J. Appl. Prob. **51,** 400–416 (2014)
Printed in England

DIRICHLET AND QUASI-BERNOULLI LAWS FOR PERPETUITIES

PAWEŁ HITCZENKO,* *Drexel University*
GÉRARD LETAC,** *Université Paul Sabatier*

Abstract

Let X, B, and Y be the Dirichlet, Bernoulli, and beta-independent random variables such that $X \sim \mathcal{D}(a_0, \dots, a_d)$, $\Pr(B = (0, \dots, 0, 1, 0, \dots, 0)) = a_i/a$ with $a = \sum_{i=0}^d a_i$, and $Y \sim \beta(1, a)$. Then, as proved by Sethuraman (1994), $X \sim X(1 - Y) + BY$. This gives the stationary distribution of a simple Markov chain on a tetrahedron. In this paper we introduce a new distribution on the tetrahedron called a quasi-Bernoulli distribution $\mathcal{B}_k(a_0, \dots, a_d)$ with k an integer such that the above result holds when B follows $\mathcal{B}_k(a_0, \dots, a_d)$ and when $Y \sim \beta(k, a)$. We extend it even more generally to the case where X and B are random probabilities such that X is Dirichlet and B is quasi-Bernoulli. Finally, the case where the integer k is replaced by a positive number c is considered when $a_0 = \dots = a_d = 1$.

Keywords: Perpetuities; Dirichlet process; Ewens' distribution; quasi-Bernoulli law; probabilities on a tetrahedron; T_c transform; stationary distribution

2010 Mathematics Subject Classification: Primary 60J05; 60E99

1. Introduction

In a recent paper Ambrus *et al.* [1] make the following observation. If V, Y, and W are independent random variables such that $V \sim (1/\pi)(\frac{1}{4} - v^2)^{-1/2}\mathbf{1}_{(-1/2,1/2)}(v)\,\mathrm{d}v$, Y is uniform on $(0, 1)$, and $\Pr(W = 1) = \Pr(W = -1) = \frac{1}{2}$, then

$$V \sim V(1 - Y) + \frac{W}{2}Y.$$

The law μ of a random variable V satisfying $V \sim VM + Q$, where the pair (M, Q) is independent of V on the right-hand side, is often called a perpetuity generated by the law ν of (M, Q). Thus, another way of stating the observation from [1] is that an arcsine random variable on $(-\frac{1}{2}, \frac{1}{2})$ is a perpetuity generated by the distribution of $(M, Q) \sim (1 - Y, WY/2)$. This property of the arcsine law is actually an instance of a much more general result due to Sethuraman (see [10] or Theorem 1.1 below) on the Dirichlet distribution.

To recall Sethuraman's result, we will need the following notation. The natural basis of $\mathbb{R}^{d+1}$ is denoted by $e_0, \dots, e_d$. The convex hull of $\{e_0, \dots, e_d\}$ is a tetrahedron that we denote by E_{d+1}. The elements of E_{d+1} are therefore the vectors $\lambda = (\lambda_0, \dots, \lambda_d)$ of $\mathbb{R}^{d+1}$ such that $\lambda_i \geq 0$ for $i = 0, \dots, d$ and such that $\lambda_0 + \dots + \lambda_d = 1$. If $p_0, \dots, p_d$ are positive numbers whose sum is equal to 1, the distribution $\sum_{i=0}^d p_i\delta_{e_i}$ of $B = (B_0, \dots, B_d) \in E_{d+1}$ is called a

Received 17 April 2012; revision received 21 January 2013.
* Postal address: Department of Mathematics, Drexel University, Philadelphia PA 19104, USA.
Partially supported by the Simons Foundation (grant number #208766).
** Postal address: Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 31062 Toulouse, France.
Email address: gerard.letac@alsatis.net
This author thanks the Fields Institute for its hospitality during the preparation of this paper.

https://doi.org/10.1239/jap/1402578633 Published online by Cambridge University Press

Bernoulli distribution. By definition, B satisfies $\Pr(B = e_i) = p_i$. If $a_0, \dots, a_d$ are positive numbers, the Dirichlet distribution $\mathcal{D}(a_0, \dots, a_d)$ of $X = (X_0, \dots, X_d) \in E_{d+1}$ is such that the law of $(X_1, \dots, X_d)$ is

$$\frac{1}{B(a_0, \dots, a_d)}(1 - x_1 - \cdots - x_d)^{a_0 - 1} x_1^{a_1 - 1} \cdots x_d^{a_d - 1} \mathbf{1}_{T_d}(x_1, \dots, x_d)\, \mathrm{d}x_1 \cdots \mathrm{d}x_d,$$

where $B(a_0, \dots, a_d) = \Gamma(a_0) \cdots \Gamma(a_d)/\Gamma(a_0 + \cdots + a_d)$ and T_d is the set of $(x_1, \dots, x_d)$ such that $x_i > 0$ for all $i = 0, 1, \dots, d$, with the convention $x_0 = 1 - x_1 \cdots - x_d$. For instance, if the real random variable X_1 follows the beta distribution

$$\beta(a_1, a_0)(\mathrm{d}x) = \frac{1}{B(a_1, a_0)} x^{a_1 - 1}(1 - x)^{a_0 - 1} \mathbf{1}_{(0,1)}(x)\, \mathrm{d}x,$$

then $(1 - X_1, X_1) \sim \mathcal{D}(a_0, a_1)$.

Theorem 1.1. ([11].) *Let $a_0, \dots, a_d$ be positive numbers. Define $a = a_0 + \cdots + a_d$. Let X, Y, and B be the Dirichlet, beta, and Bernoulli independent random variables such that $X \sim \mathcal{D}(a_0, \dots, a_d)$ and $B \sim \sum_{i=0}^{d} a_i \delta_{e_i}/a$ are valued in $\mathbb{R}^{d+1}$ and such that $Y \sim \beta(1, a)$. Then $X \sim X(1 - Y) + BY$.*

Remark. Considering each coordinate, Theorem 1.1 says that, for all $i = 0, \dots, d$, we have $X_i \sim X_i(1 - Y) + B_i Y$. Since $1 = \sum_{i=0}^{d} X_i = \sum_{i=0}^{d} B_i$, the statement for $i = 0$ is true if it is verified for $i = 1, \dots, d$. For instance, for $d = 1$, Theorem 1.1 can be reformulated as in the next result.

Corollary 1.1. *Let $a_0, a_1 > 0$. Let X_1, Y, and B_1 be three independent random variables such that $X_1 \sim \beta(a_1, a_0)$, $Y \sim \beta(1, a_0 + a_1)$, $B_1 \sim a_0\delta_0/(a_0 + a_1) + a_1\delta_1/(a_0 + a_1)$. Then $X_1 \sim X_1(1 - Y) + B_1 Y$.*

The initial remark contained in [1] is therefore a particular case of Theorem 1.1 for $d = 1$ and $a_0 = a_1 = \frac{1}{2}$. More generally, the case in which $d = 1$ and $a_0 = a_1$ covers the power semicircle distributions discussed in [2] (with $\theta = a_0 - \frac{3}{2}$). In particular, $a_0 = a_1 = \frac{3}{2}$ is the classical semicircle distribution.

Part of the reason we found Theorem 1.1 interesting is that there are relatively few examples of exact solutions to perpetuity equations in the literature.

The aim of this paper is to generalize Sethuraman's result, and our generalization (see Theorem 4.1 below) will provide more examples of the explicit generation of perpetuities. Stating Theorem 4.1 needs the introduction of a new distribution $\mathcal{B}_k(a_0, \dots, a_d)$ on the tetrahedron E_{d+1}. We call it a quasi-Bernoulli distribution of order k. It is concentrated on the faces of order less than k in a way that we will make reasonably explicit in Section 3. With these new distributions, we add a family of laws with interesting properties to the zoo of distributions on a tetrahedron.

Finally, one can prove Corollary 1.1 directly by showing that $\mathbb{E}(X_1(1 - Y) + B_1 Y)^n) = \mathbb{E}(X_1^n)$ for all integers n. Our proof of Theorem 4.1 is somewhat linked to this method of moments. It relies on the properties of the T_c transform of a distribution on the tetrahedron E_{d+1} introduced in [4] (see also [9]). We will prove several convenient properties of the T_c transform in Theorem 2.1. Theorem 5.1 extends Theorem 4.1 to random probability measures on an abstract space Ω, where the Dirichlet distribution is replaced by the Dirichlet random measure governed by the positive measure α on Ω. Surprisingly, the construction of the quasi-Bernoulli random measure of Section 5 uses Ewens' distribution.

Note that the same perpetuity can be generated by different ν. See Section 6 for a reminder of the classical link between perpetuities and the stationary distributions of the Markov chains obtained by iteration on random affine maps.

2. The T_c transform of a distribution on the tetrahedron

In the sequel if $f = (f_0, \dots, f_d)$ and $x = (x_0, \dots, x_d)$ are in $\mathbb{R}^{d+1}$, we write $\langle f, x\rangle = \sum_{i=0}^d f_i x_i$ and define $U_{d+1} = \{f = (f_0, \dots, f_d) \in \mathbb{R}^{d+1};\ f_0 > 0, \dots, f_d > 0\}$. Let $X = (X_0, \dots, X_d)$ be a random variable on E_{d+1}, and let $c > 0$. The T_c transform of X is the following function on U_{d+1}:

$$T_c(X)(f) = \mathbb{E}(\langle f, X\rangle^{-c}).$$

Its existence is clear from $T_c(X)(f) \le (\min_i f_i)^{-c} < \infty$. It satisfies

$$T_c(X)(\lambda f) = \lambda^{-c} T_c(X)(f).$$

The explicit calculation of $T_c(X)$ is easy in some rare cases, including the Dirichlet case $\mathcal{D}(a_0, \dots, a_d)$ when $c = a = a_0 + \dots + a_d$ and the Bernoulli case $\sum_{i=0}^k p_i \delta_{e_i}$. In some sense, the present paper originated from an effort to compute $T_c(X)$ when $X \sim \mathcal{D}(a_0, \dots, a_d)$ and $c = a + k$, where k is a positive integer. For $d = 1$, knowing the T_c transform is equivalent to knowing the function $t \mapsto \mathbb{E}((1 - tX)^{-c})$ on $(-\infty, 1)$ when X is a random variable valued in $[0, 1]$ since

$$T_c((1 - X, X))(1, 1 - t) = \mathbb{E}((1 - tX)^{-c}).$$

The T_c transform is a tool which is in general better adapted to the study of distributions on the tetrahedron than the Laplace transform $\mathbb{E}(\exp(-\langle f, X\rangle))$. The next theorem gathers its main properties. It shows for instance that $T_c(X)$ characterizes the distribution of X and gives in (2.4) a crucial probabilistic interpretation to the product $T_a(X_0)T_b(X_1)$ when X_0 and X_1 are independent random variables valued in E_{d+1}.

Theorem 2.1. 1. *If X and Z are random variables on E_{d+1} and if there exists $c > 0$ such that $T_c(X)(f) = T_c(Z)(f)$ for all $f \in U_{d+1}$, then $X \sim Z$.*

2. *If k is a nonnegative integer and $H = -(\partial/\partial f_0 + \dots + \partial/\partial f_d)$, then*

$$H^k T_c(X) = (c)_k T_{c+k}(X), \tag{2.1}$$

where $(c)_n$ is the Pochhammer symbol defined by $(c)_0 = 1$ and $(c)_{n+1} = (c)_n(c + n)$.

3. *If $(a_0, \dots, a_d) \in U_{d+1}$ with $a = a_0 + \dots + a_d$ and $X \sim \mathcal{D}(a_0, \dots, a_d)$, then*

$$T_a(X)(f) = f_0^{-a_0} \cdots f_d^{-a_d}. \tag{2.2}$$

4. *Suppose that $X_0, \dots, X_r, Y$ are independent random variables such that $X_i \in E_{d+1}$ for $i = 0, \dots, r$ and $Y = (Y_0, \dots, Y_r) \in E_{r+1}$ has Dirichlet distribution $\mathcal{D}(b_0, \dots, b_r)$. Then, for $b = b_0 + \dots + b_r$ and $Z = X_0Y_0 + \dots + X_rY_r$, we have, on U_{d+1},*

$$T_b(Z)(f) = T_{b_0}(X_0)(f) \cdots T_{b_d}(X_d)(f). \tag{2.3}$$

In particular, if $Y \sim \beta(b_1, b_0)$, we have

$$T_{b_0+b_1}((1 - Y)X_0 + YX_1) = T_{b_0}(X_0)T_{b_1}(X_1). \tag{2.4}$$

5. *The probability of the face $x_0 = \dots = x_k = 0$ is computable by the T_c transform:*

$$\lim_{f_0 \to \infty} T_c(X)(f_0, \dots, f_0, 1, 1, \dots, 1) = \Pr(X_0 = X_1 = \dots = X_k = 0).$$

Proof. To prove part 1, fix $g \in \mathbb{R}^{d+1}$, set $f_i = 1 - tg_i$ for small enough t, and develop $t \mapsto \mathbb{E}(\langle f, X\rangle^{-c})$ in a neighborhood of $t = 0$. Since $\langle f, X\rangle = 1 - t\langle g, X\rangle$, we have

$$T_c(X)(f) = \mathbb{E}((1 - t\langle g, X\rangle)^{-c}) = \sum_{n=0}^{\infty} \frac{(c)_n}{n!}\mathbb{E}(\langle g, X\rangle^n)t^n.$$

It follows from the hypothesis $T_c(X) = T_c(Z)$ that $\mathbb{E}(\langle g, X\rangle^n) = \mathbb{E}(\langle g, Z\rangle^n)$ for all n. Thus, $\langle g, X\rangle \sim \langle g, Z\rangle$ since both are bounded random variables with the same moments. Since this is true for all $g \in \mathbb{R}^{d+1}$, we have $X \sim Z$. Equation (2.1) is easy to obtain by induction on k using the fact that $X_0 + \cdots + X_d = 1$.

We do not give a proof of the classical formula (2.2); see Proposition 2.1 of [4], or [9] where three different proofs are discussed. Equation (2.3) follows from (2.2) by replacing $X, a_0, \ldots, a_d$ with $Y, b_0, \ldots, b_r$ and f with $\langle f, X_0\rangle, \ldots, \langle f, X_r\rangle$. Using conditioning and the independence of $X_0, \ldots, X_r$, we obtain

$$\begin{aligned} T_b(Z)(f) &= \mathbb{E}\biggl(\mathbb{E}\biggl(\biggl[\sum_{j=0}^{r} Y_j\langle f, X_j\rangle\biggr]^{-b} \,\bigg|\, X_0, \ldots, X_r\biggr)\biggr) \\ &= \mathbb{E}\biggl(\prod_{j=0}^{r}\langle f, X_j\rangle^{-b_j}\biggr) \\ &= \prod_{j=0}^{r} T_{b_j}(X_j)(f). \end{aligned}$$

Applying (2.3) to $(Y_0, Y_1) = (1 - Y, Y) \sim \mathcal{D}(b_0, b_1)$, we obtain $Z = (1 - Y)X_0 + YX_1$. This leads to (2.4). Property 5 is obvious since the events $X_0 + \cdots + X_k = 0$ and $X_0 = X_1 = \cdots = X_k = 0$ coincide.

Remark. Theorem 2.1 may be used to obtain a proof of Theorem 1.1 that is different to Sethuraman's original argument. Since it is related to our proof of Theorem 4.1 below, we briefly sketch it. Take $X_0 = X$, $X_1 = B$, $b_1 = 1$, and $b_0 = a$ in (2.4). Thus,

$$T_1(B)(f) = \frac{1}{a}\biggl(\frac{a_0}{f_0} + \cdots + \frac{a_d}{f_d}\biggr).$$

The trick for computing $T_{1+a}(X)$ is to observe from (2.1) and (2.2) that

$$T_{1+a}(X)(f) = \frac{-1}{a}\biggl(\sum_{i=0}^{d}\frac{\partial}{\partial f_i}\biggr)\prod_{i=0}^{d}\frac{1}{f_i^{a_i}} = T_a(X)(f)T_1(B)(f).$$

From (2.4) we also know that, for $Z = (1 - Y)X + YB$, we have $T_{1+a}(Z) = T_a(X)T_1(B)$. Thus, $T_{1+a}(Z) = T_{1+a}(X)$. Part 1 of Theorem 2.1 implies that $X \sim Z$.

3. The quasi-Bernoulli distributions on a tetrahedron

We begin by slightly extending the definition of a Dirichlet distribution $\mathcal{D}(a_0, \ldots, a_d)$ by allowing $a_i \geq 0$ instead of $a_i > 0$ while keeping $a = \sum_{i=0}^{d} a_i > 0$. For such a $(a_0, \ldots, a_d)$ sequence, we define the nonempty set $T = \{i; a_i > 0\}$. We say that $\mathcal{D}(a_0, \ldots, a_d)$ is the Dirichlet distribution concentrated on the tetrahedron E_T generated by $(e_i)_{i\in T}$ with parameters $(a_i)_{i\in T}$. If $X \sim \mathcal{D}(a_0, \ldots, a_d)$, the formula $\mathbb{E}(\langle f, X\rangle^{-a}) = \prod_{i=0}^{d} f_i^{-a_i}$ still holds. If T

https://doi.org/10.1239/jap/1402578633 Published online by Cambridge University Press

contains only one element i_0 then $\mathcal{D}(a_0,\dots,a_d)$ is simply $\delta_{e_{i_0}}$ and does not depend on a. Now recall a simple combinatorial formula where k is a positive integer and $a = a_0 + \cdots + a_d$:

$$\sum_{\substack{(b_0,\dots,b_d)\in\mathbb{N}^{d+1}\\ b_0+\cdots+b_d=k}} \prod_{i=0}^{d} \frac{(a_i)_{b_i}}{b_i!} = \frac{(a)_k}{k!}. \tag{3.1}$$

The proof is immediate if we use generating functions: expand $\prod_{i=0}^{d}(1-t)^{-a_i} = (1-t)^{-a}$ in a power series on both sides. We now define our new distributions.

Let $a_0,\dots,a_d > 0$ and $a = a_0+\cdots+a_d$, and let k be a positive integer. The quasi-Bernoulli distribution of order k is the distribution on the tetrahedron E_{d+1} defined as the mixing of Dirichlet distributions:

$$\mathcal{B}_k(a_0,\dots,a_d) = \frac{k!}{(a)_k} \sum_{\substack{(b_0,\dots,b_d)\in\mathbb{N}^{d+1}\\ b_0+\cdots+b_d=k}} \prod_{i=0}^{d} \frac{(a_i)_{b_i}}{b_i!} \mathcal{D}(b_0,\dots,b_d). \tag{3.2}$$

Equation (3.1) shows that (3.2) is indeed a probability on E_{d+1}. Setting $c = k$ in Theorem 4.2 below gives an explicit form of $\mathcal{B}_k(a_0,\dots,a_d)$ in the particular case $a_0 = \cdots = a_d = 1$. For the sake of simplicity in the following, define

$$\sigma_j = \sum_{i=0}^{d} \frac{a_i}{f_i^j}. \tag{3.3}$$

Proposition 3.1. *If $B \sim \mathcal{B}_k(a_0,\dots,a_d)$ then*

$$T_k(B)(f) = \frac{k!}{(a)_k} \sum_{\substack{(b_0,\dots,b_d)\in\mathbb{N}^{d+1}\\ b_0+\cdots+b_d=k}} \prod_{i=0}^{d} \frac{(a_i)_{b_i}}{b_i!\, f_i^{b_i}} \tag{3.4}$$

$$= \frac{k!}{(a)_k} \sum_{\substack{(m_1,\dots,m_k)\in\mathbb{N}^{k}\\ m_1+2m_2+\cdots+km_k=k}} \prod_{j=1}^{k} \frac{\sigma_j^{m_j}}{j^{m_j} m_j!}. \tag{3.5}$$

Proof. Equation (3.4) is obvious from the definition of $\mathcal{B}_k(a_0,\dots,a_d)$ and (2.2). To prove (3.5), denote by $k!\,B_k/(a)_k$ and by $k!\,C_k/(a)_k$ the right-hand sides of (3.4) and (3.5), respectively. Now

$$\begin{aligned}
\sum_{k=0}^{\infty} C_k t^k &= \sum_{m_1,m_2,\dots} \prod_{j\ge 1} \frac{t^{jm_j}\sigma_j^{m_j}}{j^{m_j} m_j!} \\
&= \exp\biggl(\sum_{j\ge1} \frac{t^j \sigma_j}{j}\biggr) \\
&= \exp\biggl(\sum_{j\ge1} \frac{t^j}{j} \sum_{i=0}^{d} \frac{a_i}{f_i^j}\biggr) \\
&= \prod_{i=0}^{d} \biggl(1 - \frac{t}{f_i}\biggr)^{-a_i}.
\end{aligned}$$

We compute $\sum_{k=0}^{\infty} B_k t^k$ similarly, leading to $B_k = C_k$ and (3.5).

The remainder of this section comprises several remarks on $\mathcal{B}_k(a_0, \ldots, a_d)$, with Section 4 containing further information. If $T \subset \{0, \ldots, d\}$, denote by F_T the relative interior of E_T. This set is sometimes called a face of E_{d+1}. It is equal to the relative interior of E_{d+1} if $T = \{0, \ldots, d\}$, and the family of F_Ts is a partition of E_{d+1}. Therefore, $\mathcal{B}_k(a_0, \ldots, a_d)$ is a mixing of distributions on the faces F_T which have densities $h_{k,T}$ with respect to the uniform distribution λ_T on F_T. Here $\lambda_T = \mathcal{D}(b_0, \ldots, b_d)$, where $b_i = 1$ if $i \in T$ and $b_i = 0$ otherwise. Note that if T is reduced to the point i_0 then $\lambda_T = \delta_{e_{i_0}}$, while the relative interior of E_T is empty. Observe that if $k \le d$, only faces of dimension less than k are charged by $\mathcal{B}_k(a_0, \ldots, a_d)$. To be more specific, define $a_T = \sum_{i \in T} a_i$ and $b_T = \sum_{i \in T} b_i$. When restricted to $(a_i)_{i \in T}$, (3.1) becomes

$$\sum_{\substack{(b_i)_{i \in T} \in \mathbb{N}^T \\ b_T = k}} \prod_{i \in T} \frac{(a_i)_{b_i}}{b_i!} = \frac{(a_T)_k}{k!}.$$

A probabilistic interpretation of this is

$$\mathcal{B}_k(a_0, \ldots, a_d)\Big(\bigcup_{S \subset T} F_S\Big) = \frac{(a_T)_k}{(a)_k}.$$

Since the F_S are disjoint, for $S \subset \{0, \ldots, d\}$, the weights $w_S = \mathcal{B}_k(a_0, \ldots, a_d)(F_S)$ satisfy $\sum_{S \subset T} w_S = (a_T)_k/(a)_k$. The principle of inclusion–exclusion therefore implies that $w_T = (1/(a)_k) \sum_{S \subset T} (-1)^{|T \setminus S|} (a_S)_k$. Let us introduce the symmetric polynomial

$$P_k(a_0, \ldots, a_d) = \sum_{S \subset \{0, \ldots, d\}} (-1)^{d+1-|S|} (a_S)_k.$$

Its explicit calculation is not easy. With the convention that $P_0 = 1$, we obtain the following generating function:

$$\begin{aligned}
\sum_{k=0}^{\infty} P_k(a_0, \ldots, a_d) \frac{t^k}{k!} &= \sum_{S \subset \{0, \ldots, d\}} (-1)^{d+1-|S|} \sum_{k=0}^{\infty} (a_S)_k \frac{t^k}{k!} \\
&= \sum_{S \subset \{0, \ldots, d\}} (-1)^{d+1-|S|} (1-t)^{-a_i} \\
&= \prod_{i=0}^{d} [(1-t)^{-a_i} - 1] \\
&= t^{d+1} \prod_{i=0}^{d} \frac{(1-t)^{-a_i} - 1}{t}.
\end{aligned} \tag{3.6}$$

In particular, (3.6) shows that $P_k(a_0, \ldots, a_d) = 0$ if $k \le d$ and that

$$P_{d+1}(a_0, \ldots, a_d) = (d+1)!\, a_0 \cdots a_d.$$

With this notation we have $w_T = P_k((a_i)_{i \in T})/(a)_k$ (recall that $\sum_{T \subset \{0, \ldots, d\}} w_T = 1$). Another representation of the quasi-Bernoulli distribution as a sum of mutually singular measures is

$$\mathcal{B}_k(a_0, \ldots, a_d) = \sum_{T \subset \{0, \ldots, d\}} w_T h_{k,T} \lambda_T. \tag{3.7}$$

For simplicity, denote $h_{k,T}$ by $h_{k,d}$ in the particular case $T = \{0, \dots, d\}$. Of course, it is not zero only if $k \geq d+1$. The following proposition gives a generating function for the sequence $(P_k(a_0, \dots, a_d)h_{k,d}(x))_{k \geq d+1}$ in terms of confluent hypergeometric functions.

Proposition 3.2. *For $a, b > 0$, define*

$$ {}_1F_1(a; b; z) = \sum_{n=0}^{\infty} \frac{(a)_n}{n!\,(b)_n} z^n. $$

Then

$$ \sum_{k=d+1}^{\infty} \frac{1}{(k-1)!} P_k(a_0, \dots, a_d) h_{k,d}(x_0, \dots, x_d) t^{k-d-1} = \prod_{i=0}^{d} \frac{1}{a_i} {}_1F_1(a_i + 1; 2; x_i t). $$

Proof. Restricting (3.7) to the interior of E_{d+1} we obtain, writing $n_i = b_i - 1$ and using the definition of the Dirichlet distribution,

$$ \begin{aligned} &P_k(a_0, \dots, a_d) h_{k,d}(x) \mathbf{1}_{E_{d+1}}(x)\,\mathrm{d}x \\ &\quad = \sum_{\substack{b_i > 0 \text{ for all } i \\ \sum_{i=0}^{d} b_i = k}} \left(\prod_{i=0}^{k} \frac{(a_i)_{b_i}}{b_i!} \right) \mathcal{D}(b_0, \dots, b_d)(\mathrm{d}x) \\ &\quad = (k-1)! \sum_{\substack{n_i \geq 0 \text{ for all } i \\ \sum_{i=0}^{d} n_i = k-d-1}} \left(\prod_{i=0}^{d} \frac{(a_i)_{n_i+1}}{(n_i+1)!\,n_i!} x_i^{n_i} \right) \mathbf{1}_{E_{d+1}}(x)\,\mathrm{d}x. \end{aligned} $$

Multiplying both sides by t^{k-d-1} and summing over $k = d+1, d+2, \dots$ completes the proof.

4. Perpetuities for quasi-Bernoulli

We now compute the T_k transform of a quasi-Bernoulli distribution $\mathcal{B}_k(a_0, \dots, a_d)$, deducing from it the desired extension of Theorem 1.1.

Theorem 4.1. *Let $a_0, \dots, a_d > 0$ with $a = a_0 + \dots + a_d$, and let k be a positive integer. Suppose that $X \sim \mathcal{D}(a_0, \dots, a_d)$ and $B \sim \mathcal{B}_k(a_0, \dots, a_d)$. Then*

$$ T_k(B)(f) = \frac{T_{a+k}(X)(f)}{T_a(X)(f)}. $$

In particular, if X, B, and $Y \sim \beta(k, a)$ are independent then

$$ X \sim (1 - Y)X + YB. $$

Proof. Recall the differential operator H on U_{d+1} introduced in Theorem 2.1. Consider the function $F(f) = \prod_{i=0}^{d} f_i^{-a_i} = T_a(X)(f)$. The idea of the proof is to compute $F^{-1}H^k(F)$ in two ways. A multinomial expansion shows that

$$ H^k = k! \sum_{\substack{(b_0, \dots, b_d) \in \mathbb{N}^{d+1} \\ b_0 + \dots + b_d = k}} \prod_{i=0}^{d} \frac{(-1)^{b_i}}{b_i!} \frac{\partial^{b_i}}{\partial f_i^{b_i}}. $$

We also observe that

$$\left(\prod_{i=0}^{d}(-1)^{b_i}\frac{\partial^{b_i}}{\partial f_i^{b_i}}\right)F = F\sum_{i=0}^{d}\frac{(a_i)_{b_i}}{f_i^{a_i+b_i}}.$$

Combining these two formulae with the definition of $\mathcal{B}_k(a_0,\dots,a_d)$ we obtain $F^{-1}H^k(F) = (a)_k T_k(B)$. On the other hand, by applying (2.1) to $X \sim \mathcal{D}(a_0,\dots,a_d)$ and to $c = a$, we obtain $F^{-1}H^k(F) = (a)_k T_{a+k}(X)$. Comparing these two results yields a proof of $T_a(X)T_k(B)(f) = T_{a+k}(X)$. Applying (2.4) completes the proof.

Corollary 4.1. *It holds that* $\lim_{k\to\infty}\mathcal{B}_k(a_0,\dots,a_d) = \mathcal{D}(a_0,\dots,a_d)$*, where the limit converges in the weak sense.*

Proof. If $X \sim \mathcal{D}(a_0,\dots,a_d)$, $Y_k \sim \beta(k,a)$, and $B_k \sim \mathcal{B}_k(a_0,\dots,a_d)$ are independent, Theorem 4.1 shows that $(1-Y_k)X + Y_kB_k \sim \mathcal{D}(a_0,\dots,a_d)$. Since E_{d+1} is compact, there exists a subsequence $k_n \to \infty$ as $n \to \infty$ and a probability μ on E_{d+1} such that $\mathcal{B}_{k_n}(a_0,\dots,a_d) \to \mu$ as $n \to \infty$ in the weak sense. Furthermore, the distribution of $1 - Y_k$ converges to the Dirac mass δ_0—a quick way to see this is to consider the Mellin transform for $s > 0$:

$$\mathbb{E}((1-Y_k)^s) = \frac{\Gamma(a+s)}{\Gamma(a)}\frac{\Gamma(a+k)}{\Gamma(a+k+s)} \to 0 \quad \text{as } k \to \infty.$$

Therefore, the distribution of $(1-Y_{k_n})X + Y_{k_n}B_{k_n}$ converges weakly to μ. As a consequence, $\mu = \mathcal{D}(a_0,\dots,a_d)$ and does not depend on the particular subsequence (k_n). This completes the proof.

Theorem 4.1 implies that, for all integers k and $X \sim \mathcal{D}(a_0,\dots,a_d)$, there exists a probability distribution for B such that $T_k(B) = T_{k+a}(X)/T_a(X)$. A natural question to ask is whether this statement can be extended to positive real numbers c. More specifically, does there exist a distribution $\mathcal{B}_c(a_0,\dots,a_d)$ on E_{d+1} for B such that $T_c(B) = T_{c+a}(X)/T_a(X)$? We easily observe that this cannot be true. Taking c to be a positive number, X uniform on $(0, 1)$, and $Y \sim \beta(c, 2)$ with X and Y independent. Then

- if $0 < c < 1$, it is impossible to find a distribution for a random variable B independent of X and Y such that $X \sim (1-Y)X + YB$;
- if $c \geq 1$ and $B \sim (\delta_0+\delta_1)/c + 1 + (c-1)\mathbf{1}_{(0,1)}(b)\,\mathrm{d}b/(c+1)$ is independent of (X, Y),

$$X \sim (1-Y)X + YB.$$

More generally, we prove the following result.

Theorem 4.2. *Let c be a positive number. For a nonempty set $T \subset \{0,\dots,d\}$, we denote by λ_T the uniform probability on the convex set generated by $\{e_i;\, i \in T\}$. We also introduce the uniform probability on the union of the faces of E_{d+1} of dimension k,*

$$\Lambda_k = \frac{(k+1)!\,(d-k)!}{(d+1)!}\sum_{T\subset\{0,\dots,d\},\,|T|=k+1}\lambda_T,$$

and consider the signed measure on E_{d+1} defined by

$$\nu_{c,d} = \frac{d!\,(d+1)!}{(c+1)(c+2)\cdots(c+d)}\sum_{k=0}^{d}\frac{(c-1)(c-2)\cdots(c-k)}{k!\,(k+1)!\,(d-k)!}\Lambda_k.$$

https://doi.org/10.1239/jap/1402578633 Published online by Cambridge University Press

Then, for all $f_i > 0$, $i = 0, \dots, d$,

$$\int_{E_{d+1}} \frac{\nu_{c,\mathrm{d}}(\mathrm{d}x)}{\langle f, x\rangle^c} = f_0 \cdots f_d \int_{E_{d+1}} \frac{\Lambda_d(\mathrm{d}x)}{\langle f, x\rangle^{c+d+1}}. \tag{4.1}$$

In particular, if $Y \sim \beta(c, d+1)$ *and* $X \sim \Lambda_d$ *are independent, then there exists a random variable* B *on* E_{d+1} *independent of* (Y, X) *such that* $X \sim (1-Y)X + YB$ *if and only if either* c *is a nonnegative integer or* $c > d$*. Under these conditions,* $B \sim \nu_{c,d}$.

Remarks. Note that $\Lambda_d = \mathcal{D}(1, \dots, 1)$. Therefore, Theorem 4.2 says that the quasi-Bernoulli distribution $\nu_{c,d} = \mathcal{B}_c(1, \dots, 1)$ with continuous parameter c does exist if and only if either c is an integer or $c > d$. For $d = 2$, denote by λ_{ij} the uniform distribution on the segment e_i, e_j, and by Λ_2 the uniform distribution of the triangle with vertices e_0, e_1, e_2. Then, for $c = 1$ or $c \geq 2$,

$$\begin{aligned}\nu_{c,2} &= \mathcal{B}_c(1,1,1)\\ &= \frac{1}{(c+1)(c+2)}(2(\delta_{e_0} + \delta_{e_1} + \delta_{e_2}) + 2(c-1)(\lambda_{01} + \lambda_{02} + \lambda_{12})\\ &\qquad + (c-1)(c-2)\Lambda_2).\end{aligned}$$

The proof of Theorem 4.2 is intricate enough for us to think that the existence of $\mathcal{B}_c(a_0, \dots, a_d)$ for arbitrary positive numbers $(a_0, \dots, a_d)$ is a delicate problem, even when all the a_i are equal. To illustrate this for $d = 1$, we have to determine whether there exists a positive probability $\mu(\mathrm{d}b)$ on $[0, 1]$, depending on a_0, a_1, and c, such that, for all $(f_0, f_1) = (1, 1-t)$, we have

$$\int_0^1 \frac{\mu(\mathrm{d}b)}{(1-tb)^c} = \frac{(1-t)^{a_1}}{B(a_0, a_1)} \int_0^1 \frac{x^{a_1-1}(1-x)^{a_0-1}}{(1-tx)^{a_0+a_1+c}}\,\mathrm{d}x.$$

Application of property 5 of Theorem 2.1 shows that $\mu(\mathrm{d}b)$ necessarily has atoms at 0 and 1. As shown by Proposition 4.1 below, the mass at 0 is given by the limit

$$\lim_{t\to-\infty} \frac{(1-t)^{a_1}}{B(a_0, a_1)} \int_0^1 \frac{x^{a_1-1}(1-x)^{a_0-1}}{(1-tx)^{a_0+a_1+c}}\,\mathrm{d}x = \frac{B(a_0+c, a_1)}{B(a_0, a_1)}.$$

A similar result holds for the mass at 1. However, finding the part of $\mu(\mathrm{d}b)$ concentrated on $(0, 1)$ is challenging. One can postulate that it has a density f which therefore satisfies, in terms of the Gauss hypergeometric function ${}_2F_1$,

$$\begin{aligned}\int_0^1 \frac{f(b)\,\mathrm{d}b}{(1-tb)^c} &= (1-t)^{a_1}{}_2F_1(a_0+a_1+c, a_1; a_0+a_1, t) - \frac{B(a_0+c, a_1)}{B(a_0, a_1)}\\ &\quad - \frac{B(a_0, a_1+c)}{B(a_0, a_1)}\frac{1}{(1-t)^c}.\end{aligned}$$

If a_0 and a_1 are positive integers, one can show that f is a polynomial of degree $a_0 + a_1 - 2$ with a complicated expression.

Proof of Theorem 4.2. Since all probabilities $\Lambda_0, \dots, \Lambda_d$ are mutually singular, clearly, $\nu_{c,d}$ is a positive measure if and only if either c is an integer or $c > d$. We also observe that

$$\sum_{k=0}^{d} \frac{1}{k!\,(k+1)!\,(d-k)!}(c-1)(c-2)\cdots(c-k) = \frac{1}{d!\,(d+1)!}(c+1)(c+2)\cdots(c+d).$$

(Compute the coefficient of z^{d+1} on both sides of $(1+z)^d(1+z)^c = (1+z)^{d+c}$ to see this.) This proves that the total mass of $\nu_{c,d}$ is 1. For simplicity, define $F_c(f_0, \dots, f_d) = \int_{E_{d+1}} \Lambda_d(\mathrm{d}x)/\langle f, x\rangle^{c+d+1}$. This is a symmetric function of the f_is. We now show by induction on d that

$$F_c(f_0, \dots, f_d) = \frac{d!}{(c+1)(c+2)\cdots(c+d)} \sum_{i=0}^{d} \frac{1}{f_i^{c+1}\prod_{j\neq i}(f_j - f_i)}. \tag{4.2}$$

This holds for $d = 1$ since

$$\int_0^1 \frac{\mathrm{d}x_1}{(f_0(1-x_1)+f_1x_1)^{c+2}} = \frac{1}{(c+1)(f_0-f_1)}\left(\frac{1}{f_1^{c+1}} - \frac{1}{f_0^{c+1}}\right).$$

Assuming that (4.2) holds for $d-1$ we write (recall that T_d is the tetrahedron defined in Section 1 and that its Lebesgue measure is $1/d!$)

$$\begin{aligned}
&F_c(f_0, \dots, f_d)\\
&\quad = d! \int_{T_d} \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_d}{(f_0(1-x_1-\cdots-x_d)+f_1x_1+\cdots+f_dx_d)^{c+d+1}}\\
&\quad = d! \int_{T_{d-1}} \left(\int_0^{1-x_1-\cdots-x_{d-1}} \frac{\mathrm{d}x_d}{(f_0(1-x_1-\cdots-x_d)+f_1x_1+\cdots+f_dx_d)^{c+d+1}}\right)\\
&\qquad\quad \times \mathrm{d}x_1 \cdots \mathrm{d}x_{d-1}\\
&\quad = \frac{d}{(c+d)(f_0-f_d)}(F_c(f_1, f_2, \dots, f_d) - F_c(f_0, f_1 \dots, f_{d-1})).
\end{aligned}$$

The last equality is enough to extend (4.2) from $d-1$ to d. We now apply (4.2) to the computation of $\int_{E_{d+1}} \lambda_T(\mathrm{d}x)/(\langle f, x\rangle)^c$ when $|T| = k+1$ by changing (d, c) to $(k, c-k-1)$:

$$\int_{E_{d+1}} \frac{\lambda_T(\mathrm{d}x)}{\langle f, x\rangle^c} = \frac{k!}{(c-k)(c-k+1)\cdots(c-1)} \sum_{i\in T} \frac{1}{f_i^{c-k}\prod_{j\neq i, j\in T}(f_j - f_i)}.$$

Using this result, (4.1) can be equivalently written as

$$\sum_{\varnothing\neq T\subset\{0,\dots,d\}} \sum_{i\in T} \frac{1}{f_i^{c+1-|T|}\prod_{j\neq i, j\in T}(f_j-f_i)} = f_0\cdots f_d \sum_{i=0}^{d} \frac{1}{f_i^{c+1}\prod_{j\neq i,\, j\in T}(f_j-f_i)}. \tag{4.3}$$

Interchanging the order of the summations on the left-hand side yields

$$\sum_{i=0}^{d} \frac{1}{f_i^c} \sum_{T\ni i} \prod_{j\neq i,\, j\in T} \frac{f_i}{(f_j-f_i)} = \sum_{i=0}^{d} \frac{1}{f_i^c} \prod_{j\neq i} \frac{f_j}{(f_j-f_i)}.$$

Now we easily prove that, for all $i = 0, \dots, d$,

$$\sum_{T\ni i} \prod_{j\neq i,\, j\in T} \frac{f_i}{(f_j-f_i)} = \prod_{j\neq i} \frac{f_j}{(f_j-f_i)}. \tag{4.4}$$

To see this, it is enough to prove the $i = 0$ case. Letting $X_j = f_0/(f_j - f_0)$, equality (4.4) for $i = 0$ becomes $\sum_{T\subset\{1,\dots,d\}} \prod_{j\in T} X_j = \prod_{j=1}^{d}(1 + X_j)$, which is obviously true and proves (4.1). The remainder of the theorem straightforwardly follows.

The next proposition concerns the weights of a face and a vertex for $\mathcal{B}_c(a_0, \dots, a_d)$ when this distribution exists.

Proposition 4.1. *If $B \sim \mathcal{B}_c(a_0, \dots, a_d)$ with $a = a_0 + \cdots + a_d$ and $a' = a_{k+1} + \cdots + a_d$, then*

$$\Pr(B_0 = \cdots = B_k = 0) = \frac{\Gamma(a)\Gamma(a'+c)}{\Gamma(a+c)\Gamma(a')}, \qquad \Pr(B = e_i) = \frac{\Gamma(a)\Gamma(a_i+c)}{\Gamma(a+c)\Gamma(a_i)}.$$

Proof. By definition, $T_c(B) = T_{a+c}(X)/T_a(X)$, where $X \sim \mathcal{D}(a_0, \dots, a_d)$. We use property 5 of Theorem 2.1 and consider

$$\begin{aligned} T_c(B)(f_0, \dots, f_0, 1, \dots, 1) &= \frac{f_0^{a-a'}}{B(a_0, \dots, a_d)} \int_{T_d} \frac{x_0^{a_0-1} \cdots x_{d-1}^{a_{d-1}-1}(1 - x_0 - \cdots - x_{d-1})^{a_d-1}\, \mathrm{d}x_0 \dots \mathrm{d}x_{d-1}}{((f_0-1)(x_0 + \cdots + x_k) + 1)^{a+c}} \\ &\to \frac{EF}{B(a_0, \dots, a_d)} \quad \text{as } f_0 \to \infty, \end{aligned} \tag{4.5}$$

where

$$E = \int_{(0,\infty)^{k+1}} \frac{u_0^{a_0-1} \cdots u_k^{a_k-1}\, \mathrm{d}u_0 \cdots \mathrm{d}u_k}{(1 + u_0 + \cdots + u_k)^{a+c}} = B(a_0, \dots, a_k, a'+c) \tag{4.6}$$

and

$$\begin{aligned} F &= \int_{T_{d-k-1}} x_{k+1}^{a_{k+1}-1} \cdots x_{d-1}^{a_{d-1}-1}(1 - x_{k+1} - \cdots - x_{d-1})^{a_d-1}\, \mathrm{d}x_0 \cdots \mathrm{d}x_{d-1} \\ &= B(a_{k+1}, \dots, a_d). \end{aligned}$$

Equality (4.5) is obtained by making the change of variable $u_i = f_0 x_i$ for $i = 0, \dots, k$ and taking the limit when $f_0 \to \infty$. The second equality of (4.6) can be easily proved by letting $A = 1 + u_0 + \cdots + u_k$ in

$$\frac{1}{A^{a+c}} = \int_0^\infty \mathrm{e}^{-sA} s^{a+c-1} \frac{\mathrm{d}s}{\Gamma(a+c)}.$$

Letting $k = d-1$ so that $a' = a_d$, we obtain

$$\Pr(B = e_d) = \Pr(B_0 = \cdots = B_{d-1} = 0) = \frac{\Gamma(a)\Gamma(a_d+c)}{\Gamma(a+c)\Gamma(a_d)}.$$

The general case $\Pr(B = e_i)$ follows by symmetry.

5. Quasi-Bernoulli and Dirichlet processes

Recall that if (Ω, α) is a measure space such that $\alpha(\Omega) = a$ is finite, the Dirichlet process with parameter α is a random probability $P \sim \mathcal{D}(\alpha)$ on Ω such that, for any partition $(A_0, \dots, A_d)$ of Ω,

$$(P(A_0), \dots, P(A_d)) \sim \mathcal{D}(\alpha(A_0), \dots, \alpha(A_d)).$$

While the term ‘process’ is questionable since no idea of time is involved in this concept, it is now well ingrained in the literature; the reason is that, when Ω is the interval $[0, T]$ and α is the Lebesgue measure, the distribution function $P\{[0, t]\}$ for $0 \le t \le T$ is $Y(t)/Y(T)$, where Y is the standard gamma Lévy process. A striking property of $P \sim \mathcal{D}(\alpha)$ is that it is almost surely purely atomic: if $(V_j)_{j=1}^{\infty}$ are independent and identically distributed (i.i.d.) random variables on Ω with distribution $Q = \alpha/a$, there exists random weights $(W_j)_{j=1}^{\infty}$ (that is, $W_j \geq 0$ and $\sum_{j=1}^{\infty} W_j = 1$) such that $\sum_{j=1}^{\infty} W_j \delta_{V_j} \sim \mathcal{D}(\alpha)$. Various descriptions of the distribution of $(W_j)_{j\geq 1}$ can be found in the literature, in particular in [6], but the simplest is obtained from the data of i.i.d. $(Y_j)_{j\geq 1}$ with distribution $\beta_{1,a}$ and by taking $W_j = Y_j \prod_{k=1}^{j-1}(1 - Y_k)$. A large number of papers has followed [6]. The survey by Lijoi and Prunster [10] contains a wealth of information on the Dirichlet process $P \sim \mathcal{D}(\alpha)$ and on the distributions of the functionals $\int_{\Omega} f(w)P(\mathrm{d}w)$ with numerous references. We also mention the inspiring paper by Diaconis and Kemperman [5]. Here we describe the analogous random probability $P \sim \mathcal{B}_k(\alpha)$: for any partition $(A_0, \dots, A_d)$ of Ω,

$$(P(A_0), \dots, P(A_d)) \sim \mathcal{B}_k(\alpha(A_0), \dots, \alpha(A_d)). \tag{5.1}$$

The object $\mathcal{B}_k(\alpha)$ is natural since, for $k = 1$, the random probability $P = \delta_V$, where $V \sim \alpha/a$, satisfies (5.1). Not surprisingly, we will see that random probabilities on Ω satisfying (5.1) are concentrated on at most k points $V_1, \dots, V_k$, where $V_i \sim \alpha/a$, although they will not be independent as they are in the limiting case of the Dirichlet process. Needless to say, the distribution of the random weights on these atoms will not be simpler than in the limiting case.

Before stating the theorem for general k we sketch the construction of $\mathcal{B}_2(\alpha)$. We first select $V_1 \sim \alpha/a$. Then, with probability $1/(a+1)$, we take $V_2 = V_1$ and, with probability $a/(a+1)$, we choose V_2 independently from V_1 with distribution α/a. Finally, we take W_1 uniform on $(0, 1)$ and $W_2 = 1 - W_1$, and we consider the random probability $P = W_1\delta_{V_1} + W_2\delta_{V_2}$. To prove that (5.1) is satisfied for $k = 2$, let $a_i = \alpha(A_i)$ for simplicity; we observe that the probability of $(P(A_0), \dots, P(A_d))$ equaling e_i for $i = 0, \dots, d$ is exactly

$$\Pr(V_1, V_2 \in A_i) = \frac{1}{a+1}\frac{a_i}{a} + \frac{a}{a+1}\frac{a_i^2}{a^2} = \frac{(a_i)_2}{(a)_2}.$$

For $i \neq j$, we have

$$\Pr(V_1 \in A_i,\ V_2 \in A_j) = \frac{a}{a+1}\Pr(V_1 \in A_i)\Pr(V_2 \in A_j) = \frac{a_i a_j}{(a)_2}.$$

As a consequence, the conditional distribution of $(P(A_0), \dots, P(A_d))$, with $V_1 \in A_i$ and $V_2 \in A_j$, is equal to the law of $W_1 e_i + W_2 e_j$. These two facts show that (5.1) holds for $k = 2$.

Theorem 5.1. *Let (Ω, α) be a measure space such that $\alpha(\Omega) = a$ is finite, and let k be a positive integer. We select the random variables (M, X, W) as follows.*

1. *$M = (M_1, \dots, M_k) \in \mathbb{N}^k$ are such that $M_1 + 2M_2 + \cdots + kM_k = k$, with the Ewens' distribution with parameters k and a:*

$$\Pr((M_1, \dots, M_k) = (m_1, \dots, m_k)) = C(m)\frac{a^{\sum_{j=1}^{k} m_j}}{(a)_k}.$$

   Here

$$C(m) = C(m_1, \dots, m_k) = \frac{k!}{\prod_{j=1}^{k} j^{m_j} m_j!}.$$

   We define $S_j = M_1 + \cdots + M_j$ and $B(M) = (b_t)_{t=1}^{S_k}$, where $b_t = j$ for $S_{j-1} < t \leq S_j$.

2. *Let $(X_n)_{1\leq n\leq k}$ be a sequence of i.i.d. random variables with law α/a, independent of M. Let the conditional distribution of X given M be the same as the law of $(X_1, \dots, X_{S_k})$.*

https://doi.org/10.1239/jap/1402578633 Published online by Cambridge University Press

3. *Let* $W = (W_t)_{t=1}^{k}$ *be a random vector such that the conditional law of* W *given* M *is* $\mathcal{D}(B(M))$ *for* $(W_1, \ldots, W_{S_k})$ *and* $W_j = 0$ *for* $j \in \{S_k + 1, \ldots, k\}$.

4. *When conditioned on* M*, the random variables* X *and* W *are independent.*

Then $P = \sum_{t=1}^{S_k} W_t \delta_{X_t}$ *satisfies (5.1).*

Remarks. 1. We say that $m = (m_1, \ldots, m_k)$ is the portrait of a permutation π of $\{1, \ldots, k\}$ if π has m_j cycles of order j for $j = 1, 2, \ldots, k$. Therefore, $C(m)$ is the number of permutations with portrait m. For the history and the properties of Ewens' distribution, see, e.g. Johnson *et al.* [8, Chapter 41]. Note that $m = (m_1, \ldots, m_k)$ can be seen as the coding of a partition of the integer k. For instance, if $k = 13$ and the partition is represented by the Ferrers diagram

$$\begin{array}{ccccc} \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & & & \\ \circ & \circ & & & \\ \circ & \circ & & & \\ \circ & & & & \\ \circ & & & & \end{array}$$

corresponding to the partition $1 + 1 + 2 + 2 + 2 + 5 = 13$, then $m = (2, 3, 0, 0, 1, \ldots)$, where the dots represent a sequence of eight 0s. In this example, $\sum_{i=1}^{k} m_k$, which is the height of the Ferrers diagram, is equal to 6. The sequence $(b_t)_1^{\sum_{i=1}^{k} m_k}$ mentioned in part 1 of Theorem 5.1 is also another coding of the partition and describes the lengths of the rows of the Ferrers diagram from below. In the above example, $(b_1, b_2, b_3, b_4, b_5, b_6) = (1, 1, 2, 2, 2, 5)$.

2. Let us consider Theorem 5.1 for $k = 3$. In this case

$$\Pr(M = (3, 0, 0)) = \frac{a^3}{(a)_3}, \qquad \Pr(M = (1, 1, 0)) = \frac{3a^2}{(a)_3}, \qquad \Pr(M = (0, 0, 1)) = \frac{2a}{(a)_3},$$
$$B(3, 0, 0) = (1, 1, 1), \qquad B(1, 1, 0) = (1, 2), \qquad B(0, 0, 1) = (3).$$

Therefore,

- the conditional law of P given $M = (3, 0, 0)$ is the same as $W_1\delta_{X_1} + W_2\delta_{X_2} + W_3\delta_{X_3}$ with $(W_1, W_2, W_3) \sim \mathcal{D}(1, 1, 1)$;
- the conditional law of P given $M = (1, 1, 0)$ is the same as $W_1\delta_{X_1} + W_2\delta_{X_2}$ with $(W_1, W_2) \sim \mathcal{D}(1, 2)$;
- the conditional law of P given $M = (0, 0, 1)$ is δ_{X_1}.

3. To illustrate the notation of Theorem 5.1, let us return to the case $k = 2$. Above, we informally first took $V_1 \sim \alpha/a$, then took V_2 with a mixed distribution $\delta_{V_1}/(a + 1) + a\alpha/(a + 1)a$, and finally took $P = W_1\delta_{V_1} + W_2\delta_{V_2}$. Under the new notation, M takes two values:

- $(M_1, M_2) = (0, 1)$ with probability $1/(a + 1)$—in this case $X_1 = V_1 = V_2$, $B(0, 1) = (2)$, and $P = \delta_{X_1}$;
- $(M_1, M_2) = (2, 0)$ with probability $a/(a+1)$—in this case $B(2, 0) = (1, 1)$, the random probability P has in general two atoms X_1 and X_2 (at least when α has no atoms), and $(W_1, W_2) = (W_1, 1 - W_1) \sim \mathcal{D}(1, 1)$, that is, W_1 is uniform on $(0, 1)$.

4. If we consider the particular case where $\Omega = (0,1)$ and α is the Lebesgue measure (therefore, $a = 1$), the random probability $P \sim \mathcal{B}_k(\alpha)$ on $(0, 1)$ will be computed according to Theorem 5.1 as follows. To create M, take a permutation π of $\{1, \dots, k\}$ with uniform distribution. Consider $M = (M_1, \dots, M_k)$, where $M_1, \dots, M_k$ are the numbers of cycles of π of sizes $1, \dots, k$, respectively; the sequence M induces a partition $B(M)$ of the integer k. Take i.i.d. random variables $(X_n)_{1 \le n \le k}$ uniformly distributed on $(0, 1)$ such that they are independent of M (then $X_1, \dots, X_{M_1+\cdots+M_k}$ will be the points of discontinuity of the random distribution function $F(t) = P([0, t])$). Finally, take a Dirichlet random variable $W = (W_t)_{t=1}^{M_1+\cdots+M_k} \sim \mathcal{D}(B(M))$, where W_t is the amplitude of the jump of the random process F in X_t.

5. We observe that the idea of the T_c transform extends well to the context of random probabilities on Ω. If f is a positive measurable function on Ω, $c > 0$, and P is a random probability on Ω, we define

$$T_c(P)(f) = \mathbb{E}\bigg(\bigg(\int_\Omega f(w)P(\mathrm{d}w)\bigg)^{-c}\bigg) \le \infty,$$

which is finite in particular if there exists $m > 0$ such that $f(w) \ge m$ for all $w \in \Omega$. If $P = X \sim \mathcal{D}(\alpha)$ is a Dirichlet process such that $a = \alpha(\Omega)$ then (3.1) or [4, p. 35] shows that

$$T_a(X)(f) = \mathbb{E}\bigg(\bigg(\int_\Omega f(w)X(\mathrm{d}w)\bigg)^{-a}\bigg) = \exp\bigg(-\int_\Omega \log f(w)\alpha(\mathrm{d}w)\bigg).$$

An interesting application of Proposition 3.1 gives the following when $P = B \sim \mathcal{B}_k(\alpha)$ is the quasi-Bernoulli process of Theorem 5.1. Defining $\sigma_j(f) = \int_\Omega \alpha(\mathrm{d}w)/f(w)^j$, we have the elegant result

$$T_k(B)(f) = \mathbb{E}\bigg(\bigg(\int_\Omega f(w)B(\mathrm{d}w)\bigg)^{-k}\bigg) = \frac{k!}{(a)_k} \sum_{\substack{(m_1,\dots,m_k)\in\mathbb{N}^k \\ m_1+2m_2+\cdots+km_k=k}} \prod_{j=1}^{k} \frac{\sigma_j^{m_j}}{j^{m_j} m_j!}.$$

For instance, for $k = 2$,

$$T_2(B)(f) = \frac{1}{a(a+1)}\bigg(\int_\Omega \frac{\alpha(\mathrm{d}w)}{f(w)}\bigg)^2 + \frac{1}{a(a+1)}\int_\Omega \frac{\alpha(\mathrm{d}w)}{f(w)^2}.$$

6. Needless to say, the formula which is the backbone of the paper, namely,

$$T_k(B)(f) = \frac{T_{a+k}(X)(f)}{T_a(X)(f)},$$

still holds for a Dirichlet process $X \sim \mathcal{D}(\alpha)$ and a quasi-Bernoulli process $B \sim \mathcal{B}_k(\alpha)$. Consequently, $X \sim (1 - Y)X + YB$ holds when $Y \sim \beta(k, a)$ is independent of B.

Proof of Theorem 5.1. To show (5.1), we let $a_i = \alpha(A_i)$. We compute first the distribution of $Z = (P(A_0), \dots, P(A_d))$ by conditioning with respect to M and X. Denote by $N_{i,j}$ the number of X_t such that $S_{j-1} < t \le S_j$ and $X_t \in A_i$. Note that $\sum_{i=0}^{d} N_{i,j} = M_j$. Conditionally on M, the vector $N_j = (N_{i,j})_{i=0}^{d}$ of $\mathbb{R}^{d+1}$ has a multinomial distribution

$$\Pr(N_j = (n_{0,j}, \dots, n_{d,j})) = \frac{M_j!}{n_{0,j}! \cdots n_{d,j}!} \frac{a_0^{n_{0,j}} \cdots a_d^{n_{d,j}}}{a^{M_j}},$$

where $n_{0,j} + \cdots + n_{d,j} = M_j$. Furthermore, $N_1, \dots, N_k$ are conditionally independent given M. Now we introduce the quantities $B_i = \sum_{j=1}^k jN_{i,j}$, which satisfy $\sum_{i=0}^d B_i = \sum_{i=0}^d \sum_{j=1}^k jN_{i,j} = \sum_{j=1}^k jM_j = k$. Observe that, conditionally on M and X, we have $Z \sim \mathcal{D}(B_0, \dots, B_d)$. To see this, we use the following definition of Z:

$$\begin{aligned} Z &= \Big(\sum_t W_t \mathbf{1}_{\{t;\, X_t \in A_0\}}, \dots, \sum_t W_t \mathbf{1}_{\{t;\, X_t \in A_d\}}\Big) \\ &= \Big(\sum_{j=1}^k \sum_{S_{j-1}<t\le S_j} W_t \mathbf{1}_{\{t;\, X_t \in A_0\}}, \dots, \sum_{j=1}^k \sum_{S_{j-1}<t\le S_j} W_t \mathbf{1}_{\{t;\, X_t \in A_d\}}\Big). \end{aligned}$$

A property of the Dirichlet distribution is that if $b_i = \sum_{j=1}^{k_i} a_{ij}$ with $a_{ij} \ge 0$ and $i = 0, \dots, d$,

$$(X_{ij})_{0\le i\le d,\, 1\le j\le k_i} \sim \mathcal{D}((a_{ij})_{0\le i\le d,\, 1\le j\le k_i}),$$

and $Y_i = \sum_{j=1}^{k_i} X_{ij}$, then $(Y_0, \dots, Y_d) \sim \mathcal{D}(b_0, \dots, b_d)$. A quick way to see this is to use (2.2). We apply this principle to $(X_{ij}) = (W_t)$, to $k_i = k$, to $a_t = a_{ij} = j$ when $\sum_{S_{j-1}<t\le S_j}$, and to $Y_i = P(A_i)$. We obtain

$$Z \sim \mathcal{D}\Big(\sum_{j=1}^k jN_{0,j}, \dots, \sum_{j=1}^k jN_{d,j}\Big) = \mathcal{D}(B_0, \dots, B_d).$$

The last step of the proof removes the conditioning on X and M. We have

$$\begin{aligned} \mathbb{E}\Big(\frac{1}{\langle f, Z\rangle^k}\Big) &= \mathbb{E}\Big(\mathbb{E}\Big(\frac{1}{\langle f, Z\rangle^k} \,\Big|\, M, X\Big)\Big) \\ &= \mathbb{E}\Big(\mathbb{E}\Big(\frac{1}{f_0^{B_0}\cdots f_d^{B_d}} \,\Big|\, M, X\Big)\Big) \\ &= \mathbb{E}\Big(\mathbb{E}\Big(\frac{1}{\prod_{j=1}^k f_0^{jN_{0,j}}\cdots f_d^{jN_{d,j}}} \,\Big|\, M, X\Big)\Big) \\ &= \mathbb{E}\Big(\prod_{j=1}^k \mathbb{E}\Big(\frac{1}{f_0^{jN_{0,j}}\cdots f_d^{jN_{d,j}}} \,\Big|\, M, X\Big)\Big) \\ &= \mathbb{E}\Big(\prod_{j=1}^k \frac{1}{a^{M_j}} \sigma_j^{M_j}\Big) \\ &= \frac{k!}{(a)_k} \sum_{\substack{(m_1,\dots,m_k)\in\mathbb{N}^k \\ m_1+2m_2+\cdots+km_k=k}} \prod_{j=1}^k \frac{\sigma_j^{m_j}}{j^{m_j} m_j!}, \end{aligned}$$

where in the last equality we have used the notation σ_j introduced in (3.3). The second equality follows from $Z \sim \mathcal{D}(B_0, \dots, B_d)$ when conditioned on (M, X) and from (2.2); the third equality follows from the definition of $B_0, \dots, B_d$; the fourth equality follows from the independence of the N_j; and the last equality follows from the generating function of a multinomial distribution. Equation (3.5) proves that $Z \sim \mathcal{B}_k(a_0, \dots, a_d)$.

6. A Markov chain on the tetrahedron

In this section we give an application of Theorems 1.1 and 4.1; it does not contain new results and serves as a conclusion. Consider the homogeneous Markov chain $(X(n))_{n\geq 0}$ valued in the convex hull E_{d+1} of $(e_0, \dots, e_d)$ with the following transition process. Given $X(n) \in E_{d+1}$, randomly choose a point $B(n+1) \in E_{d+1}$ such that $B(n+1) \sim \mathcal{B}_k(a_0, \dots, a_d)$ and independently a random number $Y_{n+1} \sim \beta(k, a)$. Now draw the segment $(X(n), B(n+1))$ and take the point

$$X(n+1) = X(n)(1 - Y_{n+1}) + B(n)Y_{n+1}$$

on this segment. Theorem 4.1 says that the Dirichlet distribution $\mathcal{D}(a_0, \dots, a_d)$ is a stationary distribution for the Markov chain $(X(n))_{n\geq 0}$. Recall the following principle (see [3, Proposition 1]).

Theorem 6.1. *If E is a metric space and C is the set of continuous maps $f: E \to E$, fix a probability $\nu(\mathrm{d}f)$ on C. Let $F_1, F_2, \dots$ be a sequence of independent random variables on C with the same distribution ν. Let $W_n = F_n \circ \cdots \circ F_2 \circ F_1$ and $Z_n = F_1 \circ \cdots \circ F_{n-1} \circ F_n$. Suppose that $Z = \lim_n Z_n(x)$ almost surely exists in E and does not depend on $x \in E$. Then the following assertions hold.*

1. *The distribution μ of Z is a stationary distribution of the Markov chain $(W_n(x))_{n\geq 0}$ on E.*

2. *If X and F_1 are independent and $X \stackrel{\mathrm{D}}{=} F_1(X)$, then $X \sim \mu$.*

Choose $E = E_{d+1}$. Apply Theorem 6.1 to the distribution ν of the random map F_1 on E_{d+1} defined by $F_1(x) = (1 - Y_1)x + Y_1 B(1)$, where $Y_1 \sim \beta(k, a)$ and $B(1) \sim \mathcal{B}_k(a_0, \dots, a_d)$ are independent. If the F_n defined by $F_n(x) = (1 - Y_n)x + Y_n B(n)$ are independent with distribution ν, clearly,

$$Z_n(x) = \left(\prod_{j=1}^{n}(1 - Y_j)\right)x + \sum_{k=1}^{n}\left(\prod_{j=1}^{k-1}(1 - Y_j)\right)Y_k B(k)$$

converges almost surely to the sum of the converging series

$$Z = \sum_{k=1}^{\infty}\left(\prod_{j=1}^{k-1}(1 - Y_j)\right)Y_k B(k), \tag{6.1}$$

and, therefore, the hypotheses of Theorem 6.1 are met. As a consequence, the Dirichlet law $\mathcal{D}(a_0, \dots, a_d)$ is the unique stationary distribution of the Markov chain $(X(n))_{n\geq 0}$ and is the distribution of Z defined by (6.1). Finally, recall the definition of a perpetuity [7] on an affine space $\mathcal{A}$. Let $\nu(\mathrm{d}f)$ be a probability on the space of affine transformations f mapping $\mathcal{A}$ into itself. We say that the probability μ on $\mathcal{A}$ is a perpetuity generated by ν if $X \stackrel{\mathrm{D}}{=} F(X)$ when $F \sim \nu$ and $X \sim \mu$ are independent. If the conditions of Theorem 6.1 are met for ν, there is exactly one perpetuity generated by ν. Theorems 1.1, 4.1, and 4.2 say that the Dirichlet distribution is a perpetuity for the random affine map $F(x) = (1 - Y)x + YB$ on the affine hyperplane $\mathcal{A}$ of $\mathbb{R}^{d+1}$ containing $e_0, \dots, e_d$ generated by various distributions ν of $(1-Y, YB)$. Theorem 6.1 shows that a Dirichlet process is also a perpetuity generated by the distribution of $(1 - Y, YB)$, where the set of probabilities on Ω replaces the tetrahedra with $d + 1$ vertices and where $Y \sim \beta(k, a)$ is independent of the quasi-Bernoulli process $\mathcal{B}_k(\alpha)$.

Acknowledgements

The authors are grateful to the anonymous referee for a very careful reading, as well as Persi Diaconis. We thank both for bringing the essential reference [11] to our attention.

References

[1] AMBRUS, G., KEVEI, P. AND VÍGH, V. (2012). The diminishing segment process. *Statist. Prob. Lett.* **82,** 191–195.
[2] ARIZMENDI, O. AND PÉREZ-ABREU, V. (2010). On the non-classical infinite divisibility of power semicircle distributions. *Commun. Stoch. Anal.* **4,** 161–178.
[3] CHAMAYOU, J.-F. AND LETAC, G. (1991). Explicit stationary distributions for compositions of random functions and products of random matrices. *J. Theoret. Prob.* **4,** 3–36
[4] CHAMAYOU, J.-F. AND LETAC, G. (1994). A transient random walk on stochastic matrices with Dirichlet distributions. *Ann. Prob.* **22,** 424–430.
[5] DIACONIS, P. AND KEMPERMAN, J. (1996). Some new tools for Dirichlet priors. In *Bayesian Statistics 5*, Oxford University Press, pp. 97–106.
[6] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1,** 209–230.
[7] GOLDIE, C. M. AND MALLER, R. A. (2000). Stability of perpetuities. *Ann. Prob.* **28,** 1195–1218.
[8] KOTZ, S., BALAKRISHNAN, N. AND JOHNSON, N. L. (2000). *Continuous Multivariate Distributions*, Vol. 1, 2nd edn. John Wiley, New York.
[9] LETAC, G. AND MASSAM, H. (1998). A formula on multivariate Dirichlet distribution. *Statist. Prob. Lett.* **38,** 247–253.
[10] LIJOI, A. AND PRÜNSTER, I. (2009). Distributional properties of means of random probability measures. *Statist. Surveys* **3,** 47–95.
[11] SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4,** 639–650.