Elect. Comm. in Probab. **15** (2010), 79–88

ELECTRONIC COMMUNICATIONS in PROBABILITY

A SPECIES SAMPLING MODEL WITH FINITELY MANY TYPES

ALEXANDER GNEDIN
Department of Mathematics, Utrecht University, PO Box 80010, 3508 TA Utrecht, The Netherlands
email: A.V.Gnedin@uu.nl

Submitted 14 November 2009, *accepted in final form* 6 February 2010

AMS 2000 Subject classification: 60G09, 60C05
Keywords: exchangeability, Gibbs partition, succession rule

Abstract

A two-parameter family of exchangeable partitions with a simple updating rule is introduced. The partition is identified with a randomized version of a standard symmetric Dirichlet species-sampling model with finitely many types. A power-like distribution for the number of types is derived.

1 Introduction

The Ewens-Pitman two-parameter family of exchangeable partitions $\Pi^{\alpha,\theta}$ of an infinite set has become a central model for species sampling (see [7, 18] for extensive background on exchangeability and properties of these partitions). One most attractive feature of this model is the following explicit rule of succession, which we formulate as sequential allocation of balls labelled $1, 2, \dots$ in a series of boxes. Start with box $B_{1,1}$ with a single ball 1. At step n the allocation of n balls is a certain random partition $\Pi_n^{\alpha,\theta}$ of the set of balls $[n] := \{1, \dots, n\}$ into some number K_n of nonempty boxes $B_{n,1}, \dots, B_{n,K_n}$, which we identify with their contents, and list the boxes by increase of the minimal labels of balls (which means that $\min([n] \setminus (\cup_{i=1}^{j-1} B_{n,i})) \in B_{n,j}$ for $1 \le j \le K_n$). Given at step n the number of occupied boxes is $K_n = k$, and the occupancy counts are $\#B_{n,j} = n_j$ for $1 \le j \le k$ (so $n_1 + \dots + n_k = n$), the partition $\Pi_{n+1}^{\alpha,\theta}$ of $[n+1]$ at step $n+1$ is obtained by randomly placing ball $n+1$ according to the rules

($O^{\alpha,\theta}$) : in an *old* box $B_{n+1,j} := B_{n,j} \cup \{n+1\}$ with probability

$$\omega_{n,j}^{\alpha,\theta}(k; n_1, \dots, n_k) := \frac{n_j - \alpha}{n + \theta}$$

($N^{\alpha,\theta}$) : in a *new* box $B_{n+1,k+1} := \{n+1\}$ with probability

$$\nu_n^{\alpha,\theta}(k; n_1, \dots, n_k) := \frac{\theta + k\alpha}{n + \theta}.$$

For instance, if at step $n=6$ the partition is $\{1,3\},\{2,5,6\},\{4\}$, then ball 7 is added to one of the old boxes $\{1,3\}$ or $\{2,5,6\}$ or $\{4\}$ with probabilities specified by ($\mathrm{O}^{\alpha,\theta}$), and a new box $\{7\}$ is created according to ($\mathrm{N}^{\alpha,\theta}$).
Eventually, as all balls $1,2,\dots$ get allocated in boxes $B_j := \cup_{n\in\mathbb{N}} B_{n,j}$, the collection of occupied boxes is almost surely infinite if the parameters (α,θ) are in the range $\{(\alpha,\theta): 0\le\alpha<1, \theta>-\alpha\}$. In contrast to that, the collection of occupied boxed has finite cardinality $\varkappa$ if $\alpha<0$ and $-\theta/\alpha=\varkappa$; this model, which is most relevant to the present study, has a long history going back to Fisher [4].
Exchangeable *Gibbs partitions* [18, 9] extend the Ewens-Pitman family[1]. The first rule is preserved in the sense that, given ball $n+1$ is placed in one of the old boxes, it is placed in box j with probability $\omega_{n,j}$ still proportional to $n_j-\alpha$, where $\alpha<1$ is a fixed *genus* of the partition. But the second rule allows more general functions ν_n of n and k, which agree with the first rule and the exchangeability of partition. Examples of Gibbs partitions of genus $\alpha\in(0,1)$ were studied in [9, 13, 14]; from the results of these papers one can extract complicated formulas expressing ν_n in terms of special functions. See [15] for a survey of related topics and applications of random partitions to the Bayesian nonparametric inference.
The practitioner willing to adopt a partition from the Ewens-Pitman family as a species-sampling model faces the dilemma: the total number of boxes is either a fixed finite number (Fisher's subfamily) or it is infinite. For applications it is desirable to also have tractable exchangeable partitions of $\mathbb{N}$ with finite but random number of boxes K. The present note suggests a two-parameter family of partitions of the latter kind, which are Gibbs partitions obtained by suitable mixing of Fisher's $\Pi^{-1,\theta}$-partitions over θ, where the randomized θ (which will be re-denoted $\varkappa:=\theta$) actually coincides with K. Equivalently, the partition can be generated by sampling from a mixture of symmetric Dirichlet random measures with unit weights on $\varkappa$ points.

2 Construction of the partition

A new allocation rule is as follows. Start with box $B_{1,1}$ containing a single ball 1. At step n the allocation of n balls is a certain random partition $\Pi_n=(B_{n,1},\dots,B_{n,K_n})$ of the set of balls $[n]$. Given the number of boxes is $K_n=k$, and the occupancy counts are $\#B_{n,j}=n_j$ for $1\le j\le k$, the partition of $[n+1]$ at step $n+1$ is obtained by randomly placing ball $n+1$

(O) : in an *old* box $B_{n+1,j}:=B_{n,j}\cup\{n+1\}$ with probability

$$\omega_{n,j}(k;n_1,\dots,n_k):=\frac{(n_j+1)(n-k+\gamma)}{n^2+\gamma n+\zeta},\quad j=1,\dots,k,$$

(N) : in a *new* box $B_{n+1,k+1}:=\{n+1\}$ with probability

$$\nu_n(k;n_1,\dots,n_k):=\frac{k^2-\gamma k+\zeta}{n^2+\gamma n+\zeta}.$$

To agree with the rules of probability the parameters γ and ζ must be chosen so that $\gamma\ge 0$ and (i) either $k^2-\gamma k+\zeta$ is (strictly) positive for all $k\in\mathbb{N}$, or (ii) the quadratic is positive for $k\in\{1,\dots,k_0-1\}$ and has a root at k_0. In the case (ii) the number of occupied boxes never

[1]Here we are only interested in infinite exchangeable Gibbs partitions. Finite Gibbs partitions of $[n]$ were discussed in [18, 2], but these are typically not consistent as n varies.

exceeds k_0. Part (O) is similar to the ($O^{\alpha,\theta}$)-prescription with $\alpha = -1$: given ball $n+1$ is placed in one of the old boxes, it is placed in box j with probability proportional to $n_j + 1$. But part (N) is radically different from ($N^{\alpha,\theta}$) in that the probability of creating a new box is a ratio of quadratic polynomials in k and n.

The probability of every particular partition $B_{n,1}, \dots, B_{n,K_n}$ with $K_n = k$ boxes containing $n_1, \dots, n_k$ balls is easily calculated as

$$p(n_1, \dots, n_k) = \frac{(\gamma)_{n-k} \prod_{i=1}^{k-1}(i^2 - \gamma i + \zeta)}{\prod_{m=1}^{n-1}(m^2 + \gamma m + \zeta)} \prod_{j=1}^{k} n_j! \tag{1}$$

(with $(a)_m := a(a+1)\dots(a+m-1)$), where $(n_1, \dots, n_k)$ is an arbitrary composition of integer n, that is a vector of some length $k \in \mathbb{N}$ whose components $n_j \in \mathbb{N}$ satisfy $\sum_{j=1}^{k} n_j = n$. The function p is sometimes called *exchangeable partition probability function* (EPPF) [18]. For instance, the probability that the set of balls [6] is allocated after completing step 6 in three boxes $\{1,3\}, \{2,5,6\}, \{4\}$ is equal to $p(2,3,1)$. Formula (1) is a familiar *Gibbs form* of exchangeable partition of genus $\alpha = -1$ (see [9, 18] and Section 8). An exchangeable partition Π of the infinite set of balls $\mathbb{N}$ is defined as the allocation of balls in boxes $B_j := \cup_{n \in \mathbb{N}} B_{n,j}$, with the convention $B_j = \varnothing$ in the event $K_n < j$ for all n. For $\gamma = 0$ the partition Π has only singleton boxes.

The formula for

$$v_{n,k} := \frac{(\gamma)_{n-k} \prod_{i=1}^{k-1}(i^2 - \gamma i + \zeta)}{\prod_{m=1}^{n-1}(m^2 + \gamma m + \zeta)} \tag{2}$$

can be fully split in linear factors as

$$v_{n,k} = \frac{(\gamma)_{n-k}(s_1+1)_{k-1}(s_2+1)_{k-1}}{(z_1+1)_{n-1}(z_2+1)_{n-1}},$$

by factoring the quadratics as

$$x^2 + \gamma x + \zeta = (x+z_1)(x+z_2), \qquad x^2 - \gamma x + \zeta = (x+s_1)(x+s_2),$$

for some complex z_1, z_2, s_1, s_2.

Using exchangeability and applying Equation (20) from [16], the total probability that the occupancy counts are $(n_1, \dots, n_k)$ equals

$$\begin{aligned} &\mathbb{P}(K_n = k, \#B_{n,1} = n_1, \dots, \#B_{n,K_n} = n_k) = \\ &\frac{n!}{\prod_{j=1}^{k}\{(n_j + \dots + n_k)(n_j - 1)!\}}\, p(n_1, \dots, n_k) = \\ &v_{n,k}\, n! \prod_{j=1}^{k} \frac{n_j}{n_j + \dots + n_k} \end{aligned}$$

Let $K_{n,r} = \#\{1 \le j \le K_n : \#B_{n,j} = r\}$ be the number of boxes occupied by exactly r out of n balls. By standard counting arguments, the last formula can be re-written as

$$\mathbb{P}(K_{n,r} = k_r,\ r = 1, \dots, n) = v_{n,k}\, n! \prod_{r=1}^{n} \frac{1}{k_r!}$$

for arbitrary integer vector of multiplicities $(k_1, \dots, k_n)$, with $k_r \ge 0$, $\sum_{r=1}^{n} k_r = k$ and $\sum_{r=1}^{n} r k_r = n$.

3 Mixture representation and the number of occupied boxes

Like for any Gibbs partition of genus -1, the number of occupied boxes K_n is a sufficient statistic for the finite partition Π_n, meaning that conditionally given $K_n = k$ the probability of each particular value of Π_n with occupancy counts $n_1, \dots, n_k$ equals

$$\frac{\prod_{j=1}^{k} n_j!}{d_{n,k}},$$

where the normalization constant is a Lah number [3]

$$d_{n,k} = \binom{n-1}{k-1} \frac{n!}{k!}. \tag{3}$$

The sequence $(K_n, n = 1, 2, \dots)$ is a nondecreasing Markov chain with $0-1$ increments and transition probabilities determined by the rule (N). The distribution of K_n is calculated as

$$\mathbb{P}(K_n = k) = d_{n,k} v_{n,k}. \tag{4}$$

By monotonicity, the limit $K := \lim_{n\to\infty} K_n$ exists almost surely, and coincides with the number of nonempty boxes for the infinite partition Π. Letting $n \to \infty$ in (4) and using the standard asymptotics $\Gamma(n+a)/\Gamma(n+b) \sim n^{a-b}$ we derive from (2), (3)

$$\mathbb{P}(K = \varkappa) = \frac{\Gamma(z_1+1)\Gamma(z_2+1)}{\Gamma(\gamma)} \frac{\prod_{i=1}^{\varkappa-1}(i^2 - \gamma i + \zeta)}{\varkappa!(\varkappa-1)!}. \tag{5}$$

The basic structural result about Π is the following:

Theorem 1. *Partition Π is a mixture of partitions $\Pi^{-1,\varkappa}$ over the parameter $\varkappa$, with a proper mixing distribution given by* (5).

Proof. As every other Gibbs partition of genus -1, partition Π satisfies the conditioning relation

$$\Pi \,|\, \{K = \varkappa\} \;\stackrel{d}{=} \Pi^{-1,\varkappa}, \tag{6}$$

which says that given K the partition has the same distribution as some Fisher's partition of genus -1. We only need to verify that the weights in (5) add up to the unity.
To avoid calculus, note that by the general theory [9] Π is a mixture of the $\Pi^{-1,\varkappa}$'s with $\varkappa \in \mathbb{N}$, and the trivial singleton partition. Because every $\Pi^{-1,\varkappa}$ has $\varkappa$ boxes, the probability $\mathbb{P}(K = \infty)$ is equal to the weight of the singleton component in the mixture. But for the singleton partition of $[n]$ the number of boxes is equal to the number of balls, thus it remains to check that $\mathbb{P}(K_n = n) \to 0$ as $n \to \infty$, which is easily done by inspection of the transition rule (N). □

As $\varkappa \to \infty$, the masses (5) exhibit a power-like decay,

$$\mathbb{P}(K = \varkappa) \sim \frac{c}{\varkappa^{\gamma+1}} \quad \text{with} \quad c = \frac{\Gamma(z_1+1)\Gamma(z_2+1)}{\Gamma(\gamma)\Gamma(s_1+1)\Gamma(s_2+2)}.$$

This explains, to an extent, the role of parameter γ. In particular, $\mathbb{E}K$ may be finite or infinite, depending on whether $\gamma > 1$, or $\gamma \le 1$.

4 Frequencies

Recall some standard facts about the partition $\Pi^{-1,\varkappa}$ (see [18]). This partition with $\varkappa$ boxes $B_1,\dots,B_\varkappa$ can be generated by the following steps:

(b) choose a value $(y_1,\dots,y_\varkappa)$ for the probability vector $(P_{\varkappa,1},\dots,P_{\varkappa,\varkappa})$ uniformly distributed on the $(\varkappa-1)$-simplex $\{(y_1,\dots,y_\varkappa): y_i>0, \sum_{i=1}^{\varkappa} y_i = 1\}$,

(c) allocate balls $1,2,\dots$ independently in $\varkappa$ boxes with probabilities $y_1,\dots,y_\varkappa$ of placing a ball in each of these boxes,

(d) arrange the boxes by increase of the smallest labels of balls.

The vector of *frequencies* $(\widetilde{P}_{\varkappa,1},\dots,\widetilde{P}_{\varkappa,\varkappa})$, defined through limit proportions

$$\widetilde{P}_{\varkappa,j} := \lim_{n\to\infty} \frac{\#(B_j\cap[n])}{n}, \quad 1\le j\le\varkappa, \tag{7}$$

has the same distribution as the *size-biased permutation* of $(P_{\varkappa,1},\dots,P_{\varkappa,\varkappa})$. The frequencies have a convenient stick-breaking representation

$$\widetilde{P}_{\varkappa,j} = W_j\prod_{i=1}^{j-1}(1-W_i), \quad \text{with independent } W_i \stackrel{d}{=} \text{beta}(2,\varkappa-i), \tag{8}$$

where $i=1,\dots,\varkappa$ and beta$(2,0)$ is a Dirac mass at 1. See [7] for characterizations of this and other Ewens-Pitman partitions through independence of factors in such a stick-breaking representation.

Now let us apply the above to the partition Π. The mixture representation in Theorem 1 implies that Π can be constructed by first

(a) choosing a value $\varkappa$ for K from distribution (5),

then following the above steps (b), (c) and (d). The frequencies $(\widetilde{P}_1,\dots,\widetilde{P}_K)$ of nonempty boxes $B_1,\dots,B_K$ are obtainable from (8) by mixing with weights (5).

5 Exchangeable sequences

Let $(S,\mathscr{B},\mu)$ be a Polish space with a nonatomic probability measure μ. Let Π be the partition of $\mathbb{N}$ constructed above and $T_1,T_2,\dots$ be an i.i.d. sample from $(S,\mathscr{B},\mu)$, also independent of Π. With these random objects one naturally associates an infinite exchangeable S-valued sequence $X_1,X_2,\dots$ with marginal distributions μ, as follows (see [1, 12]). Attach to every ball in box B_j the same tag T_j, for $j=1,\dots,K$. Then define $X_1,X_2,\dots$ to be the sequence of tags of balls $1,2,\dots$. Obviously, $K,T_1,\dots,T_K$ and Π can be recovered from $X_1,X_2,\dots$. Indeed, T_j is the jth distinct value in the sequence $X_1,X_2,\dots$ and $B_j=\{n: X_n=T_j\}$ for $j=1,\dots,K$. The same applies to finite partitions Π_n with $B_{n,j}=B_j\cap[n]$.

The *prediction rule* [12] associated with $X_1,X_2,\dots$ is the formula for conditional distribution

$$\mathbb{P}(X_{n+1}\in ds\,|\,X_1,\dots,X_n) = \sum_{j=1}^{K_n}\omega_{n,j}\delta_{T_j}(ds)+\nu_n\mu(ds),$$

where $T_1, \dots, T_{K_n}$ are the distinct values in $X_1, \dots, X_n$, and $\omega_{n,j}, v_n$ are the functions of the partition Π_n, as specified by the rules (O) and (N).
The random measure F in de Finetti's representation of $X_1, X_2, \dots$ is a mixture

$$F = \sum_{\varkappa=1}^{\infty} \mathbb{P}(K = \varkappa) F_\varkappa ,$$

where

$$F_\varkappa(ds) = \sum_{j=1}^{\varkappa} P_{\varkappa,j} \delta_{\hat{T}_j}(ds), \quad \varkappa \in \mathbb{N}$$

are Dirichlet$(\underbrace{1, \dots, 1}_{\varkappa})$ random measures on $(S, \mathscr{B}, \mu)$, that is the vector $(P_{\varkappa,1}, \dots, P_{\varkappa,\varkappa})$ is uniformly distributed on the $(\varkappa - 1)$-simplex and is independent of $(\hat{T}_1, \hat{T}_2, \dots)$, and the random variables $\hat{T}_j$'s are i.i.d.(μ).

6 The case $\zeta = 0$

We focus now on the case $\zeta = 0$. Then $\gamma \in [0, 1]$ is the admissible range, but we shall exclude the trivial edge cases $\gamma = 0$, respectively, $\gamma = 1$ of the singleton and single-box partitions.
Formula (2) simplifies as

$$v_{n,k} = \frac{(k-1)!(1-\gamma)_{k-1}(\gamma)_{n-k}}{(n-1)!(1+\gamma)_{n-1}},$$

and there is a further obvious cancellation of some factors. Furthermore, (5) specializes as

$$\mathbb{P}(K = \varkappa) = \frac{\gamma(1-\gamma)_{\varkappa-1}}{\varkappa!}, \quad \varkappa = 1, 2, \dots \tag{9}$$

which is a distribution familiar from the discrete renewal theory (a summary is found in [17], p. 85). The distribution has also appeared in connection with $\Pi^{\alpha,\theta}$ $(0 < \alpha < 1)$ partitions and other occupancy problems [8, 6, 10, 17].
Thinking of (9) as a prior distribution for K, the posterior distribution is found from (4), (9) and the distribution of the number of occupied boxes for the $\Pi^{-1,k}$ partition (instance of Equation (3.11) in [18]):

$$\mathbb{P}(K = \varkappa \,|\, K_n = k) = \frac{(n-1)!}{(k-1)!(\varkappa+n-1)!} \prod_{i=1}^{k-1}(\varkappa - i) \prod_{j=1}^{k}(\gamma + n - j) \prod_{l=k}^{\varkappa-1}(l - \gamma), \tag{10}$$

for $1 \le k \le n, \quad \varkappa \ge k$. Note that the conditioning here can be replaced by conditioning on an arbitrary value of the partition Π_n with k boxes.
The frequency $\widetilde{P}_1$ of box B_1 has distribution

$$\mathbb{P}(\widetilde{P}_1 \in dy) = \sum_{\varkappa=1}^{\infty} \frac{\gamma(1-\gamma)_{\varkappa-1}}{\varkappa!} \, \mathbb{P}(\widetilde{P}_{\varkappa,1} \in dy) = \gamma \delta_1(dy) + (1-\gamma)\gamma y^{\gamma-1} dy, \quad y \in (0, 1],$$

which is a mixture of Dirac mass at 1 and beta$(\gamma, 1)$ density. Interestingly, distributions of this kind have appeared in connection with other partition-valued processes [8, 11]. The distribution is useful to compute expected values of symmetric statistics of the frequencies of the kind $\sum_{j=1}^{K} f(\widetilde{P}_j)$ [18], for example

$$\mathbb{E}\left(\sum_{j=1}^{K}\widetilde{P}_j^n\right)=\mathbb{E}\left(\widetilde{P}_1^{n-1}\right)=\frac{n\gamma}{n+\gamma-1},$$

which agrees with the $\mathbb{P}(K_n=1)$ instance of (4).

7 Restricted exchangeability

It is of interest to explore a more general situation when the process starts with some initial allocation of a few balls in boxes. This can be thought of as prior information of the observer about the existing species. For simplicity we shall only consider the case $\zeta=0$.

Fix $m\geq 1$ and a partition $\mathbf{b}=(b_1,\dots,b_k)$ of $[m]$ with k positive box-sizes $\#b_j=m_j,\ j=1,\dots,k$. Let $\mathbb{P}_{\mathbf{b}}$ be the law of the infinite partition Π constructed by the rules (O) and (N) starting with the initial allocation of balls $\Pi_m=\mathbf{b}$. In particular, $\mathbb{P}=\mathbb{P}_{\{1\}}$. Note that $\mathbb{P}_{\mathbf{b}}$ is well defined for any value of the parameter in the range

$$-(m-k)<\gamma<k,$$

and for $\gamma\in(0,1)$ the measure $\Pi_{\{1\}}$, conditioned on $\{\Pi_m=\mathbf{b}\}$, coincides with $\mathbb{P}_{\mathbf{b}}$. Explicitly, under $\mathbb{P}_{\mathbf{b}}$ every value of $\Pi_n=(B_{n,1},\dots,B_{n,K_n})$ with

$$K_n=\varkappa\geq k,\ \ \#B_j=n_j,\ j=1,\dots,\varkappa;\quad n_i\geq m_i,\ i=1,\dots,k$$

has probability

$$p_{\mathbf{b}}(n_1,\dots,n_\varkappa):=\frac{p(n_1,\dots,n_\varkappa)}{p(m_1,\dots,m_k)},$$

where p is given by (1). Formula (9) for the terminal distribution of the number of boxes is still valid for the extended range of γ.

Observing that $p_{\mathbf{b}}$ is symmetric in the arguments n_j for $k\leq j\leq\varkappa$, it follows that $\mathbb{P}_{\mathbf{b}}$ is invariant under permutations of the set $\mathbb{N}\setminus[m]$. On the other hand, for every permutation $\sigma:[m]\to[m]$ we have $\mathbb{P}_{\sigma b}=\sigma\mathbb{P}_b$. Moreover, the restriction of Π on $\mathbb{N}\setminus[m]$ under $\mathbb{P}_{\mathbf{b}}$ has the same law as under $\mathbb{P}_{\sigma\mathbf{b}}$, that is the restriction depends on $\mathbf{b}$ only through $(m_1,\dots,m_k)$.

Examples Suppose $\gamma=1$. Then $\mathbb{P}_{\{1\}}(K=1)=1$ which corresponds to the trivial one-block partition, but $\mathbb{P}_{\{1\},\{2\}}(K=\varkappa)=\frac{2}{\varkappa(\varkappa+1)}$ for $\varkappa\geq 2$.

Suppose $\gamma=0$. Then Π under $\mathbb{P}_{\{1\}}$ is the trivial singleton partition, but under $\mathbb{P}_{\{1,2\}}$ we have $\mathbb{P}_{\{1,2\}}(K=\varkappa)=\frac{1}{\varkappa(\varkappa+1)}$ for $\varkappa\geq 1$.

8 General Gibbs partitions and the new family

Both the Ewens-Pitman family and the partitions introduced in this note can be constructed in a unified way, using simple algebraic identities. Recall from [9, 18] that the Gibbs form for EPPF p of genus $\alpha\in(-\infty,1)$ is [2]

$$p(n_1,\dots,n_k)=v_{n,k}\prod_{j=1}^{k}(1-\alpha)_{n_j-1},$$

[2] We omit here the case $\alpha=-\infty$.

where the triangular array $(v_{n,k})$ is nonnegative and satisfies the recursion

$$v_{n,k} = (n - k\alpha)v_{n+1,k} + v_{n+1,k+1}, \qquad 1 \le k \le n \tag{11}$$

with normalization $v_{1,1} = 1$. The recursion goes backwards, from $n+1$ to n, thus it cannot be 'solved' in a unique way and rather has a convex set of solutions, each corresponding to a distribution of some exchangeable partition.

For a Gibbs partition the number of occupied boxes $(K_n, n = 1, 2, \dots)$ is a nondecreasing Markov chain, viewed conveniently as a bivariate space-time walk (n, K_n), which has backward transition probabilities depending on α but not on $(v_{n,k})$. The backward transition probabilities are determined from the conditioning relation: given $(n, K_n) = (n, k)$, the probability of each admissible path from $(1,1)$ to (n,k) is proportional to the product of *weights* along the path, where the weight of transition $(n,k) \to (n+1,k)$ is $n - k\alpha$, and that of $(n,k) \to (n+1,k+1)$ is 1. The normalizing total sum $d_{n,k}(\alpha)$ of such products over the paths from $(1,1)$ to (n,k) is known as a generalized Stirling number [3]. Each particular solution to (11) determines the law of (K_n) via the marginal distributions $\mathbb{P}(K_n = k) = v_{n,k} d_{n,k}(\alpha)$.

Large-n properties of Gibbs partitions depend on α. In particular, there exists an almost-sure limit $K = \lim_{n\to\infty} K_n / c_n(\alpha)$, where $c_n(\alpha) = n^\alpha$ for $\alpha \in (0,1)$, $c_n(0) = \log n$ and $c_n(\alpha) \equiv 1$ for $\alpha < 0$. The law of K is characteristic for a partition of given genus. That is to say, a generic Gibbs partition is a unique mixture over $\varkappa$ of *extreme* partitions for which $K = \varkappa$ a.s. Note that K has continuous range for $\alpha \in [0,1)$, and discrete for $\alpha < 0$. For $\alpha < 0$ the extremes are Fisher's partitions $\Pi^{\alpha,-\alpha\varkappa}$. For $\alpha = 0$ the extremes are Ewens' partitions $\Pi^{0,\varkappa}$ (with $\varkappa \in [0,\infty]$). For $\alpha \in (0,1)$ Ewens-Pitman partitions are not extreme, rather the extremes are obtainable by conditioning any $\Pi^{\alpha,\theta}$ on $K = \varkappa$; the $v_{n,k}$'s for these extreme partitions were identified in [13] in terms of the generalized hypergeometric functions.

Following [5], where recursions akin to (11) were treated, one can seek for special solutions of the form

$$v_{n,k} = \frac{\prod_{i=1}^{n-k} f(i) \prod_{j=1}^{k-1} g(j)}{\prod_{m=1}^{n-1} h(m)}, \tag{12}$$

where $f, g, h : \mathbb{N} \to \mathbb{R}$ satisfy the identity

$$(n - \alpha k) f(n-k) + g(k) = h(n), \quad 1 \le k \le n,\ n \in \mathbb{N}. \tag{13}$$

Moreover, f, h must be (strictly) positive on $\mathbb{N}$, while g may be either positive on $\mathbb{N}$ or positive on some integer interval $\{1, \dots, k_0 - 1\}$ with $g(k_0) = 0$. Each such triple defines a Gibbs partition with the 'new boxes' updating rule of the form

$$v_n(k; n_1, \dots, n_k) = \mathbb{P}(K_{n+1} = k+1 \,|\, K_n = k) = \frac{g(k)}{h(n)}$$

(where $n = n_1 + \dots + n_k$), complemented by the associated version of the (O)-rule (as in Sections 1 and 2).

Now we can review two instances of (12):

- Exploiting the identity $n - \alpha k + \alpha k + \theta = n + \theta$ we may choose $f(n) \equiv 1$, $g(n) = \alpha n + \theta$ and $h(n) = n + \theta$. This yields the Ewens-Pitman partitions with the succession rule $(\mathrm{N}^{\alpha,\theta})$. Note that the admissible range for α, θ is determined straightforwardly from the positivity.

- The identity
$$(n+k)(n-k+\gamma)+k^2-\gamma k+\zeta=n^2+\gamma n+\zeta$$
is of the kind (13) with $\alpha=-1$. We choose $f(n)=n+\gamma, g(n)=n^2-\gamma n+\zeta$ and $h(n)=n^2+\gamma n+\zeta$ to arrive at the partitions itroduced in this paper.

It is natural to wonder if there are any other Gibbs partitions of the form (12).
Remark The ansatz (12) is sometimes useful to deal with the recursions similar to (11), which have weights depending in a simple way on n and k [5]. For instance, if both weights equal 1, then each solution defines the distribution of an exchangeable $0-1$ sequence (see [1]), for which the Markov chain (K_n) counts the number of 1's among the first n bits. An instructive exercise is to construct by this method of specifying the triple f,g,h two distinguished families of the exchangeable processes – the homogeneous Bernoulli processes and Pólya's urn processes with two colors.
Acknowledgement The author is indebted to J. Pitman, an associated editor and a referee for their stimulating questions and comments.

References

[1] D.J. Aldous, *Exchangeability and related topics*, *L. Notes. Math.* 1117, 1985. MR0883646

[2] N. Berestycki and J. Pitman, Gibbs distributions for random partitions generated by a fragmentation process, *J. Stat. Phys.* 127: 381-418, 2007. MR2314353

[3] C.A. Charalambides, *Combinatorial methods in discrete distributions*, Wiley, 2005. MR2131068

[4] R.A. Fisher, A.S. Corbet and C.B. Williams, The relation between the number of species and the number of individuals in a random sample from an animal population, *J. Animal. Ecol.* 12: 42-58, 1943.

[5] A. Gnedin, Boundaries from inhomogeneous Bernoulli trials, `http://arxiv.org/abs/0909.4933`, 2009

[6] A. Gnedin, B. Hansen and J. Pitman, Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws, *Probab. Surveys* 4: 146-171, 2007. MR2318403

[7] A. Gnedin, C. Haulk and J. Pitman, Characterizations of exchangeable partitions and random discrete distributions by deletion properties `http://arxiv.org/abs/0909.3642`, 2009.

[8] A. Gnedin and J. Pitman, Regenerative composition structures, *Ann. Probab.* 33: 445–479. MR2122798

[9] A. Gnedin and J. Pitman, Exchangeable Gibbs partitions and Stirling triangles, *J. Math. Sci.* 138: 5674-5685, 2006. MR2160320

[10] A. Gnedin, J. Pitman and M. Yor, Asymptotic laws for compositions derived from transformed subordinators, *Ann. Probab.* 34: 468-492, 2006. MR2223948

[11] A. Gnedin and Y. Yakubovich, On the number of collisions in exchangeable coalescents, *Elec. J. Probab.* 12: 1547-1567, 2007. MR2365877

[12] B. Hansen and J. Pitman, Prediction rules for exchangeable sequences related to species sampling, *Stat. Probab. Letters* 46: 251-256, 2000. MR1745692

[13] M.-W. Ho, L.F. James and J. W. Lau, Gibbs partitions (EPPF's) derived from a stable subordinator are Fox H and Meijer G transforms, http://arxiv.org/abs/0708.0619, 2007.

[14] A. Lijoi, I. Prünster and S.G. Walker, Bayesian nonparametric estimators derived from conditional Gibbs structures, *Ann. Appl. Probab.*, 18: 1519-1547, 2008. MR2434179

[15] A. Lijoi and I. Prünster, Models beyond the Dirichlet process, `http://www.icer.it/docs/wp2009/ICERwp23-09.pdf`, 2009.

[16] J. Pitman, Exchangeable and partially exchangeable random partitions, *Prob. Th. Rel. Fields* 102: 145-158, 1995. MR1337249

[17] J. Pitman, Partition structures derived from Brownian motion and stable subordinators, *Bernoulli* 3: 79-96, 1997. MR1466546

[18] J. Pitman, *Combinatorial stochastic processes*, *L. Notes. Math.* 1875, 2006. MR2245368