Ann Inst Stat Math (2017) 69:1–37
DOI 10.1007/s10463-015-0530-0

Extreme sizes in Gibbs-type exchangeable random partitions

Shuhei Mano[1,2]

Received: 6 November 2014 / Revised: 15 June 2015 / Published online: 29 July 2015

Abstract Gibbs-type exchangeable random partition, which is a class of multiplicative measures on the set of positive integer partitions, appear in various contexts, including Bayesian statistics, random combinatorial structures, and stochastic models of diversity in various phenomena. Some distributional results on ordered sizes in the Gibbs partition are established by introducing associated partial Bell polynomials and analysis of the generating functions. The combinatorial approach is applied to derive explicit results on asymptotic behavior of the extreme sizes in the Gibbs partition. Especially, Ewens–Pitman partition, which is the sample from the Poisson–Dirichlet process and has been discussed from rather model-specific viewpoints, and a random partition which was recently introduced by Gnedin, are discussed in the details. As by-products, some formulas for the associated partial Bell polynomials are presented.

Keywords Random partition · Extremes · Analytic combinatorics · The Bell polynomials · Gibbs partitions · The Ewens–Pitman partition · Poisson–Dirichlet process

1 Introduction

Exchangeable random partitions of a natural number appear in various contexts, including Bayesian statistics, random combinatorial structures, and stochastic models of diversity in biological, physical, and sociological phenomena (see, for example,

✉ Shuhei Mano
smano@ism.ac.jp

1 The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

2 CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

Tavaré and Ewens 1997; Aoki 2002; Arratia et al. 2003; Pitman 2006). The Gibbs-type exchangeable random partition is a class of multiplicative measures on the set of integer partitions (Gnedin and Pitman 2005). The Gibbs partition discussed in this paper (we will define precisely later) is probably the broadest class of exchangeable random partitions ever appeared in the literature; it covers many exchangeable random partitions proposed so far, including the Ewens–Pitman partition (Ewens 1972; Pitman 1995), which is the sampling formula from the Poisson–Dirichlet process (Ferguson 1973; Kingman 1975; Aldous 1985; Pitman 1995; Pitman and Yor 1997), the sampling formula from the normalized inverse-Gaussian process (Lijoi et al. 2005), the limiting conditional compound Poisson distribution (Hoshino 2009), and a random partition proposed by Gnedin (2010), which is a mixture of Dirichlet-multinomial distributions.

Let us define the Gibbs partition discussed in this paper. A partition of $[n] := \{1, 2, \dots, n\}$ into k blocks is an unordered collection of non-empty disjoint sets $\{A_1, \dots, A_k\}$ whose union is $[n]$. The multiset $\{|A_1|, \dots, |A_k|\}$ of unordered sizes of blocks of a partition π_n of $[n]$ defines a partition of integer n. The sequence of positive integer counts $(|\pi_n|_j, 1 \le j \le n)$, where $|\pi_n|_j$ is the number of blocks in π_n of size j, with

$$|\pi_n| := \sum_{j=1}^{n} |\pi_n|_j = k, \qquad \sum_{j=1}^{n} j|\pi_n|_j = n. \tag{1}$$

A random partition of $[n]$ is called exchangeable if its distribution is invariant under the permutation of $[n]$. If for each partition $\{A_1, \dots, A_k\}$ of $[n]$

$$\mathbb{P}(\Pi_n = \{A_1, \dots, A_k\}) = p_n(|A_1|, \dots, |A_k|)$$

for some symmetric function of p_n on a partition of n, the function p_n is called the exchangeable partition probability function (EPPF). Suppose we have two sequences of non-negative weights, (w_j) and $(v_{n,k})$ for $1 \le j, k \le n$. A class of EPPF on the set of integer partition is called Gibbs form (Definition 1 of Gnedin and Pitman 2005) if it admits a symmetric multiplicative representation

$$p_n(n_1, \dots, n_k) = v_{n,k} \prod_{j=1}^{k} w_{n_j}, \tag{2}$$

or with multiplicities,

$$\mathbb{P}(|\Pi_n|_j = m_j, 1 \le j \le n) = n! v_{n,k} \prod_{j=1}^{n} \left(\frac{w_j}{j!}\right)^{m_j} \frac{1}{m_j!} \tag{3}$$

for all $1 \le k \le n$ and all partitions of n, where $\sum_j n_j = n$ or $\sum_j m_j = k$ and $\sum_j j m_j = n$.

There are two natural requirements for the Gibbs partitions. Gnedin and Pitman (2005) showed that if the Gibbs form has *consistency* (Kingman 1978; Aldous 1985; Pitman 1995), which means that a random partition of size n is obtained from a random partition of size $n+1$ by discarding element $n+1$, w-weights have a form of

$$(w_\bullet) = ((1-\alpha)_{\bullet-1}), \qquad -\infty < \alpha < 1, \tag{4}$$

where for real number x and positive integer i, $(x)_i = x(x+1)\cdots(x+i-1)$ with a convention $(x)_0 = 1$. Since the consistency is a reasonable property for applications, some of recent papers called the subclass of (2) with the requirement of (4) the Gibbs partitions (for example, Griffiths and Spano 2007; Lijoi et al. 2008; Gnedin 2010). In this paper, we call the subclass the *consistent Gibbs partitions*. On the other hand, Pitman (Section 1.5 of Pitman 2006) called (2) the Gibbs partition $Gibbs_{[n]}(v_\bullet, w_\bullet)$ if the v-weights are representable as ratios

$$v_{n,k} = \frac{v_k}{B_n(v_\bullet, w_\bullet)}, \qquad B_n(v_\bullet, w_\bullet) := \sum_{k=1}^{n} v_k B_{n,k}(w_\bullet), \tag{5}$$

where

$$B_{n,k}(w_\bullet) := n! \sum_{\{m_\bullet : \sum m_j = k, \sum j m_j = n\}} \prod_{j=1}^{n} \left(\frac{w_j}{j!}\right)^{m_j} \frac{1}{m_j!}, \qquad n \geq k, \tag{6}$$

and a convention $B_{n,k}(w_\bullet) = 0$ for $n < k$, is the partial Bell polynomial in the variables $(w_\bullet)$. In addition, $Gibbs_{[n]}(v_\bullet, w_\bullet)$ is called Kolchin's model (Kolchin 1971; Kerov 1995), or $Gibbs_{[n]}(v_\bullet, w_\bullet)$ has Kolchin's representation (Pitman 2006), which is identified with the collection of terms of random sum $X_1 + \cdots + X_{|\Pi_n|}$ conditioned by $\sum_i X_i = n$ with independent and identically distributed $X_1, X_2, \ldots,$ independent of $|\Pi_n|$. As we have seen, definition of Gibbs partition depends on contexts, authors, and papers. To make our discussion have most generality, throughout this paper we call the class of exchangeable random partitions whose EPPF have the form of (2) the Gibbs partition. In this paper, we will discuss properties of the broadest class.

If a Gibbs partition $Gibbs_{[n]}(v_\bullet, w_\bullet)$ has the consistency, it reduces to the Ewens–Pitman partition (Kerov 1995; Gnedin and Pitman 2005). The Ewens–Pitman partition appears in various contexts, whose classical examples include cycle lengths in random permutation (Shepp and Lloyd 1966), a sample from the infinite many allele model in population genetics (Ewens 1972), and a sample from the Dirichlet process prior in Bayesian nonparametrics (Antoniak 1974). The v-weights have a form of

$$v_k = (\theta)_{k;\alpha}, \qquad B_n(v_\bullet, w_\bullet) = (\theta)_n, \tag{7}$$

where for real numbers x and a and positive integer i, $(x)_{i;a} = x(x+a)\cdots(x+(i-1)a)$ with a convention $(x)_{0;a} = 1$. The pair of real parameters α and θ satisfies either $0 \leq \alpha < 1$ and $\theta > -\alpha$, or $\alpha < 0$ and $\theta = -m\alpha$, $m = 1, 2, \ldots$. For $\alpha < 0$ the Ewens–Pitman partition reduces to a symmetric Dirichlet-multinomial distribution of parameter $(-\alpha)$. Another example of consistent Gibbs partitions discussed in this paper is Gnedin's partition (Gnedin 2010), which is a mixture of Dirichlet-multinomial distributions. In contrast to the Ewens–Pitman partition v-weights of Gnedin's partition are not representable as ratios; in Gnedin's partition w-weights satisfy (4) with $\alpha = -1$ and the v-weights have a form of

$$v_{n,k} = (\gamma)_{n-k} \frac{\prod_{j=1}^{k-1}(j^2 - \gamma j + \zeta)}{\prod_{j=1}^{n-1}(j^2 + \gamma j + \zeta)}, \tag{8}$$

where ζ and γ are chosen such that $\gamma \geq 0$ and $j^2 - \gamma j + \zeta > 0$ for $j \geq 1$.

In studies of exchangeable random partitions the Ewens–Pitman partition has played the central role and discussion on the generalizations has been only recently started. The Ewens–Pitman partition and the Poisson–Dirichlet process, which is closely related to the Ewens–Pitman partition, have nice properties. For example, if $\alpha = 0$ and $\theta > 0$ the Ewens–Pitman partition satisfies the *conditioning relation*:

$$(|\Pi_n|_1, \ldots, |\Pi_n|_n) \stackrel{d}{=} (Z_1, \ldots, Z_n) | \sum_{j=1}^{n} j Z_j = n, \tag{9}$$

where Z_j, $j = 1, \ldots, n$, independently follow the Poisson distribution of parameter θ/j. Moreover, the multiplicities of the small components are asymptotically independent:

$$(|\Pi_n|_1, |\Pi_n|_2, \ldots) \stackrel{d}{\to} (Z_1, Z_2, \ldots), \qquad n \to \infty. \tag{10}$$

Approaches based on the conditioning relation are very powerful if a random combinatorial structure has the property (see Arratia et al. 2003 for a comprehensive survey). Unfortunately, the asymptotic independence (10) does not hold even in the Ewens–Pitman partition with non-zero α. Studies of the Ewens–Pitman partition has been heavily dependent on properties of the Poisson–Dirichlet process (see, for example, Arratia et al. 2003; Pitman 2006). The dependency on the Poisson–Dirichlet process has made arguments model specific. The connection with Poisson–Dirichlet process might have little use in studies of general Gibbs partitions. On the other hand, in studies of random combinatorial structures analytic combinatorial approaches have been quite useful in applications to various problems (Flajolet and Odlyzko 1990; Flajolet and Sedgewick 2009). For example, Panario and Richmond (2001) showed that in a decomposition of a random permutation into cycles, which corresponds to the Ewens–Pitman partition of parameter $(\alpha, \theta) = (0, 1)$, the singularity analysis of the generating function, which is a popular tool in analytic combinatorics, yields asymptotic behavior of the ordered cycle lengths. In the present paper we will see that analytic combinatorial approaches are general enough to apply to a broader class of exchangeable random partitions, namely, the Gibbs partition.

Behavior of the extreme sizes in random partitions is a classic issue. Some examples from statistical application are an exact test for the maximum component in a periodogram by Fisher (1929), and an exact test of natural selection operating on the most frequent gene type in population genetics (Ewens 1973). In addition, as a measure of diversity, distribution of the maximum size has been discussed in, for example, population genetics (Watterson and Guess 1977) and economics (Aoki 2002). Asymptotic behavior of extreme sizes has been attracted many authors, not only by practical importance but also by mathematical interest. Asymptotic behavior of the ordered cycle length in a random permutation has been discussed in Shepp and Lloyd (1966). In the number theory, a number whose largest prime factor is not larger than x is called x-smooth number, while a number whose smallest prime factor is larger than y is called y-rough number (see, for example, Chapters III.5 and III.6 of Tenenbaum 1995). The limiting distributions of the counting functions of the smooth number and the rough number are coincidentally identical to the distribution functions of the extreme sizes in the Ewens–Pitman partition of parameters $(\alpha, \theta) = (0, 1)$ (Arratia et al. 2003). An extension to the case of $\theta \neq 1$ in the context of the number theory was also discussed in Hensley (1984). For the smallest sizes in the Ewens–Pitman partition of parameters $(\alpha, \theta) = (0, \theta)$, it is known that the probability that the smallest size is larger than $r \asymp n \to \infty$ involves a generalization of Buchstab's function in the number theory (Corollary 1) and the probability that the smallest size is larger than $r = o(n)$ follows immediately from the asymptotic independence (10), which is restricted to the case that $\alpha = 0$. In this paper, we will see how these properties change in the Ewens–Pitman partition of parameters $0 < \alpha < 1$ and $\theta > -\alpha$. We will see that we do not have "generalized Buchstab's function" (Corollary 2) and we will establish (Theorem 3) a precise asymptotic of the probability that the smallest size is larger than $r = o(n)$: $f_{\alpha,\theta}(r)n^{-\alpha-\theta}$ with some explicit function $f_{\alpha,\theta}(r)$. The asymptotic independence no longer holds, nevertheless, the singularity analysis of the generating function in analytic combinatorics gives the estimate straightforwardly.

This paper is organized as follows. In Sect. 2, we introduce associated partial Bell polynomials by restricting sizes of components in enumerating possible partitions. In Sect. 3 we obtain some distributional results of ordered sizes in the Gibbs partition in terms of the generating functions. In Sect. 4 asymptotic behavior of extreme sizes is discussed. We see explicit results on asymptotic behavior of extreme sizes in the consistent Gibbs partition, which includes the Ewens–Pitman partition and Gnedin's partition. Some of the results regarding the Ewens–Pitman partition are reproductions of known results, however, we demonstrate that our alternative derivation is simpler than approaches using model-specific properties of the Ewens–Pitman partition. In principle, the developed approach is able to apply to any type of Gibbs partition. Section 5 is devoted to the proofs.

A computer program to generate Gibbs partitions is available upon request to the author.

2 Associated partial Bell polynomials

Let us begin with a proposition on the partial Bell polynomial (6), which follows immediately from Faà di Bruno's formula (Comtet 1974; Pitman 2006):

$$B_n(v_\bullet, w_\bullet) = \left[\frac{\xi^n}{n!}\right] \breve{v}(\breve{w}(\xi)), \tag{11}$$

where $\breve{v}$ and $\breve{w}$ are the exponential generating functions:

$$\breve{v}(\eta) := \sum_{j=1}^{\infty} v_j \frac{\eta^j}{j!}, \qquad \breve{w}(\xi) := \sum_{j=1}^{\infty} w_j \frac{\xi^j}{j!},$$

and $[\xi^n/n!]f(\xi) := a_n$ for a series $f(\xi) = \sum_j a_j \xi^j / j!$.

Proposition 1 *The exponential generating functions of the partial Bell polynomials are*

$$\breve{B}_k(\xi, w_\bullet) := \sum_{n=k}^{\infty} B_{n,k}(w_\bullet)\frac{\xi^n}{n!} = \frac{(\breve{w}(\xi))^k}{k!}, \qquad k = 1, 2, \ldots \tag{12}$$

Example 1 Setting $(w_\bullet) = ((\bullet - 1)!)$ yields

$$\breve{B}_k(\xi, (\bullet - 1)!) = \frac{\{-\log(1-\xi)\}^k}{k!}, \qquad k = 1, 2, \ldots$$

where $B_{n,k}((\bullet - 1)!) \equiv |s(n,k)|$ are the signless Stirling number of the first kind defined by (Charalambides 2005)

$$B_n(\theta^\bullet, (\bullet - 1)!) = (\theta)_n = \sum_{k=0}^{n} |s(n,k)|\theta^k, \qquad n = 0, 1, \ldots \tag{13}$$

Example 2 For non-zero α, setting $(w_\bullet) = ([\alpha]_\bullet)$ yields

$$\breve{B}_k(\xi, [\alpha]_\bullet) = \frac{\{(1+\xi)^\alpha - 1\}^k}{k!}, \qquad k = 1, 2, \ldots, \tag{14}$$

where for real number x and positive integer i, $[x]_i = x(x-1)\cdots(x-i+1)$ with a convention $[x]_0 = 1$, and $B_{n,k}([\alpha]_\bullet) \equiv C(n,k;\alpha)$ are the generalized factorial coefficients introduced by (Charalambides 2005)

$$B_n([x]_\bullet, [\alpha]_\bullet) = [\alpha x]_n = \sum_{k=0}^{n} C(n,k;\alpha)[x]_k, \qquad n = 0, 1, \ldots \tag{15}$$

In general, for distinct real numbers a and b, $B_{n,k}([b-a]_{(\bullet-1);a}) \equiv S_{n,k}^{a,b}$ are generalized Stirling numbers defined by (Pitman 2006)

$$B_n([x]_{\bullet;b}, [b-a]_{(\bullet-1);a}) = [x]_{n;a} = \sum_{k=0}^{n} S_{n,k}^{a,b}[x]_{k;b}, \qquad n = 0, 1, \ldots,$$

where for real number x and a and positive integer i, $[x]_{i;a} = x(x-a)\cdots(x-(i-1)a)$ with a convention $[x]_{0;a} = 1$. For example, $|s(n,k)| = (-1)^{n+k}S_{n,k}^{1,0}$ and $C(n,k;\alpha) = \alpha^k S_{n,k}^{1,\alpha}$.

Dropping off first terms in the sequence $(w_\bullet)$ of Proposition 1 gives a modified version of Proposition 1, which introduces the *associated partial Bell polynomials*. We call the polynomials *associated* since they cover associated numbers that appear in combinatorics literature, such as the associated signless Stirling numbers of the first kind and the associated generalized factorial coefficients (see, for example, Comtet

1974; Charalambides 2005). The associated partial Bell polynomials play the central role throughout this paper.

Proposition 2 *For a sequence* (w_j), $1 \le j < \infty$, *and positive integer* r, *set* $w_{(r)j} = 0$, $1 \le j \le r-1$ *and* $w_{(r)j} = w_j$, $j \ge r$. *Define the associated partial Bell polynomials*

$$B_{n,k,(r)}(w_\bullet) := B_{n,k}(w_{(r)\bullet}) = n! \sum_{\substack{\{m_\bullet : \sum m_j = k, \sum j m_j = n, \\ m_{j<r}=0\}}} \prod_{j=1}^{n} \left(\frac{w_j}{j!}\right)^{m_j} \frac{1}{m_j!}, \qquad n \ge rk,$$

with a convention $B_{n,k,(r)}(w_\bullet) = 0$ *for* $n < rk$ *and* $B_{n,k,(1)}(w_\bullet) = B_{n,k}(w_\bullet)$. *Then, the exponential generating function of the sequence* $(w_{(r)\bullet})$, $\breve{w}_{(r)}(\xi)$, *provides the exponential generating functions of the associated partial Bell polynomials*

$$\breve{B}_{k,(r)}(\xi, w_\bullet) := \sum_{n=rk}^{\infty} B_{n,k,(r)}(w_\bullet) \frac{\xi^n}{n!} = \frac{(\breve{w}_{(r)}(\xi))^k}{k!}, \qquad k = 1, 2, \ldots \tag{16}$$

Example 3 Setting $(w_\bullet) = ((\bullet - 1)!)$ yields

$$\breve{B}_{k,(r)}(\xi, (\bullet - 1)!) = \frac{1}{k!} \left\{ -\log(1-\xi) - \sum_{j=1}^{r-1} \frac{\xi^j}{j} \right\}^k, \qquad k = 1, 2, \ldots,$$

where $B_{n,k,(r)}((\bullet - 1)!) \equiv |s_r(n,k)|$ are known as the r-associated signless Stirling numbers of the first kind (Comtet 1974; Charalambides 2005). The associated signless Stirling number of the first kind has an interpretation in terms of a decomposition of a random permutation into cycles. In decomposing a set of n elements into k cycles the number of permutations in which each length of cycle is not shorter than r is $|s_r(n,k)|$.

Example 4 For non-zero α, setting $(w_\bullet) = ([\alpha]_\bullet)$ yields

$$\breve{B}_{k,(r)}(\xi, [\alpha]_\bullet) = \frac{1}{k!} \left\{ (1+\xi)^\alpha - \sum_{j=0}^{r-1} \binom{\alpha}{j} \xi^j \right\}^k, \qquad k = 1, 2, \ldots, \tag{17}$$

where $B_{n,k,(r)}([\alpha]_\bullet) \equiv C_r(n,k;\alpha)$ are known as the r-associated generalized factorial coefficients (Charalambides 2005). Suppose that n-like balls are distributed into k distinguishable urns, each with α ($\ge n$) distinguishable cells whose capacity is limited to one ball. The enumerator for occupancy is

$$\sum_{j=1}^{\alpha} \binom{\alpha}{j} \xi^j = (1+\xi)^\alpha - 1,$$

and the generating function for occupancy of the k urns satisfies

$$\sum_{n=k}^{\alpha k} A(n,k;\alpha)\xi^n = \left\{(1+\xi)^\alpha - 1\right\}^k .$$

Comparing with the exponential generating function of the generalized factorial coefficients (14) implies that the number of different distributions of n-like balls into k distinguishable urns, each with α distinguishable cells of occupancy limited to one ball, equals $A(n,k;\alpha) = k!C(n,k;\alpha)/n!$. If each urn is occupied by at least r balls, the enumerator for occupancy of an urn is

$$\sum_{j=r}^{\alpha} \binom{\alpha}{j} \xi^j$$

and the generating function for occupancy of the k urns satisfies

$$\sum_{n=rk}^{\alpha k} A_r(n,k;\alpha)\xi^n = \left\{(1+\xi)^\alpha - \sum_{j=0}^{r-1}\binom{\alpha}{j}\xi^j\right\}^k .$$

Comparing with the exponential generating function of the associated generalized factorial coefficients (17) implies that the number of different distributions of n-like balls into k distinguishable urns, each with α distinguishable cells of occupancy limited to one ball, so that each urn is occupied by at least r balls equals $A_r(n,k;\alpha) = k!C_r(n,k;\alpha)/n!$.

When the sequence $(w_\bullet)$ is truncated we have another modified version of Proposition 1. The following proposition provides another kind of associated partial Bell polynomials. The author is unaware of the literature where this type of associated combinatorial numbers are discussed, but they will play important roles in this paper. The enumerating interpretations are similar to those in Examples 3 and 4.

Proposition 3 *For a sequence $(w_\bullet)$, and positive integer r, set $w_j^{(r)} = w_j$, $1 \le j \le r$ and $w_j^{(r)} = 0$, $j \ge r+1$. Define the associated partial Bell polynomials*

$$B_{n,k}^{(r)}(w_\bullet) := B_{n,k}(w_\bullet^{(r)}) = n! \sum_{\substack{\{m_\bullet: \sum m_j = k, \sum j m_j = n,\\ m_{j>r}=0\}}} \prod_{j=1}^{n} \left(\frac{w_j}{j!}\right)^{m_j} \frac{1}{m_j!}, \qquad k \le n \le rk,$$

with a convention $B_{n,k}^{(r)}(w_\bullet) = 0$ for $n < k$ and $n > rk$. We have $B_{n,k}^{(r)}(w_\bullet) = B_{n,k}(w_\bullet)$, $r \ge n-k+1$. Then, the exponential generating function of the sequence $(w_\bullet^{(r)})$, $\breve{w}^{(r)}(\xi)$, provides the exponential generating functions of the associated partial Bell polynomials

$$\breve{B}_k^{(r)}(\xi, w_\bullet) := \sum_{n=k}^{rk} B_{n,k}^{(r)}(w_\bullet)\frac{\xi^n}{n!} = \frac{(\breve{w}^{(r)}(\xi))^k}{k!}, \qquad k = 1,2,\ldots \tag{18}$$

In applications, especially for large n, recurrence relations are inevitable to compute the associated partial Bell polynomials introduced above. We provide some recurrence relations for the associated partial Bell polynomials in Appendix A.

Further modification of Proposition 1 provides another kind of associated partial Bell polynomials. As a natural extension of Propositions 2 and 3 is consideration of a set of n elements into k blocks so that the size of the ith largest block is not larger than r. Following proposition gives the extension. The proof is provided in Sect. 5.1.

Proposition 4 *Let the exponential generating functions $\breve{w}_{(r)}$ and $\breve{w}^{(r)}$ be defined as Propositions 2 and 3. Let us define associated Bell polynomials by*

$$B_{n,k}^{(r),(i)}(w_\bullet) := n! \sum_{\substack{\{m_\bullet : \sum m_j = k, \sum j m_j = n,\\ m_{r+1}+\cdots+m_n < i\}}} \prod_{j=1}^{r} \left(\frac{w_j}{j!}\right)^{m_j} \frac{1}{m_j!}, \qquad n \geq k,$$

with a convention $B_{n,k}^{(r),(i)}(w_\bullet) = 0$, $n < k$, for $2 \leq i \leq k$, and $B_{n,k}^{(r),(1)}(w_\bullet) = B_{n,k}^{(r)}(w_\bullet)$. Then, the exponential generating function is given by

$$\breve{B}_k^{(r),(i)}(\xi, w_\bullet) := \sum_{n=k}^{\infty} B_{n,k}^{(r),(i)}(w_\bullet) \frac{\xi^n}{n!} = \sum_{j=0}^{i-1} \breve{B}_{j,(r+1)}(\xi, w_\bullet) \breve{B}_{k-j}^{(r)}(\xi, w_\bullet).$$

Proposition 4 means that an associated partial Bell polynomials $B_{n,k}^{(r),(i)}(w_\bullet)$ is representable as quadratic polynomial in the associated partial Bell polynomials $B_{n,k,(r)}(w_\bullet)$ and $B_{n,k}^{(r)}(w_\bullet)$. Moreover, the next proposition, whose proof is in Sect. 5.1, implies that $B_{n,k,(r)}(w_\bullet)$ and $B_{n,k}^{(r)}(w_\bullet)$ can be expressed in terms of the partial Bell polynomials $B_{n,k}(w_\bullet)$. Therefore, in principle all associated partial Bell polynomials introduced in this paper can be expressed in terms of the partial Bell polynomials.

Proposition 5 *For positive integer r and k, the associated partial Bell polynomials, $B_{n,k}^{(r)}(w_\bullet)$, satisfy, if $r + k \leq n \leq rk$,*

$$B_{n,k}^{(r)}(w_\bullet) = B_{n,k}(w_\bullet) + \sum_{l=1}^{\lfloor (n-k)/r \rfloor} \frac{(-1)^l}{l!} \sum_{\substack{i_1,\ldots,i_l \geq r+1,\\ i_1+\cdots+i_l \leq n-k+l}} B_{n-(i_1+\cdots+i_l),k-l}(w_\bullet)[n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!} \tag{19}$$

and $B_{n,k}^{(r)}(w_\bullet) = B_{n,k}(w_\bullet)$ if $k \leq n \leq r + k - 1$. For positive integer r and k, the associated partial Bell polynomials, $B_{n,k,(r)}(w_\bullet)$, satisfy, if $n \geq rk$,

$$B_{n,k,(r)} = B_{n,k}(w_\bullet) + \sum_{l=1}^{k-1} \frac{(-1)^l}{l!} \sum_{\substack{1\le i_1,\dots,i_l\le r-1,\\ i_1+\cdots+i_l\le n-k+l}} B_{n-(i_1+\cdots+i_l),k-l}(w_\bullet)[n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!}. \tag{20}$$

3 Ordered sizes in the Gibbs partition

It is straightforward to obtain some explicit distributional results on the ordered sizes in the Gibbs partition of the form (2) in terms of the associated partial Bell polynomials. Denote the descending order statistics of the block sizes by $|A_{(1)}|, \dots, |A_{(|\Pi_n|)}|$, where $|A_{(1)}| \ge |A_{(2)}| \ge \cdots \ge |A_{(|\Pi_n|)}|$. The distribution of the number of blocks in (3) follows immediately (Gnedin and Pitman 2005):

$$\mathbb{P}(|\Pi_n| = k) = \sum_{\{m_\bullet:\sum m_j=k,\sum jm_j=n\}} \mathbb{P}(|\Pi_n|_j = m_j, 1\le j\le n) = v_{n,k}B_{n,k}(w_\bullet), \qquad 1\le k\le n.$$

The conditional distribution given the number of blocks is

$$\mathbb{P}(|\Pi_n|_j = m_j, 1\le j\le n||\Pi_n|=k) = \frac{n!}{B_{n,k}(w_\bullet)} \prod_{j=1}^{n} \left(\frac{w_j}{j!}\right)^{m_j} \frac{1}{m_j!}, \qquad 1\le k\le n. \tag{21}$$

In statistical mechanics this is a microcanonical Gibbs distribution function whose number of microstates of a block of size j is w_j (Berestycki and Pitman 2007). For Gibbs partitions the number of blocks is the sufficient statistics for v-weights. By virtue of the sufficiency, discussion on the ordered sizes reduces to enumeration of microstates of the microcanonical Gibbs distribution which fulfills a given condition. The definitions of the associated partial Bell polynomials introduced in the previous section were defined by such enumeration. For example, the distribution of the largest size conditioned by the number of blocks is

$$\mathbb{P}(|A_{(1)}| \le r||\Pi_n| = k) = \frac{B^{(r)}_{n,k}(w_\bullet)}{B_{n,k}(w_\bullet)}, \qquad 1\le k\le n, \quad n/k\le r\le n, \tag{22}$$

and $\mathbb{P}(|A_{(1)}| \le r||\Pi_n| = k) = 0$ for $1 \le r < n/k$. Note that the associated partial Bell polynomial, $B^{(r)}_{n,k}(w_\bullet)$, is the number of microstates in the microcanonical Gibbs distribution of the form (21) whose largest size is not larger than r. The marginal distributions of the ordered sizes have following representation.

Lemma 1 *In a Gibbs partition of the form* (2) *the marginal distributions of the ordered sizes are*

$$\mathbb{P}(|A_{(1)}| \le r) = \sum_{k=\lceil n/r \rceil}^{n} v_{n,k} B_{n,k}^{(r)}(w_\bullet), \tag{23}$$

$$\mathbb{P}(|A_{(i)}| \le r) = \sum_{k=1}^{i-1} v_{n,k} B_{n,k}(w_\bullet) + \sum_{k=i}^{n} v_{n,k} B_{n,k}^{(r),(i)}(w_\bullet), \qquad 2 \le i \le n,$$

and

$$\mathbb{P}(|A_{(|\Pi_n|)}| \ge r) = \sum_{k=1}^{\lfloor n/r \rfloor} v_{n,k} B_{n,k,(r)}(w_\bullet), \tag{24}$$

for $1 \le r \le n$*, where* $|A_{(i)}| = 0$ *if* $i > |\Pi_n|$.

Hence, discussion on the ordered sizes reduces to analysis of the partial Bell polynomials and the mixtures of them by v-weights. Distributions of the extremes are representable as composition of the exponential generating functions; substituting the exponential generating functions (18) and (16) into (23) and (24), respectively, yields

$$\mathbb{P}(|A_{(1)}| \le r) = \left[\frac{\xi^n}{n!}\right] \breve{v}_n(\breve{w}^{(r)}(\xi)), \tag{25}$$

and

$$\mathbb{P}(|A_{(|\Pi_n|)}| \ge r) = \left[\frac{\xi^n}{n!}\right] \breve{v}_n(\breve{w}_{(r)}(\xi)), \tag{26}$$

where $\breve{v}_n$ is the exponential generating function of $(v_{n,j})$-weights:

$$\breve{v}_n(\xi) := \sum_{j=1}^{\infty} v_{n,j} \frac{\xi^j}{j!}.$$

Remark 1 A Gibbs partition $Gibbs_{[n]}(v_\bullet, w_\bullet)$, in which v-weights are representable as ratios (5), has Kolchin's representation (Pitman 2006), which is identified with the collection of terms of random sum $X_1 + \cdots + X_{|\Pi_n|}$ conditioned on $\sum_i X_i = n$ with independent and identically distributed $X_1, X_2, \ldots$ independent of $|\Pi_n|$ (Kerov 1995). Here, the (ordinary) probability generating function of $X_\bullet$ is $\breve{w}(\xi)/\breve{w}(1)$ and the probability generating function of the random sum $X_1 + \cdots + X_{|\Pi_n|}$ is $\breve{v}(\breve{w}(\xi))/\breve{v}(\breve{w}(1))$. In Propositions 2 and 3 $\breve{w}_{(r)}$ and $\breve{w}^{(r)}$ are introduced by dropping terms from the sequence of w-weights. Therefore, $\breve{w}_{(r)}$ and $\breve{w}^{(r)}$ are the probability generating functions of defective distributions induced by a proper probability mass function $(w_\bullet/\bullet!)$. Suppose independent and identically distributed random variables $X_{(r)1}, X_{(r)2}, \ldots$, whose probability generating function are $w_{(r)}$, and $X_1^{(r)}, X_2^{(r)}, \ldots$ whose probability generating function are $w^{(r)}$. Then,

$$\mathbb{P}(|A_{(1)}| \le r) = \frac{\mathbb{P}(X_1^{(r)} + \cdots + X_{|\Pi_n|}^{(r)} = n)}{\mathbb{P}(X_1 + \cdots + X_{|\Pi_n|} = n)},$$

and

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) = \frac{\mathbb{P}(X_{(r)1} + \cdots + X_{(r)|\Pi_n|} = n)}{\mathbb{P}(X_1 + \cdots + X_{|\Pi_n|} = n)}.$$

Hence, distributions of the extreme sizes are the ratios of the probability mass of random sum at n in the defective distribution to that in the proper distribution.

4 Asymptotic behavior of extreme sizes

Asymptotic behavior of the extreme sizes in the Ewens–Pitman partition has been discussed in various contexts (see, for example, Shepp and Lloyd 1966; Watterson 1976; Griffiths 1988; Arratia and Tavare 1992; Pitman and Yor 1997; Panario and Richmond 2001; Arratia et al. 2003; Handa 2009). In this section we discuss asymptotic behavior of the extreme sizes in general Gibbs partitions of the form (2). The developed Lemma 1 is useful for keeping generality of our discussion, since it holds in any Gibbs partition of the form (2). Then, some explicit results for the consistent Gibbs partition, which is a class of Gibbs partitions whose w-weights have a form of (4), are presented. Subsequently, further explicit results are presented for two specific examples of consistent Gibbs partitions, the Ewens–Pitman partition and Gnedin's partition.

Explicit asymptotic forms appear in this section involve an extension of incomplete Dirichlet integrals, which involves a *Dirichlet distribution with negative parameters*. Let us prepare some notations. The probability density of a Dirichlet distribution of $b+1$ variables parametrized by two parameters $\rho > 0$ and $\nu > 0$ is

$$p(y_1, y_2, \ldots, y_{b+1}) = \frac{\Gamma(\rho + b\nu)}{\Gamma(\rho)\Gamma(\nu)^b} y_{b+1}^{\rho-1} \prod_{j=1}^{b} y_j^{\nu-1}, \qquad \sum_{j=1}^{b+1} y_j = 1,$$

whose support is the b-dimensional simplex $\Delta_b := \{y_i : 0 < y_i, 1 \leq i \leq b, \sum_{j=1}^{b} y_j < 1\}$. Incomplete Dirichlet integrals are usually defined in this setting (Sobel 1977). But let us introduce an integral with non-zero real parameters ρ and ν:

$$\mathcal{I}_{p,q}^{(b)}(\nu; \rho) := \frac{\Gamma(\rho + b\nu)}{\Gamma(\rho)\Gamma(\nu)^b} \int_{\Delta_b(p,q)} y_{b+1}^{\rho-1} \prod_{j=1}^{b} y_j^{\nu-1} dy_j,$$

with a convention

$$\mathcal{I}_{p,q}^{(b)}(0; \rho) := \int_{\Delta_b(p,q)} y_{b+1}^{\rho-1} \prod_{j=1}^{b} y_j^{-1} dy_j$$

and $\mathcal{I}_{p,q}^{(0)}(\nu; \rho) = 1$, where

$$\Delta_b(p,q) := \left\{ y_i : p < y_i, 1 \leq i \leq b; \sum_{j=1}^{b} y_j < 1 - q \right\}, \quad 0 < q < 1, \quad 0 < p < \frac{1-q}{b}.$$

Of course, when either of ρ and ν is negative the integral over the simplex Δ_b does not exist. But throughout this paper integrals involving this extension of incomplete Dirichlet integrals are well defined since the domain of integration, $\Delta_b(p,q)$, is appropriately chosen.

4.1 General distributional results

Let us begin with seeing asymptotic behavior of the smallest sizes in the Gibbs partition of the form (2). Lemma 1 and the Cauchy–Goursat theorem provide a way to evaluate it in terms of a contour integral. This kind of method to evaluate asymptotics, which is called the singularity analysis of generating functions in analytic combinatorics literature, has been quite popular in studies of random combinatorial structures (see, for example, Flajolet and Sedgewick 2009). Noting the expression (26) the distribution of the smallest size can be evaluated as

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) = \frac{n!}{2\pi\sqrt{-1}} \oint \frac{\breve{v}_n(\breve{w}_{(r)}(\xi))}{\xi^{n+1}} \mathrm{d}\xi, \qquad n \to \infty, \ r = o(n). \tag{27}$$

It is interesting to see the event that the smallest size is extremely large. To see the asymptotic behavior we need asymptotic forms of the v-weights and the associated partial Bell polynomials, $B_{n,k,(r)}(w_\bullet)$, in $n, r \to \infty$ with $r \asymp n$ and fixed k. Explicit results are immediately deduced by substituting these asymptotic forms into (24).

Asymptotic behavior of the largest size in the Gibbs partition of the form (2) can be discussed similarly. The expression (25) leads

$$\mathbb{P}(|A_{(1)}| \leq r) = \frac{n!}{2\pi\sqrt{-1}} \oint \frac{\breve{v}_n(\breve{w}^{(r)}(\xi))}{\xi^{n+1}} \mathrm{d}\xi, \qquad n \to \infty, \ r = o(n). \tag{28}$$

It is also interesting to see the event that the largest size is extremely small. Following lemma, whose proof is in Sect. 5.2, provides asymptotic expressions of the marginal distribution of the largest size in terms of the partial Bell polynomials.

Lemma 2 *In a Gibbs partition of the form* (2) *whose weights and induced partial Bell polynomials satisfy*

$$\frac{w_n}{n!} = O(n^{-1-\eta_1}), \quad n! v_{n,k} = O(n^{1-\eta_2(k)}), \qquad \frac{B_{n,k}(w)}{n!} = O(n^{-1-\eta_3(k)}), \ n \to \infty,$$

for fixed positive integer k, the largest size satisfies

$$\begin{aligned}\mathbb{P}(|A_{(1)}| \leq r) &= 1 + \sum_{l=1}^{\lfloor (n-\lceil n/r \rceil)/r \rfloor} \frac{(-1)^l}{l!} \sum_{\substack{i_1,\ldots,i_l \geq r+1, \\ i_1+\cdots+i_l \leq n-\lceil n/r \rceil + l}} [n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!} \\ &\quad \times \sum_{k=0}^{n-(i_1+\cdots+i_l)} v_{n,k+l} B_{n-(i_1+\cdots+i_l),k}(w_\bullet) + o(1), \quad n, r \to \infty, \ r \asymp n,\end{aligned} \tag{29}$$

if $l\eta_1+\eta_2(k)>0$ *and* $\eta_2(k)+\eta_3(k)>0$ *hold for* $1\le l\le\lfloor n/r\rfloor, 1\le k\le\lceil n/r\rceil-1$, *and* $\breve{w}(1)<\infty$.

4.2 The consistent Gibbs partition

Some explicit results are available for the consistent Gibbs partition, which is a class of Gibbs partitions whose w-weights have a form of (4). Let us begin with seeing asymptotic behavior of the smallest size conditioned by the number of blocks in the consistent Gibbs partition. For the case that the size is extremely large, $O(n)$, asymptotic forms of the associated partial Bell polynomials, $B_{n,k,(r)}(w_\bullet)$ with w-weights (4), in $n,r\to\infty$ with $r\asymp n$ and fixed k yield the distribution of the smallest size conditioned by the number of blocks immediately. The asymptotic forms are developed in Appendix B.

Proposition 6 *In a consistent Gibbs partition, which has the form of* (2) *with w-weights satisfies* (4), *the smallest size conditioned by the fixed number of blocks satisfies*

$$\mathbb{P}(|A_{(|\Pi_n|)}|\ge r||\Pi_n|=k)\sim\tilde{\omega}_\alpha(x,k),\qquad n,r\to\infty,\ r\sim xn,$$

where if $x^{-1}\ge k$ *then*

$$\tilde{\omega}_\alpha(x,k)=\frac{\Gamma(-\alpha)}{\Gamma(-k\alpha)}\frac{(-1)^{k-1}}{k}\mathcal{I}^{(k-1)}_{x,x}(-\alpha;-\alpha)n^{-(k-1)\alpha},\qquad 0<\alpha<1,$$
$$\tilde{\omega}_0(x,k)=\frac{(\log n)^{1-k}}{k}\mathcal{I}^{(k-1)}_{x,x}(0;0),\qquad \alpha=0,$$
$$\tilde{\omega}_\alpha(x,k)=\mathcal{I}^{(k-1)}_{x,x}(-\alpha;-\alpha),\qquad \alpha<0.$$

If $x^{-1}<k$, $\tilde{\omega}_\alpha(x,k)=0$ *for* $-\infty<\alpha<1$.

To derive explicit expressions of the marginal distributions asymptotic forms of the v-weights are needed. Substituting the asymptotic forms presented in Propositions 16 and 17 in Appendix B into (24) provides following Proposition.

Proposition 7 *In a consistent Gibbs partition, which has the form of* (2) *with w-weights satisfying* (4) *and v-weights satisfying $n!v_{n,k}=f_kO(n^{1-\eta_2(k)})$, $n\to\infty$, for fixed positive integer k, the smallest size satisfies*

$$\mathbb{P}(|A_{(|\Pi_n|)}|\ge r)\sim\sum_{k=1}^{\lfloor x^{-1}\rfloor}f_k\frac{n^{-\eta_2(k)}}{k!}\mathcal{I}^{(k-1)}_{x,x}(0;0),\qquad n,r\to\infty,\ r\sim xn,\ \alpha=0,$$

and

$$\mathbb{P}(|A_{(|\Pi_n|)}|\ge r)\sim\sum_{k=1}^{\lfloor x^{-1}\rfloor}f_k\frac{n^{-\eta_2(k)-k\alpha}}{(-\alpha)^k\Gamma(-k\alpha)k!}\mathcal{I}^{(k-1)}_{x,x}(-\alpha;-\alpha),\quad n,r\to\infty,\ r\sim xn,$$

for $\alpha\ne0$.

Asymptotic behavior of the largest size also admits some explicit expressions. Recall following theorem on the number of blocks in the Ewens–Pitman partition (Korwar and Hollander 1973; Pitman 1997), which is a member of the consistent Gibbs partitions. If v-weights of a consistent Gibbs partition are representable as (5), it coincides with the Ewens–Pitman partition (Gnedin and Pitman 2005). The EPPF satisfies

$$\mathbb{P}(|\Pi_n|_j = m_j, 1 \le j \le n) = \frac{(-1)^n}{(-\alpha)^k}\frac{(\theta)_{k;\alpha}}{(\theta)_n} n! \prod_{j=1}^{n} \binom{\alpha}{j}^{m_j} \frac{1}{m_j!}. \tag{30}$$

Theorem 1 (Korwar and Hollander 1973; Pitman 1999) *For* $0 < \alpha < 1$ *and* $\theta > -\alpha$ *the number of blocks in the Ewens–Pitman partition of the form* (30) *satisfies*

$$\frac{|\Pi_n|}{n^\alpha} \to S_\alpha, \qquad n \to \infty,$$

in almost surely and pth mean for every $p > 0$. *Here*, S_α *has the probability density*

$$\mathbb{P}(\mathrm{d}s) = \frac{\Gamma(1+\theta)}{\Gamma(1+\theta/\alpha)} s^{\frac{\theta}{\alpha}} g_\alpha(s)\mathrm{d}s,$$

where $g_\alpha(s)$ *is the probability density of the Mittag-Leffler distribution* (Pitman 2006). *For* $\alpha = 0$ *and* $\theta > 0$,

$$\frac{|\Pi_n|}{\log n} \to \theta, \qquad a.s., \ n \to \infty.$$

For $\alpha < 0$ *and* $\theta = -m\alpha$, $m = 1, 2, \ldots$, $|\Pi|_n = m$ *for all sufficiently large n almost surely.*

Theorem 1 provides us idea how the number of blocks should be scaled with n to see proper conditional distribution of the largest size given the number of blocks in general consistent Gibbs partitions. In fact, Proposition 5 and the asymptotic forms of the partial Bell polynomials with w-weights (4) given in Appendix B yield following results.

Proposition 8 *In a consistent Gibbs partition, which has the form of* (2) *with* w*-weights satisfying* (4) *and* $\alpha < 0$, *the largest size conditioned by the number of blocks satisfies*

$$\mathbb{P}(|A_{(1)}| \le r \,||\Pi_n| = k) \sim \tilde{\rho}_\alpha(x, k), \qquad n, r \to \infty, \ r \sim xn,$$

for fixed positive integer k, where if $x^{-1} \le k$ *then*

$$\tilde{\rho}_\alpha(x, k) = \sum_{0 \le j < x^{-1}} (-1)^j \binom{k}{j} \mathcal{I}_{x,0}^{(j)}(-\alpha; (j-k)\alpha),$$

and if $x^{-1} > k$, $\tilde{\rho}_\alpha(x, k) = 0$.

Remark 2 It may be natural to ask similar expressions for the case that $0 < \alpha < 1$ with $k = O(n^\alpha)$, but the author do not know such expressions. Substituting (19) into (22) yields

$$\begin{aligned}\mathbb{P}(|A_{(1)}| \le r||\Pi_n| = k) &= 1 + \sum_{l=1}^{\lfloor (n-k)/r \rfloor} \frac{\alpha^l}{l!} \sum_{\substack{i_1,\ldots,i_l \ge r+1,\\ i_1+\cdots+i_l \le n-k+l}} \\ &\quad \times \frac{C(n-(i_1+\cdots+i_l), k-l;\alpha)}{C(n,k;\alpha)(\Gamma(1-\alpha))^l} \\ &\quad \times (-1)^{i_1+\cdots+i_l}[n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{\Gamma(i_j-\alpha)}{\Gamma(i_j+1)}. \end{aligned} \tag{31}$$

The asymptotic form (41) in Appendix B of the generalized factorial coefficients for $n \to \infty$, $k \sim sn^\alpha$, and fixed l yields

$$\frac{C(n-(i_1+\cdots+i_l), k-l;\alpha)}{C(n,k;\alpha)} (-1)^{i_1+\cdots+i_l}[n]_{i_1+\cdots+i_l} \sim (1-k)_l \left(1 - \frac{i_1+\cdots+i_l}{n}\right)^{-1-\alpha}, \tag{32}$$

as long as $n - (i_1 + \cdots + i_l) \asymp n$. Substituting (32) into (31) yields an expression

$$\tilde{\rho}_\alpha(x,k) \sim \sum_{0 \le l < x^{-1}} \frac{\Gamma(-\alpha)}{\Gamma(-(l+1)\alpha)} \frac{s^l}{l!} \mathcal{I}^{(l)}_{x,0}(-\alpha;-\alpha), \quad n, r \to \infty,\ r \sim xn,\ k \sim sn^\alpha,$$

but the incomplete Dirichlet integrals are divergent. This is because (32) does not hold in the whole domain of the summation in (31). We have similar observation for $\alpha = 0$ with $k = O(\log n)$, where (42) in Appendix B is employed.

Remark 3 It is worth mentioned that these expressions give asymptotic forms of the associated partial Bell polynomials, $B^{(r)}_{n,k}((1-\alpha)_{\bullet-1})$, with $n, r \to \infty$, $r \asymp n$, and appropriately scaled k.

4.3 The Ewens–Pitman partition

If v-weights are specified in a consistent Gibbs partition further explicit results are available. In this subsection the v-weights of the form (7) or the Ewens–Pitman partition, is discussed. The Ewens–Pitman partition has nice properties and appears in various contexts (see, for example, Tavaré and Ewens 1997; Arratia et al. 2003; Pitman 2006).

4.3.1 The smallest size

In the Ewens–Pitman partition with $\alpha = 0$ and $\theta > 0$ multiplicities of the small components are asymptotically independent (10) and this property immediately leads following theorem on asymptotic behavior of the smallest size in the Ewens–Pitman partition (Arratia and Tavare 1992). But evaluating (27) also provides the theorem. We omit the proof because it is similar to the proof of Theorem 3.

Theorem 2 (Arratia and Tavare 1992) *In the Ewens–Pitman partition of the form* (30) *with* $\alpha = 0$ *and* $\theta > 0$ *the smallest size satisfies*

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim e^{-\theta h_{r-1}}, \qquad n \to \infty, \ r = o(n),$$

where $h_r = \sum_{k=1}^{r} k^{-1}$ *with a convention* $h_0 = 0$. *Moreover,* $\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim r^{-\theta} e^{-\gamma\theta}$, $n, r \to \infty$, $r = o(n)$, *where* γ *is the Euler–Mascheroni constant.*

A necessary condition of the asymptotic independence (10) is the conditioning relation (9) with the *logarithmic condition*:

$$j\mathbb{P}(Z_j = 1) \to \theta, \qquad j\mathbb{E}(Z_j) \to \theta, \ j \to \infty,$$

for some $\theta > 0$ (Arratia et al. 2003). In the Ewens–Pitman partition the logarithmic condition holds only if $\alpha = 0$. Therefore, application of the asymptotic independence (10) is restricted to the case that $\alpha = 0$. On the other hand, evaluating (27) is possible for non-zero α and gives the following theorem, whose proof is in Sect. 5.3.

Theorem 3 *In the Ewens–Pitman partition of the form* (30) *with* $0 < \alpha < 1$ *and* $\theta > -\alpha$ *the smallest size satisfies*

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)} \left\{ \sum_{j=1}^{r-1} p_\alpha(j) \right\}^{-1-\frac{\theta}{\alpha}} n^{-\theta-\alpha}, \qquad n \to \infty,$$

for $r = o(n)$, $r \geq 2$, *where*

$$p_\alpha(j) = \binom{\alpha}{j} (-1)^{j+1}, \qquad j = 1, 2, \ldots \tag{33}$$

Moreover,

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)} n^{-\theta-\alpha}, \qquad n, r \to \infty, \ r = o(n).$$

Remark 4 In the case $0 < \alpha < 1$ (33) is a probability mass function. Devroye called it Sibuya's distribution (Sibuya 1979; Devroye 1993).

Remark 5 Theorem 3 gives a convergence result:

$$|A_{(|\Pi_n|)}| = 1 + O_p(n^{-\theta-\alpha}), \qquad n \to \infty. \tag{34}$$

In addition, since

$$\mathbb{P}(|A_{(|\Pi_n|)}| = n) \sim \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)} n^{-\theta-\alpha}, \qquad n \to \infty,$$

we have $\mathbb{P}(r \le |A_{(|\Pi_n|)}| < n) = o(n^{-\theta-\alpha})$ as $n, r \to \infty, r \asymp n$. Therefore, apart from the mass at $|A_{(|\Pi_n|)}| = n$ probability mass concentrates around one. Study of asymptotic behavior of small sizes in an infinite exchangeable random partition goes back to works by Karlin (1967) and Rouault (1978). See also Yamato and Sibuya (2000), Pitman (2006). For an infinite exchangeable random partition $T_\alpha := \lim_{n\to\infty} |\Pi_n|/n^\alpha$, where $0 < \alpha < 1$, is an almost sure and strictly positive limit if and only if the ranked frequencies, $P_{(j)}, j = 1, 2, \ldots$ satisfies $P_{(j)} \sim Zj^{-1/\alpha}$ as $j \to \infty$ with $0 < Z < \infty$. In that case $Z^{-\alpha} = \Gamma(1-\alpha)T_\alpha$ and $|\Pi_n|_j \sim p_\alpha(j) T_\alpha n^\alpha$ as $n \to \infty$ for each j. The Poisson–Dirichlet process and the Ewens–Pitman partition with $0 < \alpha < 1$ and $\theta > -\alpha$ satisfy these conditions and T_α has the probability density of S_α defined in Theorem 1. By noting that $\{|\Pi_n|_1 = 0\} = \{|A_{(|\Pi_n|)}| > 1\}$ it can be seen that $|\Pi_n|_1 \sim \alpha S_\alpha n^\alpha$, which is consistent with (34).

The next proposition, whose proof is given in Sect. 5.3, implies that for $\alpha < 0$ and $\theta = -m\alpha, m = 1, 2, \ldots$, the smallest size is $\Omega(n)$ in probability.

Proposition 9 *In the Ewens–Pitman random partition of the form* (30) *with* $\alpha < 0$ *and* $\theta = -m\alpha, m = 2, 3, \ldots$ *the smallest size satisfies*

$$\mathbb{P}(|A_{(|\Pi_n|)}| \ge r) \sim 1 + \frac{m\Gamma(-m\alpha)}{\Gamma((1-m)\alpha)} \left\{ \sum_{j=1}^{r-1} p_\alpha(j) \right\} n^\alpha, \qquad n \to \infty,$$

for $r = o(n), r \ge 2$, *where* $p_\alpha(\bullet)$ *is defined by* (33). *Moreover*, $\mathbb{P}(|A_{(|\Pi_n|)}| \ge r) \sim 1 - O((n/r)^\alpha), n, r \to \infty, r = o(n)$.

Let us see a large deviation, where the smallest size is extremely large. Using the conditioning relation (9), Arratia et al. (2003) established following assertion for the Ewens–Pitman partition with $\alpha = 0$ and $\theta > 0$. But the assertion is the direct consequence of Proposition 7.

Corollary 1 (Arratia et al. 2003) *In the Ewens–Pitman partition of the form* (30) *with* $\alpha = 0$ *and* $\theta > 0$ *the smallest size satisfies*

$$\mathbb{P}(|A_{(|\Pi_n|)}| \ge r) \sim \Gamma(\theta)(xn)^{-\theta}\omega_\theta(x), \qquad n, r \to \infty, \ r \sim xn,$$

where

$$\omega_\theta(x) = x^\theta \sum_{k=1}^{\lfloor x^{-1} \rfloor} \frac{\theta^k}{k!} \mathcal{I}_{x,x}^{(k-1)}(0; 0).$$

In addition, Proposition 7 immediately gives asymptotic behavior of the smallest size in the Ewens–Pitman partition for non-zero α.

Corollary 2 *In the Ewens–Pitman partition of the form* (30) *with* $0 < \alpha < 1$ *and* $\theta > -\alpha$ *the smallest size satisfies asymptotically*

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)} n^{-\theta-\alpha}, \qquad n, r \to \infty, \; r \sim xn.$$

For $\alpha < 0$ *and* $\theta = -m\alpha$, $m = 1, 2, \ldots$,

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim \mathcal{I}_{x,x}^{(m-1)}(-\alpha; -\alpha), \qquad x^{-1} \geq m,$$

and $\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) = O(n^{(m-\lfloor x^{-1} \rfloor)\alpha})$ *for* $x^{-1} < m$.

Remark 6 In Corollary 1 the function $\omega_1(x^{-1})$ is known as Buchstab's function for the frequency of rough numbers in the number theory (Buchstab 1937; Tenenbaum 1995). Corollary 2 shows that we do not have "generalized Buchstab's function" for $0 < \alpha < 1$ and $\theta > -\alpha$.

The factorial moments of the smallest size satisfy

$$\mathbb{E}([|A_{(|\Pi_n|)}|]_i) = i \sum_{j=i}^{n} [j-1]_{i-1} \mathbb{P}(|A_{(|\Pi_n|)}| \geq j), \qquad i = 1, 2, \ldots$$

Theorem 3, Corollaries 1 and 2 provide an implication on the factorial moments of the smallest size, which might be important in statistical applications. The proof is given in Sect. 5.3.

Theorem 4 *In the Ewens–Pitman partition of the form* (30) *with* $\alpha \geq 0$ *and* $\theta > -\alpha$ *the* i*th factorial moments of the smallest size exist if and only if* $\theta \geq i - \alpha$, $i = 1, 2, \ldots$ (*for* $\alpha = 0$, $\theta > i$). *Moreover, if* $\alpha > 0$ *and* $\theta > i - \alpha$,

$$\mathbb{E}([|A_{(|\Pi_n|)}|]_i) \sim \delta_{i,1}, \qquad i = 1, 2, \ldots$$

For $\alpha < 0$ *and* $\theta = m\alpha$, $m = 1, 2, \ldots$,

$$\mathbb{E}\left(\frac{[|A_{(|\Pi_n|)}|]_i}{n^i}\right) \sim i \int_0^{m^{-1}} x^{i-1} I_{x,x}^{(m-1)}(-\alpha; -\alpha) \mathrm{d}x, \qquad n \to \infty, \; i = 1, 2, \ldots$$

Remark 7 In applications it might be interesting to consider a test of the hypothesis "$\alpha > 0$", since properties of the Ewens–Pitman partition crucially depend on sign of α. For example, for $\alpha \leq 0$ the partition is a one-parameter family, but it is not for $\alpha > 0$. According to Theorems 2 and 3 if $\alpha > 0$ and $\theta > -\alpha$, $|A_{(|\Pi_n|)}| \xrightarrow{p} 1$, $n \to \infty$, while if $\alpha = 0$ and $\theta > 0$, $\mathbb{P}(|A_{(|\Pi_n|)}| > 1) \sim e^{-\theta}$. In addition, Proposition 9 implies that $\mathbb{P}(|A_{(|\Pi_n|)}| > 1) \to 1$, $n \to \infty$, for $\alpha < 0$, $\theta = -m\alpha$, $m = 1, 2, \ldots$ Therefore, we can reject the hypothesis "$\alpha > 0$" by the event $\{|A_{(|\Pi_n|)}| > 1\}$. This test seems powerful, however, might be unstable since power of the test increases with decreasing θ, while the moments exist only for large θ. For an illustration Table 1

Table 1 Simulation results for the number of the event $\{|A_{(|\Pi_n|)}| > 1\}$ occurred in the Ewens–Pitman partition based on 10,000 trials with $n = 100$

θ	α						
	0.9	0.5	0.1	0	−0.1	−0.5	−1
−0.01	15	1162	8699	–	–	–	–
0	12	1103	8551	–	–	–	–
0.01	15	1029	8416	9909	–	–	–
0.1	14	785	7358	9042	10,000	–	–
0.5	1	161	4082	5961	7661	10,000	–
1	0	43	1003	3610	5391	9426	10,000
5	0	0	21	82	244	3372	8238

displays simulation results for the number of the event $\{|A_{(|\Pi_n|)}| > 1\}$ occurred in 10,000 trials with $n = 100$.

4.3.2 The largest size

Asymptotic behavior of the marginal distribution of the largest size in the Ewens–Pitman partition follows immediately from Lemma 2. For $0 \le \alpha < 1$ and $\theta > -\alpha$ the distribution is identical with the marginal distribution of the first component of the Poisson–Dirichlet distribution, and the assertions of the following corollary have been established in studies of the Poisson–Dirichlet process (see, for example, Watterson 1976; Griffiths 1988; Pitman and Yor 1997; Arratia and Tavare 1992; Arratia et al. 2003; Handa 2009). Our proof, which is provided in Sect. 5.3, seems simpler than the treatments of the Poisson–Dirichlet process.

Corollary 3 *In the Ewens–Pitman partition of the form* (30) *the largest size satisfies*

$$\mathbb{P}(|A_{(1)}| \le r) \sim \rho_{\alpha,\theta}(x), \qquad n, r \to \infty, \qquad r \sim xn,$$

where

$$\rho_{\alpha,\theta}(x) = \sum_{k=0}^{\lfloor x^{-1} \rfloor} \frac{(\theta)_{k;\alpha}}{\alpha^k k!} \mathcal{I}_{x,0}^{(k)}(-\alpha; k\alpha + \theta), \qquad 0 < \alpha < 1,\ \theta > -\alpha,$$

$$\rho_{0,\theta}(x) = \sum_{k=0}^{\lfloor x^{-1} \rfloor} \frac{(-\theta)^k}{k!} \mathcal{I}_{x,0}^{(k)}(0; \theta), \qquad \alpha = 0,\ \theta > 0.$$

For $\alpha < 0$ and $\theta = -m\alpha$, $m = 1, 2, \ldots,$

$$\rho_{\alpha,(-m\alpha)}(x) = \sum_{k=0}^{\lfloor x^{-1} \rfloor} (-1)^k \binom{m}{k} \mathcal{I}_{x,0}^{(k)}(-\alpha; (k-m)\alpha), \qquad \lceil x^{-1} \rceil \le m$$

and $\rho_{\alpha,(-m\alpha)}(x) = 0$, $\lceil x^{-1} \rceil > m$.

Remark 8 The function $\rho_{0,1}(x^{-1})$ is known as Dickman's function for the frequency of smooth numbers in the number theory (Dickman 1930; Tenenbaum 1995).

Let us move to the situation that the largest size is extremely small. For $\alpha < 0$ and $\theta = -m\alpha$, $m = 1, 2, \ldots$, the largest size is $\Omega(n)$ almost surely, because the expression (25) implies

$$\mathbb{P}(|A_{(1)}| \le r) = \sum_{k=1}^{m} \frac{[m]_k}{(-m\alpha)_n} \left[\frac{\xi^n}{n!}\right] \left\{ \sum_{j=0}^{r} \binom{\alpha}{j} (-\xi)^j \right\}^k = 0, \qquad r < \frac{n}{m}. \tag{35}$$

For the case that $0 < \alpha < 1$ and $\theta > -\alpha$ evaluation of (28) leads following theorem, whose proof is given in Sect. 5.3. According to the theorem in the Ewens–Pitman partition with $0 < \alpha < 1$ the probability that the largest size is $o(n)$ decays exponentially.

Theorem 5 *In the Ewens–Pitman partition of the form* (30) *with* $0 < \alpha < 1$ *and* $\theta > -\alpha$ *the largest size satisfies*

$$\mathbb{P}(|A_{(1)}| \le r) \sim \frac{\Gamma(\theta)}{\Gamma\left(\frac{\theta}{\alpha}\right)} \left\{-\rho_r f_r'(\rho_r)\right\}^{-\frac{\theta}{\alpha}} \rho_r^{-n-1} n^{\frac{\theta}{\alpha}-\theta}, \qquad n \to \infty, \ r = o(n),$$

where

$$f_r(\xi) = \sum_{j=0}^{r} \binom{\alpha}{j} (-\xi)^j,$$

$f_r'(\xi) = \mathrm{d} f_r(\xi)/\mathrm{d}\xi$ *and* ρ_r *is the unique real positive root of the equation* $f_r(\xi) = 0$. *Moreover,*

$$\mathbb{P}(|A_{(1)}| \le r) \sim \frac{\Gamma(\theta)}{\Gamma\left(\frac{\theta}{\alpha}\right)} \left(\frac{\alpha}{\Gamma(2-\alpha)}\right)^{-\frac{\theta}{\alpha}} e^{-\frac{1-\alpha}{\alpha}\frac{n}{r}} \left(\frac{n}{r}\right)^{\frac{\theta}{\alpha}-\theta}, \quad n, r \to \infty, \ r = o(n).$$

Although less explicit, a similar result is available for the case that $\alpha = 0$. The proof is given in Sect. 5.3.

Proposition 10 *In the Ewens–Pitman partition of the form* (30) *with* $\alpha = 0$ *and* $\theta > 0$ *the probability that the largest size is* $o(n)$ *is exponentially small in* n.

The factorial moments of the largest size satisfy

$$\mathbb{E}([|A_{(1)}|]_i) = [n]_i - i \sum_{j=i}^{n} [j-1]_{i-1} \mathbb{P}(|A_{(1)}| \le j-1), \qquad i = 1, 2, \ldots$$

Corollary 3 readily gives explicit expressions of asymptotic forms of the factorial moments of the largest size.

Corollary 4 *In the Ewens–Pitman partition of the form* (30) *the largest size satisfies*

$$\mathbb{E}\left(\frac{[|A_{(1)}|]_i}{n^i}\right) \sim 1 - i\mu_{\alpha,\theta}^{(i-1)}, \qquad n \to \infty,\ i = 1, 2, \ldots,$$

where $\mu_{\alpha,\theta}^{(i)}$ *is the* i*th moments of* $\rho_{\alpha,\theta}(x)$*, which is defined in Corollary* 3 *and an explicit expression is given in Proposition 17 of* Pitman and Yor (1997).

4.4 Gnedin's partition

The partition proposed by Gnedin (2010) is a randomized version of a symmetric Dirichlet-multinomial distribution. Gnedin's partition is consistent, but the v-weights, which are given in (8), are not representable by ratios (5). In applications it is desirable to have exchangeable random partitions of integer with finite but random number of blocks. The partition is a two-parameter family obtained by mixing of the Ewens–Pitman partition of parameter $(\alpha, \theta) = (-1, m)$ over m. Here, the Ewens–Pitman partition of parameter $(\alpha, \theta) = (-1, m)$ is the Dirichlet-multinomial distribution over the m-dimensional simplex. Gnedin (2010) showed that Gnedin's partition is a mixture with a mixing distribution of $|\Pi_\infty| := \lim_{n\to\infty} |\Pi_n|$, where

$$\mathbb{P}(|\Pi_\infty| = m) = B(z_1, z_2)\frac{(s_1)_m(s_2)_m}{m!(m-1)!}, \qquad m = 1, 2, \ldots \tag{36}$$

with $z_1z_2 = s_1s_2 = \zeta$ and $z_1+z_2 = -(s_1+s_2) = \gamma$. It is possible to obtain asymptotic behavior of Gnedin's partition by by mixing asymptotic forms for the Ewens–Pitman partition of $(\alpha, \theta) = (-1, m)$, $m = 1, 2, \ldots$, over m, while asymptotic behavior can be addressed directly by analyzing the generating functions. According to Proposition 9 and (35) the extreme sizes in Gnedin's partition are asymptotically $\Omega(n)$. For the smallest size, the following corollary is a direct consequence of Proposition 7.

Corollary 5 *In Gnedin's partition, whose* w*-weights satisfy* (4) *with* $\alpha = -1$ *and* v*-weights satisfy* (8)*, the smallest size satisfies*

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim B(z_1, z_2)\sum_{m=1}^{\lfloor x^{-1}\rfloor}\frac{(s_1)_m(s_2)_m}{m!(m-1)!}(1-mx)^{m-1},\ n, r \to \infty,\ r \sim xn,$$

where $z_1z_2 = s_1s_2 = \zeta$ *and* $z_1 + z_2 = -(s_1 + s_2) = \gamma$*. The smallest size satisfies*

$$\begin{aligned}\mathbb{E}\left(\frac{[|A_{(|\Pi_n|)}|]_j}{n^j}\right) &\sim j!B(z_1, z_2)\sum_{m=1}^{\infty}\frac{(s_1)_m(s_2)_m}{m!(m+j-1)!m^j}\\ &= \mathbb{E}\left\{|\Pi_\infty|^{-j}\binom{|\Pi_\infty|+j-1}{j}^{-1}\right\},\end{aligned}$$

for $n \to \infty$, $j = 1, 2, \ldots$

For the largest size, Corollary 3 immediately gives following Corollary.

Corollary 6 *In Gnedin's partition, whose w-weights satisfy* (4) *with* $\alpha = -1$ *and v-weights satisfy* (8), *the largest size satisfies*

$$\mathbb{P}(|A_{(1)}| \le r) \sim B(z_1, z_2) \sum_{m=\lceil x^{-1}\rceil}^{\infty} \frac{(s_1)_m (s_2)_m}{m!(m-1)!} \rho_{-1,m}(x), \qquad n, r \to \infty, \ r \sim xn,$$

where $z_1 z_2 = s_1 s_2 = \zeta$, $z_1 + z_2 = -(s_1 + s_2) = \gamma$, and $\rho_{-1,m}(x)$ is defined in Corollary 3. The largest size satisfies

$$\mathbb{E}\left(\frac{[|A_{(1)}|]_j}{n^j}\right) \sim 1 - B(z_1, z_2) j \sum_{m=2}^{\infty} \frac{(s_1)_m (s_2)_m}{m!(m-1)!} \mu_{-1,m}^{(j-1)}, \qquad n \to \infty, \ j = 1, 2, \ldots$$

where $\mu_{-1,m}^{(j)}$ are the jth moments of $\rho_{-1,m}(x)$.

5 Proofs

5.1 Proofs for the associated partial Bell polynomials

Proof of Proposition 4 The event that the ith largest size is not larger than r consists of the disjoint events that all sizes are equal to or smaller than r, and the j sizes with sum m are larger than $r+1$ and remaining sizes are equal to or smaller than r, where $1 \le j \le i-1$. Consequently, we have

$$\begin{aligned} B_{n,k}^{(r),(i)}(w_\bullet) &= B_{n,k}^{(r)}(w_\bullet) \\ &\quad + \sum_{j=1}^{i-1} \sum_{\substack{m=(r+1)j \\ \vee\{n-r(k-j)\}}}^{n-k+j} \binom{n}{m} B_{m,j,(r+1)}(w_\bullet) B_{n-m,k-j}^{(r)}(w_\bullet). \end{aligned}$$

Summing up both hand sides of the equation in n with multiplying $u^n/n!$ the first term in the right-hand side gives $(\breve{w}^{(r)}(u))^k/k!$. The second term in the right-hand side is

$$\begin{aligned} &\sum_{j=1}^{i-1} \sum_{n=k}^{\infty} \sum_{\substack{m=(r+1)j \\ \vee\{n-r(k-j)\}}}^{n-k+j} B_{m,j,(r+1)}(w_\bullet) \frac{u^m}{m!} B_{n-m,k-j}^{(r)}(w_\bullet) \frac{u^{n-m}}{(n-m)!} \\ &= \sum_{j=1}^{i-1} \sum_{m=(r+1)j}^{\infty} \sum_{l=k-j}^{r(k-j)} B_{m,j,(r+1)}(w_\bullet) \frac{u^m}{m!} B_{l,k-j}^{(r)}(w_\bullet) \frac{u^l}{l!} \\ &= \sum_{j=1}^{i-1} \frac{(\breve{w}_{(r+1)}(u))^j}{j!} \frac{(\breve{w}^{(r)}(u))^{k-j}}{(k-j)!}, \end{aligned}$$

where the indexes are changed as $l = n - m$. □

Proof of Proposition 5 The binomial expansion of the left hand side of (18) yields

$$B_{n,k}^{(r)}(w_\bullet) = [u^n]\frac{n!}{k!}(\breve{w}^{(r)}(u))^k = [u^n]\frac{n!}{k!}\left(\breve{w}(u) - \sum_{j=r+1}^{\infty} w_i \frac{u^i}{i!}\right)^k$$
$$= B_{n,k}(w_\bullet) + n! \sum_{l=1}^{\lfloor (n-k)/r \rfloor} \frac{(-1)^l}{l!} \sum_{\substack{i_1,\ldots,i_l \geq r+1, \\ i_1+\cdots+i_l \leq n-k+l}} [u^{n-(i_1+\cdots+i_l)}] \frac{(\breve{w}(u))^{k-l}}{(k-l)!} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!}.$$

But noting that the exponential generating function (12) gives

$$[u^{n-(i_1+\cdots+i_l)}]\frac{(\breve{w}(u))^{k-l}}{(k-l)!} = \frac{B_{n-(i_1+\cdots+i_l),k-l}(w_\bullet)}{(n-(i_1+\cdots+i_l))!}$$

for $n-(i_1+\cdots+i_l) \geq k-l$, we establish the expression (19). The expression (20) can be established in the same manner. □

5.2 Proof for the Gibbs partition

Proof of Lemma 2 By virtue of the identity (19), (23) yields

$$\mathbb{P}(|A_{(1)}| \leq r) = \sum_{k=\lceil n/r \rceil}^{n} v_{n,k} B_{n,k}(w_\bullet) + \sum_{k=\lceil n/r \rceil}^{n-r} v_{n,k} \sum_{l=1}^{\lfloor ((n-k)/r \rfloor}$$
$$\times \frac{(-1)^l}{l!} \sum_{\substack{i_1,\ldots,i_l \geq r+1, \\ i_1+\cdots+i_l \leq n-k+l}} B_{n-(i_1+\cdots+i_l),k-l}(w_\bullet)[n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!}.$$

By changing order of the summations we have

$$\mathbb{P}(|A_{(1)}| \leq r) = 1 + \sum_{l=1}^{\lfloor (n-\lceil n/r \rceil)/r \rfloor} \frac{(-1)^l}{l!} \sum_{\substack{i_1,\ldots,i_l \geq r+1, \\ i_1+\cdots+i_l \leq n-\lceil n/r \rceil+l}} [n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!}$$
$$\times \sum_{k=0}^{n-(i_1+\cdots+i_l)} v_{n,k+l} B_{n-(i_1+\cdots+i_l),k}(w_\bullet) - R_1 - R_2,$$

where $R_1 := \sum_{k=1}^{\lceil n/r \rceil - 1} v_{n,k} B_{n,k}(w_\bullet)$ and

$$R_2 := \sum_{l=1}^{\lfloor (n-\lceil n/r \rceil)/r \rfloor} \frac{(-1)^l}{l!} \sum_{\substack{i_1,\ldots,i_l \geq r+1, \\ i_1+\cdots+i_l \leq n-\lceil n/r \rceil+l}} [n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!}$$
$$\sum_{k=0}^{\lceil n/r \rceil - l - 1} v_{n,k+l} B_{n-(i_1+\cdots+i_l),k}(w_\bullet).$$

$R_1 = o(1)$ follows immediately. For R_2, let us consider the series

$$\tilde{R}_2 := \sum_{\substack{i_1,\ldots,i_l \ge r+1,\\ i_1+\cdots+i_l \le n-\lceil n/r \rceil + l}} [n]_{i_1+\cdots+i_l} \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!} \sum_{k=0}^{\lceil n/r \rceil - l - 1} v_{n,k+l} B_{n-(i_1+\cdots+i_l),k}(w_\bullet).$$

Note that all terms are non-negative. By the assumption for the weights, we can take some positive real number c such that

$$\tilde{R}_2 \le c n^{-l(1+\eta_1)+1} \sum_{\substack{i_1,\ldots,i_l \ge r+1,\\ i_1+\cdots+i_l \le n-\lceil n/r \rceil + l}} \sum_{k=0}^{\lceil n/r \rceil - l - 1} n^{-\eta_2(k+l)} \frac{B_{n-(i_1+\cdots+i_l),k}(w_\bullet)}{\{n-(i_1+\cdots+i_l)\}!}.$$

The first summation can be indexed by $m = n - (i_1 + \cdots + i_l)$ and the right-hand side is bounded by

$$c n^{-l(1+\eta_1)+1} \sum_{m=\lceil n/r \rceil - l}^{n-(r+1)l} \binom{n-m-rl-1}{l-1} \sum_{k=0}^{\lceil n/r \rceil - l - 1} n^{-\eta_2(k+l)} \frac{B_{m,k}(w_\bullet)}{m!}$$
$$< c' n^{-l\eta_1} \sum_{k=0}^{\lceil n/r \rceil - l - 1} n^{-\eta_2(k+l)} \sum_{m=\lceil n/r \rceil - l}^{n-(r+1)l} \frac{B_{m,k}(w_\bullet)}{m!} < c' n^{-l\eta_1} \sum_{k=l}^{\lceil n/r \rceil - 1} n^{-\eta_2(k)} \frac{(\breve{w}(1))^{k-l}}{(k-l)!},$$

where c' is a positive real number. Using assumptions for η_1, $\eta_2(k)$, and $\breve{w}(1)$, we establish $R_2 = o(1)$ and the assertion follows. □

5.3 Proofs for the Ewens–Pitman partition

Proof of Theorem 3 Let us evaluate (27):

$$\mathbb{P}(|A_{(|\Pi_n|)}| \ge r) = \frac{n!}{(\theta)_n} \frac{1}{2\pi\sqrt{-1}} \oint \frac{\{h_r(\xi)\}^{-\frac{\theta}{\alpha}}}{\xi^{n+1}} \mathrm{d}\xi,$$

where

$$h_r(\xi) := (1-\xi)^\alpha + 1 - f_{r-1}(\xi), \qquad f_r(\xi) := \sum_{j=0}^{r} \binom{\alpha}{j} (-\xi)^j.$$

There is no root of the equation $h_r(\xi) = 0$ in $|\xi| \le 1$, since

$$|h_r(\xi) - 1| \le \sum_{j=r}^{\infty} \binom{\alpha}{j} (-1)^{j-1} \le \sum_{j=2}^{\infty} \binom{\alpha}{j} (-1)^{j-1} = 1 - \alpha < 1,$$

for $|\xi| \leq 1$. The contour of the Cauchy integral is the contour introduced in the proof of Theorem 5 with replacing the branch at $\xi = \rho_r$ by the branch at $\xi = 1$. As does in the proof of Theorem 5 contribution to the integral comes from the integral along a contour $\mathcal{H}$ with changing the variable $\xi = 1 + t/n$. We have

$$\begin{aligned}
&\oint_{\mathcal{H}} \left(1+\frac{t}{n}\right)^{-n-1} \left\{\left(-\frac{t}{n}\right)^{\alpha} - \sum_{j=1}^{r-1} \binom{\alpha}{j} \left(-1-\frac{t}{n}\right)^{j}\right\}^{-\frac{\theta}{\alpha}} \frac{\mathrm{d}t}{n} \\
&= (1-f_{r-1}(1))^{-\frac{\theta}{\alpha}} \\
&\quad \times \oint_{\mathcal{H}} e^{-t} \left\{1 - \frac{\theta}{\alpha}(1-f_{r-1}(1))^{-1}\left(-\frac{t}{n}\right)^{\alpha} + O(n^{(-1)\vee(-2\alpha)})\right\} \frac{\mathrm{d}t}{n} \\
&= (1-f_{r-1}(1))^{-1-\frac{\theta}{\alpha}} \left(-\frac{\theta}{\alpha}\right) n^{-1-\alpha} \oint_{\mathcal{H}} e^{-t}(-t)^{\alpha}\mathrm{d}t + O(n^{(-2)\vee(-1-2\alpha)}),
\end{aligned}$$

where the first term of the integrand in the second line vanishes. Extending the rectilinear part of the contour $\mathcal{H}$ towards $+\infty$ gives a new contour $\mathcal{H}'$, and the process introduces only exponentially small terms in the integral. Using the Hankel representation of the gamma function:

$$\frac{1}{2\pi\sqrt{-1}} \oint_{\mathcal{H}'} e^{-x}(-x)^{-z}\mathrm{d}x = \frac{1}{\Gamma(z)}, \tag{37}$$

the first assertion follows. The second assertion follows immediately since $p_\alpha(j)$, $j = 1, 2, \ldots$, is a probability mass function. □

Proof of Proposition 9 The identity (20) yields

$$\begin{aligned}
&C_r(n,k;\alpha) - C(n,k;\alpha) \\
&= n! \sum_{l=1}^{k-1} \frac{(-1)^l}{l!} \sum_{\substack{1\leq i_1,\ldots,i_l\leq r-1,\\ i_1+\cdots+i_l\leq n-k+l}} \frac{C(n-(i_1+\cdots+i_l),k-l;\alpha)}{(n-(i_1+\cdots+i_l))!} \prod_{j=1}^{l} \binom{\alpha}{i_j}.
\end{aligned}$$

Substituting the asymptotic form presented in Propositions 15 in Appendix B into the right-hand side yields

$$\frac{C_r(n,k;\alpha) - C(n,k;\alpha)}{n!} \sim \frac{(-1)^n n^{-1-(k-1)\alpha}}{\Gamma((1-k)\alpha)(k-1)!} \sum_{j=1}^{r-1} p_\alpha(j), \quad n\to\infty,\ r = o(n).$$

Substituting this expression into (24) and using the identity (15),

$$\mathbb{P}(|A_{(|\Pi_n|)}| \geq r) \sim 1 + \frac{m\Gamma(-m\alpha)}{\Gamma((1-m)\alpha)} \left\{\sum_{j=1}^{r-1} p_\alpha(j)\right\} n^{\alpha}, \qquad n\to\infty,\ r = o(n),$$

which establishes the first assertion. For sufficiently large r_0, let

$$\sum_{j=1}^{r-1} p_\alpha(j) = \sum_{j=1}^{r_0-1} p_\alpha(j) + \sum_{j=r_0}^{r-1} p_\alpha(j). \tag{38}$$

The first sum is bounded as

$$\left|\sum_{j=1}^{r_0-1} p_\alpha(j)\right| = \sum_{j=1}^{r_0-1}\prod_{k=1}^{j}\left(1-\frac{\alpha+1}{k}\right) \leq \sum_{j=1}^{r_0-1}\left\{\max_{k=1,\dots,j}\left(1-\frac{\alpha+1}{k}\right)\right\}^j,$$

where the maximum is less than 1 for $-1<\alpha<0$ and $(-\alpha)$ for $\alpha\leq -1$. The second sum is

$$\sum_{j=r_0}^{r-1} p_\alpha(j) = -\frac{1}{\Gamma(-\alpha)}\sum_{j=r_0}^{r-1}\frac{\Gamma(j-\alpha)}{\Gamma(j+1)} = -\frac{r^{-\alpha}}{\Gamma(1-\alpha)}(1+O(r_0^{-1})),$$

where the last equality follows by $\Gamma(j-\alpha)/\Gamma(j+1)\sim j^{-\alpha-1}$ for $j\geq r_0\to\infty$. It is possible to take the limit $r>r_0\to\infty$ such that the second sum in (38) dominates and the second assertion follows. □

Proof of Theorem 4 The assertion for $\alpha<0$ and $\theta=-m\alpha$, $m=1,2,\dots$ follows immediately from Corollary 2. Let us consider the case that $\alpha>0$ and $\theta>-\alpha$. The expectation satisfies

$$1+(n-1)\mathbb{P}(|A_{(|\Pi_n|)}|=n) < \mathbb{E}(|A_{(|\Pi_n|)}|) < 1+(n-1)\mathbb{P}(|A_{(|\Pi_n|)}|\geq 2).$$

Theorem 3 implies $\mathbb{E}(|A_{(|\Pi_n|)}|) = 1+O(n^{-\theta-\alpha+1})$ as $n\to\infty$. The $i\,(\geq 2)$th moments satisfy

$$i\sum_{j=i}^{n}[j-1]_{i-1}\mathbb{P}(|A_{(|\Pi_n|)}|=n) < \mathbb{E}([|A_{(|\Pi_n|)}|]_i) < i\sum_{j=i}^{n}[j-1]_{i-1}\mathbb{P}(|A_{(|\Pi_n|)}|\geq 2).$$

Since $i\sum_{j=i}^n[j-1]_{i-1}=[n]_i$, $\mathbb{E}([|A_{(|\Pi_n|)}|]_i)=O(n^{-\theta-\alpha+i})$ as $n\to\infty$ and the assertion follows. For the case that $\alpha=0$ and $\theta>0$, $\mathbb{E}(|A_{(|\Pi_n|)}|)=\sum_{j=1}^n\mathbb{P}(|A_{(|\Pi_n|)}|\geq j)$. Corollary 1 provides an estimate

$$\mathbb{E}(|A_{(|\Pi_n|)}|) = \sum_{j=1}^{n}\mathbb{P}(|A_{(|\Pi_n|)}|\geq j) \sim n^{1-\theta}\Gamma(\theta)\int_1^n u^{\theta-2}\omega_\theta(u)du, \qquad n\to\infty,$$

where $u=n/j$. Since the generalized Buchstab's function satisfies (Arratia et al. 2003)

$$\frac{d}{du}\{u^\theta\omega_\theta(u)\} = \theta(u-1)^{\theta-1}\omega_\theta(u-1), \qquad u>2,$$

$\omega_\theta(u) \sim c$ as $u \to \infty$ for some $c > 0$. If $\theta \geq 1$ the integral grows as $O(n^{\theta-1})$, while if $\theta < 1$ the integral converges. Therefore, $\mathbb{E}(|A_{(|\Pi_n|)}|) < \infty$ if $\theta > 1$. The assertion for the i (≥ 2)th moments can be established in the similar manner to the argument for the case that $\alpha > 0$ and $\theta > -\alpha$. □

Proof of Corollary 3 For $0 < \alpha < 1$ and $\theta > -\alpha$ the assumptions of Lemma 2 are satisfied since $\eta_1 = \alpha$, $\eta_2(k) = \theta$, $\eta_3(k) = \alpha$ by the asymptotic form in Proposition 15 in Appendix B and $\breve{w}(1) = 1/\alpha$. We have

$$\begin{aligned}
&\sum_{k=0}^{n-(i_1+\cdots+i_l)} v_{n,k+l} B_{n-(i_1+\cdots+i_l),k}(w_\bullet) \prod_{j=1}^{l} \frac{w_{i_j}}{i_j!} \\
&= \frac{(-1)^n}{(-\alpha)^l} \frac{(\theta)_{l;\alpha}}{(\theta)_n} \prod_{i=1}^{l} \binom{\alpha}{i_j} \sum_{k=0}^{n-(i_1+\cdots+i_l)} \left[-\frac{\theta}{\alpha} - l\right]_k C(n-(i_1+\cdots+i_l), k; \alpha) \\
&= \frac{(-1)^n}{(-\alpha)^l} \frac{(\theta)_{l;\alpha}}{(\theta)_n} \prod_{i=1}^{l} \binom{\alpha}{i_j} [-\theta - \alpha l]_{n-(i_1+\cdots+i_l)},
\end{aligned}$$

where in the last equality (15) is used. Substituting this expression into (29) and taking the limit $n, r \to \infty$ with $i_j \to y_j$, $j = 1, \ldots, l$, and $r \sim xn$, the assertion for $0 < \alpha < 1$ and $\theta > -\alpha$ follows. Then, assume $\alpha = 0$ and $\theta > 0$. $\eta_1 = 0$, $\eta_2(k) = \theta$. For positive fixed integer k the signless Stirling numbers of the first kind $|s(n,k)|$ satisfies asymptotically (Jordan 1947)

$$\frac{|s(n,k)|}{n!} \sim \frac{1}{(k-1)!} \frac{(\log n)^{k-1}}{n}, \qquad n \to \infty.$$

For $\alpha = 0$ and $\theta > 0$ slight modification of Lemma 2 gives the assertion with $\eta_3(k) = 0$. Finally, for $\alpha < 0$ and $\theta = -m\alpha$, $m = 1, 2, \ldots$, $\rho_{\alpha,(-m\alpha)}(x) = 0$ for $x^{-1} > m$ since the support of $(v_{n,k})$ is $1 \leq k \leq m$. Since $\eta_1 = \alpha$, $\eta_2(k) = \theta$, $\eta_3(k) = k\alpha$ by the asymptotic form in Proposition 15 in Appendix B and $\breve{w}(1) = 1/\alpha$, the assumptions of Lemma 2 are satisfied and the assertion follows. □

Let us prepare a lemma for the proof of Theorem 5.

Lemma 3 *For* $0 < \alpha < 1$ *let*

$$f_r(\xi) := \sum_{j=0}^{r} \binom{\alpha}{j} (-\xi)^j. \tag{39}$$

The equation $f_r(\xi) = 0$ *has a real positive root. Moreover, it is the unique root of* $f_r(\xi) = 0$ *in* $|\xi| \leq \rho_r$, *where* ρ_r *is the real positive root satisfying*

$$\rho_r = 1 + \frac{1-\alpha}{\alpha r} + O(r^{-2}), \qquad r \to \infty. \tag{40}$$

Proof Let us show the existence of the real positive root of the equation $f_r(x) = 0$, $x \in \mathbb{R}$. It is straightforward to see that $f_r(1) > 0$ and $f_r(x)$ is a monotonically and strictly decreasing in $x > 0$, and $f_r(\infty) = -\infty$. Hence, there exists $L > 1$ such that $f_r(L) < 0$. According to the intermediate value theorem, the real-valued continuous function $f_r(x)$, $x \in (0, L)$ there exists the unique positive real root $\rho_r > 1$ such that $f_r(\rho_r) = 0$. Let $g_r(\xi) := 1 - f_r(\xi)$. Because

$$|g_r(\xi)| \leq \sum_{j=1}^{r} \binom{\alpha}{j} (-1)^{j+1} |\xi|^j < 1, \qquad |\xi| < \rho_r,$$

$f_r(\xi) = 0$ has no root in the open disk $|\xi| < \rho_r$. If $\rho_r e^{\sqrt{-1}\phi}$, $0 \leq \phi < 2\pi$, is another root of $f_r(\xi) = 0$,

$$\sum_{j=1}^{r} \binom{\alpha}{j} (-1)^{j+1} \rho_r^j \cos(j\phi) = 1$$

and $\phi = 0$ is obvious. Therefore, ρ_r is the unique root of $f_r(\xi) = 0$ in the closed disk $|\xi| \leq \rho_r$. Since the series (39) converges in $|\xi| \leq 1$ and $f_\infty(1) = 0$, let $\rho_r = 1 + y$, $y = o(1), r \to \infty$. Using the Taylor expansion, it can be seen that $y = -f_r(1)/f_r'(1) + O(y^2)$. Since $f_\infty(1) = 0$, similar argument to the evaluation of the second sum in (38) provides $f_r(1) = r^{-\alpha}/\Gamma(1-\alpha) + O(r^{-\alpha-1})$. $f_r'(1)$ is obtained in the similar manner and the assertion is established. □

The following proof for the first assertion is similar to the proof of Theorem 3.A in Flajolet and Odlyzko (1990).

Proof of Theorem 5 Let us evaluate (28):

$$\mathbb{P}(|A_{(1)}| \leq r) = \frac{n!}{(\theta)_n} \frac{1}{2\pi\sqrt{-1}} \oint \frac{\{f_r(\xi)\}^{-\frac{\theta}{\alpha}}}{\xi^{n+1}} d\xi.$$

Consider the Cauchy integral takes a contour (see Fig. 1) $\mathcal{C} = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, where

$$\gamma_1 = \left\{ \xi = \rho_r - \frac{t}{n};\ t = e^{\sqrt{-1}\theta},\ \theta \in \left[\frac{\pi}{2}, -\frac{\pi}{2}\right] \right\},$$

$$\gamma_2 = \left\{ \xi = \rho_r + \frac{\eta t + \sqrt{-1}}{n};\ t \in [0, n] \right\},$$

$$\gamma_3 = \left\{ \xi;\ |\xi| = \sqrt{(\rho_r + \eta)^2 + \frac{1}{n^2}};\ \Re(\xi) \leq \rho_r + \eta \right\},$$

$$\gamma_4 = \left\{ \xi = \rho_r + \frac{\eta t - \sqrt{-1}}{n};\ t \in [n, 0] \right\}.$$

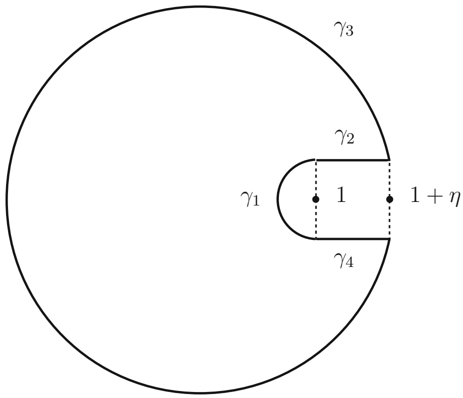

Fig. 1 The contour $\mathcal{C}$ used in the proof of Theorem 5

According to Lemma 3 we can take $\eta > 0$ such that no root of $f_r(\xi) = 0$ exists in the closed disk $|\xi| \leq \rho_r + \eta$ except ρ_r. The integrand is holomorphic in the disk with the single singularity at the origin with the cut along the real line $[\rho_r, \infty)$. The contribution of γ_3, which is $O((\rho_r + \eta)^{-n})$ with $\rho_r + \eta > 1$, is exponentially small. Changing the variable $\xi = \rho_r + t/n$ and letting $\mathcal{H}$ be the contour on which t varies when u varies on the rest of the contour, $\gamma_4 \cup \gamma_1 \cup \gamma_2$, yields

$$\begin{aligned}&\int_{\mathcal{H}} (\rho_r + t/n)^{-n-1} \{f_r(\rho_r + t/n)\}^{-\frac{\theta}{\alpha}} \frac{\mathrm{d}t}{n}\\&= \{-\rho_r f_r'(\rho_r))\}^{-\frac{\theta}{\alpha}} \rho_r^{-n-1} n^{\frac{\theta}{\alpha}-1} \int_{\mathcal{H}} e^{-\frac{t}{\rho_r}} \left(-\frac{t}{\rho_r}\right)^{-\frac{\theta}{\alpha}} \mathrm{d}t + O(\rho_r^{-n-1} n^{\frac{\theta}{\alpha}-2}).\end{aligned}$$

Using the Hankel representation of the gamma function (37) the first assertion is established. For the second assertion, let us evaluate

$$-\rho_r f_r'(\rho_r) = \alpha\rho_r \sum_{j=0}^{r-1} \binom{\alpha-1}{j} (-\rho_r)^j.$$

For sufficiently large r_0, it can be seen that

$$\sum_{j=r_0}^{r-1} \binom{\alpha-1}{j} (-\rho_r)^j = \frac{r^{1-\alpha}}{\Gamma(2-\alpha)} + O(r_0^{1-\alpha}), \qquad \sum_{j=0}^{r_0-1} \binom{\alpha-1}{j} (-\rho_r)^j < \sum_{j=0}^{r_0-1} \rho_r^j.$$

Substituting (40) and taking the limit $r, r_0 \to \infty$ with keeping $r_0 = o(r^{1-\alpha})$ the first sum dominates and the second assertion follows. □

Proof of Proposition 10 Let us evaluate (28):

$$\mathbb{P}(|A_{(1)}| \leq r) = \frac{n!}{(\theta)_n} \frac{1}{2\pi\sqrt{-1}} \oint e^{(n+1) f_{r,n}(\xi)} \mathrm{d}\xi,$$

where

$$f_{r,n}(\xi) := \frac{\theta}{n+1}\sum_{j=1}^{r}\frac{\xi^j}{j} - \log\xi.$$

The saddle points of $f_{r,n}(\xi)$ are

$$\rho_{r,n,j} = \left(\frac{n}{\theta}\right)^{\frac{1}{r}} e^{2\pi\sqrt{-1}j/r} - \frac{1}{r} + O(n^{-\frac{1}{r}}) =: \rho_j e^{\sqrt{-1}\varphi_j}, \qquad j = 0, 1, \ldots, r-1.$$

Taylor's expansions of $f_{r,n}(\xi)$ around the saddle points yields

$$f_{r,n}(\rho_{r,n,j} + \xi_j e^{\sqrt{-1}\eta_j}) = f_{r,n}(\rho_{r,n,j}) + \frac{1}{2}\left(\frac{\xi_j}{\rho_j}\right)^2$$
$$\left[1 + O(n^{-\frac{2}{r}})\right] e^{2\sqrt{-1}(\eta_j - \varphi_j)} + O\left(\frac{\xi_j}{\rho_j}\right)^3,$$

for $j = 0, 1, \ldots, r-1$, and thus the direction of the steepest descent of the ith saddle point is $\eta_j = \varphi_j + \pi/2$. The contour can be deformed such that it goes through each saddle point along the direction of the steepest descent without changing the value of the Cauchy integral. The value is evaluated as

$$\frac{1}{2\pi\sqrt{-1}}\oint e^{(n+1)f_{r,n}(\xi)}\mathrm{d}\xi \sim \frac{1}{\sqrt{2\pi n}}\sum_{j=0}^{r-1}\rho_{r,n,j}^{-n}\exp\left(\theta\sum_{k=1}^{r}\frac{\rho_{r,n,j}^k}{k} + \sqrt{-1}\varphi_j\right), \; n \to \infty,$$

and the assertion is established. □

Appendix A: Recurrence relations

We provide some recurrence relations for the associated partial Bell polynomials introduced in this paper.

Proposition 11 *The associated partial Bell polynomials, $B_{n,k(r)}(w_\bullet)$, for fixed positive integer r, satisfy the recurrence relation*

$$B_{n+1,k,(r)}(w_\bullet) = \sum_{j=r-1}^{n-r(k-1)}\binom{n}{j} w_{j+1} B_{n-j,k-1,(r)}(w_\bullet),$$

for $n = rk-1, rk, \ldots, k = 1, 2, \ldots, B_{0,0,(r)}(w_\bullet) = 1, B_{j,0,(r)}(w_\bullet) = 0, j = 1, 2, \ldots$

Proof Let

$$f_{r,k}(u) = \sum_{n=rk}^{\infty} B_{n,k,(r)}(w_\bullet)\frac{u^n}{n!}.$$

Differentiating the middle and the rightmost-hand sides of (16) yields

$$\sum_{n=rk}^{\infty} B_{n,k(r)}(w_\bullet)\frac{\xi^{n-1}}{(n-1)!} = \breve{B}_{k-1,(r)}(\xi, w_\bullet)\sum_{j=r}^{\infty} w_j \frac{\xi^{j-1}}{(j-1)!}$$
$$= \sum_{j=r-1}^{\infty}\sum_{m=r(k-1)}^{\infty} \frac{w_{j+1}}{j!} B_{m,k-1,(r)}(w_\bullet)\frac{\xi^{m+j}}{m!}$$
$$= \sum_{n=rk-1}^{\infty}\sum_{j=r-1}^{n-r(k-1)} \frac{w_{j+1}}{j!} B_{n-j,k-1,(r)}(w_\bullet)\frac{\xi^{n}}{(n-j)!},$$

where the indexes are changed as $m = n - j$. Equating the coefficients of $\xi^n/n!$ in the leftmost and the rightmost hand sides yields the recurrence relation. □

The next proposition holds in the same manner so we omit the proof.

Proposition 12 *The associated partial Bell polynomials, $B^{(r)}_{n,k}(w_\bullet)$, for fixed positive integer r, satisfy the recurrence relation*

$$B^{(r)}_{n+1,k}(w_\bullet) = \sum_{j=0\vee(n-rk+r)}^{(r-1)\wedge(n-k+1)} \binom{n}{j} w_{j+1} B^{(r)}_{n-j,k-1}(w_\bullet),$$

for $n = k-1, \ldots, rk-1, k = 1, 2, \ldots$ with $B^{(r)}_{0,0}(w_\bullet) = 1$, $B^{(r)}_{j,0}(w_\bullet) = 0$, $j = 1, 2, \ldots$

Proposition 13 *The associated partial Bell polynomials, $B_{n,k,(r)}(w_\bullet)$, for positive integer r, satisfy the recurrence relation*

$$B_{n,k,(r+1)}(w_\bullet) = \sum_{j=0}^{k} \frac{[n]_{rj}}{j!}\left(-\frac{w_r}{r!}\right)^j B_{n-rj,k-j,(r)}(w_\bullet)$$

for $n = k, k+1, \ldots, (r+1)k$, $k = 0, 1, \ldots$, with $B_{0,0,(r)}(w_\bullet) = 1$, $B_{j,0,(r)}(w_\bullet) = 0$, $j = 1, 2, \ldots$

Proof We have

$$\breve{B}_{k,(r+1)}(\xi, w_\bullet) = \frac{1}{k!}\left(\sum_{j=r}^{\infty} w_j\frac{\xi^j}{j!} - w_r\frac{\xi^r}{r!}\right)^k = \sum_{j=0}^{k}\left(-\frac{w_r}{r!}\right)^j \frac{\xi^{rj}}{j!}\breve{B}_{k-j,(r)}(\xi, w_\bullet),$$

whose expansion into power series of ξ yields

$$\sum_{n=(r+1)k}^{\infty} B_{n,k,(r+1)}(w_\bullet)\frac{\xi^n}{n!} = \sum_{j=0}^{k}\sum_{m=r(k-j)}^{\infty}\left(-\frac{w_r}{r!}\right)^j B_{m,k-j,(r)}(w_\bullet)\frac{\xi^{m+rj}}{j!m!}$$

$$= \sum_{n=rk}^{\infty}\sum_{j=0}^{k}\left(-\frac{w_r}{r!}\right)^j B_{n-rj,k-j,(r)}(w_\bullet)\frac{\xi^n}{j!(n-rj)!},$$

where the indexes are changed as $m = n - rj$. Equating the coefficients of $\xi^n/n!$ yields the recurrence relation. □

The next proposition holds in the same manner so we omit the proof.

Proposition 14 *The associated partial Bell polynomials, $B^{(r)}_{n,k}(w_\bullet)$, for positive integer r, satisfy the recurrence relation*

$$B^{(r+1)}_{n,k}(w_\bullet) = \sum_{j=0\vee(n-rk)}^{\lfloor(n-k)/r\rfloor}\frac{[n]_{(r+1)j}}{j!}\left(\frac{w_{r+1}}{(r+1)!}\right)^j B^{(r)}_{n-j(r+1),k-j}(w_\bullet)$$

for $n = k, k+1, \ldots, (r+1)k$, $k = 0, 1, \ldots$, with $B^{(r)}_{0,0}(w_\bullet) = 1$, $B^{(r)}_{j,0}(w_\bullet) = 0$, $j = 1, 2, \ldots$

Appendix B: Asymptotic forms

Asymptotic forms of the generalized factorial coefficients are given in the next proposition. The assertion for positive α appears in Charalambides (2005) as an exercise.

Proposition 15 *For non-zero α and fixed positive integer k the generalized factorial coefficients, $C(n,k;\alpha)$, satisfy asymptotically*

$$\frac{C(n,k;\alpha)}{n!} \sim \frac{(-1)^{n+k-1}}{\Gamma(-\alpha)(k-1)!}n^{-1-\alpha}, \qquad n\to\infty,\ \alpha>0$$

and

$$\frac{C(n,k;\alpha)}{n!} \sim \frac{(-1)^{n}}{\Gamma(-k\alpha)k!}n^{-1-k\alpha}, \qquad n\to\infty,\ \alpha<0.$$

Proof Applying the generalized binomial theorem to (14) yields

$$\frac{C(n,k;\alpha)}{n!} = \frac{1}{k!}[u^n]((1+u)^\alpha-1)^k = \frac{1}{k!}[u^n]\sum_{j=0}^{k}\binom{k}{j}(1+u)^{j\alpha}(-1)^{k-j}$$

$$= \frac{1}{k!}\sum_{j=1}^{k}\binom{k}{j}\binom{j\alpha}{n}(-1)^{k-j} = \sum_{j=1}^{k}\frac{\Gamma(n-j\alpha)}{\Gamma(-j\alpha)\Gamma(n+1)}\frac{(-1)^{k+n-j}}{j!(k-j)!}$$

$$= \frac{1}{n}\sum_{j=1}^{k}\frac{n^{-j\alpha}}{\Gamma(-j\alpha)}\frac{(-1)^{k+n-j}}{j!(k-j)!}(1+O(n^{-1})),$$

where the last equality follows by $\Gamma(n-j\alpha)/\Gamma(n+1)\sim n^{-j-1}$ as $n\to\infty$. □

A more general result is available. For positive α Pitman (1999) showed that

$$\frac{C(n,k;\alpha)}{n!}\sim\frac{(-1)^{n+k}}{(k-1)!}\alpha g_\alpha(s)n^{-1-\alpha}, \qquad n\to\infty,\ k\sim sn^\alpha, \tag{41}$$

where $g_\alpha(s)$ is the probability density of the Mittag-Leffler distribution (Pitman 2006). For the signless Stirling number of the first kind Hwang (1995) showed that

$$\frac{|s(n,k)|}{n!}\sim\frac{(\log n)^{k-1}}{(k-1)!n}\left[\left\{\Gamma\left(1+\frac{k-1}{\log n}\right)\right\}^{-1}+O\left(\frac{k}{(\log n)^2}\right)\right], \quad n\to\infty, \quad 2\le k\le s\log n, \tag{42}$$

and a precise local limit theorem for k around $\log n+\gamma+1/(2n)+O(n^{-2})$ is available, where γ is the Euler–Mascheroni constant (Louchard 2010).

Then, let us develop asymptotic forms of the associated signless Stirling number of the first kind, $|s_r(n,k)|$, and the associated generalized factorial coefficients, $C_r(n,k;\alpha)$. The author is unaware of the literature in which these asymptotics are discussed.

Proposition 16 *For non-zero α and integer k with $1\le k<n/r$ the r-associated generalized factorial coefficients, $C_r(n,k;\alpha)$, satisfy*

$$\frac{C_r(n,k;\alpha)}{n!}\sim\frac{(-1)^n}{\Gamma(-k\alpha)k!}\mathcal{I}^{(k-1)}_{x,x}(-\alpha;-\alpha)n^{-1-k\alpha}, \qquad n,r\to\infty, \qquad r\sim xn. \tag{43}$$

For integer $k=n/r\ge 2$, $C_r(n,k;\alpha)/n!=O(n^{-k(1+\alpha)})$.

Proof Since the assertion is trivial for $k=1$, assume $k\ge 2$. The exponential generating function (17) yields

$$\frac{C_r(n,k;\alpha)}{n!}=\frac{1}{k!}\sum_{\substack{i_j\ge r;j=1,\ldots,k\\ i_1+\cdots+i_k=n}}\prod_{j=1}^{k}\binom{\alpha}{i_j}=\frac{1}{k!}\frac{(-1)^n}{\Gamma(-\alpha)^k}\sum_{\substack{i_j\ge r;j=1,\ldots,k\\ i_1+\cdots+i_k=n}}\prod_{j=1}^{k}\frac{\Gamma(i_j-\alpha)}{\Gamma(i_j+1)}$$

$$=\frac{n^{-k(1+\alpha)}}{k!}\frac{(-1)^n}{\Gamma(-\alpha)^k}\sum_{\substack{i_j\ge r;j=1,\ldots,k\\ i_1+\cdots+i_k=n}}\prod_{j=1}^{k}\left(\frac{i_j}{n}\right)^{-1-\alpha}(1+O(n^{-1})),$$

where the last equality follows by $\Gamma(i_j-\alpha)/\Gamma(i_j+1)\sim i_j^{-1-\alpha}$ for $i_j\geq r\to\infty$. The assertion for $k=n/r\geq 2$ follows immediately. For $1\leq k<n/r$,

$$\sum_{\substack{i_j\geq r;\, j=1,\ldots,k\\ i_1+\cdots+i_k=n}}\prod_{j=1}^{k}\left(\frac{i_j}{n}\right)^{-1-\alpha}\to\frac{\Gamma(-\alpha)^k}{\Gamma(-k\alpha)}\mathcal{I}_{x,x}^{(k-1)}(-\alpha;-\alpha)n^{k-1},\quad n,r\to\infty,\ r\sim xn.$$

□

For the associated signless Stirling numbers of the first kind similar expression are available.

Proposition 17 *For integer k with $2\leq k<n/r$ the r-associated signless Stirling numbers of the first kind, $|s_r(n,k)|$, satisfy*

$$\frac{|s_r(n,k)|}{n!}\sim\frac{1}{k!n}\mathcal{I}_{x,x}^{(k-1)}(0;0),\qquad n,r\to\infty,\ r\sim xn.$$

For integer $k=n/r\geq 2$, $|s_r(n,k)|/n!=O(n^{-k})$.

Acknowledgements The author thanks Akinobu Shimizu for comments in connection with Sect. 2 and Hsien-Kuei Hwang, Masaaki Sibuya, and Hajime Yamato for discussions and comments in connection with Sect. 4.

References

Aldous, D. J. (1985). *Exchangeability and related topics. Lecture notes in mathematics* (Vol. 1117). Berlin: Springer.

Antoniak, C. E. (1974). Mixtures of Dirichlet process with applications to Bayesian nonparametric problems. *Annals of Statistics*, *2*, 1152–1174.

Aoki, M. (2002). *Modeling aggregate behavior and fluctuations in economics*. Cambridge: Cambridge University Press.

Arratia, R., Tavaré, S. (1992). Limit theorem for combinatorial structures via discrete process approximations. *Random Structures and Algorithms*, *3*, 321–345.

Arratia, R., Barbour, A. D., Tavaré, S. (2003). *Logarithmic combinatorial structures: A probabilistic approach*. Zurich: European Mathematical Society.

Berestycki, N., Pitman, J. (2007). Gibbs distributions for random partitions generated by a fragmentation process. *Journal of Statistical Physics*, *127*, 381–418.

Buchstab, A. A. (1937). An asymptotic estimation of a general number-theoretic function. *Mathematicheskiĭ Sbornik*, *44*, 1239–1246.

Charalambides, C. A. (2005). *Combinatorial methods in discrete distributions*. New York: Wiley.

Comtet, L. (1974). *Advanced combinatorics: The art of finite and infinite expansions*. Dordrecht, Holland: D. Reidel.

Devroye, L. (1993). A triptych of discrete distributions related to the stable law. *Statistics and Probability Letters*, *18*, 349–351.

Dickman, K. (1930). On the frequency of numbers containing prime factors of a certain relative magnitude. *Arkiv för Matematik, Astronomi och Fysik*, *22*, 1–44.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, *3*, 87–112. erratum. ibid. 3 (1972), 240, 376.

Ewens, W. J. (1973). Testing for increased mutation rate for neutral alleles. *Theoretical Population Biology*, *4*, 251–258.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, *1*, 109–230.
Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London*, *125*, 54–59.
Flajolet, P., Odlyzko, A. (1990). Singularity analysis of generating functions. *SIAM Journal of Discrete Mathematics*, *3*, 216–240.
Flajolet, P., Sedgewick, R. (2009). *Analytic combinatorics*. New York: Cambridge University Press.
Gnedin, A. (2010). A species sampling model with finitely many types. *Electronic Communications in Probability*, *15*, 79–88.
Gnedin, A., Pitman, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zapiski Nauchnykh Seminar POMI*, *325*, 83–102.
Griffiths, R. C. (1988). On the distribution of points in a Poisson Dirichlet process. *Journal of Applied Probability*, *25*, 336–345.
Griffiths, R. C., Spanò, D. (2007). Record indices and age-ordered frequencies in exchangeable Gibbs partition. *Electronic Journal of Probability*, *12*(40), 1101–1130.
Handa, K. (2009). The two-parameter Poisson–Dirichlet point process. *Bernoulli*, *15*, 1082–1116.
Hensley, D. (1984). The sum of $\alpha^{\Omega(n)}$ over integers $n \leq x$ with all prime factors between α and y. *Journal of Number Theory*, *18*, 206–212.
Hoshino, N. (2009). The quasi-multinomial distribution as a tool for disclosure risk assessment. *Journal of Official Statistics*, *25*, 269–291.
Hwang, H.-K. (1995). Asymptotic expansions for Stirling's number of the first kind. *Journal of Combinatorial Theory, Series A*, *71*, 343–351.
Jordan, C. (1947). *The calculus of finite differences* (2nd ed.). New York: Chelsea.
Karlin, S. (1967). Central limit theorems for certain infinite urn schemes. *Journal of Mathematics and Mechanics*, *17*, 373–401.
Kerov, S. V. (2005). Coherent random allocations and the Ewens–Pitman formula. *Zapiski Nauchnykh Seminar POMI*, *325*, 127–145.
Kingman, J. F. C. (1975). Random discrete distribution. *Journal of the Royal Statistical Society, Series B*, *37*, 1–15.
Kingman, J. F. C. (1978). The representation of partition structures. *Journal of London Mathematical Society*, *18*, 374–380.
Kolchin, V. F. (1971). A certain problem of the distribution of particles in cells, and cycles of random permutations. *Teoriya Veroyatnostej i Ee Primeneniya*, *16*, 67–82.
Korwar, R. M., Hollander, M. (1973). Contribution to the theory of Dirichlet process. *Annals of Probability*, *1*, 705–711.
Lijoi, A., Mena, R. H., Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse Gaussian priors. *Journal of the American Statistical Association*, *100*, 1278–1291.
Lijoi, A., Prünster, I., Walker, S. G. (2008). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Annals of Applied Probability*, *18*, 1519–1547.
Louchard, G. (2010). Asymptotics of the Stirling number of the first kind revisited: A saddle point approach. *Discrete Mathematics and Theoretical Computer Science*, *12*, 167–184.
Panario, D., Richmond, B. (2001). Smallest components in decomposable structures: Exp-log class. *Algorithmica*, *29*, 205–226.
Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, *102*, 145–158.
Pitman, J. (1997). Partition structures derived from Brownian motion and stable subordinators. *Bernoulli*, *3*, 79–96.
Pitman, J. (1999). Brownian motion, bridge, excursion and meander characterized by sampling at independent uniform times. *Electronic Journal of Probability*, *4*(11), 1–33.
Pitman, J. (2006). *Combinatorial stochastic processes. Lecture notes in mathematics* (Vol. 1875). Berlin: Springer.
Pitman, J., Yor, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, *25*, 855–900.
Rouault, A. (1978). Lois de Zipf et sources markoviennes. *Annales de l'Institute Henri Poincaré Section B*, *14*, 169–188.
Shepp, L. A., Lloyd, S. P. (1966). Ordered cycle length in random permutation. *Transactions of the American Mathematical Society*, *121*, 340–357.

Sibuya, M. (1979). Generalized hypergeometric, digamma and trigamma distributions. *Annals of the Institute of Statistical Mathematics*, *31*, 373–390.
Sobel, M., Uppuluri, V. R. R., Frankowski, K. (1977). *Selected tables in mathematical statistics* (Vol. IV). Providence, RI: American Mathematical Society.
Tavaré, S., Ewens, W. J. (1997). The Ewens sampling formula. In N. L. Johnson, S. Kotz, N. Balakrishnan (Eds.), *Multivariate discrete distributions* (pp. 1–20). New York: Wiley.
Tenenbaum, G. (1995). *Introduction to analytic and probabilistic number theory*. New York: Cambridge University Press.
Watterson, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *Journal of Applied Probability*, *13*, 639–651; correction. ibid. 14, 897 (1976).
Watterson, G. A., Guess, H. A. (1977). Is the most frequent allele the oldest? *Theoretical Population Biology*, *11*, 141–160.
Yamato, H., Sibuya, M. (2000). Moments of some statistics of Pitman sampling formula. *Bulletin of Informatics and Cybernetics*, *32*, 1–10.