Check for updates

J. Appl. Probab. **63,** 501–535 (2026)
*doi:*10.1017/jpr.2025.10042

INFINITELY RESIZABLE URN MODELS

NOBUAKI HOSHINO,* *Kanazawa University*

Abstract

Any margin of the multinomial distribution is multinomially distributed. Retaining this closure property, a family of generalized multinomial distributions is proposed. This family is characterized within multiplicative probability measures, using the Bell polynomial. The retained closure property simplifies marginal properties such as moments. The family can be obtained by conditioning independent infinitely divisible distributions on the total and also by mixing the multinomial distribution with the normalized infinitely divisible distribution. The closure property justifies a stochastic process of the family by Kolmogorov's extension theorem. Over time, Gibbs partitions of a positive integer appear as the limiting distributions of the family.

Keywords: Bell polynomial; conditional compound Poisson distribution; infinite divisibility; overdispersed multinomial distribution; partition of an integer

2020 Mathematics Subject Classification: Primary 62H05
Secondary 62H17; 60E07

1. Introduction

The random structure considered in this paper is well represented by an urn model. We randomly distribute indistinguishable balls over distinguishable urns. Each urn has a size, which controls the probability of a ball put into the urn. An urn is divided if it is replaced by two urns whose total size equals that of the divided urn. Conversely, two urns are merged if they are replaced by one urn whose size equals the total size of merged urns. Any combination of division and merger is called resizing.

The distribution of balls over urns is closed under resizing if it is unchanged after resizing. For example, let us regard the marginal parameter (cell probability) of the multinomial distribution as the size of an urn. Then the multinomial distribution is closed under infinite resizing, as we will see.

The multinomial distribution possesses many good features due to this closure property. However, it cannot control second moments, which is a classical motivation to generalize the multinomial distribution. We will see some existing approaches for the generalization in subsequent sections, yet unlike them, our study generalizes the multinomial distribution so that the closure property remains.

The present article is organized as follows. Section 1.1 introduces notation and the Bell polynomials of [11]. Section 2 characterizes a family of generalized multinomial distribution

Received 26 July 2025 UTC; accepted 26 July 2025 UTC.
* Postal address: School of Economics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. Email address: hoshino@kenroku.kanazawa-u.ac.jp

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

closed under resizing. This family is called the Bell polynomial distribution (BPD) because its probability mass is expressed using the Bell polynomial. The BPD can be obtained by conditioning random frequencies that have an infinitely divisible distribution; a distribution is infinitely divisible if and only if it equals the distribution of the sum of infinitely many independent and identically distributed (i.i.d.) random variables. Section 3 derives the marginal properties of the BPD such as moments, exploiting the closure property. Section 4 considers the case of random sizes, which includes the mixture of the BPD with the Dirichlet distribution, with implications to Bayesian inference. An important generalization of the Dirichlet distribution is the normalized infinitely divisible (NID) distribution proposed by [21]. It turns out that the mixed multinomial distribution with a NID distribution equals the BPD. Section 5 examines two examples of the BPD, one of which is the Dirichlet multinomial mixture. Section 6 shows the existence of a stochastic process such that the finite-dimensional distribution is the BPD. This fact leads to a limiting argument to obtain a random partitioning distribution of a positive integer. The present study ends with Section 7. Proofs are collected in the Appendix.

1.1. Notation and Bell polynomials

This subsection sets up notation and summarizes some properties of the Bell polynomials to be used. For a recent review of Bell polynomials, see [45].

We denote the set of positive integers and the set of non-negative integers by $\mathbb{N}$ and $\mathbb{N}_0$, respectively. The set of positive real numbers is denoted by $\mathbb{R}_+$. Similarly, $\mathbb{R}_-$ denotes the set of negative real numbers, and $\mathbb{R}$ denotes the set of real numbers. Also, $\mathbb{R}_{0+} := \{0\} \cup \mathbb{R}_+$ and $\mathbb{R}_{0-} := \{0\} \cup \mathbb{R}_-$.

For $n \in \mathbb{N}$, let $\boldsymbol{w}_n := (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$. Then partial Bell polynomials are defined for $k \in [n] := \{1, 2, \ldots, n\}$ as

$$B_{n,k}(\boldsymbol{w}_n) := n! \sum_{(s_1,\ldots,s_n)\in\mathcal{P}_{n,k}} \prod_{i=1}^{n} \left(\frac{w_i}{i!}\right)^{s_i} \frac{1}{s_i!},$$

where

$$\mathcal{P}_{n,k} := \left\{ (s_1, \ldots, s_n) \colon s_i \in \mathbb{N}_0,\ i \in \mathbb{N},\ \sum_{i=1}^{n} i s_i = n, \sum_{i=1}^{n} s_i = k \right\}.$$

Considering this definition, $B_{n,1}(\boldsymbol{w}_n) = w_n$, and $B_{n,n}(\boldsymbol{w}_n) = w_1^n$. In fact, $B_{n,k}(\boldsymbol{w}_n)$ does not depend on $w_i,\ i > n - k + 1$.

The partial Bell polynomial is sometimes called the Bell polynomial of the second kind. The first kind or the total Bell polynomial $B_n(\boldsymbol{w}_n)$ is defined as the sum of partial Bell polynomials over k. Namely,

$$B_n(\boldsymbol{w}_n) := n! \sum_{(s_1,\ldots,s_n)\in\mathcal{P}_n} \prod_{i=1}^{n} \left(\frac{w_i}{i!}\right)^{s_i} \frac{1}{s_i!},$$

where $\mathcal{P}_n := \bigcup_{k=1}^{n} \mathcal{P}_{n,k}$. It is important that $B_0(\cdot) = B_{0,0}(\cdot) \equiv 1$; see e.g. [46]. Conventionally, $B_{n,0}(\boldsymbol{w}_n) = 0$ for $n \in \mathbb{N}$.

For future use, we cite equation 11.3 of [8], namely

$$B_n(cw_1, c^2 w_2, \ldots, c^n w_n) = c^n B_n(\boldsymbol{w}_n). \tag{1.1}$$

In particular,

$$B_n(-w_1, (-1)^2 w_2, \ldots, (-1)^n w_n) = (-1)^n B_n(\boldsymbol{w}_n). \tag{1.2}$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Also, from equation 11.15 of [8],

$$B_n(\lambda, \boldsymbol{w}_n) := B_n(\lambda w_1, \lambda w_2, \ldots, \lambda w_n) = \sum_{k=1}^{n} \lambda^k B_{n,k}(\boldsymbol{w}_n). \tag{1.3}$$

Equation (1.3) implies that

$$B_n(\lambda, \boldsymbol{w}_n) = \lambda w_n + \mathrm{O}(\lambda^2), \quad \lambda \to 0. \tag{1.4}$$

Also

$$B_n(\lambda, \boldsymbol{w}_n) = \lambda^n w_1^n + \mathrm{O}(\lambda^{n-1}), \quad \lambda \to \infty. \tag{1.5}$$

In the following, the variables of the Bell polynomial are written as $\boldsymbol{w} := (w_i)_{i\in\mathbb{N}}$ for simplicity and dealing with the sequence of Bell polynomials. We regard $B_{n,k}(\boldsymbol{w})$ as $B_{n,k}(\boldsymbol{w}_n)$ so that $B_n(w) = B_n(\boldsymbol{w}_n)$.

Concerning an urn model, we denote the deterministic number of urns by $J \in \mathbb{N}$. The total number of balls is deterministic and denoted by $n \in \mathbb{N}_0$. The random number of balls in the jth urn is denoted by $F_j \in \mathbb{N}_0$, $j \in [J]$. The space of these frequencies $\boldsymbol{F}_J := (F_1, F_2, \ldots, F_J)$ is denoted by

$$\mathcal{F}_{n,J} := \left\{ \boldsymbol{f}_J := (f_1, f_2, \ldots, f_J) \colon f_j \in \mathbb{N}_0,\ j \in [J],\ \sum_j f_j = n \right\}.$$

The size of the jth urn is denoted by λ_j. The vector of the sizes is denoted by $\boldsymbol{\lambda}_J := (\lambda_1, \lambda_2, \ldots, \lambda_J)$. The total of the sizes is denoted by $\lambda_\bullet = \sum_{j=1}^{J} \lambda_j$; the subscript '$\bullet$' expresses summing up over the index throughout the manuscript.

2. Generalization of the multinomial distribution

2.1. Construction

Vershik [55] defines a multiplicative probability measure on $\mathcal{P}_n$ for a sequence of functions $(\tilde{h}_i \colon \mathbb{N}_0 \mapsto \mathbb{R}_+)_{i=1}^{\infty}$ by the following probability mass function (PMF):

$$p(\boldsymbol{s}_n) = \prod_{i=1}^{n} \tilde{h}_i(s_i)/Q_n,\ \boldsymbol{s}_n \in \mathcal{P}_n,$$

where $Q_n = \sum_{\boldsymbol{s}_n \in \mathcal{P}_n} \prod_{i=1}^{n} \tilde{h}_i(s_i)$ is a normalizing constant. In this family, a margin is not independent from other margins, but the multiplicative form of the probability mass simplifies the combinatorial structure of the family.

Analogously, we can define a multiplicative probability measure on $\mathcal{F}_{n,J}$ for a sequence of functions $(\tilde{h}_j \colon \mathbb{N}_0 \mapsto \mathbb{R}_+)_{j=1}^{\infty}$ by the following PMF:

$$p(\boldsymbol{f}_J) = \prod_{j=1}^{J} \tilde{h}_j(f_j)/Q_{n,J}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}, \tag{2.1}$$

where $Q_{n,J} = \sum_{\boldsymbol{f}_J \in \mathcal{F}_{n,J}} \prod_{j=1}^{J} \tilde{h}_j(f_j)$ is a normalizing constant.

For example, if $\tilde{h}_j(f_j) = \lambda_j^{f_j}/f_j!$ then $Q_{n,J} = \lambda_\bullet^n/n!$ in (2.1). This case

$$p(\boldsymbol{f}_J) = \frac{n!}{\lambda_\bullet^n} \prod_{j=1}^{J} \frac{\lambda_j^{f_j}}{f_j!}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}, \tag{2.2}$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

defines the multinomial distribution, which is usually parameterized by a cell probability or $\pi_j := \lambda_j/\lambda_\bullet,\ j \in [J]$. Because $\lambda_J = \lambda_\bullet - \sum_{j=1}^{J-1} \lambda_j$, this reparameterization is expressed as

$$\boldsymbol{\lambda}_J \mapsto (\pi_1, \pi_2, \ldots, \pi_{J-1}, \lambda_\bullet). \tag{2.3}$$

However, for notational convenience, we denote the transformed parameters by $(\boldsymbol{\pi}_J; \lambda_\bullet)$, where $\boldsymbol{\pi}_J := (\pi_1, \pi_2, \ldots, \pi_J)$. The $(J-1)$-dimensional multinomial distribution (2.2) reparameterized by (2.3) is referred to by $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$, which does not depend on $\lambda_\bullet$.

We regard the space of cell probabilities as the $(J-1)$-dimensional unit simplex

$$\triangle^{J-1} := \left\{ \boldsymbol{\pi}_J \colon \sum_{j=1}^{J} \pi_j = 1,\ \pi_j \in \mathbb{R}_+,\ j \in [J] \right\}.$$

It is important that π_j is strictly positive in this definition. If $\pi_j = 0$ then the observed F_j is almost surely zero, which is called structural zero in statistics. By excluding structural zeros, we avoid indeterminate J. Also, as a result, the support of $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$ does not degenerate from $\mathcal{F}_{n,J}$. Namely,

$$\text{for all } \boldsymbol{f}_J \in \mathcal{F}_{n,J}, \quad p(\boldsymbol{f}_J) > 0. \tag{2.4}$$

We fix the space of cell probabilities to $\triangle^{J-1}$. Then the corresponding space of $\boldsymbol{\lambda}_J$ can only be $\mathbb{R}_+^J$ or $\mathbb{R}_-^J$. However, the negative size of an urn is difficult to interpret. Hence, for the time being, we assume that $\lambda_j \in \mathbb{R}_+, j \in [J]$.

The multinomial distribution $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$ is the n-fold convolution of i.i.d. samples of $\mathtt{Mn}_J(1, \boldsymbol{\pi}_J)$. Consequently, (2.2) is invariant under the permutation of n samples. This property is called the exchangeability of balls.

We note that (2.2) is also invariant under the permutation of urns. The jth urn has two attributes: a size λ_j and the number of balls f_j in the urn. Permuting urns exchanges $(f_j; \lambda_j)$ with $(f_i; \lambda_i)$, $i \neq j$, by which the right-hand side of (2.2) is unchanged. We call this property the exchangeability of urns.

It is noteworthy that exchangeability often refers to that of arguments of a function. For example, if all λ_j are equal then $p(\ldots, f_i, \ldots, f_j, \ldots) = p(\ldots, f_j, \ldots, f_i, \ldots)$, $i \neq j$, for (2.2). This $p(\cdot)$ is thus called exchangeable, but the exchangeability of urns is a different concept since it holds for unequal λ_j.

We would like to generalize the multinomial distribution so that the closure property under resizing remains. Since resizing operates on urns, the closure property seems relevant to the exchangeability of urns. Hence we retain it but neglect the exchangeability of balls.

We note another characteristic of (2.2), that the normalizing constant $\lambda_\bullet^n/n!$ has the same functional form as that of the marginal numerator $\lambda_j^{f_j}/f_j!$. This symmetry between the numerator and the denominator is actually the source of the closure property.

Consequently, we consider a subfamily of (2.1) with constraints $\tilde{h}_j(f_j) = h(f_j; \lambda_j)/f_j!$ and $Q_{n,J} = h(n; \lambda_\bullet)/n!$. Namely, $h(\cdot)$ is restricted to satisfy

$$\sum_{\boldsymbol{f}_J \in \mathcal{F}_{n,J}} p(\boldsymbol{f}_J) = 1 \tag{2.5}$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

under

$$p(\boldsymbol{f}_J) = \frac{n!}{h(n;\lambda_\bullet)} \prod_{j=1}^{J} \frac{h(f_j;\lambda_j)}{f_j!}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}. \tag{2.6}$$

To determine the form of $h(\cdot)$, we additionally assume continuity of $h(\,\cdot\,;\lambda)$ as follows.

Definition 2.1. The PMF (2.6) that satisfies the following three conditions for $n \in \mathbb{N}_0$, $J \in \mathbb{N}$, and $\boldsymbol{\lambda}_J \in \mathbb{R}_+^J$ defines a family of distributions denoted by $\mathcal{M}$.

(i) *Positivity.* The PMF is strictly positive as in (2.4).

(ii) *Unity.* The total of probabilities is unity as in (2.5).

(iii) *Continuity.* $h(f;\lambda)$ is continuous in λ for $f \in \mathbb{N}_0$.

Theorem 2.1 below shows that $h(f;\lambda)$ is uniquely determined by the sequence of $h(f;1) = h_f$, $f \in \mathbb{N}$, for a member of $\mathcal{M}$. Therefore $(h_f)_{f\in\mathbb{N}}$ can be regarded as a parameter vector. It is noteworthy that $h(0;1)$ leads to 1.

Theorem 2.1. *For $(h_f)_{f\in\mathbb{N}} \in \mathbb{R}_+^\infty$, define $h(f;\lambda)$ by equating the coefficient of s^f on both sides of*

$$\sum_{f=0}^{\infty} \frac{s^f}{f!} h(f;\lambda) = \left(1 + \sum_{f=1}^{\infty} \frac{s^f}{f!} h_f\right)^{\lambda}. \tag{2.7}$$

Then the family of (2.6) *is equivalent to $\mathcal{M}$.*

2.2. Characterization

The family $\mathcal{M}$ of Definition 2.1 is not very treatable when it is parameterized with $(h_f)_{f\in\mathbb{N}}$. Hence we reparameterize it using Bell polynomials.

Theorem 2.2. *Let $h_f = B_f(\boldsymbol{w})$, $f \in \mathbb{N}$. Then, for all $n \in \mathbb{N}_0$,* (2.6) *equals*

$$p(\boldsymbol{f}_J) = \frac{n!}{B_n(\lambda_\bullet, \boldsymbol{w})} \prod_{j=1}^{J} \frac{B_{f_j}(\lambda_j, \boldsymbol{w})}{f_j!}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}. \tag{2.8}$$

As a distribution on $\mathcal{F}_{n,J}$, the parameter vector of (2.8) is $(\boldsymbol{\lambda}_J, \boldsymbol{w}_n)$, where $\boldsymbol{w}_f$ consists of the first f entries of $\boldsymbol{w}_n$. Nevertheless, the three conditions of Definition 2.1 must hold for all $n \in \mathbb{N}_0$. Theorem 2.3 thus clarifies the space of $\boldsymbol{w}$.

Theorem 2.3. *The distribution* (2.8) *satisfies* (2.4) *for $n \in \mathbb{N}_0$, $J \in \mathbb{N}$, and $\boldsymbol{\lambda}_J \in \mathbb{R}_+^J$ if and only if $\boldsymbol{w}$ belongs to either $\mathbb{R}_+ \times \mathbb{R}_{0+}^\infty =: \Omega_1$ or $\mathbb{R}_- \times \prod_{i=2}^\infty (-1)^i \mathbb{R}_{0+} =: \Omega_2$, where $(-1)\mathbb{R}_{0+} = \mathbb{R}_{0-}$.*

When $\boldsymbol{\lambda}_J \in \mathbb{R}_-^J$ in Theorem 2.3, $\boldsymbol{w}$ belongs to either $\mathbb{R}_- \times \mathbb{R}_{0-}^\infty$ or $\mathbb{R}_+ \times \prod_{i=2}^\infty (-1)^i \mathbb{R}_{0-}$ due to (1.2). Therefore only four parameter spaces of $(\boldsymbol{\lambda}_J, \boldsymbol{w})$ exist, but the distributional property of (2.8) is equivalent among them due to (1.2). Hence we restrict our attention to the most intuitive space $\mathbb{R}_+^J \times \Omega_1$ as in Definition 2.2.

Definition 2.2. Let $\boldsymbol{w} \in \Omega_1$. For each $n \in \mathbb{N}_0$, the distribution that has the PMF (2.8) for $J \in \mathbb{N}$ and $\boldsymbol{\lambda}_J \in \mathbb{R}_+^J$ is called the Bell polynomial distribution (BPD) of the nth order, which is referred to by $\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$. The sequence of the BPD of finite orders generated by $\boldsymbol{w}$ constitutes a family $\{(\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w}))_{n\in\mathbb{N}_0} : \boldsymbol{w} \in \Omega_1\}$ denoted by $\mathcal{BPD}$.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

In the following, we do not restate the spaces of n, J, $\boldsymbol{\lambda}_J$, and $\boldsymbol{w}$, when they are the same as those in Definition 2.2.

Theorem 2.1 shows an expression of $\mathcal{M}$ parameterized by $(h_f)_{f\in\mathbb{N}}$. The reparameterization of it with $\boldsymbol{w}$ in Theorem 2.2 is bijective in fact. Hence we have the following characterization.

Theorem 2.4. $\mathcal{BPD} = \mathcal{M}$.

2.3. Generating sequence of the Bell polynomial distribution

We note that different sequences $\boldsymbol{w}$ can generate the same BPD as in Theorem 2.5, which immediately results from (1.1) and (1.3).

Theorem 2.5. *For $c \in \mathbb{R}_+$, the following two equations hold:*

$$\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w}) = \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (w_i c^i)_{i\in\mathbb{N}}), \tag{2.9}$$

$$\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (w_i c)_{i\in\mathbb{N}}) = \mathtt{BP}_J(n, (\lambda_j c)_{j\in[J]}; \boldsymbol{w}). \tag{2.10}$$

Also, observing (2.8), the BPD of the nth order does not depend on w_i, $i > n$. This fact delivers Theorem 2.6.

Theorem 2.6. *Let $\boldsymbol{v}$ and $\boldsymbol{w}$ belong to Ω_1. If the first n entries of $\boldsymbol{v}$ and $\boldsymbol{w}$ are the same, then* $(\mathtt{BP}_J(i, \boldsymbol{\lambda}_J; \boldsymbol{v}))_{i\in[n]} = (\mathtt{BP}_J(i, \boldsymbol{\lambda}_J; \boldsymbol{w}))_{i\in[n]}$.

The following corollary to Theorem 2.6 is the special case of $n = 1$; it is easy to verify that

$$\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (1, 0, 0, \dots)) = \mathtt{Mn}_J(n, \boldsymbol{\pi}_J).$$

Then, from (2.9),

$$\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (w_1, 0, 0, \dots)) = \mathtt{Mn}_J(n, \boldsymbol{\pi}_J) \quad \text{for } w_1 \in \mathbb{R}_+.$$

Corollary 2.1. *Irrespective of $\boldsymbol{w}$,* $\mathtt{BP}_J(1, \boldsymbol{\lambda}_J; \boldsymbol{w}) = \mathtt{Mn}_J(1, \boldsymbol{\pi}_J)$.

Any BPD can be parameterized with cell probabilities by (2.3). Therefore we interchange $\boldsymbol{\lambda}_J$ of $\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$ with $(\boldsymbol{\pi}_J; \lambda_\bullet)$ without reservation. However, not every BPD depends on $\lambda_\bullet$, as the multinomial distribution demonstrates. To discriminate between them, define

$$\mathcal{W} := \{\boldsymbol{w} \colon (BP_J(n, (\pi_J; \lambda_\bullet); \boldsymbol{w}))_{n\in\mathbb{N}_0} \text{ depends on } \lambda_\bullet\}$$

and

$$\mathcal{W}^c := \Omega_1 \setminus \mathcal{W}. \tag{2.11}$$

For example, $(1, 0, 0, \dots) \in \mathcal{W}^c$. Theorem 4.10 will exemplify $\mathcal{W}^c$ more generally.

In $\mathcal{W}$, the result of taking $\lambda_\bullet \to \infty$ for the BPD is simply Theorem 2.7, because $B_f(\lambda_j, \boldsymbol{w}) \approx (\pi_j \lambda_\bullet w_1)^f$ as $\lambda_\bullet \to \infty$; see (1.5). This limiting argument is widely known for the Dirichlet multinomial mixture (e.g. [53]), which is the quintessence of the BPD examined in Section 5.1.

Theorem 2.7. *Let $\boldsymbol{\pi}_J \in \triangle^{J-1}$ and $\boldsymbol{w} \in \mathcal{W}$. Then*

$$\mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \lambda_\bullet); \boldsymbol{w}) \to \mathtt{Mn}_J(n, \boldsymbol{\pi}_J) \quad \text{as } \lambda_\bullet \to \infty.$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

2.4. Infinite divisibility and continuity

The family of $\mathcal{BPD}$ overlaps the family of the conditional compound Poisson (CCP) distribution of [30] as shown below.

The compound Poisson distribution over non-negative integers is defined by the probability generating function (PGF) of the following form:

$$G(s) = \mathrm{e}^{\lambda(\sum_{i=1}^{\infty} s^i q_i - 1)}, \quad \lambda \in \mathbb{R}_+, \tag{2.12}$$

where $(q_i)_{i\in\mathbb{N}} =: \boldsymbol{q}$ belongs to $\blacktriangle := \{\boldsymbol{q}\colon\ \sum_{i=1}^{\infty} q_i = 1,\ q_i \in \mathbb{R}_{0+},\ i \in \mathbb{N}\}$ in the canonical form; see e.g. [36, p. 188]. The distribution of (2.12) is referred to as $\mathtt{CP}(\lambda, \boldsymbol{q})$.

The notion of the CCP distribution arises from conditioning independent compound Poisson variables on the total. Let us denote $\boldsymbol{F}_J$ constrained so that $\sum_{j=1}^{J} F_j = n$ by $(\boldsymbol{F}_J \mid \sum_{j=1}^{J} F_j = n)$. This notation is used similarly below. The distribution of a random vector (variable) follows after '$\sim$' henceforth.

Theorem 2.8. *Let $F_j \sim \mathtt{CP}(\lambda_j, \boldsymbol{q})$, $j \in [J]$, independently. If $\boldsymbol{q} \in (\blacktriangle \cap \Omega_1)$ then*

$$\left(\boldsymbol{F}_J \,\middle|\, \sum_{j=1}^{J} F_j = n\right) \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (q_i i!)_{i\in\mathbb{N}}) \quad \textit{for all } n \in \mathbb{N}_0.$$

Definition 2.3. In Theorem 2.8, the conditional distribution of $(\boldsymbol{F}_J \mid \sum_{j=1}^{J} F_j = n)$ defines the nth CCP distribution generated by $\boldsymbol{q}$.

The restriction of $\boldsymbol{q} \in (\blacktriangle \cap \Omega_1)$ in the theorem above excludes the case of $q_1 = 0$, where the conditional distribution may not exist. For instance, let $q_2 = 1$ and $q_i = 0$, $i \neq 2$. Then $\mathbb{P}(\sum_{j=1}^{J} F_j = n) = 0$ for odd n.

When $q_1 = 1$ and $q_i = 0$, $i \geq 2$, Theorem 2.8 reproduces a known fact that the multinomial distribution is the conditional distribution of independent Poisson variates; see [35, p. 32].

Conversely to Theorem 2.8, let us consider the sequence of CCP distributions equivalent to $(\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w}))_{n\in\mathbb{N}_0}$. If $\sum_{i=1}^{\infty} w_i/i! < \infty$, then putting $q_i = w_i/(i!\ \sum_{j=1}^{\infty} w_j/j!)$ lets $\boldsymbol{q} \in \blacktriangle$. For this $\boldsymbol{q}$, assuming that $F_j \sim \mathtt{CP}(\lambda_j \sum_{i=1}^{\infty} w_i/i!, \boldsymbol{q})$ independently,

$$\left(\boldsymbol{F}_J \,\middle|\, \sum_{j=1}^{J} F_j = n\right) \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$$

by (2.10). This equation fails when $\sum_{i=1}^{\infty} w_i/i!$ diverges, but we can equate $\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$ with the nth CCP distribution generated by $q_i = w_i/(i!\ \sum_{j=1}^{n} w_j/j!)$, $i \in [n]$, and $q_i = 0$, $i > n$; see Theorem 2.6.

Let us examine the case of $w_i = i!$, $i \in \mathbb{N}$. Although $\sum_{i=1}^{\infty} w_i/i!$ diverges, the PMF of $\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (i!)_{i\in\mathbb{N}})$ is explicitly written as

$$p(\boldsymbol{f}_J) = \frac{n!}{\sum_{k=1}^{n} \lambda_\bullet^k \binom{n-1}{k-1} \frac{n!}{k!}} \prod_{j=1}^{J} \frac{\sum_{k=1}^{f_j} \lambda_j^k \binom{f_j - 1}{k-1} \frac{f_j!}{k!}}{f_j!}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}, \tag{2.13}$$

since

$$B_{n,k}(1!, 2!, \dots) = \binom{n-1}{k-1} \frac{n!}{k!},$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

which is called the Lah number; see [45, eq. 1.7] and references therein. By (2.9),

$$\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (i!)_{i\in\mathbb{N}}) = \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (i!\,\theta^i)_{i\in\mathbb{N}}), \quad \theta \in \mathbb{R}_+. \tag{2.14}$$

We note that $\sum_{i=1}^{\infty} \theta^i = (1-\theta)^{-1}$ for $\theta \in (0, 1)$. Then, if $F_j \sim \mathtt{CP}(\lambda_j(1-\theta)^{-1}, (i!\,\theta^i(1-\theta))_{i\in\mathbb{N}})$ independently, the PMF of $(\boldsymbol{F}_J \mid \sum_{j=1}^{J} F_j = n)$ is (2.13). Because the conditional distribution is free of θ, $\sum_{j=1}^{J} F_j$ is sufficient for θ called a power parameter. See [30, Theorem 2]. However, in (2.14), $\theta \in \mathbb{R}_+ \supset (0, 1)$.

The compound Poisson distribution is tantamount to the infinitely divisible (ID) distribution over non-negative integers; see e.g. [52, Theorem 3.2]. Therefore the CCP distribution results from conditioning the ID distribution. Section 4 revisits this view in comparison with normalizing the ID distribution over non-negative real numbers.

We did not assume that $h(f; \lambda)$ is a polynomial in λ, but a polynomial sequence $(h(f; \lambda))_{f=0}^{\infty}$ is called binomial if it satisfies two conditions: (i) $h(0; \lambda) = 1$, and (ii) when $J = 2$, the unity (2.5) holds for all $(\lambda_1, \lambda_2) \in \mathbb{R}^2$ on (2.6).

The fact that $h(0; \lambda) = 1$ enables the construction of the BPD in Theorem 2.8 because the distribution of F_j cannot be ID if $\mathbb{P}(F_j = 0) = 0$; see e.g. [52, p. 23].

When $J = 2$, the unity condition (2.5) on (2.8) is rewritten as

$$\sum_{f=0}^{n} \frac{B_f(\lambda_1, \boldsymbol{w})}{f!} \frac{B_{n-f}(\lambda_2, \boldsymbol{w})}{(n-f)!} = \frac{B_n(\lambda_1 + \lambda_2, \boldsymbol{w})}{n!}, \tag{2.15}$$

which is a fundamental property of the Bell polynomial; see e.g. [8, Exercises 11.9.1]. This closure property under convolution (2.15) actually holds for all $(\lambda_1, \lambda_2) \in \mathbb{R}^2$.

Since $B_0(\cdot) = 1$, the sequence of Bell polynomials is binomial. In fact, the Bell polynomial is known to be equivalent to a polynomial sequence of binomial type; see e.g. [2, Theorem 3.59].

Theorem 2.4, or more directly (A.5), shows equivalence between the sequence of continuous but not necessarily polynomial functions $(h(f;\ \lambda))_{f=0}^{\infty}$ and the sequence of Bell polynomials. We note that continuous functions include polynomials.

The generating function of $h(\cdot)$ is not a PGF. Hence we cannot exploit the continuity theorem of a PGF. Alternatively, we have assumed the continuity of $h(\cdot)$, by which the space of $\boldsymbol{w}$ has been determined; see the proof of Theorem 2.3. The continuity of $h(\cdot)$ seems inevitable.

3. Marginal properties of the Bell polynomial distribution

3.1. Marginal distributions

The binomial property (2.15) of the Bell polynomial directly shows the following Theorem 3.1, which generalizes the fact that the marginal distribution of the multinomial distribution is again the multinomial distribution.

Theorem 3.1. *Suppose that* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$. *Let*

$$\mathcal{J}_m := \{j_1, j_2, \ldots, j_m\} \subset [J].$$

Then

$$\left(F_{j_1}, F_{j_2}, \ldots, F_{j_m}, n - \sum_{k=1}^{m} F_{j_k}\right) \sim \mathtt{BP}_{m+1}\left(n, \left(\lambda_{j_1}, \lambda_{j_2}, \ldots, \lambda_{j_m}, \lambda_\bullet - \sum_{k=1}^{m} \lambda_{j_k}\right); \boldsymbol{w}\right).$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Using Theorem 3.1, we can show Theorem 3.2, which generalizes a known fact on the multinomial distribution; see e.g. [35, Sec. 3, Chap. 35].

Theorem 3.2. *Under the assumption of Theorem 3.1,*

$$(F_l,\ l \in [J] \setminus \mathcal{J}_m \mid F_j = f_j,\ j \in \mathcal{J}_m) \sim \mathtt{BP}_{J-m}\left(n - \sum_{k=1}^{m} f_{j_k},\ (\lambda_l)_{l \in [J]\setminus \mathcal{J}_m};\ \boldsymbol{w}\right).$$

In particular, the conditional distribution of the BPD given the jth margin does not depend on λ_j, as

$$(F_l,\ l \in [J] \setminus \{j\} \mid F_j = f_j) \sim \mathtt{BP}_{J-1}(n - f_j,\ (\lambda_l)_{l \in [J]\setminus \{j\}};\ \boldsymbol{w})$$

due to Theorem 3.2. Hence the following corollary to Theorem 3.2 holds.

Corollary 3.1. *If* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$*, then* F_j *is a sufficient statistic for* λ_j*.*

3.1.1. *Application: random sampling.* We have seen two closure properties of the BPD, which are useful for sampling from the BPD. First, we can decompose multivariate sampling into sequential sampling of univariate margins. This simplification is called the conditional distribution method; see [18]. Second, sampling from the BPD is closed under multi-stage sampling: partial frequencies are generated given their sum, which is generated in an earlier stage.

The conditional distribution method exploits the following general relationship:

$$\mathbb{P}(\boldsymbol{F}_J = \boldsymbol{f}_J) = \prod_{j=2}^{J} \mathbb{P}(F_j = f_j \mid \boldsymbol{F}_{j-1} = \boldsymbol{f}_{j-1})\, \mathbb{P}(F_1 = f_1). \tag{3.1}$$

Equation (3.1) implies that the J-dimensional sampling of $\boldsymbol{F}_J$ is accomplished by sequential sampling of univariate F_j conditionally given $\boldsymbol{F}_{j-1}$, beginning with F_1.

Owing to Theorem 3.2, the univariate conditional BPD is $\mathtt{BP}_2$. The initial unconditional univariate BPD is also $\mathtt{BP}_2$ by Theorem 3.1. Accordingly, Proposition 3.1 below determines the conditional distribution method of sampling from the BPD.

Proposition 3.1. *Suppose that* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$*. Then*

$$\left(F_j,\ n - \sum_{k=1}^{j-1} f_k - F_j \,\middle|\, \boldsymbol{F}_{j-1} = \boldsymbol{f}_{j-1}\right) \sim \mathtt{BP}_2\left(n - \sum_{k=1}^{j-1} f_k,\ \left(\lambda_j,\ \sum_{k=j+1}^{J} \lambda_k\right);\ \boldsymbol{w}\right)$$

for $j \geq 2$*, and* $(F_1, n - F_1) \sim \mathtt{BP}_2(n, (\lambda_1, \lambda_\bullet - \lambda_1); \boldsymbol{w})$*.*

Hence sampling from the BPD is feasible if we can sample from $\mathtt{BP}_2$. The inversion method enables us to do so. This method maps a uniform random number on [0,1] by the inverse of the (cumulative) distribution function; see e.g. [18]. Because the PMF of $\mathtt{BP}_2$ is explicit, the distribution function of $\mathtt{BP}_2$ can be computationally obtained, although it can be vulnerable to rounding errors.

Ho *et al.* [28] considered the conditional distribution method for the multinomial distribution, where cell probabilities are sorted in descending order since sampling stops when a remaining frequency becomes zero. The same argument applies to the BPD, and thus we should sort the λ_j in descending order to curtail the time of sampling. It is important that Proposition 3.1 holds even after sorting due to the exchangeability of urns.

Next, we note that the conditional distribution of partial frequencies given their sum is again BP distributed by Theorem 3.2. In particular, the following fact holds.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Proposition 3.2. *If* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$*, then* $(\boldsymbol{F}_m \mid \sum_{l=1}^m F_l = \tilde{n}) \sim \mathtt{BP}_m(\tilde{n}, \boldsymbol{\lambda}_m; \boldsymbol{w})$*, where* $m \in [J]$ *and* $\tilde{n} \in [n]$.

Proposition 3.2 justifies two-stage sampling from the BPD. In its simplest form, the first stage generates the aggregated frequency of $\tilde{n} = \sum_{l=1}^m F_l$ and $n - \tilde{n} = \sum_{l=m+1}^J F_l$ as

$$(\tilde{n}, n - \tilde{n}) \sim \mathtt{BP}_2\left(n, \left(\sum_{l=1}^m \lambda_l, \sum_{l=m+1}^J \lambda_l\right); \boldsymbol{w}\right),$$

and the second stage generates

$$\left(\boldsymbol{F}_m \,\middle|\, \sum_{l=1}^m F_l = \tilde{n}\right) \quad \text{and} \quad \left(F_{m+1}, F_{m+2}, \ldots, F_J \,\middle|\, \sum_{l=m+1}^J F_l = n - \tilde{n}\right),$$

using Proposition 3.2. Then we have $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$. A recursive argument shows that the BPD is generally closed under multi-stage sampling.

Malefaki and Iliopoulos [39] provide empirical evidence that multinomial sampling of a large scale can be made faster by introducing stages. In [33], it is argued that rejection sampling from the quasi-multinomial distribution (5.6) can be made faster by multi-stage sampling because it decreases a rejection rate.

3.1.2. *Application: information matrix.* In practice, a frequency vector is usually observed only once. Then we cannot estimate both $\boldsymbol{\lambda}_J$ and $\boldsymbol{w}$ of the BPD without constraints. The primary concern tends to lie in $\boldsymbol{\lambda}_J$, and thus we derive the Fisher information matrix of $\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$, given $\boldsymbol{w}$.

Theorem 3.3. *Suppose that* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J; \boldsymbol{w})$*, where its likelihood function is denoted by* $p(\boldsymbol{f}_J; \boldsymbol{\lambda}_J)$*. Then the diagonal element of the Fisher information matrix is, for* $j \in [J]$*,*

$$\mathbb{E}\left[-\frac{\partial^2 \log p(\boldsymbol{F}_J;\boldsymbol{\lambda}_J)}{(\partial\lambda_j)^2}\right] = \mathbb{E}\left[\frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial\lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right]^2 - \left(\frac{\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial\lambda_j}{B_n(\lambda_\bullet, \boldsymbol{w})}\right)^2 \tag{3.2}$$

$$= \frac{\partial^2 B_n(\lambda_\bullet, \boldsymbol{w})/(\partial\lambda_j)^2}{B_n(\lambda_\bullet, \boldsymbol{w})} - \left(\frac{\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial\lambda_j}{B_n(\lambda_\bullet, \boldsymbol{w})}\right)^2$$

$$- \mathbb{E}\left[\frac{\partial}{\partial\lambda_j}\left(\frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial\lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right)\right], \tag{3.3}$$

and the off-diagonal element is, for $i \in [J]$*,* $j \in [J]$*,* $i \neq j$*,*

$$\mathbb{E}\left[-\frac{\partial^2 \log p(\boldsymbol{F}_J;\boldsymbol{\lambda}_J)}{\partial\lambda_i\partial\lambda_j}\right] = \frac{\partial^2 B_n(\lambda_\bullet, \boldsymbol{w})/(\partial\lambda_i\partial\lambda_j)}{B_n(\lambda_\bullet, \boldsymbol{w})} - \frac{(\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial\lambda_i)(\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial\lambda_j)}{B_n(\lambda_\bullet, \boldsymbol{w})^2}. \tag{3.4}$$

Observing Theorem 3.3, we only need univariate marginal distributions to evaluate the information matrix. This marginalization corresponds to the sufficiency of F_j for λ_j; see Corollary 3.1. The factorization theorem of sufficiency suggests that this marginalization is the property of a multiplicative probability measure (2.1), and a distribution on $\mathcal{F}_{n,J}$ outside (2.1) seems unable to enjoy this marginalization. For example, Morel and Nagaraj [40] consider a simple finite mixture of multinomial distributions, which causes computational issues on the information matrix as discussed by Neerchal and Morel [42].

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Paul, Balasooriya, and Banerjee [43] derive the information matrix of the Dirichlet multinomial mixture that is parameterized by $(\boldsymbol{\pi}_J; \lambda_\bullet)$. They exploit the marginalization of the BPD.

3.2. Marginal moments

The marginal moments of the BPD are the same as those of the CCP distribution because of Theorem 2.8. Theorem 1 of [30] shows the univariate marginal expectation of the CCP distribution. Second moments are newly derived below.

Theorem 3.4. *Suppose that* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}; \boldsymbol{w})$. *Then*

$$\mathbb{E}(F_j) = n\pi_j, \quad j \in [J], \tag{3.5}$$

$$\mathrm{Var}\,(F_j) = n\pi_j(1-\pi_j)\phi(n, \lambda_\bullet, \boldsymbol{w}), \quad j \in [J], \tag{3.6}$$

and

$$\mathrm{Cov}\,(F_i, F_j) = -n\pi_i\pi_j\phi(n, \lambda_\bullet, \boldsymbol{w}), \quad i \in [J],\ j \in [J],\ i \neq j, \tag{3.7}$$

where for $n \geq 2$,

$$\phi(n, \lambda_\bullet, \boldsymbol{w}) = 1 + \frac{\lambda_\bullet (n-1)!}{B_n(\lambda_\bullet, \boldsymbol{w})} \sum_{i=0}^{n-2} \frac{B_i(\lambda_\bullet, \boldsymbol{w}) w_{n-i}}{i!\,(n-i-2)!}$$

and $\phi(1, \lambda_\bullet, \boldsymbol{w}) := 1$.

Observing Theorem 3.4, the covariance matrix of the BPD is that of $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$ times $\phi(n, \lambda_\bullet, \boldsymbol{w})$. Hence the following corollary to Theorem 3.4 holds.

Corollary 3.2. *Any BPD with the same* $(n, \boldsymbol{\pi}_J)$ *has the same correlation matrix as that of* $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$, *irrespective of* $\boldsymbol{w}$.

On the multiplication function $\phi(n, \lambda_\bullet, \boldsymbol{w})$, we have the following bounds.

Theorem 3.5. *Suppose that* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}; \boldsymbol{w})$ *and* $n \geq 2$. *Then*

$$\phi(n, \lambda_\bullet, \boldsymbol{w}) \geq 2 - \frac{B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1}{B_n(\lambda_\bullet, \boldsymbol{w})} \geq 1 \tag{3.8}$$

and

$$\phi(n, \lambda_\bullet, \boldsymbol{w}) \leq 1 + (n-1)\left\{1 - \frac{B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1}{B_n(\lambda_\bullet, \boldsymbol{w})}\right\} < n. \tag{3.9}$$

The lower bound of (3.8) is attained, that is, $\phi(n, \lambda_\bullet, \boldsymbol{w}) = 1$ *if and only if* $w_i = 0$ *for* $i \geq 2$. *Also, additionally when* $\boldsymbol{w} \in \mathcal{W}$,

$$\phi(n, \lambda_\bullet, \boldsymbol{w}) = 1 + \mathrm{O}(\lambda_\bullet^{-1}), \quad \lambda_\bullet \to \infty, \tag{3.10}$$

and

$$\phi(n, \lambda_\bullet, \boldsymbol{w}) = n - \mathrm{O}(\lambda_\bullet), \quad \lambda_\bullet \to 0. \tag{3.11}$$

When $w_i = 0$ for $i \geq 2$, the BPD reduces to the multinomial distribution. Hence Theorem 3.5 implies that the marginal variance of the multinomial distribution is the smallest among the BPD, given cell probabilities. It corresponds to the known fact that the compound Poisson

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

distribution is overdispersed, i.e. of greater variance than the Poisson distribution; see [47]. Accordingly, (3.10) validates Theorem 2.7, where the BPD converges in distribution to the multinomial distribution as $\lambda_\bullet \to \infty$.

Remark 3.1. As $\phi(n, \lambda_\bullet, \boldsymbol{w}) \to n$, $\mathrm{Var}\,(F_j/n) \to \pi_j(1-\pi_j)$. This limit corresponds to the binomial sampling of $\{\mathbb{P}(F_j = n) = \pi_j, \mathbb{P}(F_j = 0) = 1 - \pi_j\}$.

If a distribution over $\mathcal{F}_{n,J}$ has the same mean vector as (3.5) and its covariance matrix is that of $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$ times $\{1 + \rho^2(n-1)\}$, then it is called an overdispersed multinomial distribution with an overdispersion parameter ρ by [42]. Theorem 3.5 shows that $1 \le \phi(n, \lambda_\bullet, \boldsymbol{w}) < n$. Therefore a BPD is an overdispersed multinomial distribution with $\rho^2 \in [0, 1)$. In [27], an overdispersed multinomial distribution is discriminated from the standard quasi-likelihood approach to overdispersion, in which the covariance matrix of $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$ is multiplied by a constant independent of n.

An important implication of Theorem 3.4 is that if an observed covariance matrix is not proportional to that of the multinomial distribution of the same cell probabilities, then the observation does not follow the BPD.

Testing the proportionality of covariance matrices has been motivated by, for example, genetic experiments [22]. Recently, the high-dimensional setting of $n \ll J$ has renewed an interest in the test. See [9], [54], and their reviews of the literature, where a continuous distribution is typically assumed, unlike our study.

4. Random cell probabilities

4.1. Bayesian inference

A Bayesian assumes a prior distribution for cell probabilities of the multinomial distribution. It is typically the Dirichlet distribution of the following density:

$$f(\boldsymbol{\pi}_J; \boldsymbol{\alpha}_J) = \Gamma(\alpha_\bullet) \prod_{j=1}^{J} \pi_j^{\alpha_j - 1} / \Gamma(\alpha_j), \qquad \boldsymbol{\pi}_J \in \triangle^{J-1}, \tag{4.1}$$

where $\boldsymbol{\alpha}_J := (\alpha_1, \alpha_2, \ldots, \alpha_J) \in \mathbb{R}_+^J$. The distribution (4.1) is referred to by $\mathtt{Dir}_J(\boldsymbol{\alpha}_J)$.

In the case of the BPD, we first rewrite the PMF of $\mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \lambda_\bullet); \boldsymbol{w})$ as

$$p(\boldsymbol{f}_J) = \binom{n}{\boldsymbol{f}_J} \frac{1}{B_n(\lambda_\bullet, \boldsymbol{w})} \sum_{l_1=0}^{f_1} \sum_{l_2=0}^{f_2} \cdots \sum_{l_J=0}^{f_J} \prod_{j=1}^{J} \pi_j^{l_j} \lambda_\bullet^{l_j} B_{f_j, l_j}(\boldsymbol{w}). \tag{4.2}$$

The right-hand side of (4.2) is the weighted sum of $\prod_{j=1}^{J} \pi_j^{l_j}$, which is proportional to the PMF of $\mathtt{Mn}_J(l_\bullet, \boldsymbol{\pi}_J)$. This view leads to a Bayesian argument of Theorem 4.1, whose proof is straightforward and thus omitted.

Theorem 4.1. *Let $\lambda_\bullet \in \mathbb{R}_+$. Suppose that $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \lambda_\bullet); \boldsymbol{w})$, and $\boldsymbol{\pi}_J \sim \mathtt{Dir}_J(\boldsymbol{\alpha}_J)$. Then the marginal distribution of $\boldsymbol{F}_J$ has the following PMF:*

$$p(\boldsymbol{f}_J) = \binom{n}{\boldsymbol{f}_J} \frac{\Gamma(\alpha_\bullet)}{B_n(\lambda_\bullet, \boldsymbol{w})} \sum_{l_1=0}^{f_1} \cdots \sum_{l_J=0}^{f_J} \frac{1}{\Gamma(\alpha_\bullet + l_\bullet)} \prod_{j=1}^{J} \lambda_\bullet^{l_j} B_{f_j, l_j}(\boldsymbol{w}) \frac{\Gamma(\alpha_j + l_j)}{\Gamma(\alpha_j)}, \tag{4.3}$$

and the posterior distribution of $\boldsymbol{\pi}_J$ given $\boldsymbol{f}_J$ is the finite mixture of

$$\mathtt{Dir}_J(l_1 + \alpha_1, l_2 + \alpha_2, \ldots, l_J + \alpha_J)$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

with the following density:

$$f(\boldsymbol{\pi}_J \mid \boldsymbol{f}_J) = \sum_{l_1=0}^{f_1} \cdots \sum_{l_J=0}^{f_J} \Gamma(l_\bullet + \alpha_\bullet) \prod_{j=1}^{J} \frac{\pi_j^{l_j+\alpha_j-1}}{\Gamma(l_j+\alpha_j)} \left(\frac{\prod_{j=1}^{J} \lambda_\bullet^{l_j} B_{f_j,l_j}(\boldsymbol{w})}{\prod_{j=1}^{J} B_{f_j}(\lambda_\bullet, \boldsymbol{w})} \right). \tag{4.4}$$

We note that (4.4) reduces to (4.1) when $\boldsymbol{f}_J = (0, 0, \ldots, 0)$. If $n \geq 1$ and $\boldsymbol{w} \in \mathcal{W}$ then taking $\lambda_\bullet \to \infty$ results in another simplification: the BPD converges in distribution to the multinomial distribution by Theorem 2.7, and thus the posterior (4.4) should converge in distribution to $\mathtt{Dir}_J(\alpha_1 + f_1, \alpha_2 + f_2, \ldots, \alpha_J + f_J)$. This fact can be confirmed by regarding (4.4) as a polynomial in $\lambda_\bullet$. The highest degree of this polynomial is $\sum_{j=1}^{J} f_j$, which is the case of $l_j = f_j$ for all $j \in [J]$.

Remark 4.1. The finite mixture of Dirichlet distributions is a conjugate prior of cell probabilities for the BPD, which results from Theorem 4.1. In particular, the Dirichlet distribution is widely known as a conjugate prior for the multinomial distribution.

4.2. Mixing

Mixing of a distribution integrates out its parameters that are distributed. For example, mixing $\boldsymbol{\pi}_J$ of $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J)$ with $\mathtt{Dir}_J(\boldsymbol{\alpha}_J)$ is denoted by $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J) \wedge_{\boldsymbol{\pi}_J} \mathtt{Dir}_J(\boldsymbol{\alpha}_J)$, which results in the Dirichlet multinomial mixture (5.1). The following arguments consider mixing $\boldsymbol{\pi}_J$ of $\mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \lambda_\bullet); \boldsymbol{w})$ with some distribution on a unit simplex.

Lemma 4.1. *Let* $\boldsymbol{P}_J := (P_1, P_2, \ldots, P_J)$ *be distributed over* $\triangle^{J-1}$. *Given* $\boldsymbol{P}_J = \boldsymbol{\pi}_J$, *suppose that* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \lambda_\bullet); \boldsymbol{w})$. *Then*

$$\begin{aligned} \mathbb{E}(F_j) &= n\mathbb{E}(P_j), \quad j \in [J], \\ \mathrm{Var}\,(F_j) &= n\phi(n, \lambda_\bullet, \boldsymbol{w})\mathbb{E}(P_j(1-P_j)) + n^2 \,\mathrm{Var}\,(P_j), \quad j \in [J], \end{aligned}$$

and

$$\begin{aligned} \mathrm{Cov}\,(F_i, F_j) = &-n\phi(n, \lambda_\bullet, \boldsymbol{w})\mathbb{E}(P_i)\mathbb{E}(P_j) \\ &+ \{n^2 - n\phi(n, \lambda_\bullet, \boldsymbol{w})\}\,\mathrm{Cov}\,(P_i, P_j), \quad i \in [J],\ j \in [J],\ i \neq j. \end{aligned}$$

Correlations are immediate from Lemma 4.1, and the proof of Theorem 4.2 is thus omitted.

Theorem 4.2. *In Lemma 4.1, if* $\phi(n, \lambda_\bullet, \boldsymbol{w}) = \mathrm{o}(n)$ *then* $\lim_{n\to\infty} \mathrm{Cor}\,(F_i, F_j) = \mathrm{Cor}\,(P_i, P_j)$.

On the multinomial distribution, $\phi(n, \lambda_\bullet, \boldsymbol{w}) = 1 = \mathrm{o}(n)$. Hence, by Theorem 4.2, the correlation between any two mixed multinomial variates, F_i and F_j, whatever the mixing distribution, converges to $\mathrm{Cor}\,(P_i, P_j)$ as $n \to \infty$, as pointed out by Mosimann [41, p. 72].

If the moments of $\boldsymbol{P}_J$ are like (4.6), (4.7), and (4.8), then $\mathrm{Cor}\,(F_i, F_j) = \mathrm{Cor}\,(P_i, P_j)$ for all n under the assumption of Lemma 4.1; see Theorem 4.7.

Aitchison [3] calls data over a simplex compositional data and provides a good introduction to compositional data analysis; see [20] for a recent review. In this literature, the merger of margins is called amalgamation.

The Dirichlet distribution is closed under amalgamation because it equals the conditional distribution of independent gamma distributions given their total, where the gamma distribution is ID. This structure parallels that of the CCP distribution; see Section 2.4.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Although the CCP distribution depends on a total frequency n, the Dirichlet distribution does not depend on the sum of gamma variates. This property indeed characterizes the Dirichlet distribution.

Let X_j, $j \in [J]$, be independent non-degenerate and positive random variables. Write $X_\bullet = \sum_{j=1}^J X_j$ and let $P_j = X_j/X_\bullet$. Then Lukacs [38] shows for $J = 2$ that each P_j is independent of $X_\bullet$ if and only if all X_j are gamma-distributed with the same scale parameter. This result is generalized in [41, p. 75], for all J.

If each X_j is independently gamma-distributed with the same scale parameter, then $\boldsymbol{P}_J$ is Dirichlet-distributed independently from $X_\bullet$. Conversely, if $\boldsymbol{P}_J$ is independent of $X_\bullet$ then, by definition, each P_j is independent of $X_\bullet$, which is equivalent to all X_j being independently gamma-distributed with the same scale parameter. Therefore $\boldsymbol{P}_J$ is independent of $X_\bullet$ if and only if X_j are independently gamma-distributed with the same scale parameter.

For example, the gamma distribution is a special case of the generalized inverse Gaussian (GIG) distribution. Barndorff-Nielsen and Jørgensen [5] consider the conditional distribution of independent GIG-distributed variables given their total. This total remains as a parameter, except for the case of the Dirichlet distribution.

Thus the conditional distribution of $\boldsymbol{P}_J$ given $X_\bullet$ generally depends on $X_\bullet$. By contrast, the marginal distribution of $\boldsymbol{P}_J$ is independent of $X_\bullet$, which is integrated out. Favaro, Hadjicharalambous, and Prünster [21] consider the family of the marginal distribution of $\boldsymbol{P}_J$, where each X_j has an independent ID distribution over $\mathbb{R}_+$. This family is called the normalized ID (NID) distribution, which includes the Dirichlet distribution, the normalized half stable distribution [7], and the normalized IG distribution [37], to name a few.

To be precise, let a random variable X be ID over $\mathbb{R}_+$. Write the Lévy–Khintchine representation of the Laplace transform of X as

$$\mathbb{E}(\mathrm{e}^{-uX}) = \exp(-\psi(u)) = \exp\left\{-\int_0^\infty (1-\mathrm{e}^{-ux})\nu(\mathrm{d}x)\right\}, \quad u \geq 0,$$

where ψ is called a Laplace exponent and ν is the Lévy measure. It is known that $\mathbb{P}(X = 0) = 0$ is equivalent to

$$\int_{(0,\infty)} \nu(\mathrm{d}x) = \infty; \tag{4.5}$$

see e.g. [52, Proposition 4.4]. We assume (4.5) for Lévy measures below.

We construct a subfamily of the NID distribution that is closed under amalgamation, based on the following fact. If $\exp(-\psi(u))$ is the Laplace transform of an ID distribution over $\mathbb{R}_+$ then, for all $\mu > 0$, $\exp(-\mu\psi(u))$ is also the Laplace transform of an ID distribution over $\mathbb{R}_+$; see [52, Proposition 6.1(i)].

Definition 4.1. Let $\boldsymbol{\mu}_J := (\mu_1, \mu_2, \ldots, \mu_J) \in \mathbb{R}_+^J$. Let $\exp(-\psi(u))$ be the Laplace transform of an ID distribution over $\mathbb{R}_+$ characterized by a Lévy measure ν. Let an independent random variable X_j, $j \in [J]$, have the Laplace transform $\exp(-\mu_j\psi(u))$. Then the distribution of $(X_1/X_\bullet, X_2/X_\bullet, \ldots, X_J/X_\bullet)$ defines $\mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu)$.

NID is closed under amalgamation, which simplifies its marginal moments. Similarly to a cell probability, define $\rho_j := \mu_j/\mu_\bullet$.

Theorem 4.3. *Suppose that* $\boldsymbol{P}_J \sim \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu)$. *Write the Laplace exponent as*

$$\psi(u) = \int_0^\infty (1-\mathrm{e}^{-ux})\nu(\mathrm{d}x).$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Then

$$\mathbb{E}(P_j) = \rho_j, \quad j \in [J], \tag{4.6}$$

$$\mathrm{Var}\,(P_j) = \rho_j(1-\rho_j)\varphi(\mu_\bullet, \psi), \quad j \in [J], \tag{4.7}$$

$$\mathrm{Cov}\,(P_i, P_j) = -\rho_i\rho_j\varphi(\mu_\bullet, \psi), \quad i \in [J],\ j \in [J],\ i \neq j, \tag{4.8}$$

and

$$\mathrm{Cor}\,(P_i, P_j) = -\sqrt{\frac{\rho_i\rho_j}{(1-\rho_i)(1-\rho_j)}}, \quad i \in [J],\ j \in [J],\ i \neq j, \tag{4.9}$$

where

$$\varphi(\mu_\bullet, \psi) = 1 - \mu_\bullet^2 \int_0^\infty u\left(\frac{\mathrm{d}\psi(u)}{\mathrm{d}u}\right)^2 \exp\left(-\mu_\bullet\psi(u)\right)\mathrm{d}u. \tag{4.10}$$

The moments shown in Theorem 4.3 are parallel to those of the BPD given by Theorem 3.4. We can regard $\varphi(\mu_\bullet, \psi)$ as an index of overdispersion similarly to $\phi(n, \lambda_\bullet, \boldsymbol{w})$ of Theorem 3.4. Theorem 4.4 below shows the interval of $\varphi(\mu_\bullet, \psi)$, which is comparable to $1 \leq \phi(n, \mu_\bullet, \boldsymbol{w}) < n$; see Theorem 3.5.

Theorem 4.4. *In Theorem 4.3,* $0 \leq \varphi(\mu_\bullet, \psi) < 1$.

Theorem 4.4 holds because the elements of the second term of the right-hand side of (4.10) are strictly positive due to (4.5) and the variance (4.7) has to be non-negative. It is noteworthy that, by (A.18), $\varphi(\mu_\bullet, \psi) = 0$ when $\psi(u) = uc$ for some positive c, where $\mathbb{P}(X_j = c\mu_j) = 1$.

The parallelism in overdispersion between NID and the BPD will be justified by Theorem 4.8, based on the following result.

Theorem 4.5. *Suppose that* $\boldsymbol{\mu}_J \in \mathbb{R}_+^J$. *Write the Laplace exponent of a Lévy measure* ν *as*

$$\psi(u) = \int_0^\infty (1 - \mathrm{e}^{-ux})\nu(\mathrm{d}x).$$

Let

$$w_i = -\mathrm{d}^i\psi(1-s)/(\mathrm{d}s)^i|_{s=0}, \quad i \in \mathbb{N}. \tag{4.11}$$

Then

$$\mathtt{BP}_J(n, \boldsymbol{\mu}_J; \boldsymbol{w}) = \mathtt{Mn}_J(n, \boldsymbol{\pi}_J) \underset{\boldsymbol{\pi}_J}{\wedge} \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu). \tag{4.12}$$

Not every BPD can be expressed as mixed multinomial with NID because $\boldsymbol{w}$ may not define the Laplace exponent of a proper distribution, by (4.11). Correspondingly, not every compound Poisson distribution is mixed Poisson with an ID distribution; see [52, Example 12.2].

The right-hand side of (4.12) implies i.i.d. sampling given π_J. Accordingly, the law of large numbers yields Theorem 4.6, where '$\xrightarrow{d}$' denotes convergence in distribution.

Theorem 4.6. *Let* $(F_{1,n}, F_{2,n}, \ldots, F_{J,n}) \sim \mathtt{BP}_J(n, \boldsymbol{\mu}_J; \boldsymbol{w})$. *Suppose that the Laplace exponent of a Lévy measure* ν *satisfies* (4.11). *Then*

$$(F_{1,n}/n, F_{2,n}/n, \ldots, F_{J,n}/n) \xrightarrow{d} \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu) \quad \text{as } n \to \infty.$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Now we combine Theorem 4.3 with Lemma 4.1 to obtain the marginal moments of a mixed BPD with NID. The proof of Theorem 4.7 is omitted.

Theorem 4.7. *Suppose that* $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \lambda_\bullet), \boldsymbol{\nu}) \wedge_{\boldsymbol{\pi}_J} \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu)$, *where* $\lambda_\bullet \in \mathbb{R}_+$, $\boldsymbol{\nu} \in \Omega_1$. *Then*

$$\begin{aligned}
\mathbb{E}(F_j) &= n\rho_j, \quad j \in [J], \\
\mathrm{Var}\,(F_j) &= n\rho_j(1-\rho_j)\{\phi(n, \lambda_\bullet, \boldsymbol{\nu}) - \phi(n, \lambda_\bullet, \boldsymbol{\nu})\varphi(\mu_\bullet, \psi) + n\varphi(\mu_\bullet, \psi)\}, \quad j \in [J], \\
\mathrm{Cov}\,(F_i, F_j) &= -n\rho_i\rho_j\{\phi(n, \lambda_\bullet, \boldsymbol{\nu}) - \phi(n, \lambda_\bullet, \boldsymbol{\nu})\varphi(\mu_\bullet, \psi) + n\varphi(\mu_\bullet, \psi)\},
\end{aligned}$$

where $i \in [J]$, $j \in [J]$, $i \neq j$, *and* $\mathrm{Cor}\,(F_i, F_j)$ *is the same as* $\mathrm{Cor}\,(P_i, P_j)$ *of* (4.9).

Theorem 4.7 actually implies (4.13), to be used for more results.

Theorem 4.8. *In Theorem 4.5,*

$$1 + \varphi(\mu_\bullet, \psi)(n-1) = \phi(n, \mu_\bullet, \boldsymbol{w}). \tag{4.13}$$

By (4.13), the known interval of $1 \le \phi(n, \mu_\bullet, \boldsymbol{w}) < n$ is equivalent to that of $0 \le \varphi(\mu_\bullet, \psi) < 1$, which is an alternative proof of Theorem 4.4. Also (4.13), together with (3.10) and (3.11), results in Theorem 4.9 below.

Theorem 4.9. *Suppose that* $\varphi(\mu, \psi)$ *of* (4.10) *depends on* μ. *Then* $\varphi(\mu, \psi) = \mathrm{O}(\mu^{-1})$ *as* $\mu \to \infty$, *and* $\varphi(\mu, \psi) = 1 - \mathrm{O}(\mu)$ *as* $\mu \to 0$.

Let us consider the stable distribution over $\mathbb{R}_+$; see [52, Example 4.9]. In this case, $\psi(u) = u^\gamma$ for $\gamma \in (0, 1]$, where we can obtain $\boldsymbol{w}$ explicitly by (4.11) as $(\gamma \prod_{j=1}^{i-1}(j-\gamma))_{i=1}^\infty$. Then, by (4.10),

$$\varphi(\mu, \cdot^\gamma) = 1 - \gamma, \tag{4.14}$$

which is free of μ. The special case of $\gamma = 1$ has already appeared after Theorem 4.4 as an example for which $\varphi(\mu_\bullet, \psi) = 0$ when $\psi(u) = u$. By (4.13) and (4.14), we have

$$\phi\left(n, \mu, \left(\gamma \prod_{j=1}^{i-1}(j-\gamma)\right)_{i=1}^\infty\right) = 1 + (n-1)(1-\gamma), \tag{4.15}$$

which is also free of μ.

Let us recall definition (2.11) of $\mathcal{W}^c$, in which $\mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \mu); \boldsymbol{w})$ is free of μ. By Theorem 2.7, $\phi(n, \mu, \boldsymbol{w})$ depends on μ when $\mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \mu); \boldsymbol{w})$ depends on μ. Hence the following result holds.

Theorem 4.10. *For* $\gamma \in (0, 1]$, $(\gamma \prod_{j=1}^{i-1}(j-\gamma))_{i=1}^\infty \in \mathcal{W}^c$.

The Laplace transform of the stable distribution is generally expressed as $\exp(-\lambda u^\gamma)$. If the mean parameter of the Poisson distribution follows this stable distribution, then the PGF of the mixed Poisson distribution results in $G(z) = \exp(-\lambda(1-z)^\gamma)$, as pointed out by Devroye [19]; see also Lemma A.1 in the Appendix. This $G(z)$ defines the discrete stable distribution [51]. It is also a compound Poisson distribution by rewriting that $G(z) = \exp(\lambda(g(z) - 1))$, where $g(z) = 1 - (1-z)^\gamma$. This $g(z)$ defines the Sibuya distribution [49], whose PMF is

$$p(i) = \gamma \prod_{j=1}^{i-1}(j-\gamma)/i!, \quad i \in \mathbb{N}.$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

The CCP distribution generated by the discrete stable distribution can be expressed as

$$\mathtt{BP}_J\left(n, (\boldsymbol{\pi}_J; \lambda_\bullet), \left(\gamma \prod_{j=1}^{i-1} (j-\gamma)\right)_{i=1}^{\infty}\right).$$

The second moment formulae, seemingly new, of this distribution are immediate from (4.15).

5. Examples

5.1. Negative hypergeometric distribution

Suppose that $w_i = (i-1)!$. Then $B_{n,k}(\boldsymbol{w})$ is an unsigned Stirling number of the first kind $|s(n,k)|$, and $B_n(\lambda, \boldsymbol{w}) = \lambda(\lambda+1)\cdots(\lambda+n-1)$; see [44, eq. 1.54] for example. Consequently, $\mathtt{BP}_J(n, \boldsymbol{\alpha}_J; (0!, 1!, 2!, \dots))$ has the following PMF:

$$p(\boldsymbol{f}_J) = \binom{n}{\boldsymbol{f}_J} \frac{\Gamma(\alpha_\bullet)}{\Gamma(\alpha_\bullet + n)} \prod_{j=1}^{J} \frac{\Gamma(\alpha_j + f_j)}{\Gamma(\alpha_j)}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}. \tag{5.1}$$

This PMF (5.1) defines the negative (multivariate) hypergeometric distribution [10].

The marginal moments of this distribution can be found in, for example, [35, pp. 81–82]. When the distribution of $\boldsymbol{F}_J$ is that of (5.1),

$$\mathbb{E}(F_j) = n\alpha_j/\alpha_\bullet, \tag{5.2}$$

$$\mathrm{Var}\,(F_j) = n\alpha_j/\alpha_\bullet(1-\alpha_j/\alpha_\bullet)(1+(n-1)/(\alpha_\bullet+1)), \tag{5.3}$$

and

$$\mathrm{Cov}\,(F_i, F_j) = -n\alpha_i/\alpha_\bullet\alpha_j/\alpha_\bullet(1+(n-1)/(\alpha_\bullet+1)).$$

These moments are consistent with Theorem 3.4. Comparing (5.3) with (3.6), the overdispersion index $\phi(n, \alpha_\bullet, (0!, 1!, 2!, \dots))$ equals $1+(n-1)/(\alpha_\bullet+1)$, which is equivalent to Proposition 5.1 below. As a check of Theorem 3.4, a direct proof of Proposition 5.1 is provided in the Appendix. Also, it is straightforward to confirm Theorem 3.5 on $\phi(n, \alpha_\bullet, (0!, 1!, 2!, \dots))$.

Proposition 5.1. *For $n \in \mathbb{N} \setminus \{1\}$ and $\alpha \in \mathbb{R}_+$,*

$$\frac{(n-1)!\,\alpha}{\Gamma(\alpha+n)} \sum_{i=0}^{n-2} \frac{(n-i-1)\Gamma(\alpha+i)}{\Gamma(i+1)} = \frac{n-1}{\alpha+1}.$$

The negative hypergeometric distribution (5.1) is the mixed multinomial distribution with the Dirichlet distribution (4.1) as discussed by Ishii and Hayakawa [34]. This view corresponds to the case of $\boldsymbol{w} = (1, 0, 0, \dots)$ in Lemma 4.1. When $\boldsymbol{P}_J \sim \mathtt{Dir}_J(\boldsymbol{\alpha}_J)$,

$$\mathbb{E}(P_j) = \alpha_j/\alpha_\bullet, \tag{5.4}$$

$$\mathbb{E}({P_j}^2) = \alpha_j(1+\alpha_j)/(\alpha_\bullet(\alpha_\bullet+1)),$$

and

$$\mathrm{Var}\,(P_j) = \frac{\alpha_j}{\alpha_\bullet}\left(1-\frac{\alpha_j}{\alpha_\bullet}\right)\frac{1}{1+\alpha_\bullet}. \tag{5.5}$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

By putting these moments into the moment formulae in Lemma 4.1 with $\phi(n, \lambda_\bullet, \boldsymbol{w}) = 1$, we can reproduce (5.2) and (5.3).

The Dirichlet distribution (4.1) is an instance of $\mathtt{NID}$, where $\nu(\mathrm{d}x) = x^{-1}\exp(-\lambda x)\,\mathrm{d}x$. In this case, $\psi(u) = \log(\lambda + u) - \log\lambda$. By putting $\mu_j = \alpha_j$, Theorem 4.3 validates (5.4). On the variance (5.5), we note an alternative expression (A.18) of

$$\varphi(\alpha_\bullet, \log(\lambda + \cdot) - \log\lambda) = \alpha_\bullet \lambda^{\alpha_\bullet} \int_0^\infty u(\lambda + u)^{-\alpha_\bullet - 2}\,\mathrm{d}u.$$

Further,

$$\int_0^\infty u(\lambda + u)^{-\alpha_\bullet - 2}\,\mathrm{d}u = (\lambda^{\alpha_\bullet}\alpha_\bullet(\alpha_\bullet + 1))^{-1},$$

using integration by parts. Hence $\varphi(\alpha_\bullet, \log(\lambda + \cdot) - \log\lambda) = (\alpha_\bullet + 1)^{-1}$, which validates (4.7). It follows that $\lim_{\alpha_\bullet \to \infty} \varphi(\alpha_\bullet, \log(\lambda + \cdot) - \log\lambda) = 0$ and $\lim_{\alpha_\bullet \to 0} \varphi(\alpha_\bullet, \log(\lambda + \cdot) - \log\lambda) = 1$. These results are consistent with Theorem 4.9.

Also, we can confirm Theorem 4.5 as $-\psi(1-s) = \log(1 - 1/(\lambda+1)) - \log(1 - s/(\lambda+1))$. Let $1/(\lambda + 1) = \theta$. Then

$$-\psi(1 - s) = \log(1 - \theta) + \sum_{i=1}^{\infty} (s\theta)^i / i.$$

Therefore

$$-\mathrm{d}^i\psi(1 - s)/(\mathrm{d}s)^i|_{s=0} = \theta^i(i - 1)! \quad \text{or} \quad w_i = \theta^i(i - 1)!.$$

By (1.1), $B_n(\boldsymbol{w}) = \theta^n B_n(0!, 1!, 2!, \ldots)$. Then θ cancels out, and thus

$$\mathtt{BP}_J(n, \boldsymbol{\alpha}_J; (0!, 1!, 2!, \ldots)) = \mathtt{Mn}_J(n, \boldsymbol{\pi}_J) \underset{\boldsymbol{\pi}_J}{\wedge} \mathtt{NID}_J(\boldsymbol{\alpha}_J \times \nu).$$

We verify the simplest case of Theorem 4.6. If $X \sim \mathtt{BP}_2(n, (\alpha_1, \alpha_2); (0!, 1!, 2!, \ldots))$, then the moment generating function (MGF) of X is $M(z) = {}_2F_1(-n, \alpha_1; \alpha_1 + \alpha_2; 1 - \mathrm{e}^z)$; see e.g. [36, eq. 6.10]. The MGF of X/n is

$$\begin{aligned} M(z/n) &= \sum_{i=0}^{\infty} \frac{\Gamma(n+1)\Gamma(\alpha_1 + i)\Gamma(\alpha_1 + \alpha_2)}{\Gamma(n - i + 1)\Gamma(\alpha_1)\Gamma(\alpha_1 + \alpha_2 + i)\Gamma(i + 1)} \left\{(z/n)^i + \mathrm{O}(n^{-i-1})\right\} \\ &\to \sum_{i=0}^{\infty} \frac{\Gamma(\alpha_1 + i)\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_1 + \alpha_2 + i)\Gamma(i + 1)} z^i \end{aligned}$$

as $n \to \infty$. The last expression equals the MGF of the beta distribution or $\mathtt{Dir}_2(\alpha_1, \alpha_2)$. Hence X/n converges weakly to $\mathtt{NID}_2((\alpha_1, \alpha_2) \times \nu)$.

Theorem 4.1 arranges the Dirichlet Dirichlet-multinomial mixture so that

$$\boldsymbol{F}_J \mid \boldsymbol{P}_J \sim \mathtt{Mn}_J(n, \boldsymbol{P}_J), \quad \boldsymbol{P}_J \mid \boldsymbol{\lambda}_J \sim \mathtt{Dir}_J(\boldsymbol{\lambda}_J) \quad \text{and} \quad \boldsymbol{\lambda}_J/\lambda_\bullet = \boldsymbol{\pi}_J \sim \mathtt{Dir}_J(\boldsymbol{\alpha}_J).$$

In this case

$$\boldsymbol{F}_J \mid \boldsymbol{\lambda}_J \sim \mathtt{BP}_J(n, \boldsymbol{\lambda}_J, (0!, 1!, 2!, \ldots)),$$

and the marginal distribution of $\boldsymbol{F}_J$ is (4.3), where

$$B_n(\lambda_\bullet, \boldsymbol{w}) = \Gamma(n + \lambda_\bullet)/\Gamma(\lambda_\bullet) \quad \text{and} \quad B_{f,l}(\lambda_\bullet, \boldsymbol{w}) = \lambda_\bullet^l |s(f, l)|.$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Also, the posterior of $\boldsymbol{\pi}_J$ is (4.4).

The Dirichlet multinomial mixture is the conditional distribution of independent negative-binomial-distributed variables given the total; see e.g. [32]. The Dirichlet distribution is the conditional distribution of independent gamma-distributed variables given the total, as discussed in Section 4.2. Hence the Dirichlet Dirichlet-multinomial mixture is the multivariate and conditional (given n) case of the gamma-negative-binomial mixture discussed in [56, Section 5].

5.2. Quasi-multinomial distribution

An interesting instance of the BPD arises when $w_i = i^{i-1}$. In this case, $B_{n,k}(\boldsymbol{w}) = \binom{n-1}{k-1} n^{n-k}$, and $B_n(\lambda, \boldsymbol{w}) = \lambda(\lambda + n)^{n-1}$; see [44, eq. 1.56], for example. The PMF of $\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (1^0, 2^1, 3^2, \dots))$ is expressed as

$$p(\boldsymbol{f}_J) = \binom{n}{\boldsymbol{f}_J} \frac{1}{\lambda_\bullet (\lambda_\bullet + n)^{n-1}} \prod_{j=1}^{J} \lambda_j (\lambda_j + f_j)^{f_j - 1}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}. \tag{5.6}$$

This expression is the canonical form of the quasi-multinomial distribution [15] as a BPD. The special case of $J = 2$, called the quasi-binomial distribution (type 2), is proposed in [14].

Since $B_{n,k}(1^0, 2^1, 3^2, \dots)$ is explicit, we can express the Dirichlet quasi-multinomial mixture in a closed form by Theorem 4.1, which is left for future work.

As noted in [30], the quasi-multinomial distribution is a CCP distribution by regarding the Lagrangian Poisson distribution [13] as a compound Poisson distribution. The PGF of the Lagrangian Poisson distribution is expressed implicitly; see e.g. [12, p. 166]. Its expression as a mixed Poisson distribution is also implicit. Therefore we do not discuss relationships between the quasi-multinomial distribution and the NID distribution in the present article.

A more intuitive parameterization of (5.6) with cell probabilities is utilized in [30]. Let $\lambda_\bullet = 1/\beta$ in (5.6). Then $\mathtt{BP}_J(n, (\boldsymbol{\pi}_J; 1/\beta); (1^0, 2^1, 3^2, \dots))$ is

$$p(\boldsymbol{f}_J) = \binom{n}{\boldsymbol{f}_J} \frac{1}{(1 + n\beta)^{n-1}} \prod_{j=1}^{J} \pi_j (\pi_j + f_j \beta)^{f_j - 1}, \quad \boldsymbol{f}_J \in \mathcal{F}_{n,J}, \tag{5.7}$$

which reduces to the PMF of the multinomial distribution when $\beta = 0$.

The marginal moments of the quasi-multinomial distribution are derived by Consul and Mittal [15] up to second order. Equivalent expressions to their results under the parameterization of (5.7) are $\mathbb{E}(F_j) = n\pi_j$,

$$\mathrm{Var}\,(F_j) = n\pi_j \left[\left\{ \frac{(n-1)!}{1 + n\beta} \sum_{i=2}^{n} \frac{\pi_j + i\beta}{(n-i)!} \left(\frac{\beta}{1 + n\beta} \right)^{i-2} \right\} + 1 - n\pi_j \right], \tag{5.8}$$

and

$$\mathrm{Cov}\,(F_i, F_j) = -\frac{n\pi_i\pi_j(1 + n^2\beta)}{1 + n\beta} + \frac{n(n-1)\pi_i\pi_j}{1 + n\beta} \sum_{i=1}^{n-2} \frac{(n-2)!}{(n-2-i)!} \left(\frac{\beta}{1 + n\beta} \right)^{i}. \tag{5.9}$$

We rewrite (5.9) as

$$\mathrm{Cov}\,(F_i, F_j) = -n\pi_i\pi_j \left\{ n - \frac{(n-1)!}{1 + n\beta} \sum_{i=0}^{n-2} \frac{1}{(n-2-i)!} \left(\frac{\beta}{1 + n\beta} \right)^{i} \right\} \tag{5.10}$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

so that cell probabilities separate from the dispersion parameter β. Also, Theorem 3.4 shows that the π_j are separable from β in (5.8), which leads to Proposition 5.2 below.

Proposition 5.2. *For $\beta > 0$, (5.8) can be rewritten as*

$$\mathrm{Var}\,(F_j) = n\pi_j(1-\pi_j)\phi(n, 1/\beta, (1^0, 2^1, 3^2, \dots)), \tag{5.11}$$

where

$$\begin{aligned}\phi(n, 1/\beta, (1^0, 2^1, 3^2, \dots)) &= 1 + \frac{(n-1)!}{(1+n\beta)^{n-1}} \sum_{i=0}^{n-2} \frac{(1+i\beta)^{i-1}((n-i)\beta)^{n-i-1}}{i!\,(n-i-2)!} \\ &= n - \frac{(n-1)!}{1+n\beta} \sum_{i=0}^{n-2} \frac{1}{(n-2-i)!} \left(\frac{\beta}{1+n\beta}\right)^i. \end{aligned} \tag{5.12}$$

Equation (5.12) follows from (5.10) due to (3.7). The equivalence between (5.8) and (5.11) does not seem straightforward.

6. Infinite-dimensional Bell polynomial distributions

Let us consider distributing n balls over urns sequentially so that the frequency vector of the first t urns, $\boldsymbol{F}_t$, has the BPD. An important question is the existence of a stochastic process $\boldsymbol{F} := (F_j)_{j=1}^{\infty}$ defined by the finite-dimensional distribution of such $\boldsymbol{F}_t$. An affirmative answer is obtained by Kolmogorov's extension theorem, as below.

Theorem 6.1. *Let $\lambda_j \in \mathbb{R}_+$, $j \in \mathbb{N}$, so that $\lambda_\bullet = \sum_{j=1}^{\infty} \lambda_j < \infty$. Then there exists a process $\boldsymbol{F}$ such that $\boldsymbol{F}_t \sim \mathtt{BP}_{t+1}(n, (\boldsymbol{\lambda}_t, \lambda_\bullet - \sum_{j=1}^{t} \lambda_j); \boldsymbol{w})$, $t \in \mathbb{N}$.*

In practice, the sample path of $\boldsymbol{F}$ defined in Theorem 6.1 can be generated by the conditional distribution method discussed in Section 3.1.1. As in (3.1),

$$(F_t \mid \boldsymbol{F}_{t-1} = \boldsymbol{f}_{t-1}) = \left(F_t, n - \sum_{j=1}^{t-1} f_j - F_t\right) \sim \mathtt{BP}_2\left(n - \sum_{j=1}^{t-1} f_j, \left(\lambda_t, \lambda_\bullet - \sum_{j=1}^{t} \lambda_j\right); \boldsymbol{w}\right) \tag{6.1}$$

by Proposition 3.1. If $N_t := \sum_{j=1}^{t} F_j$ equals n then no ball remains to be put into remaining urns. Hence the sampling procedure (6.1) stops at

$$T := \min\{t \colon N_t = n\}. \tag{6.2}$$

We can derive this stopping time by noting that $T = t$ when $N_t = n$ and $N_{t-1} < n$. That is, $\mathbb{P}(T = t) = \mathbb{P}(N_t = n) - \mathbb{P}(N_{t-1} = n)$. Because

$$N_t \sim \mathtt{BP}_2\left(n, \left(\sum_{j=1}^{t} \lambda_t, \lambda_\bullet - \sum_{j=1}^{t} \lambda_t\right); \boldsymbol{w}\right)$$

due to the closure property under the merger of urns, Theorem 6.2 below holds.

Theorem 6.2. *Suppose that T of (6.2) is defined on $\boldsymbol{F}$ of Theorem 6.1. Then*

$$\mathbb{P}(T = t) = \left\{B_n\left(\sum_{j=1}^{t} \lambda_j, \boldsymbol{w}\right) - B_n\left(\sum_{j=1}^{t-1} \lambda_j, \boldsymbol{w}\right)\right\} / B_n(\lambda_\bullet, \boldsymbol{w})$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

or, equivalently,

$$\mathbb{P}(T \le t) = \frac{B_n(\sum_{j=1}^{t} \lambda_j, \boldsymbol{w})}{B_n(\lambda_\bullet, \boldsymbol{w})}.$$

Next, for a random vector $(F_{1,J}, F_{2,J}, \ldots, F_{J,J}) \in \mathcal{F}_{n,J}$, define

$$S_{i,J} := \sum_{j=1}^{J} I(F_{j,J} = i), \quad i \in [n].$$

This statistic is called the frequency of frequencies [25] or a size index [50]. The random vector of $(S_{1,J}, S_{2,J}, \ldots, S_{n,J}) =: \boldsymbol{S}_{n,J}$ is equivalent to the order statistics of positive frequencies among $(F_{1,J}, F_{2,J}, \ldots, F_{J,J})$, and thus it does not diverge when $J \to \infty$. The sample space of $\boldsymbol{S}_{n,J}$ is $\bigcup_{k=1}^{J} \mathcal{P}_{n,k}$, which is a subset of $\mathcal{P}_n$ when $J < n$. However, as $J \to \infty$, $\boldsymbol{S}_{n,J}$ is distributed over the whole of $\mathcal{P}_n$.

A distribution over $\mathcal{P}_n$ randomly partitions a positive integer n. Theorem 6.3 presents a limiting argument to obtain a random partitioning distribution from the BPD. The same limiting argument has been applied to the CCP distribution; see [29, Construction 1].

Theorem 6.3. *Let*

$$(F_{1,J}, F_{2,J}, \ldots, F_{J,J}) \sim \mathtt{BP}_J(n, (\lambda_{1,J}, \lambda_{2,J}, \ldots, \lambda_{J,J}); \boldsymbol{w}).$$

Suppose that $\lambda_{j,J} = \mu/J + \mathrm{o}(J^{-1})$, $j \in [J]$, *where* $\mu \in \mathbb{R}_+$. *Then the limiting distribution of* $\boldsymbol{S}_{n,J}$ *as* $J \to \infty$ *has the following PMF:*

$$p(\boldsymbol{s}_n) = \frac{n!}{B_n(\mu, \boldsymbol{w})} \prod_{i=1}^{n} \left(\frac{\mu w_i}{i!} \right)^{s_i} \frac{1}{s_i!}, \quad \boldsymbol{s}_n \in \mathcal{P}_n. \tag{6.3}$$

The resulting family (6.3) is called assemblies; see [4, eq. 2.2]. Also (6.3) is a special case of Gibbs partitions in the sense of [44, Sec. 1.5].

For example, (6.3) is called the Ewens distribution when $w_i = (i-1)!$; see the review in [16]. Equation (4.12) can explain this case by regarding $\mathtt{BP}_J(n, (\mu/J, \mu/J, \ldots, \mu/J); ((i-1)!)_{i=1}^{\infty})$ as i.i.d. sampling given $\boldsymbol{\pi}_J \sim \mathtt{Dir}_J(\mu/J, \mu/J, \ldots, \mu/J)$.

The limiting distribution of the order statistics of such $\boldsymbol{\pi}_J$ as $J \to \infty$ is known as the Poisson–Dirichlet distribution. The size indices of i.i.d. samples from the realization of the Poisson–Dirichlet distribution are known to have the Ewens distribution; see e.g. [23]. Because the size indices of i.i.d. samples are the same as those of the order statistics of the samples, the limiting distribution of the Dirichlet multinomial mixture is the Ewens distribution.

In Theorem 6.3, the convergence of $\lambda_{j,J}$ can be replaced by the convergence in probability, '$\xrightarrow{p}$', of a cell probability as in the following corollary.

Corollary 6.1. *Let* $\boldsymbol{P}_J = (P_{1,J}, P_{2,J}, \ldots, P_{J,J})$ *be a random vector over* $\triangle^{J-1}$ *such that* $P_{j,J} \xrightarrow{p} 1/J$, $j \in [J]$, *for all* $J \in \mathbb{N}$. *Suppose that* $(\boldsymbol{F}_J \mid \boldsymbol{P}_J = \boldsymbol{\pi}_J) \sim \mathtt{BP}_J(n, (\boldsymbol{\pi}_J; \mu); \boldsymbol{w})$, *where* $\mu \in \mathbb{R}_+$. *Then* $\boldsymbol{S}_{n,J}$ *converges in distribution to* (6.3) *as* $J \to \infty$.

For example, let ν_s be the Lévy measure of the stable distribution that $\psi(u) = u^{\gamma}$. If all μ_j are equal then

$$\boldsymbol{P}_J \sim \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu_s) \xrightarrow{p} (1/J, 1/J, \ldots, 1/J) \quad \text{as } \gamma \to 1$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

because $\mathrm{Var}\,(P_{j,J}) \to 0$ by (4.7) and (4.14); using Chebyshev's inequality, $P_{j,J} \xrightarrow{p} \mathbb{E}(P_{j,J}) = 1/J$ by (4.6). Corollary 6.1 covers this case.

If a random vector $(S_1, S_2, \ldots, S_n)$ is distributed as (6.3), then

$$\mathbb{P}\left(\sum_{i=1}^{n} S_i = k\right) = \mu^k B_{n,k}(\boldsymbol{w})/B_n(\mu, \boldsymbol{w}), \quad k \in [n].$$

Hence $\lim_{\mu\to\infty} \mathbb{P}(\sum_{i=1}^{n} S_i = n) = 1$. Equivalently, $\mathbb{P}(S_1 = n) = 1$, which implies $\mathbb{P}(F_{j,J} \geq 2) = 0$. Observing (4.12), the underlying fact is that

$$\boldsymbol{P}_J \sim \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu) \xrightarrow{p} (1/J, 1/J, \ldots, 1/J) \quad \text{as } \mu \to \infty$$

when $\varphi(\mu, \psi)$ depends on μ; from Theorem 4.9, $\varphi(\mu, \psi) = \mathrm{O}(\mu^{-1})$, by which

$$\lim_{\mu\to\infty} \mathrm{Var}\,(P_{j,J}) = 0.$$

Under i.i.d. sampling from uniform J cells, it is obvious that $\mathbb{P}(F_{j,J} \geq 2) \to 0$ as $J \to \infty$.

7. Concluding remarks

The family $\mathcal{M}$ of Definition 2.1 is constructed to be closed under resizing, but further analysis of its properties should be difficult under this parameterization. The equivalence to the BPD given by Theorem 2.4 has initiated an unexpectedly long investigation of the family.

The moment expressions of the BPD naturalize cell probabilities. This finding has led us to consider random cell probabilities, which greatly advance our understanding. In particular, the mixed multinomial expression of the BPD with the NID distribution presented by Theorem 4.5 seems insightful.

Section 4 on random cell probabilities, or equivalently a random probability measure, is motivated by Bayesian non-parametrics; see e.g. [24]. Together with the infinite-dimensional extension of the BPD shown by Theorem 6.1, the BPD should open up myriads of research opportunities.

BPD sampling can control the Fisher information by changing $\boldsymbol{w}$, as revealed by Theorem 3.3. We can apply this property to information protection; see [31] for this context of data privacy.

The present article has been too long to additionally demonstrate the applicability of the BPD. Hence the author will show in a separate paper that the BPD is useful.

Appendix A. Proofs

Proof of Theorem 2.1. If (2.5) holds for $J = 2$, then (2.6) is rewritten as

$$\sum_{f=0}^{n} \frac{h(f; \lambda_1)}{f!} \frac{h(n-f; \lambda_2)}{(n-f)!} = \frac{h(n; \lambda_1 + \lambda_2)}{n!}. \tag{A.1}$$

Conversely, if (A.1) holds, then (2.5) holds for $J \geq 3$ on (2.6) by recursion. Hence (A.1) is necessary and sufficient for (2.6) to satisfy (2.5) for all J.

Let us define a formal generating function by

$$G(s; \lambda) := \sum_{f=0}^{\infty} \frac{s^f}{f!} h(f; \lambda).$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Then (A.1) holds for all $n \in \mathbb{N}_0$ if and only if

$$\log G(s;\lambda_1) + \log G(s;\lambda_2) = \log G(s;\lambda_1+\lambda_2). \tag{A.2}$$

As noted by a referee, (A.2) is an instance of Cauchy's functional equation, whose solution is $G(s;\lambda) = G(s;1)^{\lambda}$ under the continuity of $h(\,\cdot\,;\lambda)$; see e.g. [1, Theorem 1, p. 34]. Noting that $h(f;1) = h_f$, this solution is equivalent to (2.7). Hence the unity and continuity conditions of $\mathcal{M}$ are equivalent to (2.7).

On the positivity condition, the spaces of $(h_f)_{f\in\mathbb{N}}$ are exhausted by Theorem 2.3 because of (A.5). The representative space of h_f is $\mathbb{R}_+$ when $\boldsymbol{w} \in \mathbb{R}_+ \times \mathbb{R}_{0+}^{\infty}$. □

Proof of Theorem 2.2. Define

$$G(s) = 1 + \sum_{f=1}^{\infty} \frac{s^f}{f!} h_f,$$

and expand it as

$$\log G(s) = \sum_{i=0}^{\infty} \frac{s^i}{i!} w_i.$$

Then we express h_f as the function of the w_i, comparing the coefficients of s between the two sides of

$$\exp(w_0) \exp\left(\sum_{i=1}^{\infty} \frac{s^i}{i!} w_i \right) = 1 + \sum_{f=1}^{\infty} \frac{s^f}{f!} h_f. \tag{A.3}$$

Equating the coefficient of s^0 in (A.3), we find that $1 = \exp(w_0)$ or $w_0 = 0$. Thus $G(s) = \exp(\gamma(s))$, where $\gamma(s) := \sum_{i=1}^{\infty} \frac{s^i}{i!} w_i$. For $f \in \mathbb{N}$, the coefficient of s^f of $\gamma^k(s)/k!$ is

$$\sum_{\sum i k_i = f, \sum k_i = k} \frac{(w_i/i!)^{k_i}}{k!} \cdot \binom{k}{k_1 k_2 k_3 \cdots k_f} = \frac{B_{f,k}(\boldsymbol{w})}{f!}.$$

Hence the coefficient of s^f of

$$\exp(\gamma(s)) = \sum_{k=0}^{\infty} \gamma(s)^k / k!$$

is

$$\sum_{k=1}^{f} B_{f,k}(\boldsymbol{w})/f! = B_f(1, \boldsymbol{w})/f!.$$

Namely,

$$\exp\left(\sum_{i=1}^{\infty} \frac{s^i}{i!} w_i \right) = 1 + \sum_{f=1}^{\infty} \frac{s^f}{f!} B_f(1, \boldsymbol{w}). \tag{A.4}$$

Comparing it with (A.3), we have $h_f = B_f(1, \boldsymbol{w})$, $f \in \mathbb{N}$. This argument completes bijective reparameterization between $(h_f)_{f\in\mathbb{N}}$ and $\boldsymbol{w}$.

Let us note that

$$G(s)^{\lambda} = \exp\left(\sum_{i=1}^{\infty} \frac{s^i}{i!} \lambda w_i \right)$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

determines $h(f;\lambda)$ in the proof of Theorem 2.1. Therefore, using (1.3), we have

$$h(f;\lambda) = B_f(\lambda, \boldsymbol{w}). \tag{A.5}$$

Putting this expression into (2.6), we obtain (2.8). □

Proof of Theorem 2.3. We begin by proving the 'if' part. Since $B_0(\boldsymbol{w}) \equiv 1$ irrespective of $\boldsymbol{w}$, we consider the case of $n \in \mathbb{N}$. By definition (1.3),

$$B_n(\lambda, \boldsymbol{w}) = \sum_{k=1}^{n-1} \lambda^k B_{n,k}(\boldsymbol{w}) + \lambda^n B_{n,n}(\boldsymbol{w}) = \sum_{k=1}^{n-1} \lambda^k B_{n,k}(\boldsymbol{w}) + \lambda^n w_1^n.$$

Because $B_{n,k}(\boldsymbol{w})$ is a weighted sum of the w_i with non-negative weights, $\sum_{k=1}^{n-1} \lambda^k B_{n,k}(\boldsymbol{w}) \geq 0$ when $\boldsymbol{w} \in \Omega_1$ and $\lambda \in \mathbb{R}_{0+}$. We observe that $\lambda^n w_1^n > 0$ when $\lambda \in \mathbb{R}_+$ and $w_1 \in \mathbb{R}_+$. Therefore $B_n(\lambda, \boldsymbol{w})$ is strictly positive for $n \in \mathbb{N}_0$ when $\lambda \in \mathbb{R}_+$ and $\boldsymbol{w} \in \Omega_1$. Consequently, (2.8) satisfies (2.4) for $n \in \mathbb{N}_0$, $J \in \mathbb{N}$, and $\boldsymbol{\lambda}_J \in \mathbb{R}_+^J$ if $\boldsymbol{w} \in \Omega_1$. When $\boldsymbol{w} \in \Omega_2$, using (1.2) results in the same argument because $(-1)^{\sum_{j=1}^J f_j} = (-1)^n$.

Next, the 'only if' part follows. If the sign of $\lim_{l \to +0} B_f(l, \boldsymbol{w})$ is different from that of $\lim_{l \to +\infty} B_f(l, \boldsymbol{w})$, then there exists $\lambda \in \mathbb{R}_+$ such that $B_f(\lambda, \boldsymbol{w}) = 0$ since $B_f(l, \boldsymbol{w})$ is a continuous function of l. Therefore the condition of $p(\boldsymbol{f}_J) > 0$ excludes $\boldsymbol{w}$ that satisfies

$$\operatorname{sgn}\Big(\lim_{l \to +0} B_f(l, \boldsymbol{w})\Big) \neq \operatorname{sgn}\Big(\lim_{l \to +\infty} B_f(l, \boldsymbol{w})\Big).$$

By (1.4) and (1.5), we thus discard $\boldsymbol{w}$ such that $\operatorname{sgn}(l w_f) \neq \operatorname{sgn}(l^f w_1^f)$ for some $f \in \mathbb{N}_0$. Accordingly, $\operatorname{sgn}(w_1)$ determines $\operatorname{sgn}(w_i)$ for all $i \geq 2$. Hence either $\boldsymbol{w} \in \Omega_1$ or $\boldsymbol{w} \in \Omega_2$ is allowed when $\lambda_j \in \mathbb{R}_+$. The case of $w_1 = 0$ is excluded because then $B_{1,1}(\boldsymbol{w}) = 0$, which contradicts (2.4). □

Proof of Theorem 2.8. Rewrite (2.12) as

$$G(s) = \exp\left(\sum_{i=1}^{\infty} \frac{s^i}{i!}(\lambda q_i\, i!)\right) \exp(-\lambda).$$

Then, from (A.4),

$$\exp(\lambda) G(s) = 1 + \sum_{f=1}^{\infty} \frac{s^f}{f!} B_f(\lambda, (q_i\, i!)_{i \in \mathbb{N}}).$$

Therefore

$$\mathbb{P}(F_j = f) = \exp(-\lambda_j) B_f(\lambda_j, (q_i\, i!)_{i \in \mathbb{N}})/f!.$$

Also,

$$\mathbb{P}\left(\sum_{j=1}^{J} F_j = n\right) = \exp(-\lambda_\bullet) B_n(\lambda_\bullet, (q_i\, i!)_{i \in \mathbb{N}})/n!,$$

since the PGF of $\sum_{j=1}^J F_j$ is

$$\exp\left(\lambda_\bullet \left(\sum_{i=1}^{\infty} s^i q_i - 1\right)\right).$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Combining these results, the conditional distribution of $\boldsymbol{F}_J$ given $\sum_{j=1}^{J} F_j = n$ is

$$\mathtt{BP}_J(n, \boldsymbol{\lambda}_J; (q_i\, i!)_{i\in\mathbb{N}}).$$

Then $\mathbb{P}(\sum_{j=1}^{J} F_j = n) > 0$ for all n is equivalent to $B_n(\lambda_\bullet, (q_i\, i!)_{i\in\mathbb{N}}) > 0$ for all n. We have already proved in Theorem 2.3 that this positivity is equivalent to $(q_i\, i!)_{i\in\mathbb{N}} \in \Omega_1$ or $\boldsymbol{q} \in \Omega_1$. □

Proof of Theorem 3.3. First, write the score function as

$$\frac{\partial \log p(\boldsymbol{F}_J; \boldsymbol{\lambda}_J)}{\partial \lambda_j} = \frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})} - \frac{\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial \lambda_j}{B_n(\lambda_\bullet, \boldsymbol{w})}. \tag{A.6}$$

The expectation of the score function is zero because

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right] &= \sum_{f_J} \frac{n!}{B_n(\lambda_\bullet, \boldsymbol{w})} \prod_{i=1}^{J} \frac{B_{f_i}(\lambda_i, \boldsymbol{w})}{f_i!} \frac{\partial B_{f_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{B_{f_j}(\lambda_j, \boldsymbol{w})} \\
&= \frac{n!}{B_n(\lambda_\bullet, \boldsymbol{w})} \sum_{f_j=0}^{n} \frac{B_{n-f_j}(\lambda_\bullet - \lambda_j, \boldsymbol{w}) \partial B_{f_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{(n-f_j)!\, f_j!} \\
&= \frac{n!}{B_n(\lambda_\bullet, \boldsymbol{w})} \frac{\partial}{\partial \lambda_j} \sum_{f_j=0}^{n} \frac{B_{n-f_j}(\lambda_\bullet - \lambda_j, \boldsymbol{w}) B_{f_j}(\lambda_j, \boldsymbol{w})}{(n-f_j)!\, f_j!} \\
&= \frac{\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial \lambda_j}{B_n(\lambda_\bullet, \boldsymbol{w})}.
\end{aligned} \tag{A.7}$$

Replacing $(\partial/\partial\lambda_j)$ with $(\partial^2/(\partial\lambda_j)^2)$ in (A.7), we also obtain

$$\mathbb{E}\left[\frac{\partial^2 B_{F_j}(\lambda_j, \boldsymbol{w})/(\partial \lambda_j)^2}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right] = \frac{\partial^2 B_n(\lambda_\bullet, \boldsymbol{w})/(\partial \lambda_j)^2}{B_n(\lambda_\bullet, \boldsymbol{w})}. \tag{A.8}$$

Noting that

$$\frac{\partial}{\partial \lambda_j} \frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})} = \frac{\partial^2 B_{F_j}(\lambda_j, \boldsymbol{w})/(\partial \lambda_j)^2}{B_{F_j}(\lambda_j, \boldsymbol{w})} - \left(\frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right)^2, \tag{A.9}$$

the differentiation of (A.6) yields

$$\begin{aligned}
\frac{\partial^2 \log p(\boldsymbol{F}_J; \boldsymbol{\lambda}_J)}{(\partial \lambda_j)^2} = &\frac{\partial^2 B_{F_j}(\lambda_j, \boldsymbol{w})/(\partial \lambda_j)^2}{B_{F_j}(\lambda_j, \boldsymbol{w})} - \left(\frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right)^2 \\
&- \frac{\partial^2 B_n(\lambda_\bullet, \boldsymbol{w})/(\partial \lambda_j)^2}{B_n(\lambda_\bullet, \boldsymbol{w})} + \left(\frac{\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial \lambda_j}{B_n(\lambda_\bullet, \boldsymbol{w})}\right)^2.
\end{aligned}$$

Then taking expectation on both sides results in (3.2) because of (A.8).

Next, (3.2) equals

$$\mathbb{E}\left[\frac{\partial^2 B_{F_j}(\lambda_j, \boldsymbol{w})/(\partial \lambda_j)^2}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right] - \mathbb{E}\left[\frac{\partial}{\partial \lambda_j}\left(\frac{\partial B_{F_j}(\lambda_j, \boldsymbol{w})/\partial \lambda_j}{B_{F_j}(\lambda_j, \boldsymbol{w})}\right)\right] - \left(\frac{\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial \lambda_j}{B_n(\lambda_\bullet, \boldsymbol{w})}\right)^2$$

because of (A.9). This expression reduces to (3.3) by (A.8). Therefore the diagonal elements have been shown to be (3.2) or (3.3).

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Next we differentiate (A.6) as

$$\frac{\partial^2 \log p(\boldsymbol{F}_J;\boldsymbol{\lambda}_J)}{\partial\lambda_i\partial\lambda_j} = -\frac{\partial}{\partial\lambda_i}\frac{\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial\lambda_j}{B_n(\lambda_\bullet, \boldsymbol{w})},$$

which equals

$$-\frac{\partial^2 B_n(\lambda_\bullet, \boldsymbol{w})/(\partial\lambda_i\partial\lambda_j)}{B_n(\lambda_\bullet, \boldsymbol{w})} + \frac{(\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial\lambda_i)(\partial B_n(\lambda_\bullet, \boldsymbol{w})/\partial\lambda_j)}{B_n(\lambda_\bullet, \boldsymbol{w})^2}.$$

Then taking expectation on both sides produces (3.4). □

Proof of Theorem 3.4. Without loss of generality, we consider the two-dimensional case that $(F_1, F_2) \sim \mathtt{BP}_2(n, \lambda_1, \lambda_2; \boldsymbol{w})$, and the moments of F_1 are shown.

Define

$$\eta(s) := \sum_{i=0}^{\infty} \frac{s^i}{i!} w_i.$$

Then

$$G(s) = \sum_{i=0}^{\infty} \exp(w_0\lambda_\bullet)\frac{B_i(\lambda_\bullet, \boldsymbol{w})}{i!}s^i = \exp(\lambda_\bullet\eta(s)).$$

We denote $G_1(s) = \exp(\lambda_1\eta(s))$ and $G_2(s) = \exp(\lambda_2\eta(s))$. Write $\mathrm{d}G_1(s)/\mathrm{d}s$ as

$$G_1'(s) = \sum_{i=0}^{\infty} \exp(w_0\lambda_1)\frac{B_i(\lambda_1, \boldsymbol{w})}{(i-1)!}s^{i-1} = G_1(s)\lambda_1\eta'(s),$$

where $\eta'(s) = \mathrm{d}\eta(s)/\mathrm{d}s$. Then

$$G_1'(s)G_2(s) = \exp(w_0(\lambda_1+\lambda_2)) \sum_{n=0}^{\infty}\sum_{i=0}^{n} \frac{iB_i(\lambda_1, \boldsymbol{w})B_{n-i}(\lambda_2, \boldsymbol{w})}{i!\,(n-i)!}s^{n-1}. \tag{A.10}$$

Also, since

$$\mathrm{d}G(s)/\mathrm{d}s = G'(s) = G(s)(\lambda_1+\lambda_2)\mathrm{d}\eta(s)/\mathrm{d}s,$$

we obtain

$$\frac{\lambda_1}{\lambda_1+\lambda_2}G'(s) = \lambda_1 G(s)\mathrm{d}\eta(s)/\mathrm{d}s = G_1'(s)G_2(s),$$

using $G(s) = G_1(s)G_2(s)$. Then we equate the coefficient of s^{n-1} between (A.10) and

$$\frac{\lambda_1}{\lambda_1+\lambda_2}G'(s) = \frac{\lambda_1}{\lambda_1+\lambda_2}\sum_{n=0}^{\infty}\exp(w_0(\lambda_1+\lambda_2))\frac{B_n((\lambda_1+\lambda_2), \boldsymbol{w})}{(n-1)!}s^{n-1}.$$

It follows that

$$\sum_{i=0}^{n}\frac{iB_i(\lambda_1, \boldsymbol{w})B_{n-i}(\lambda_2, \boldsymbol{w})}{i!\,(n-i)!} = \frac{B_n((\lambda_1+\lambda_2), \boldsymbol{w})}{n!}\frac{n\lambda_1}{\lambda_1+\lambda_2}.$$

Dividing both sides by $B_n((\lambda_1+\lambda_2), \boldsymbol{w})/n!$, we obtain $\mathbb{E}(F_1) = n\lambda_1/(\lambda_1+\lambda_2)$, which shows (3.5).

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

On the variance, we use the second derivative:

$$G_1''(s) = \lambda_1\{G_1'(s)\eta'(s) + G_1(s)\eta''(s)\} = \lambda_1^2 G_1(s)\eta'(s)\eta'(s) + \lambda_1 G_1(s)\eta''(s).$$

Hence

$$G_1''(s)G_2(s) = \lambda_1^2 G(s)\eta'(s)\eta'(s) + \lambda_1 G(s)\eta''(s).$$

By $G(s)\eta'(s) = G'(s)/(\lambda_1+\lambda_2)$,

$$\begin{aligned} G_1''(s)G_2(s) &= \frac{\lambda_1^2}{\lambda_1+\lambda_2}\left\{G'(s)\eta'(s) + \frac{\lambda_1+\lambda_2}{\lambda_1}G(s)\eta''(s)\right\} \\ &= \frac{\lambda_1^2}{\lambda_1+\lambda_2}\left\{\frac{G''(s)}{\lambda_1+\lambda_2} + \frac{\lambda_2}{\lambda_1}G(s)\eta''(s)\right\} \\ &= \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^2 G''(s) + \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}G(s)\eta''(s). \end{aligned}$$

Then, noting that

$$G(s)\eta''(s) = \sum_{i=0}^{\infty} \mathrm{e}^{w_0(\lambda_1+\lambda_2)}\frac{B_i((\lambda_1+\lambda_2),\boldsymbol{w})}{i!}s^i \times \sum_{i=0}^{\infty}\frac{w_i}{i!}i(i-1)s^{i-2},$$

which equals

$$\sum_{n=0}^{\infty}\sum_{i=0}^{n} \mathrm{e}^{w_0(\lambda_1+\lambda_2)}\frac{B_i((\lambda_1+\lambda_2),\boldsymbol{w})}{i!}\frac{w_{n-i}}{(n-i)!}(n-i)(n-i-1)s^{n-2},$$

we obtain

$$\begin{aligned} &G_1''(s)G_2(s) \\ &\quad= \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^2 \sum_{i=0}^{\infty} \mathrm{e}^{w_0(\lambda_1+\lambda_2)}\frac{i(i-1)B_i((\lambda_1+\lambda_2),\boldsymbol{w})}{i!}s^{i-2} \\ &\qquad+ \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\sum_{n=0}^{\infty}\sum_{i=0}^{n} \mathrm{e}^{w_0(\lambda_1+\lambda_2)}\frac{B_i((\lambda_1+\lambda_2),\boldsymbol{w})}{i!}w_{n-i}(n-i)(n-i-1)s^{n-2}. \end{aligned} \tag{A.11}$$

Alternatively,

$$\begin{aligned} G_1''(s)G_2(s) &= \mathrm{e}^{w_0(\lambda_1+\lambda_2)}\sum_{n=0}^{\infty}\sum_{i=0}^{n}\frac{i(i-1)B_i(\lambda_1,\boldsymbol{w})B_{n-i}(\lambda_2,\boldsymbol{w})}{i!\,(n-i)!}s^{n-2} \\ &= \mathrm{e}^{w_0(\lambda_1+\lambda_2)}\sum_{n=0}^{\infty}\mathbb{E}(F_1(F_1-1))\frac{B_n((\lambda_1+\lambda_2),\boldsymbol{w})}{n!}s^{n-2}. \end{aligned} \tag{A.12}$$

Therefore, equating the coefficient of s^{n-2} between (A.11) and (A.12), we have

$$\mathbb{E}(F_1(F_1-1)) = \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^2 n(n-1) + \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\frac{n!}{B_n((\lambda_1+\lambda_2),\boldsymbol{w})}\sum_{i=0}^{n}\frac{B_i((\lambda_1+\lambda_2),\boldsymbol{w})w_{n-i}}{i!\,(n-i-2)!}.$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Since $\mathrm{Var}\,(F_1) = \mathbb{E}(F_1(F_1-1)) + \mathbb{E}(F_1) - \mathbb{E}(F_1)^2$, the variance (3.6) follows immediately.

On the covariance, we use $\mathrm{Var}\,(F_i + F_j) = \mathrm{Var}\,(F_i) + \mathrm{Var}\,(F_j) + 2\,\mathrm{Cov}\,(F_i, F_j)$. Because of the closure property of marginal variables,

$$\mathrm{Var}\,(F_i + F_j) = n(\pi_i + \pi_j)(1 - (\pi_i + \pi_j))\phi(n, \lambda_\bullet, \boldsymbol{w}).$$

Then we have (3.7). □

Proof of Theorem 3.5. For $n \in \mathbb{N}$,

$$B_n(\lambda, \boldsymbol{w}) = \sum_{i=0}^{n-1} \binom{n-1}{i} B_i(\lambda, \boldsymbol{w}) \lambda w_{n-i};$$

see e.g. [8, eq. 11.10]. It implies for $n \geq 2$ that

$$B_n(\lambda_\bullet, \boldsymbol{w}) - B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1 = \lambda_\bullet (n-1)! \sum_{i=0}^{n-2} \frac{B_i(\lambda_\bullet, \boldsymbol{w}) w_{n-i}}{i!\,(n-i-1)!}. \tag{A.13}$$

Because $B_n(\lambda_\bullet, \boldsymbol{w}) > 0$ in the parameter space, we divide the both sides of (A.13) by $B_n(\lambda_\bullet, \boldsymbol{w})$. Consequently,

$$1 - \frac{B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1}{B_n(\lambda_\bullet, \boldsymbol{w})} = \frac{\lambda_\bullet (n-1)!}{B_n(\lambda_\bullet, \boldsymbol{w})} \sum_{i=0}^{n-2} \frac{B_i(\lambda_\bullet, \boldsymbol{w}) w_{n-i}}{i!\,(n-i-1)!}. \tag{A.14}$$

Therefore, for $n \geq 2$,

$$1 - \frac{B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1}{B_n(\lambda_\bullet, \boldsymbol{w})} \leq \frac{\lambda_\bullet (n-1)!}{B_n(\lambda_\bullet, \boldsymbol{w})} \sum_{i=0}^{n-2} \frac{B_i(\lambda_\bullet, \boldsymbol{w}) w_{n-i}}{i!\,(n-i-2)!} = \phi(n, \lambda_\bullet, \boldsymbol{w}) - 1.$$

Hence the first inequality of (3.8) holds.

The middle term of (3.8) takes its minimum when (A.14) is minimized, where $w_i = 0$, $i \geq 2$, since all other elements of (A.14) are positive. Then the middle term of (3.8) equals one.

The following proof of (3.9) parallels that of (3.8).

Because of (A.13), it follows for $n \geq 2$ that

$$\begin{aligned}(n-1)\{B_n(\lambda_\bullet, \boldsymbol{w}) - B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1\} &\geq \lambda_\bullet (n-1)! \sum_{i=0}^{n-2} \frac{B_i(\lambda_\bullet, \boldsymbol{w}) w_{n-i}}{i!\,(n-i-2)!} \\ &= B_n(\lambda_\bullet, \boldsymbol{w})(\phi(n, \lambda_\bullet, \boldsymbol{w}) - 1).\end{aligned}$$

Because $B_n(\lambda_\bullet, \boldsymbol{w}) > 0$ in the parameter space, we divide the both sides of the above inequality by $B_n(\lambda_\bullet, \boldsymbol{w})$. Consequently,

$$(n-1)(1 - B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1 / B_n(\lambda_\bullet, \boldsymbol{w})) \geq \phi(n, \lambda_\bullet, \boldsymbol{w}) - 1 \quad \text{for } n \geq 2.$$

Hence the first inequality of (3.9) holds.

We observe in the parameter space that $B_{n-1}(\lambda_\bullet, \boldsymbol{w})\lambda_\bullet w_1 / B_n(\lambda_\bullet, \boldsymbol{w}) > 0$. Therefore the middle term of (3.9) is strictly smaller than n.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

As $\lambda_\bullet \to \infty$,

$$\phi(n, \lambda_\bullet, \boldsymbol{w}) - 1 = \frac{\lambda_\bullet (n-1)!}{\lambda_\bullet^n w_1^n + \mathrm{O}(\lambda_\bullet^{n-1})} \left\{ \frac{B_{n-2}(\lambda_\bullet, \boldsymbol{w}) w_2}{(n-2)!} + \mathrm{O}(\lambda_\bullet^{n-3}) \right\} = \mathrm{O}(\lambda_\bullet^{-1})$$

due to (1.5). Therefore (3.10) holds.

As $\lambda_\bullet \to 0$,

$$\phi(n, \lambda_\bullet, \boldsymbol{w}) - 1 = \frac{\lambda_\bullet (n-1)!}{\lambda_\bullet w_n + \mathrm{O}(\lambda_\bullet^2)} \left\{ \frac{B_0(\lambda_\bullet, \boldsymbol{w}) w_n}{(n-2)!} + \mathrm{O}(\lambda_\bullet) \right\} = (n-1) + \mathrm{O}(\lambda_\bullet)$$

due to (1.4). Therefore (3.11) holds. □

Proof of Lemma 4.1. Generally, $\mathbb{E}(F_j) = \mathbb{E}_\pi(\mathbb{E}(F_j | \boldsymbol{\pi}_J))$, where $\mathbb{E}_\pi(\cdot)$ implies that the expectation is taken under random $\boldsymbol{\pi}_J$. Similarly,

$$\mathrm{Var}\,(F_j) = \mathbb{E}_\pi(\mathrm{Var}\,(F_j \mid \boldsymbol{\pi}_J)) + \mathrm{Var}_\pi\,(\mathbb{E}(F_j \mid \boldsymbol{\pi}_J))$$

and

$$\mathrm{Cov}\,(F_i, F_j) = \mathbb{E}_\pi(\mathrm{Cov}\,(F_i, F_j \mid \boldsymbol{\pi}_J)) + \mathrm{Cov}_\pi\,(\mathbb{E}(F_i \mid \boldsymbol{\pi}_J), \mathbb{E}(F_j \mid \boldsymbol{\pi}_J)).$$

The conditional moments of F_j given $\boldsymbol{\pi}_J$ are provided in Theorem 3.4. Using these and taking the expectation under random $\boldsymbol{\pi}_J$, the formulae are obtained. □

Proof of Theorem 4.3. Write

$$G(u) = \exp(-\mu_\bullet \psi(u)) = \mathbb{E}(\mathrm{e}^{-uX_\bullet}).$$

Denote $\mathrm{d}G(u)/\mathrm{d}u$ by $G'(u)$, and other derivatives are similarly denoted.

Because $\mathbb{P}(X_\bullet = 0) = 0$ is assumed, $G(\infty) = 0$ and $G(0) = 1$. Also

$$uG'(u) = u \frac{\mathrm{d}}{\mathrm{d}u} \int_0^\infty \mathrm{e}^{-ut} F_{X_\bullet}(\mathrm{d}t) = u \int_0^\infty \frac{\mathrm{d}\mathrm{e}^{-ut}}{\mathrm{d}u} F_{X_\bullet}(\mathrm{d}t) = \mathbb{E}\left(-\frac{uX_\bullet}{\mathrm{e}^{uX_\bullet}} \right),$$

which implies that

$$[uG'(u)]_0^\infty = 0. \tag{A.15}$$

Proposition 2 of [21] implies that

$$\mathbb{E}(P_j) = \int_0^\infty \mu_j \psi'(u) \exp(-\mu_\bullet \psi(u))\, \mathrm{d}u.$$

Noting that $\psi'(u) \exp(-\mu_\bullet \psi(u)) = G'(u)/(-\mu_\bullet)$, we have

$$\mathbb{E}(P_j) = -\rho_j \int_0^\infty G'(u)\, \mathrm{d}u = -\rho_j [G(u)]_0^\infty = \rho_j.$$

Hence (4.6) is proved.

Equation (7) of [21] implies that

$$\mathbb{E}(P_j^2) = \int_0^\infty u \left(\frac{\mathrm{d}^2 \exp(-\mu_j \psi(u))}{\mathrm{d}u^2} \right) \exp(-(\mu_\bullet - \mu_j)\psi(u))\, \mathrm{d}u. \tag{A.16}$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Then, by

$$(\mathrm{d}^2/\mathrm{d}u^2)\exp(-\mu\psi(u)) = (-\mu)\psi''(u)\exp(-\mu\psi(u)) + \mu^2(\psi'(u))^2\exp(-\mu\psi(u)),$$

(A.16) can be rewritten as

$$\begin{aligned}\mathbb{E}(P_j^2) &= \int_0^\infty u\{(-\mu_j)\psi''(u)G(u) + {\mu_j}^2(\psi'(u))^2G(u)\}\,\mathrm{d}u \\ &= \int_0^\infty u\{(-\mu_j)\psi''(u)G(u) + {\mu_j}^2\psi'(u)G'(u)/(-\mu_\bullet)\}\,\mathrm{d}u\end{aligned}$$

since $\psi'(u)G(u) = G'(u)/(-\mu_\bullet)$. Further, $\psi'(u)G'(u) = G''(u)/(-\mu_\bullet) - \psi''(u)G(u)$, and thus

$$\begin{aligned}\mathbb{E}(P_j^2) &= \int_0^\infty u\{({\mu_j}^2/\mu_\bullet - \mu_j)\psi''(u)G(u) + (\mu_j/\mu_\bullet)^2G''(u)\}\,\mathrm{d}u \\ &= \int_0^\infty \{\lambda_\bullet\rho_j(\rho_j - 1)u\psi''(u)G(u) + \pi_j^2uG''(u)\}\,\mathrm{d}u.\end{aligned}$$

Employing integration by parts,

$$\int_0^\infty uG''(u)\,\mathrm{d}u = [uG'(u)]_0^\infty - \int_0^\infty G'(u)\,\mathrm{d}u = 1 \tag{A.17}$$

because of (A.15). Hence

$$\mathbb{E}(P_j^2) = \rho_j(1-\rho_j)(-\mu_\bullet)\int_0^\infty u\psi''(u)G(u)\,\mathrm{d}u + \rho_j^2.$$

Because $\mathrm{Var}(P_j) = \mathbb{E}(P_j^2) - \mathbb{E}(P_j)^2$,

$$\varphi(\mu_\bullet) = -\mu_\bullet\int_0^\infty u\psi''(u)G(u)\,\mathrm{d}u. \tag{A.18}$$

Then, by

$$\psi''(u)G(u) = G''(u)/(-\mu_\bullet) - \psi'(u)G'(u) = G''(u)/(-\mu_\bullet) + \mu_\bullet(\psi'(u))^2G(u),$$

we have

$$\varphi(\mu_\bullet) = \int_0^\infty uG''(u)\,\mathrm{d}u - \mu_\bullet^2\int_0^\infty u(\psi'(u))^2G(u)\,\mathrm{d}u. \tag{A.19}$$

The first term of the right-hand side of (A.19) is unity because of (A.17), which shows (4.7).

The covariance (4.8) can be shown by noting that

$$\mathrm{Var}(P_i + P_j) = (\pi_i + \pi_j)(1 - (\pi_i + \pi_j))\eta(\lambda_\bullet) = \mathrm{Var}(P_i) + \mathrm{Var}(P_j) + 2\,\mathrm{Cov}(P_i, P_j);$$

the first equation holds because $\mathbb{E}(\mathrm{e}^{-u(X_i+X_j)}) = \exp(-(\mu_i+\mu_j)\psi(u))$.

The correlation is derived using (4.7) and (4.8). □

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Two lemmas for Theorem 4.5

Lemma A.1. *Let ν be a Lévy measure, and $\mu \in \mathbb{R}_+$. Write the Laplace transform of an $\mathbb{R}_+$-valued random variable M as $\exp(-\mu\psi(u))$, where $\psi(u) = \int (1 - \mathrm{e}^{-ux})\nu(\mathrm{d}x)$. Then a random variable satisfies $F \sim \mathtt{Po}(M)$ if and only if $\mathbb{E}(s^F) = \exp(-\mu\psi(1-s))$.* □

Proof of Lemma A.1. This link between the PGF of a mixed Poisson distribution and the Laplace transform of its mixing distribution is widely known; see [26, Proposition 2.1(vi)]. □

Lemma A.2. *In Lemma A.1,*

$$\mathbb{P}(F = f) = \exp(-\mu\psi(1))B_f(\mu, \boldsymbol{w})/f!,$$

where $\boldsymbol{w}$ satisfies (4.11).

Proof of Lemma A.2. Because $\mathbb{E}(s^F) = \sum_{f=0}^{\infty} \mathbb{P}(F = f)s^f$, we obtain $\mathbb{P}(F = f)$ by differentiating $\exp(-\mu\psi(1-s))$. First, expand

$$-\psi(1-s) = \sum_{i=0}^{\infty} \frac{s^i}{i!} w_i.$$

Then $w_0 = -\psi(1)$, and $\boldsymbol{w}$ satisfies (4.11). Let

$$G(s) = \exp\left(\sum_{i=0}^{\infty} \frac{s^i}{i!} w_i\right) = \sum_{f=0}^{\infty} \frac{s^f}{f!} h_f.$$

By (A.4), $h_f = \exp(w_0)B_f(\boldsymbol{w})$. Noting that $\mathbb{E}(s^F) = G(s)^\mu$, we obtain $\mathbb{P}(F = f)$. □

Proof of Theorem 4.5. Suppose that $F_j \sim \mathtt{Po}(M_j)$, $j \in [J]$, independently, where the Laplace transform of M_j is $\exp(-\mu_j\psi(u))$. Write

$$\boldsymbol{M}_J := (M_1, M_2, \ldots, M_J) \quad \text{and} \quad \boldsymbol{m}_J := (m_1, m_2, \ldots, m_J).$$

Then

$$\mathbb{P}\left(\boldsymbol{F}_J = \boldsymbol{f}_J \,\middle|\, \sum_{j=1}^{J} F_j = n, \boldsymbol{M}_J = \boldsymbol{m}_J\right) = \binom{n}{\boldsymbol{f}_J} \prod_{j=1}^{J} \left(\frac{m_j}{m_\bullet}\right)^{f_j}.$$

Taking the expectation over $\boldsymbol{M}_J$,

$$\mathbb{P}\left(\boldsymbol{F}_J = \boldsymbol{f}_J \,\middle|\, \sum_{j=1}^{J} f_j = n\right) = \int_{\mathbb{R}_+^J} \binom{n}{\boldsymbol{f}_J} \prod_{j=1}^{J} \left(\frac{m_j}{m_\bullet}\right)^{f_j} \mathrm{d}F(\boldsymbol{m}_J).$$

Next we transform variables so that $p_j = m_j/m_\bullet$, $j \in [J]$, and integrate out $m_\bullet$. Write $\boldsymbol{p}_{J-1} := (p_1, p_2, \ldots, p_{J-1})$, and the Jacobian of $\boldsymbol{m}_J \mapsto (\boldsymbol{p}_{J-1}, m_\bullet)$ is $m_\bullet^{J-1}$. Therefore

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{F}_J = \boldsymbol{f}_J \,\middle|\, \sum_{j=1}^{J} F_j = n\right) &= \int_{\triangle^{J-1} \times \mathbb{R}_+} \binom{n}{\boldsymbol{f}_J} \prod_{j=1}^{J} p_j^{f_j} m_\bullet^{J-1} \,\mathrm{d}F(\boldsymbol{p}_{J-1}, m_\bullet) \\ &= \int_{\triangle^{J-1}} \binom{n}{\boldsymbol{f}_J} \prod_{j=1}^{J} p_j^{f_j} \,\mathrm{d}F(\boldsymbol{p}_{J-1}). \end{aligned} \tag{A.20}$$

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

The right-hand side of (A.20) equals $\mathtt{Mn}(n, \boldsymbol{\pi}_J) \wedge_{\boldsymbol{\pi}_J} \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu)$. On the other hand, each

$$\mathbb{P}(F_j = f_j) = \exp(-\mu_j \psi(1)) B_{f_j}(\mu_j, \boldsymbol{w})/f_j!$$

by Lemma A.2. Then, similarly to Theorem 2.8, the left-hand side of (A.20) equals the PMF of $\mathtt{BP}_J(n, \boldsymbol{\mu}_J; \boldsymbol{w})$. Hence (4.12) holds.

Proof of Theorem 4.6. By Theorem 4.5, if $\boldsymbol{w}$ satisfies (4.11) of a proper distribution ν, then the PGF of $(F_{1,n}, F_{2,n}, \dots, F_{J,n})$ equals that of $\mathtt{Mn}_J(n, \boldsymbol{\pi}_J) \wedge_{\boldsymbol{\pi}_J} \mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu)$ for all $n \in \mathbb{N}$. If so,

$$(F_{1,n}, F_{2,n}, \dots, F_{J,n} \mid \boldsymbol{\pi}_J) \sim \mathtt{Mn}_J(n, \boldsymbol{\pi}_J).$$

Then the multivariate Lindeberg–Lévy central limit theorem applies, and

$$(F_{1,n}/n, F_{2,n}/n, \dots, F_{J,n}/n \mid \boldsymbol{\pi}_J) \xrightarrow{p} \boldsymbol{\pi}_J \quad \text{as } n \to \infty;$$

see [48, Theorem 2.7 (p. 109) and Application C (p. 26)]. Therefore the conditional PGF of $(F_{1,n}/n, F_{2,n}/n, \dots, F_{J,n}/n)$ given $\boldsymbol{\pi}_J$ converges as $n \to \infty$:

$$\mathbb{E}\left(\prod_{j=1}^{J} z_j^{F_{j,n}/n} \mid \boldsymbol{\pi}_J \right) \to \prod_{j=1}^{J} z_j^{\pi_j}.$$

Hence the PGF of $(F_{1,n}/n, F_{2,n}/n, \dots, F_{J,n}/n)$ unconditionally converges to $\mathbb{E}_{\boldsymbol{\pi}_J}(\prod_{j=1}^{J} z_j^{\pi_j})$, which is the PGF of $\mathtt{NID}_J(\boldsymbol{\mu}_J \times \nu)$. □

Proof of Theorem 4.8. Suppose that $\boldsymbol{v} = (1, 0, 0, \dots)$ in Theorem 4.7. Then we obtain $\boldsymbol{F}_J \sim \mathtt{BP}_J(n, \boldsymbol{\mu}_J; \boldsymbol{w})$ by Theorem 4.5. The marginal variance of this distribution is given by (3.6). This expression has to equal the special case of the variance formula of Theorem 4.7 with $\phi(n, \lambda_\bullet, \boldsymbol{v}) = 1$. Hence (4.13) holds. □

Direct proof of Proposition 5.1. By induction, it holds that

$$\sum_{i=0}^{n} \Gamma(\alpha + i)/\Gamma(1 + i) = \Gamma(1 + n + \alpha)/a\Gamma(1 + n).$$

Using this equation, we have

$$\begin{aligned}
\frac{(n-1)!\,\alpha}{\Gamma(\alpha+n)} \sum_{i=0}^{n-2} \frac{(n-i-1)\Gamma(\alpha+i)}{\Gamma(i+1)} &= \frac{(n-1)!\,\alpha}{\Gamma(\alpha+n)} \left\{ (n-1) \sum_{i=0}^{n-2} \frac{\Gamma(\alpha+i)}{\Gamma(i+1)} - \sum_{i=1}^{n-2} \frac{\Gamma(\alpha+i)}{\Gamma(i)} \right\} \\
&= \frac{(n-1)!\,\alpha}{\Gamma(\alpha+n)} \left\{ (n-1) \frac{\Gamma(n-1+\alpha)}{\alpha\Gamma(n-1)} - \frac{\Gamma(n-1+\alpha)}{(\alpha+1)\Gamma(n-2)} \right\} \\
&= \frac{(n-1)!\,\alpha}{\Gamma(\alpha+n)} \left\{ \frac{\Gamma(n+\alpha)}{\alpha(\alpha+1)\Gamma(n-1)} \right\} \\
&= \frac{n-1}{\alpha+1}.
\end{aligned}$$

□

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

Proof of Theorem 6.1. Corollary 2.20 of [6] states conditions of such $\boldsymbol{F}$ to exist, which are confirmed below: it is apparent that $p_t(\cdot) \geq 0$ and $\sum_{f_1} p_2(f_1, n-f_1) = 1$. The key condition is that

$$\sum_{f_t} p_{t+1}\left(\boldsymbol{f}_t, n - \sum_{j=1}^{t} f_j\right) = p_t\left(\boldsymbol{f}_{t-1}, n - \sum_{j=1}^{t-1} f_j\right).$$

This equation holds because Theorem 3.1 ensures that

$$\left(\boldsymbol{F}_{t-1}, n - \sum_{j=1}^{t-1} F_j\right) \sim \mathtt{BP}_t\left(n, \left(\boldsymbol{\lambda}_{t-1}, \lambda_\bullet - \sum_{j=1}^{t-1} \lambda_j\right); \boldsymbol{w}\right)$$

when

$$\left(\boldsymbol{F}_t, n - \sum_{j=1}^{t} F_j\right) \sim \mathtt{BP}_{t+1}\left(n, \left(\boldsymbol{\lambda}_t, \lambda_\bullet - \sum_{j=1}^{t} \lambda_j\right); \boldsymbol{w}\right).$$

These conditions hold within the parameter space of the BPD. □

Proof of Theorem 6.3. The PGF of $\boldsymbol{S}_{n,J}$ is expressed as

$$\mathbb{E}\left(\prod_{i=1}^{n} z_i^{S_{i,J}}\right) = \prod_{j=1}^{J}\left\{1 + \sum_{i=1}^{n} \frac{z_i}{i!} \sum_{k=1}^{i} {\lambda_{j,J}}^k B_{i,k}(\boldsymbol{w})\right\} \frac{n!}{B_n(\lambda_\bullet, \boldsymbol{w})}.$$

Then the assumption that $\lambda_{j,J} = \mu/J + \mathrm{o}(J^{-1})$ lets the right-hand side be

$$\left\{1 + \sum_{i=1}^{n} \frac{z_i}{i!}\left(\frac{\mu w_i}{J} + \mathrm{o}(J^{-1})\right)\right\}^J \frac{n!}{B_n(\lambda_\bullet, \boldsymbol{w})} \to \exp\left(\mu \sum_{i=1}^{n} \frac{z_i}{i!} w_i\right) \frac{n!}{B_n(\mu, \boldsymbol{w})} \quad \text{as } J \to \infty.$$

The last expression is the PGF of (6.3). □

Acknowledgements

The present article was prepared during the author's visit to the Institute of Statistical Mathematics, Japan. The author expresses his sincere gratitude to Professors Satoshi Kuriki and Shuhei Mano for accepting his visit. Also the author is very grateful to anonymous referees for their constructive comments on the long draft.

Funding information

The author was supported by a JSPS KAKENHI grant.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Aczél, J. (2006). *Lectures on Functional Equations and Their Applications*. Dover, Mineola, NY.
[2] Aigner, M. (1997). *Combinatorial Theory*. Springer, Berlin and Heidelberg.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

[3] AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.
[4] ARRATIA, R., BARBOUR, A. D. AND TAVARÉ, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach*. European Mathematical Society, Zürich.
[5] BARNDORFF-NIELSEN, O. E. AND JØRGENSEN, B. (1991). Some parametric models on the simplex. *J. Multivariate Anal.* **39**, 106–116.
[6] BREIMAN, L. (1992). *Probability*. SIAM, Philadelphia.
[7] CARLTON, M. A. (2002). A family of densities derived from the three-parameter Dirichlet process. *Adv. Appl. Prob.* **39**, 764–774.
[8] CHARALAMBIDES, C. A. (2002). *Enumerative Combinatorics*. Chapman & Hall/CRC, Boca Raton, FL.
[9] CHEG, G., LIU, B., PENG, L., ZHANG, B. AND ZHENG, S. (2019). Testing the equality of two high-dimensional spatial sign covariance matrices. *Scandinavian J. Statist.* **46**, 257–271.
[10] CHENG PING (1964). Minimax estimates of parameters of distributions belonging to the exponential family. *Acta Math. Sinica* **5**, 277–299.
[11] COMTET, L. (1974). *Advanced Combinatorics*. Reidel, Boston.
[12] CONSUL, P. C. AND FAMOYE, F. (2006). *Lagrangian Probability Distributions*. Birkhäuser, Boston.
[13] CONSUL, P. C. AND JAIN, G. C. (1973). A generalization of the Poisson distribution. *Technometrics* **15**, 791–799.
[14] CONSUL, P. C. AND MITTAL, S. P. (1975). A new urn model with predetermined strategy. *Biometrische Zeitschrift* **17**, 67–75.
[15] CONSUL, P. C. AND MITTAL, S. P. (1977). Some discrete multinomial probability models with predetermined strategy. *Biometrical J.* **19**, 161–173.
[16] CRANE, H. (2016). The ubiquitous Ewens sampling formula. *Statist. Sci.* **31**, 1–19.
[17] DAVIS, C. S. (1993). The computer generation of multinomial random variates. *Comput. Statist. Data Anal.* **16**, 205–217.
[18] DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.
[19] DEVROYE, L. (1993). A triptych of discrete distributions related to the stable law. *Statist. Prob. Lett.* **18**, 349–351.
[20] EGOZCUE, J. J. AND PAWLOWSKY-GLAHN, V. (2019). Compositional data: The sample space and its structure. *TEST* **28**, 599–638.
[21] FAVARO, S, HADJICHARALAMBOUS, G. AND PRÜNSTER, I. (2011). On a class of distributions on the simplex. *J. Statist. Planning Infer.* **141**, 2987–3004.
[22] FEDERER, W. T. (1951). Testing proportionality of covariance matrices. *Ann. Math. Statist.* **22**, 102–106.
[23] FENG, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics*. Springer, Heidelberg.
[24] GHOSH, J. K. AND RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics*. Springer, New York.
[25] GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–264.
[26] GRANDELL, J. (1997) *Mixed Poisson Processes*. Chapman & Hall, London.
[27] HINDE, J. AND DEMÉTRIO, C. G. B. (1998). Overdispersion: Models and estimation. *Comput. Statist. Data Anal.* **27**, 151–170.
[28] HO, F. C. M., GENTLE, J. E. AND KENNEDY, W. J. (1979). Generation of random variates from the multinomial distribution. *Proc. Amer. Statist. Assoc. Statist. Comput. Sec*. 336–339.
[29] HOSHINO, N. (2006). A discrete multivariate distribution resulting from the law of small numbers. *J. Appl. Prob.* **43**, 852–866.
[30] HOSHINO, N. (2009). The quasi-multinomial distribution as a tool for disclosure risk assessment. *J. Official Statist.* **25**, 269–291.
[31] HOSHINO, N. (2020). A firm foundation for statistical disclosure control. *Japan. J. Statist. Data Sci.* **3**, 721–746.
[32] HOSHINO, N. AND TAKEMURA, A. (1998). Relationship between logarithmic series model and other superpopulation models useful for microdata disclosure risk assessment. *J. Japan Statist. Soc.* **28**, 125–134.
[33] HU, J. AND HOSHINO, N. (2018). The quasi-multinomial synthesizer for categorical data. *Privacy in Statistical Databases (PSD 2018)* (Lecture Notes in Computer Science **11126**), eds J. Domingo-Ferrer and F. Montes, pp. 75–91. Springer, Cham.
[34] ISHII, G. AND HAYAKAWA, R. (1960). On the compound binomial distribution. *Ann. Inst. Statist. Math.* **12**, 69–80. Errata, **12**, 208.
[35] JOHNSON, N. L., KOTZ, S. AND BALAKRISHNAN, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.
[36] JOHNSON, N. L., KOTZ, S. AND KEMP, A. W. (1993). *Univariate Discrete Distributions*, 2nd edn. Wiley, New York.
[37] LIJOI, A., MENA, I. R. H. AND PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100**, 1278–1291.
[38] LUKACS, E. (1955). A characterization of the gamma distribution. *Ann. Math. Statist.* **26**, 319–324.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press

[39] MALEFAKI, S. AND ILIOPOULOS, G. (2007). Simulating from a multinomial distribution with large number of categories. *Comput. Statist. Data Anal.* **51**, 5471–5476.
[40] MOREL, J. G. AND NAGARAJ, N. K. (1993). A finite mixture distribution for modelling multinomial extra variation. *Biometrika* **80**, 363–371.
[41] MOSIMANN, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* **49**, 65–82.
[42] NEERCHAL, N. K. AND MOREL, J. G. (2005). An improved method for the computation of maximum likelihood estimates for multinomial overdispersion models. *Comput. Statist. Data Anal.* **49**, 33–43.
[43] PAUL, S. R., BALASOORIYA, U. AND BANERJEE, T. (2005). Fisher information matrix of the Dirichlet-multinomial distribution. *Biometrical J.* **47**, 230–236.
[44] PITMAN, J. (2006). *Combinatorial Stochastic Processes* (Lecture Notes in Mathematics **1875**). Springer, New York.
[45] QI, F., NIU, D. W., LIM, D. AND YAO, Y. H. (2020). Special values of the Bell polynomials of the second kind for some sequences and functions. *J. Math. Anal. Appl.* **491**, 124382.
[46] RIORDAN, J. (1968). *Combinatorial Identities*. Wiley, New York.
[47] SATTERTHWAITE, F. E. (1942). Generalized Poisson distributions. *Ann. Math. Statist.* **13**, 410–417.
[48] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
[49] SIBUYA, M. (1979). Generalized hypergeometric, digamma and trigamma distributions. *Ann. Inst. Statist. Math.* **31**, 373–390.
[50] SIBUYA, M. (1993). A random clustering process. *Ann. Inst. Statist. Math.* **45**, 459–465.
[51] STEUTEL, F. W. AND VAN HARN, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Prob.* **1**, 893–899.
[52] STEUTEL, F. W. AND VAN HARN, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*. Dekker, New York.
[53] TAKEMURA, A. (1999). Some superpopulation models for estimating the number of population uniques. In *Proceedings of the Statistical Data Protection Conference*, pp. 45–58. Office for Official Publications of the European Communities, Luxembourg.
[54] TSUKUDA, K. AND MATSUURA, S. (2019). High-dimensional testing for proportional covariance matrices. *J. Multivariate Anal.* **171**, 412–420.
[55] VERSHIK, A. M. (1996). Statistical mechanics of combinatorial partitions, and their limit shapes. *Functional Analysis and its Applications*, **30**, 90–105.
[56] ZHOU, M. AND CARIN, L. (2015). Negative binomial process count and mixture modeling. *IEEE Trans. Pattern Anal. Mach. Intellig.* **37**, 307–320.

https://doi.org/10.1017/jpr.2025.10042 Published online by Cambridge University Press