EXCHANGEABLE GIBBS PARTITIONS AND STIRLING TRIANGLES

A. Gnedin† and J. Pitman‡

UDC 519.217.72, 519.217.74

For two collections of nonnegative and suitably normalized weights \( W = (W_j) \) and \( V = (V_{\alpha,k}) \), a probability distribution on the set of partitions of the set \( \{1, \ldots, n\} \) is defined by assigning to a generic partition \( \{A_i, j \leq k\} \) the probability \( V_{\alpha,k} W_{|A_1|} \cdots W_{|A_k|} \), where \( |A_j| \) is the number of elements of \( A_j \). We impose constraints on the weights by assuming that the resulting random partitions \( \Pi_n \) of \( [n] \) are consistent as \( n \) varies, meaning that they define an exchangeable partition of the set of all natural numbers. This implies that the weights \( W \) must be of a very special form depending on a single parameter \( \alpha \in (-\infty, 1] \). The case \( \alpha = 1 \) is trivial, and for each value of \( \alpha \neq 1 \) the set of possible \( V \)-weights is an infinite-dimensional simplex. We identify the extreme points of the simplex by solving the boundary problem for a generalized Stirling triangle. In particular, we show that the boundary is discrete for \( -\infty \leq \alpha < 0 \) and continuous for \( 0 \leq \alpha < 1 \). For \( \alpha \leq 0 \) the extremes correspond to the members of the Ewens–Pitman family of random partitions indexed by \((\alpha, \theta)\), while for \( 0 < \alpha < 1 \) the extremes are obtained by conditioning an \((\alpha, \theta)\)-partition on the asymptotics of the number of blocks of \( \Pi_n \) as \( n \) tends to infinity. Bibliography: 29 titles.

1. Introduction

By a random partition of the set of natural numbers \( \mathbb{N} \) we mean a consistent sequence \( \Pi = (\Pi_n) \) of random partitions of finite sets \( [n] := \{1, \ldots, n\} \). For each \( n \) the range of the random variable \( \Pi_n \) is the set of all partitions of \( [n] \) into some number of disjoint nonempty blocks, and the consistency means that \( \Pi_n \) is obtained from \( \Pi_{n+1} \) by discarding the element \( n+1 \). A random partition \( \Pi \) is exchangeable if for each \( n \) the probability distribution of \( \Pi_n \) is invariant under all permutations of \( [n] \).

Let \( \{A_j, 1 \leq j \leq k\} \) denote a generic partition of the set \( [n] \), with the blocks \( A_j \) labelled in the increasing order of their least elements. The exchangeability of \( \Pi \) means that

\[
P(\Pi_n = \{A_1, \ldots, A_k\}) = p(|A_1|, \ldots, |A_k|)
\]

for some nonnegative function

\[
p(\lambda) := p(\lambda_1, \ldots, \lambda_k)
\]

of compositions \( \lambda = (\lambda_1, \ldots, \lambda_k) \) of \( n \) such that \( p \) is symmetric in the arguments \( \lambda_1, \ldots, \lambda_k \) for each \( k \), \( p \) is normalized by the condition \( p(1) = 1 \), and \( p \) satisfies the addition rule

\[
p(\lambda) = \sum_{\mu: \mu \preceq \lambda} p(\mu), \tag{1}
\]

where the sum is over compositions \( \mu \) derived from \( \lambda \) by either increasing a part by one or by appending 1 at the end of the sequence \( \lambda \). For instance, if \( \lambda = (3, 2, 2) \), then \( \mu \) assumes the values \((4, 2, 2), (3, 3, 2), (3, 2, 3), \) and \((3, 2, 2, 1)\), and (1) specializes to

\[
p(3, 2, 2) = p(4, 2, 2) + 2p(3, 3, 2) + p(3, 2, 2, 1).
\]

A function \( p \) with these properties is known as an exchangeable partition probability function (EPPF). Such a function uniquely determines the probability law of the corresponding exchangeable random partition \( \Pi \).

According to Kingman’s paintbox representation [22, 9, 10], every such exchangeable partition has the same distribution as the partition \( \Pi \) constructed from some random closed set \( Z \subset [0, 1] \) as follows: let \( u_1, u_2, \ldots \) be independent uniform \( [0, 1] \) variables independent of \( Z \), and let distinct integers \( i \) and \( j \) belong to the same block of \( \Pi \) if and only if \( u_i \) and \( u_j \) fall in the same open interval component of \( [0, 1] \setminus Z \).

†Utrecht University, The Netherlands, e-mail: gnedin@math.uu.nl.
‡University of California at Berkeley, USA, e-mail: pitman@stat.berkeley.edu.


5674 1072-3374/06/1383-5674 ©2006 Springer Science+Business Media, Inc.
A distinguished class of exchangeable partitions is the two-parameter family with EPPF

\[ p_{\alpha, \theta}(\lambda_1, \ldots, \lambda_k) := \frac{(\theta + \alpha)^{k-1} \alpha}{(\theta + 1)^{n-1}} \prod_{j=1}^{k} (1 - \alpha)^{\lambda_j - 1}, \]

where \( n = \sum \lambda_j \) and

\[ (x)_{m|\beta} := \prod_{j=1}^{m} (x + (j-1)\beta), \quad (x)_{m|} = (x)_{m|1} \]

are rising factorials, with the convention that \((x)_{0|\beta} := 1\). The possible values of the parameters \((\alpha, \theta)\) are either \(-\infty \leq \alpha < 0\) and \(\theta = m|\alpha\) for some \(m = 1, 2, \ldots, \infty\); or \(0 \leq \alpha \leq 1\) and \(\theta \geq -\alpha\), with a proper understanding of \((2)\) in some limiting cases. See [24] for a detailed exposition of the general theory of exchangeable partitions and features of the \((\alpha, \theta)\)-family.

In this paper, we are interested in a special class of Gibbs partitions, which generalise \((2)\) as follows.

**Definition 1.** An exchangeable random partition \(\Pi\) of the set of natural numbers is said to be of Gibbs form if for some nonnegative weights \(W = (W_j)\) and \(V = (V_{n,k})\) the EPPF of \(\Pi\) satisfies

\[ p(\lambda_1, \ldots, \lambda_k) = V_{n,k} \prod_{j=1}^{k} W_{\lambda_j} \]

for all \(1 \leq k \leq n\) and all compositions \((\lambda_1, \ldots, \lambda_k)\) of \(n\).

For fixed \(n\) we can choose arbitrary nonnegative weights \(V_1, \ldots, V_n\) and \(W_1, \ldots, W_n\) that are not identically zero, and use \((3)\) to define a random partition of \([n]\) by setting \(V_{n,k} = V_k/c_n\) for \(c_n\) a suitable normalization constant (see [29] for another version of the Gibbs formalism). The block sizes of such a Gibbs partition can be realized by Kolchin’s model, that is, they can be identified with the collection of terms of a random sum \(S = X_1 + \ldots + X_K\) conditioned on \(S = n\), with independent identically distributed \(X_1, X_2, \ldots\) independent of \(K\). For integer weights, this is the distribution on partitions of \([n]\) induced by the components of a random composite structure built over partitions of \([n]\), when there are \(W_j\) possible configurations associated with every subset of \([n]\) with \(j\) elements, \(V_k\) possible configurations associated with every collection of \(k\) blocks, and a uniform distribution is assigned to all possible composite structures subject to these constraints. For instance, if \(W_j = (j-1)!\) and \(V_k = \theta^k\) (with \(\theta \in \mathbb{N}\)), then the product \(V_k W_{\lambda_1} \cdots W_{\lambda_k}\) counts the number of colored permutations of \([n]\) with cycle sizes \((\lambda_1, \ldots, \lambda_k)\) and one of \(\theta\) possible colors assigned to each of the cycles; then \((3)\) reduces to \((2)\) with \(\alpha = 0, \theta > 0\). Thus in this case, there is an infinite exchangeable partition \(\Pi\) whose restrictions \(\Pi_n\) are all of Gibbs form \((3)\). Many other combinatorially interesting examples of Gibbs partitions \(\Pi_n\) can be given using the prescription \((3)\) for each fixed \(n\): see, for instance, [3, 24]. But, typically, the distributions of these combinatorially defined \(\Pi_n\) are not consistent as \(n\) varies, so they are not realizable as the sequence of restrictions to \([n]\) of an infinite Gibbs partition.

The special case of \(V\)-weights representable as ratios \(V_{n,k} = V_k/c_n\) was studied by Kerov [13] in the framework of Kolchin’s model. In this case, one assumes a single infinite sequence of weights \((V_k)\), and the \(c_n\)’s appear as normalization constants. Kerov [13] established that the Gibbs partitions of this type are precisely the members of the two-parameter family \((2)\).

We will show that in the more general setting \((3)\), allowing an arbitrary triangular array \(V_{k,n}\), the \(W\)-weights must be still as in \((2)\), with a single parameter \(\alpha \in [-\infty, 1]\) defining the type of a Gibbs partition. The case \(\alpha = 1\) is trivial. For each nontrivial type \(\alpha < 1\), the set of all possible \(V\)-weights is an infinite simplex \(V_\alpha\). We identify the extreme elements of \(V_\alpha\) by solving a boundary problem for an instance of the generalized Stirling triangle, as introduced in another paper by Kerov [12] (see also [14, Chap. I]). It turns out that the nature of the extremal set depends substantially on the type. According to our main result, stated more formally in Theorem 12, there are three qualitatively different ranges of \(\alpha\). For \(\alpha \in (-\infty, 0]\), the extremal set is discrete and corresponds to the members of the \((\alpha, \theta)\)-family. For \(\alpha = 0\), this set is continuous and still corresponds to the members of the \((\alpha, \theta)\)-family (Ewens partitions). For \(\alpha \in [0, 1]\), the \((\alpha, \theta)\)-partitions are not extreme, rather the extremes of \(V_\alpha\) comprise a continuous \((\alpha|s)\)-family (with parameter \(s \in [0, \alpha]\)) that appears by conditioning the \((\alpha, \theta)\)-partitions on the asymptotics of the number of blocks. In [21], the \((\alpha|s)\)-partitions were derived from their Kingman representation, with the random closed set \(Z\) being the scaled range of an \(\alpha\)-stable subordinator conditioned on its value at a fixed time. This identification of extreme elements of \(V_\alpha\) for \(0 \leq \alpha < 1\) was indicated without proof in [21, Theorem 8].