}(X_1,X_2,\ldots)$ conditional on $T = t$ is the Poisson–Dirichlet distribution, that is, for any $t \in (0,1]$,*

$$\mathscr{L}(V_1,V_2,\ldots) = \mathscr{L}(t^{-1}(X_1,X_2,\ldots) \,|\, T = t). \tag{3.3}$$

Hence, by mixing with respect to the distribution of T conditional on the event $T \leqslant 1$, we have the relation which involves elementary conditioning:

$$\mathscr{L}(V_1,V_2,\ldots) = \mathscr{L}(T^{-1}(X_1,X_2,\ldots) \,|\, T \leqslant 1). \tag{3.4}$$

Proof. For $0 < t \leqslant 1$, (3.3) follows from (3.1) just by scale invariance and the independence of $\mathscr{M}$ on disjoint intervals. In detail, the event $T = t$ is the intersection of the events that $T(t) = t$ and that $\mathscr{M}$ restricted to $(t,1]$ has no points. By the independence of the restrictions of the Poisson process $\mathscr{M}$ to the intervals $(0,t]$ and $(t,1]$, conditioning on $T = t$ is the same as conditioning $\mathscr{M}$ restricted to $(0,t]$ on having $T(t) = t$, together with conditioning $\mathscr{M}$ restricted to $(t,1]$ on having no points. By the scale invariance of $\mathscr{M}$, the restriction to $(0,t]$, conditioned on $T(t) = t$, and then scaled up by dividing the location of every point by t, is equal in distribution to $\mathscr{M}$ restricted to $(0,1]$ and conditioned on $T = 1$. □

Having identified what happens to the scale-invariant Poisson process restricted to $(0, 1]$, conditional on $T = t$ for $0 < t \leqslant 1$, it is natural to ask what happens when $t > 1$. The following extends Theorem 3.1 in the opposite direction from the extension of Corollary 3.1.

Corollary 3.2. *For $t \geqslant 1$, the distribution of $t^{-1}(X_1, X_2, \ldots)$ conditional on $T = t$ is the Poisson–Dirichlet distribution conditional on its first component being at most $1/t$, that is, for any $t \geqslant 1$,*

$$\mathscr{L}((V_1, V_2, \ldots) \mid V_1 \leqslant t^{-1}) = \mathscr{L}(t^{-1}(X_1, X_2, \ldots) \mid T = t). \tag{3.5}$$

Proof. Our proof consists of the following chain of equalities.

$$\begin{aligned}
&\mathscr{L}((V_1, V_2, \ldots) \mid V_1 \leqslant t^{-1}) \\
&\quad = \mathscr{L}((X_1, X_2, \ldots) \mid X_1 \leqslant t^{-1}, X_1 + X_2 + \cdots = 1) \\
&\quad = \mathscr{L}(t^{-1}(tX_1, tX_2, \ldots) \mid tX_1 \leqslant 1, tX_1 + tX_2 + \cdots = t) \\
&\quad = \mathscr{L}(t^{-1}(X_1, X_2, \ldots) \mid T = t).
\end{aligned}$$

The first equality above holds for any $t > 0$, by (3.1), as does the second, by simple algebra. The final equality requires $t \geqslant 1$, and uses scale invariance, that $t\mathscr{M} \stackrel{\mathrm{d}}{=} \mathscr{M}$. The subtlety is in the labelling convention (2.1) needed in (2.3). We have for any $t > 0$ that $t\mathscr{M} \stackrel{\mathrm{d}}{=} \mathscr{M}$, but $tX_1, tX_2, \ldots$ is the list of points, in decreasing order, of $t\mathscr{M}$ restricted to $(0, t]$ rather than to $(0, 1]$. We need $t \geqslant 1$ to conclude that $(0, 1] \subset (0, t]$, so that conditioning first on $tX_1 \leqslant 1$ is just conditioning on $t\mathscr{M} \cap (1, t] = \emptyset$; it leaves the distribution of $t\mathscr{M}$ restricted to (0,1] unchanged, and guarantees that the sum $tX_1 + tX_2 + \cdots$ of locations of points of $t\mathscr{M}$ in $(0, t]$ equals the sum of locations of points of $t\mathscr{M}$ in $(0, 1]$. □

Note that the density of V_1 is strictly positive everywhere in $(0, 1)$. This implies that the Poisson–Dirichlet distribution in (3.3), and the conditioned Poisson–Dirichlet distributions in (3.5) for various $t > 1$, are all distinct, because any two of the distributions have, for sufficiently small ϵ, different values for the probability that the first component is less than ϵ. The same reasoning shows that the conditioning $T \leqslant 1$ in (3.4) cannot be omitted, and in fact cannot be replaced by conditioning on $T \leqslant c$ for any choice $c \in (1, \infty]$.

4. Total variation distance

Can the Poisson–Dirichlet process be distinguished from the scale-invariant Poisson process if one only observes the small coordinates? As a consequence of Theorem 3.1 it is possible to give a precise answer in a relatively simple formula.

4.1. A general lemma on preserving the total variation distance

One reason that the total variation distance is a useful metric is that inequalities for the total variation distance are preserved by arbitrary functionals: if X, Y are random elements of a measurable space $(S, \mathscr{S})$, and $h : (S, \mathscr{S}) \to (T, \mathscr{T})$ is any measurable map, then

$$d_{TV}(h(X), h(Y)) \leqslant d_{TV}(X, Y).$$

When can the above inequality be replaced by equality? For the discrete case, a necessary and sufficient condition [7] is that $h(a) \neq h(b)$ whenever $a, b \in S$ with $\mathbb{P}(X = a) > \mathbb{P}(Y = a)$ and $\mathbb{P}(X = b) < \mathbb{P}(Y = b)$. Lemma 4.1 gives the corresponding necessary and sufficient condition for the general measurable case, written in terms of the distributions μ, ν of the random elements X and Y discussed above.

Lemma 4.1. *Let $\mu, \nu \in \mathscr{P}(S, \mathscr{S})$, let $h : (S, \mathscr{S}) \to (T, \mathscr{T})$, and let $\mu' = \mu h^{-1}$, $\nu' = \nu h^{-1}$. Let $\gamma = (\mu+\nu)/2$ and $\gamma' = (\mu'+\nu')/2$, so that μ and ν are absolutely continuous with respect to γ, likewise for μ', ν', γ'. Let L be any version of the Radon–Nikodym derivative $d\mu/d\gamma$, and similarly let $L' = d\mu'/d\gamma'$. Consider the hypotheses*

(i) $L' \geqslant 1$ *on* $B \in \mathscr{T}$ *implies* $L \geqslant 1$ *(a.e. γ) on* $h^{-1}(B)$;
(ii) $L' \leqslant 1$ *on* $B \in \mathscr{T}$ *implies* $L \leqslant 1$ *(a.e. γ) on* $h^{-1}(B)$.

Then $d_{TV}(\mu, \nu) = d_{TV}(\mu', \nu')$ if and only if (i) *and* (ii).

Proof. Assume first that (i) and (ii) hold. Let $B_1 := \{t \in T : L' \geqslant 1\}$ and $B_2 := T \setminus B_1$ so that $B_1, B_2 \in \mathscr{T}$, and (i) applies to B_1, and (ii) applies to B_2. Let $A_1 = h^{-1}B_1$. Note $L \geqslant 1$ (a.e. γ) on A_1 using (i) and $L \leqslant 1$ (a.e. γ) on $S \setminus A_1$ using (ii). Now $d_{TV}(\mu', \nu') = \mu'(B_1) - \nu'(B_1) = \mu(A_1) - \nu(A_1) = d_{TV}(\mu, \nu)$.

For the opposite implication, we prove the contrapositive. Assume that (i) or (ii) does not hold. Without loss of generality we assume that (i) does *not* hold. Thus for B_1, A_1 as above there exists $A_2 \subset A_1$ with $A_2 \in \mathscr{S}$ and $\gamma(A_2) > 0$ and $L < 1$ everywhere on A_2. Hence for some $\epsilon, a > 0$ there exists $A_3 \subset A_2$ with $A_3 \in \mathscr{S}, \gamma(A_3) \geqslant a$, and $L < 1 - \epsilon$ on A_3. Thus $\mu(A_3) - \nu(A_3) \leqslant -2\epsilon a$ (because $L = d\mu/d\gamma$, so $2 - L = d\nu/d\gamma$ and $d(\mu - \nu)/d\gamma = -2(1 - L)$). Consider $A := A_1 \setminus A_3$. We have $d_{TV}(\mu, \nu) \geqslant \mu(A) - \nu(A) = \mu(A_1) - \nu(A_1) - (\mu(A_3) - \nu(A_3)) \geqslant \mu(A_1) - \nu(A_1) + 2\epsilon a = \mu'(B_1) - \nu'(B_1) + 2\epsilon a = d_{TV}(\mu', \nu') + 2\epsilon a$. □

Diaconis and Pitman [8] view 'sufficiency' as the unifying concept in explaining equalities for total variation distance, and indeed, for all *natural* examples encountered so far, sufficiency is present when equality holds. Recall that h is a 'sufficient statistic' for comparing the distributions of X and Y if the likelihood ratio factors through h. (In place of the usual likelihood ratio $R = d\mu/d\nu$ we have used $L = 2d\mu/d(\mu + \nu)$ as a device to avoid dividing by zero; the relations are $L = 2R/(1 + R)$, $R = L/(2 - L)$.)

Corollary 4.1. *Sufficiency is sufficient to preserve d_{TV}.*

Proof. Assume that h is sufficient, so that some version of the likelihood L as in Lemma 4.1 factors through h, that is, with $\mathscr{B}$ denoting the Borel sigma algebra on the $\mathbb{R}$, there is a function $f : (T, \mathscr{T}) \to (\mathbb{R}, \mathscr{B})$ such that $L = f \circ h$ is a version of $d\mu/d\gamma$. In this situation, we can take $L' = f$, that is, f is a version of $d\mu'/d\gamma'$. For this pair L, L' condition (i) simply says, 'for $B \in \mathscr{S}$, $f \geqslant 1$ on B implies $f \circ h \geqslant 1$ on $h^{-1}(B)$', which is obviously true; similarly for condition (ii). □

4.2. Poisson–Dirichlet versus scale-invariant Poisson

For any $\theta > 0$, we can view the scale-invariant Poisson process $\mathscr{M}$ with intensity θ/x as a random subset of $(0, \infty)$, and the Poisson–Dirichlet process with parameter θ as a random subset $\mathscr{PD} = \{V_1, V_2, \ldots\}$ of $(0, 1]$. Theorem 3.1 shows that the difference between the distributions of $\mathscr{M}_1 = \mathscr{M} \cap (0, 1]$ and $\mathscr{PD}$ lies only in conditioning on $T = 1$. This suggests that, if attention is restricted to $(0, \beta]$ for $\beta \leqslant 1$, the distributions should be closer, and progressively so as $\beta \to 0$. Theorem 4.1 below reduces the total variation distance between the two processes to a simpler total variation distance between two random variables.

We denote this simpler distance by $H_\theta(\beta)$. It is defined for $\theta > 0$ and $\beta \in [0, 1]$ by

$$H_\theta(\beta) := d_{TV}(\mathscr{L}(T(\beta)), \mathscr{L}(T(\beta)|T = 1)).$$

We review the formula for H and its derivation, taken from [20]. For $0 < \beta < 1$, consider the distributions of $T(\beta)$ and $T - T(\beta)$, which are independent of one another. Because $T(\beta) \stackrel{\mathrm{d}}{=} \beta T$ by scale invariance, its density $g_{\theta,\beta}$ is given in terms of the density g_θ of T by

$$g_{\theta,\beta}(x) = \beta^{-1} g_\theta(x/\beta).$$

For $\beta \in (0, 1]$, the distribution of $T - T(\beta)$ has an atom at zero, corresponding to no points of $\mathscr{M}$ in $(\beta, 1]$:

$$\mathbb{P}(T - T(\beta) = 0) = \mathbb{P}(\mathscr{M} \cap (\beta, 1] = \emptyset) = \beta^\theta.$$

For $\beta \in [0, 1)$, the distribution of $T - T(\beta)$ has a continuous part, with density $h_{\theta,\beta}$ satisfying $h_{\theta,\beta}(x) = 0$ for $x < \beta$, and, for all $x > 0$,

$$h_{\theta,\beta}(x) = \frac{\theta}{x}\left(\beta^\theta \mathbb{1}(\beta \leqslant x \leqslant 1) + \int_{x-1}^{x-\beta} h_{\theta,\beta}(u)du\right). \tag{4.1}$$

An analysis of differential-difference equations related to (4.1) is carried out in [12, 13].

It follows that the total variation distance between the distributions of $T(\beta)$ and the conditional distribution of $T(\beta)$ given $T = 1$ is given by

$$\begin{aligned} 2H_\theta(\beta) &= \int_0^1 g_{\theta,\beta}(x)\left|\frac{h_{\theta,\beta}(1-x)}{g_\theta(1)} - 1\right| dx + \beta^\theta \frac{g_{\theta,\beta}(1)}{g_\theta(1)} + \int_1^\infty g_{\theta,\beta}(x)dx \\ &= \int_0^1 g_{\theta,\beta}(x)\left|\frac{h_{\theta,\beta}(1-x)}{g_\theta(1)} - 1\right| dx + \beta^{\theta-1} \frac{g_\theta(1/\beta)}{g_\theta(1)} + \mathbb{P}(T > 1/\beta). \end{aligned} \tag{4.2}$$

Theorem 4.1. *For any $\theta > 0$, view the scale-invariant Poisson process $\mathscr{M}$ with intensity θ/x as a random subset of $(0, \infty)$ and the Poisson–Dirichlet process with parameter θ as a random subset $\mathscr{PD} := \{V_1, V_2, \ldots\}$ of $(0, 1]$. For every $\beta \in [0, 1]$,*

$$d_{TV}(\mathscr{M} \cap [0, \beta], \mathscr{PD} \cap [0, \beta]) = d_{TV}(T(\beta), (T(\beta)|T = 1)). \tag{4.3}$$

Proof. For any countable collection of points $x = \{x_1, x_2, \ldots\}$ satisfying $1 > x_1 > x_2 > \cdots$ and, with only finitely many in any interval (a, b) with $0 < a < b < 1$, let $x^{(\beta)}$ denote x

restricted to $(0,\beta]$. Then, by Theorem 3.1 and the independence of $T(\beta)$ and $T - T(\beta)$,

$$\frac{d\mathscr{L}(\mathscr{PD}\cap[0,\beta])}{d\mathscr{L}(\mathscr{M}\cap[0,\beta])}(x^{(\beta)}) = \begin{cases} h_{\theta,\beta}(1-t_\beta(x))/g_\theta(1), & \text{if } t_\beta(x) < 1, \\ \infty, & \text{if } t_\beta(x) = 1, \\ 0, & \text{if } t_\beta(x) > 1, \end{cases}$$

is a function of $t_\beta(x) = \sum_{j\geqslant 1} x_j \mathbb{1}(x_j \leqslant \beta)$ alone. The theorem follows now from Corollary 4.1. □

In the case $\theta = 1$, the limit $H_1(\beta)$ was specified in [6], with a heuristic argument that it would give the limit for total variation distance between the cycle structure of random permutations on n objects, and an initial segment of the corresponding independent limit process, observing cycles of size i for all $i \leqslant \beta n$. Stark [20] proved this limit for total variation distance for permutations, together with extensions to various random 'assemblies' attracted to the Poisson–Dirichlet with parameter θ for general $\theta > 0$, including in particular random mappings, for which $\theta = 1/2$. Convergence to a Poisson–Dirichlet distribution for the large components of such random combinatorial structures in general was proved by Hansen [11]; see also [4]. In the special case $\theta = 1$, the expression (4.2) for H_1 can be expressed entirely in terms of Dickman's function ρ and Buchstab's function ω, and indeed [5] and [22] show that the function H_1 appears in a variant of Kubilius' fundamental lemma concerning the small prime factors of a random integer chosen uniformly from 1 to n.

5. Connecting the two Poisson representations

In this paper we have given a representation of the Poisson–Dirichlet process based on the scale-invariant Poisson process $\mathscr{M}$ with intensity θ/x. The earlier Gamma representation uses the Poisson process $\mathscr{N}$ with intensity $\theta e^{-x}/x$. The relation between these two representations has its root in combinatorics.

Shepp and Lloyd [19] analysed random permutations of n objects by applying Tauberian analysis to the following setup. Consider independent Poisson random variables Z_i with $\mathbb{E}Z_i = \theta z^i/i$ for any $z \in (0,1)$ and $\theta > 0$, and let $T_\infty := \sum_{i\geqslant 1} iZ_i$. It requires $z < 1$ to conclude that $\mathbb{E}T_\infty < \infty$ and T_∞ is almost surely finite; if $z \geqslant 1$ then $T_\infty = \infty$ almost surely. For $\theta = 1$, conditional on the event $T_\infty = n$, the joint distribution of $(Z_1, Z_2, \ldots)$ is the distribution of counts of cycles of lengths $1, 2, \ldots$ in a random permutation of n objects. Vershik and Shmidt [23] show that the process listing the longest, second longest, … cycle lengths, rescaled by n, converges in distribution to the Poisson–Dirichlet (with parameter $\theta = 1$). It is easy to show that, for any fixed $\theta, c > 0$, using $z = z(n) = e^{-c/n}$, the point processes having mass Z_i at i/n converge to the Poisson process with intensity $\theta e^{-cx}/x$. Thus, with $c = 1$, we see that the Shepp and Lloyd method corresponds to the Gamma representation (1.5), using $s = 1$. Note that the sum of locations of all points, which is T_∞/n for the discrete processes, converges to the Gamma-distributed limit S in (1.3).

Arratia and Tavaré [6, 7] modified this by considering $T_n := \sum_{1\leqslant i\leqslant n} iZ_i$ in place of T_∞. The cycle structure of a random permutation is given by the joint distribution of $(Z_1, Z_2, \ldots, Z_n)$ conditional on $T_n = n$ for $\theta = 1$ and *any* $z > 0$, including $z = 1$, in

$\mathbb{E}Z_i := \theta z^i/i$. This allows one to take the limit directly: $\mathbb{E}Z_i = 1/i$, setting $z = 1$ in place of using $z(n) \nearrow 1$. The point processes with mass Z_i at i/n, using $\mathbb{E}Z_i = \theta/i$, converge to the scale-invariant Poisson process of Section 2, and the sum of the locations of the points in $(0, 1]$, which is T_n/n for the discrete processes, converges to the limit random variable T in (2.3).

Now the continuum analogue of replacing T_∞ by T_n and replacing $z(n) = e^{-c/n}$ for $c = 1$ by $z = 1$ is exactly replacing S, the sum of locations of points in the Poisson process on $(0, \infty)$ with intensity $\theta e^{-cx}/x$, by T, the sum of locations of points in $(0, 1]$ in the Poisson process on $(0, \infty)$ with intensity θ/x. This analogy suggests the following alternative proof of Theorem 3.1 and Corollary 3.1.

Proof. Compare S, the sum of locations of all points of $\mathcal{N}$ defined in (1.3), with $S_1 := \sum_{i\geqslant 1} \sigma_i \mathbb{1}(\sigma_i \leqslant 1)$, the sum of locations of points in the Poisson process $\mathcal{N}_1$ with intensity $\theta e^{-cx}/x$ *restricted* to $(0, 1]$. Write $\mathcal{M}_1$ for the Poisson process with intensity θ/x restricted to $(0, 1]$, and recall that T is the sum of the locations of the points of $\mathcal{M}_1$. For a configuration $(x_1, x_2, \ldots)$ with $1 \geqslant x_1 > x_2 > \cdots > x_k \geqslant \beta > x_{k+1} > 0$ and $x_1 + x_2 + \cdots + x_k = s$, the likelihood ratio for the restrictions of $\mathcal{N}$ and $\mathcal{M}$ to $[\beta, 1]$ is $e^{-cs} \exp(\theta \int_\beta^1 (1 - e^{-cx})/x \, dx)$, where the second factor corresponds to the requirement of no points in $[\beta, 1]$ other than $x_1, \ldots, x_k$. Thus, for an infinite configuration of points at $1 \geqslant x_1 > x_2 > \cdots > 0$ with $s = x_1 + x_2 + \cdots$, the likelihood ratio for $\mathcal{N}_1$ versus $\mathcal{M}_1$ is $e^{-cs} \exp(\theta \int_0^1 (1 - e^{-cx})/x \, dx)$. It follows that for any $s > 0$, $\mathcal{N}_1$ conditional on $S_1 = s$ has the same distribution as $\mathcal{M}_1$ conditional on $T = s$. We need $0 < s \leqslant 1$ so that $S = s$ implies $S = S_1$ and $\mathcal{N} = \mathcal{N}_1$. □

References

[1] Arratia, R. (1996) Independence of small prime factors: total variation and Wasserstein metrics, insertions and deletions, and the Poisson–Dirichlet process. Draft, available from rarratia@math.usc.edu

[2] Arratia, R., Barbour, A. D. and Tavaré, S. (1999) *Logarithmic Combinatorial Structures*. Monograph, in preparation.

[3] Arratia, R., Barbour, A. D. and Tavaré, S. (1997) Random combinatorial structures and prime factorizations. *Notices of the AMS* **44** 903–910.

[4] Arratia, R., Barbour, A. D. and Tavaré, S. (1999) On Poisson–Dirichlet limits for random decomposable combinatorial structures. *Combinatorics, Probability and Computing* **8** 193–208.

[5] Arratia, R. and Stark, D. (1999) A total variation distance invariance principle for primes, permutations, and the Poisson–Dirichlet process. In preparation.

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

[7] Arratia, R. and Tavaré, S. (1994) Independent process approximations for random combinatorial structures. *Adv. Math.* **104** 90–154.

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

[9] Dickman, K. (1930) On the frequency of numbers containing prime factors of a certain relative magnitude. *Ark. Math. Astr. Fys.* **22** 1–14.

[10] Donnelly, P. and Joyce, P. (1989) Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stoch. Proc. Appl.* **31** 89–103.

[11] Hansen, J. C. (1994) Order statistics for decomposable combinatorial structures. *Rand. Struct. Alg.* **5** 517–533.
[12] Hildebrand, A. (1990) The asymptotic behavior of the solutions of a class of differential-difference equations. *J. London Math. Soc.* **42** 11–31.
[13] Hildebrand, A. and Tenenbaum, G. (1993) On a class of differential-difference equations arising in number theory. *J. d'Analyse* **61** 145–179.
[14] Ignatov, T. (1982) On a constant arising in the asymptotic theory of symmetric groups, and on Poisson–Dirichlet measures. *Theory Probab. Appl.* **27** 136–147.
[15] Kingman, J. F. C. (1975) Random discrete distributions. *J. Royal Statist. Soc. B* **37** 1–22.
[16] Kingman, J. F. C. (1993) *Poisson Processes*, Oxford University Press, Oxford.
[17] Perman, M. (1993) Order statistics for jumps of normalized subordinators. *Stoch. Proc. Appl.* **46** 267–281.
[18] Pitman, J. and Yor, M. (1997) The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900.
[19] Shepp, L. and Lloyd, S. P. (1966) Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.* **121** 340–357.
[20] Stark, D. (1997) Explicit non-zero limits of total variation distance in independent Poisson approximations of logarithmic combinatorial assemblies. *Combinatorics, Probability and Computing* **6** 87–106.
[21] Tenenbaum, G. (1995) *Introduction to Analytic and Probabilistic Number Theory*, Vol. 46 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press.
[22] Tenenbaum, G. (1999) Crible d'Ératosthène et modèle de Kubilius. In *Proceedings of the International Meeting in Honor of A. Schinzel, Zakopane 1997*, de Gruyter. To appear.
[23] Vershik, A. M. and Shmidt, A. A. (1977) Limit measures arising in the theory of groups, I. *Theory Probab. Appl.* **22** 79–85.
[24] Vervaat, W. (1972) Success epochs in Bernoulli trials with applications in number theory. *Mathematical Center Tracts* No. 42, Mathematisch Centrum, Amsterdam.
[25] Watterson, G. A. (1976) The stationary distribution of the infinitely-many-alleles diffusion model. *J. Appl. Probab.* **13** 639–651.