Multivariate Discrete Distributions via Sums and Shares

by M.C. Jones
The Open University, U.K.
and Éric Marchand
Université de Sherbrooke, Canada

Abstract

In this article, we develop a sum and share decomposition to model multivariate discrete distributions, and more specifically multivariate count data that can be divided into a number of distinct categories. From a Poisson mixture model for the sum and a multinomial mixture model for the shares, a rich ensemble of properties, examples and relationships arises. As a main example, a seemingly new multivariate model involving a negative binomial sum and Pólya shares is considered, previously seen only in the bivariate case. For other choices of the distribution of the sum, natural but novel discrete multivariate Liouville distributions emerge; an important special case of these is that of Schur constant distributions. A new variance-mean inequality for univariate discrete distributions with decreasing probability mass functions ensues. Analogies and interactions with related continuous multivariate distributions are to the fore throughout.

Keywords: Liouville distribution; Multinomial mixture; Poisson mixture; Pólya distribution; Schur constant distribution; Variance-mean inequality.

Subject classification: 62E10.

Running head: Discrete sums and shares.

1. Introduction

This article is concerned with hierarchical constructions for multivariate count data, thinking of total counts as *sums* and their separation into distinct categories as *shares* of those sums. Such representations are plentiful in practice, consisting, for instance, of events (accidents, insurance claims, occurrences of diseases, presence of a member of a species, etc.) falling into different geographical locations, types, time periods, etc.

Formally, let $\mathbb{N}_0$ denote the set of non-negative integers and d denote dimensionality. We are concerned with joint distributions for random variables $M_1, ..., M_d \in \mathbb{N}_0^d$. For convenience, write $\mathcal{M}_q = (M_1, ..., M_q)$, $q = d-1$ or d. Our starting point is to transform linearly from $\mathcal{M}_d$ to $(\mathcal{M}_{d-1}, T)$ where $T = M_1 + \cdots + M_d$ is the sum of the random variables. Then, this article is concerned with the construction of multivariate discrete distributions in the following manner:

- let the sum T have a distribution with probability mass function (p.m.f.) $p_T(t)$, $t \in \mathbb{N}_0$;
- conditionally on $T = t$, share t out between values for $\mathcal{M}_d$, that is, let $\mathcal{M}_{d-1}|T = t$ have a distribution with p.m.f. $b_{[t]}(m_1, ..., m_{d-1})$ on the discrete simplex defined by $\mathcal{M}_{d-1} \in \{0, 1, ..., t\}^{d-1}$ such that $M_1 + \cdots + M_{d-1} \in \{0, 1, ..., t\}$.

Of course, we have just rewritten the joint p.m.f., $p(m_1, ..., m_d)$, of any $\mathcal{M}_d \in \mathbb{N}_0^d$ in the equivalent form

$$p(m_1, ..., m_d) = b_{[m_1+\cdots+m_d]}(m_1, ..., m_{d-1})\, p_T(m_1 + \cdots + m_d), \qquad (1)$$

rather than making any reduction in generality.

Our aim in this article is to investigate certain families of multivariate discrete distributions which are especially natural and/or attractive to define through this 'sum and share' construction. Let us cut straight to the chase. Since T is a count random variable, the Poisson distribution is a natural first choice for p_T; for a first choice of distribution with p.m.f. $b_{[t]}(m_1, ..., m_{d-1})$ on the unit simplex, the multinomial distribution springs to mind. It is easy to see that the resulting joint distribution is that of d independent Poisson random variables with parameters $r_i \equiv \lambda u_i$, $i = 1, ..., d$, where λ is the parameter of the Poisson distribution and $u_1, ..., u_d$ are the parameters of the multinomial distribution. For greater generality and to induce correlation, we consider instead mixing these distributions over distributions for $\Lambda > 0$ and for $0 < U_1, ..., U_d < 1$ such that $U_1 + \cdots + U_d = 1$. The resulting joint distributions are considered in general terms in Section 2.

We then specialise again by making the natural choices of Λ following a gamma distribution (so that T is negative binomial) and of $U_1, ..., U_d$ following a Dirichlet distribution (so that $\mathcal{M}_{d-1}|T = t$ follows a Dirichlet-multinomial, or multivariate Pólya, distribution). The resulting joint distribution is a multivariate extension of what in the bivariate case Laurent (2012) called the Bailey distribution. It is a focus of this article, and is considered in detail in Section 3. Its two main special cases are included in Section 3.1. Likelihood inference is straightforward for sum-and-share models when the sum and share distributions have no parameters in common; this is the case for the distribution of Section 3, as described in Section 3.2.

In Section 4, we look rather briefly at a different 'super case' of the distribution of Section 3, what we call the multivariate discrete Liouville distribution. Prominent among this class of distributions are the Schur constant distributions of Castañer et al. (2015), discussed in Section 4.1. We finish the article in Section 5 with further brief discussion.

2. Poisson Mixtures for Sums and Multinomial Mixtures for Shares

Let us first consider taking $b_{[t]}(m_1, ..., m_{d-1})$ to be the p.m.f. of a *multinomial mixture* distribution:

$$\mathcal{M}_{d-1}|T = t, U_1 = u_1, ..., U_{d-1} = u_{d-1} \sim \text{Multinomial}(t, u_1, ..., u_{d-1})$$

where

$$U_1, ..., U_{d-1} \sim H \;\text{ on }\; 0 < u_1 + \cdots + u_{d-1} < 1, \;\text{independent of}\; T;$$

and to take $p_T(t)$ to be the p.m.f. of a *Poisson mixture* distribution:

$$T|\Lambda = \lambda \sim \text{Poisson}(\lambda)\;,\; \Lambda \sim L \;\text{ on }\; (0, \infty),\; \text{independent of}\; U_1, ..., U_{d-1}.$$

In this case, (1) becomes for absolutely continuous densities h for $(U_1, \ldots, U_{d-1})$ and ℓ for Λ,

$$\begin{aligned} p(m_1, \ldots, m_d) \;&= \underset{0<u_1+\cdots+u_{d-1}<1}{\int \cdots \int} \frac{(m_1 + \cdots + m_d)!}{m_1! \cdots m_d!} u_1^{m_1} \cdots u_{d-1}^{m_{d-1}} (1 - u_1 - \cdots - u_{d-1})^{m_d} \\ &\qquad \times h(u_1, ..., u_{d-1}) du_1 \cdots du_{d-1} \\ &\qquad \times \int_0^\infty \frac{e^{-\lambda}\lambda^{m_1+\cdots+m_d}}{(m_1 + \cdots + m_d)!} \ell(\lambda) d\lambda, \qquad m_1, ..., m_d \in \mathbb{N}_0^d. \end{aligned} \tag{2}$$

Alternatively, one can think of (2) as the result of mixing independent Poisson distributions with parameters $r_i = \lambda u_i$, $i = 1, ..., d$, over the distribution of $R_i = \Lambda\, U_i$, $i = 1, ..., d$, where $(R_1, ..., R_d)$ follow the continuous analogue of (1) in which $\Lambda = R_1 + \cdots + R_d$ plays the role of T and $(R_1, ..., R_{d-1})$ plays the role of $\mathcal{M}_{d-1}$:

$$f(r_1, ..., r_d) = \frac{1}{(r_1 + \cdots + r_d)^{d-1}}\, h\left(\frac{r_1}{r_1 + \cdots + r_d}, \cdots, \frac{r_{d-1}}{r_1 + \cdots + r_d}\right) \ell(r_1 + \cdots + r_d), \tag{3}$$

$r_1, ..., r_d > 0$.

Marginal distributions for M_i, $i = 1, ..., d$, are Poisson mixture distributions but the marginal distributions of R_i, $i = 1, ..., d$, and hence of M_i, $i = 1, ..., d$, are not tractable in general. The moments of M_i, $i = 1, ..., d$, are readily available in terms of those of Λ and U_i, $i = 1, ..., d$, however. In particular, $E(M_i) = E(\Lambda)E(U_i)$ and

$$V(M_i) = E(\Lambda^2)V(U_i) + V(\Lambda)\{E(U_i)\}^2 + E(\Lambda)E(U_i).$$

Recall that Poisson mixture distributions are necessarily overdispersed, as is reflected in these formulae. Covariances simplify because $\text{Cov}(M_i, M_j|R_1, ..., R_d) = 0$ for any $i \neq j$ so that

$$\text{Cov}(M_i, M_j) = \text{Cov}(R_i, R_j).$$

Therefore, for $1 \leq (i \neq j) \leq d$,

$$\text{Cov}(M_i, M_j) = V(\Lambda)E(U_iU_j) + \{E(\Lambda)\}^2\text{Cov}(U_i, U_j).$$

In particular, for the bivariate version of (2),

$$\text{Cov}(M_1, M_2) = V(\Lambda)E\{U(1 - U)\} - \{E(\Lambda)\}^2V(U)$$

where $U \equiv U_1 \sim H$ on $(0, 1)$. Clearly, this covariance is always negative for degenerate Λ (i.e. T is Poisson distributed) and non-degenerate U. On the other hand, the covariance is always positive for degenerate U (i.e. $M_1 = 1 - M_2$ is binomially distributed) and non-degenerate Λ.

3. Negative Binomial Sums and Pólya Shares

The most natural mixing distributions to employ for U and Λ would seem to be

$$U \sim \text{Dirichlet}(\alpha_1, ..., \alpha_d), \qquad \Lambda \sim \text{Gamma}(a, b),$$

$\alpha_1, ..., \alpha_d, a, b > 0$; write $\alpha_\bullet = \alpha_1 + \cdots + \alpha_d > 0$ and $0 < \theta = b/(1+b) < 1$. This gives a $(d+2)$-parameter family whose p.m.f. is readily seen to be

$$p(m_1, ..., m_d) = \frac{(a)_{m_1+\cdots+m_d}}{m_1! \cdots m_d!} \frac{\prod_{i=1}^d (\alpha_i)_{m_i}}{(\alpha_\bullet)_{m_1+\cdots+m_d}} \theta^a (1-\theta)^{m_1+\cdots+m_d}, \tag{4}$$

$m_1, ..., m_d \in \mathbb{N}_0^d$. Here, $(\alpha)_m$ denotes the ascending factorial $\Gamma(\alpha + m)/\Gamma(\alpha)$. By construction, we have

$$\mathcal{M}_{d-1}|T = t \sim \text{Pólya}(t; \alpha_1, ..., \alpha_d) \;\; \text{and} \;\; T \sim \text{NegativeBinomial}\,(a, \theta)\,.$$

When $d = 2$, this is the Bailey distribution of Laurent (2012) and the distribution in (4) thus represents its multivariate extension.

The $(R_1, ..., R_d)$ distribution associated with the distribution with p.m.f. (4) has density

$$f(r_1, ..., r_d) \;=\; \frac{b^a \Gamma(\alpha_\bullet)}{\Gamma(a) \prod_{i=1}^d \Gamma(\alpha_i)} \left(\prod_{i=1}^d r_i^{\alpha_i - 1} \right) (r_1 + \cdots + r_d)^{a - \alpha_\bullet} \, e^{-b(r_1 + \cdots + r_d)}, \qquad (5)$$

$r_1, ..., r_d \;>\; 0$. Of course, $R_1, ..., R_d | \Lambda \;=\; \lambda \;\sim\; \lambda \times \text{Dirichlet}(\alpha_1, ..., \alpha_d)$ and $\Lambda = R_1 + \cdots + R_d \sim \text{Gamma}(a, b)$. This 'Dirichlet-gamma' distribution has been used in low-dimensional cases as a prior distribution in Bayesian analysis, in the guise of a 'beta-gamma' distribution by e.g. Bhattacharya et al. (2014) when $d = 2$ and by Peña & Gupta (1990) when $d = 3$. It is a particular continuous multivariate Liouville distribution (Gupta & Richards, 1987).

Moments are readily available. Inserting the moments of the gamma and Dirichlet distributions into the formulae in Section 2, we find that, for $i = 1, ..., d$,

$$E(M_i) = \frac{a\,\alpha_i}{b\,\alpha_\bullet},$$

$$V(M_i) = \frac{a\alpha_i}{b^2\,\alpha_\bullet^2 (1 + \alpha_\bullet)} \left[\alpha_\bullet \{ a + 1 + (1 + \alpha_\bullet) b \} + (\alpha_\bullet - a) \alpha_i \right]$$

and, for $1 \le (i \ne j) \le d$,

$$\text{Cov}(M_i, M_j) = \frac{a(\alpha_\bullet - a)\alpha_i \alpha_j}{b^2\,\alpha_\bullet^2 (1 + \alpha_\bullet)}. \qquad (6)$$

The signs of the covariances are the same for all i, j and depend directly on the sign of $\alpha_\bullet - a$. Covariances are zero when $\alpha_\bullet = a$ which corresponds to independence: (5) reduces to the distribution of independent $\text{Gamma}(\alpha_i, b)$ random variables, $i = 1, ..., d$, and hence (4) to the distribution of independent $\text{NegativeBinomial}(\alpha_i, \theta)$ random variables, $i = 1, ..., d$.

In fact, independence holds in (4) (equivalently in (5)) if and only if $\alpha_\bullet = a$. This can be seen directly from the density functions or from the product moments. For general product moments, let $K = k_1 + \cdots + k_d$. Then, we have

$$E\left(\prod_{i=1}^d R_i^{k_i} \right) = E\left(\Lambda^K\right) E\left(\prod_{i=1}^d U_i^{k_i} \right) = \frac{(a)_K \prod_{i=1}^d (\alpha_i)_{k_i}}{b^K (\alpha_\bullet)_K} = E\left\{ \prod_{i=1}^d (M_i - k_i + 1)_{k_i} \right\}.$$

The final equality holds because the kth descending factorial moment of the Poisson distribution with parameter λ is λ^k. Inter alia, this gives a formula for marginal binomial moments:

$$E\left\{ \binom{M_i}{k_i} \right\} = \frac{(a)_{k_i} (\alpha_i)_{k_i}}{k_i!\, b^{k_i} (\alpha_\bullet)_{k_i}}, \quad i = 1, ..., d.$$

Marginal distributions are a little more tractable than they were in the more general case of Section 2. In Appendix A, we show that the p.m.f. of M_1 can be written

$$\mathbb{P}(M_i = m_i) \;=\; \frac{b^a}{(1+b)^{a+m_i}} \frac{(a)_{m_i} (\alpha_i)_{m_i}}{(\alpha_\bullet)_{m_i}\, m_i!} \, {}_2F_1\left(a + m_i, \alpha_\bullet - \alpha_i; \alpha_\bullet + m_i; \frac{1}{1+b} \right) .$$

Here, ${}_2F_1$ is the Gauss hypergeometric function. A closely related expression was given when $d = 2$ in Laurent (2012).

For a pair of random variables, alongside the covariance and correlation, it is also of interest to consider *local* dependence which, in the case of discrete models, is measured by the set of log cross-product ratios of adjacent 2×2 cells in a (usually infinitely) large ordinal contingency table (Yule & Kendall, 1950, Goodman, 1969, 1985): for $(m_1, m_2) \in \mathbb{N}_0^2$, define

$$\theta(m_1, m_2) \equiv \log \left\{ \frac{p(m_1, m_2) p(m_1 + 1, m_2 + 1)}{p(m_1 + 1, m_2) p(m_1, m_2 + 1)} \right\} .$$

When $d = 2$, (4) becomes

$$p(m_1, m_2) = \frac{(\alpha_1)_{m_1} (\alpha_2)_{m_2}}{m_1!\, m_2!} \frac{(a)_{m_1+m_2}}{(\alpha_1 + \alpha_2)_{m_1+m_2}} \, \theta^a \, (1 - \theta)^{m_1+m_2}$$

and hence the distribution has local dependence function

$$\theta(m_1, m_2) = \log \left\{ \frac{(\alpha_1 + \alpha_2 + m_1 + m_2)(a + m_1 + m_2 + 1)}{(\alpha_1 + \alpha_2 + m_1 + m_2 + 1)(a + m_1 + m_2)} \right\} .$$

This is a function only of $t = m_1 + m_2$, $\alpha_\bullet = \alpha_1 + \alpha_2$ and a; for fixed t, it is increasing in $\alpha_\bullet$ and decreasing in a. It is necessarily zero for all t if and only if $\alpha_\bullet = a$, previously established to be the case of independence. The local dependence function is positive (negative) for all t when $\alpha_0 > (<)\, a$ (like the correlation). It is largest in absolute value when $t = 0$ where it takes the value $\log[\alpha_\bullet (a + 1)/\{a(\alpha_\bullet + 1)\}]$ and tends to zero as $t \to \infty$.

3.1 Special Cases

1: $\alpha_1 = \cdots = \alpha_d = 1$. The Dirichlet mixing distribution reduces to the continuous uniform distribution on the simplex and so the Dirichlet-multinomial distribution reduces to the discrete uniform distribution on the discrete simplex. We then have the 'discrete Schur-constant' distribution of Castañer et al. (2015) with negative binomial T, which has

$$p(m_1, ..., m_d) = \frac{(a)_{m_1+\cdots+m_d}}{(d)_{m_1+\cdots+m_d}} \, \theta^a \, (1 - \theta)^{m_1+\cdots+m_d} \, , \qquad (7)$$

$m_1, ..., m_d \in \mathbb{N}_0^d$. Notice that (7) depends on $m_1, ..., m_d$ only through the sum $m_1 + \cdots + m_d$. Similarly, the underlying distribution of $(R_1, ..., R_d)$ has density

$$f(r_1, ..., r_d) \;=\; \frac{b^a \Gamma(d)}{\Gamma(a)} (r_1 + \cdots + r_d)^{a-d} \, e^{-b(r_1+\cdots+r_d)}, \tag{8}$$

$r_1, ..., r_d > 0$, which depends on $r_1, ..., r_d$ only through $r_1+\cdots+r_d$. Immediately, if $a = d$, (7) reduces to the distribution of d independent geometric(θ) random variables and (8) to the distribution of d independent exponential(b) random variables, respectively. Distribution (8) underlies the 'gamma-simplex copula' of McNeil & Nešlehová (2010) in a sense to be explained in a more general discussion of distributions with uniform shares in Section 4.1 to follow.

2: $\alpha_1, ..., \alpha_d \to \infty$ such that $\alpha_i/(\alpha_1+\cdots+\alpha_d) \to \phi_i$, $i = 1, ..., d$. The Dirichlet mixing distribution becomes degenerate at values $0 < \phi_1, ..., \phi_d < 1$ such that $\sum_{i=1}^d \phi_i = 1$, so the Dirichlet-multinomial distribution reduces to the multinomial distribution and we have

$$p(m_1, ..., m_d) = \frac{(a)_{m_1+\cdots+m_d}}{m_1! \cdots m_d!} \left(\prod_{i=1}^d \phi_i^{m_i} \right) \theta^a (1-\theta)^{m_1+\cdots+m_d}, \tag{9}$$

$m_1, ..., m_d \in \mathbb{N}_0^d$. In this case, it is not difficult to show that, for $1 \le (i \ne j) \le d$,

$$\mathrm{Corr}(M_i, M_j) = \sqrt{\frac{\phi_i}{\phi_i + b} \, \frac{\phi_j}{\phi_j + b}} \;>\; 0$$

(where $b = \theta/(1-\theta)$). The amount of correlation varies monotonically from 0 to 1 as b decreases from ∞ to 0, or equivalently as θ decreases from 1 to 0. When $d = 2$,

$$p(m_1, m_2) = \frac{(a)_{m_1+m_2}}{m_1! \, m_2!} \, \theta^a q_1^{m_1} q_2^{m_2}, \qquad m_1, m_2 \in \mathbb{N}_0^2,$$

where $q_1 = \phi_1(1-\theta)$, $q_2 = (1-\phi_1)(1-\theta)$. This has marginals that are Negative-Binomial$(a, \theta/(\theta + q_i))$, $i = 1, 2$, and local dependence function

$$\theta(m_1, m_2) = \log(a + m_1 + m_1 + 1) - \log(a + m_1 + m_2) > 0.$$

The latter depends only on a rather than on b and ϕ_1.

3.2 Likelihood Inference

Let $(m_{1j}, ..., m_{dj})$, $j = 1, ..., n$, be a sample of independent observations taken from the distribution with density (4); also let $t_j = m_{1j} + \cdots + m_{dj}$, $j = 1, ..., n$, be the sample totals. Likelihood inference for the parameters a, θ associated with the distribution of

Λ and for the parameters $\alpha_1, ..., \alpha_d$ associated with the distribution of U proceeds as two separate problems. The $(d+2) \times (d+2)$ Fisher information matrix associated with distribution (4) will therefore be in two blocks, one of size 2×2, the other $d \times d$, and maximum likelihood (ML) estimators of a and θ will be asymptotically independent of ML estimators of $\alpha_1, ..., \alpha_d$.

The problem of estimating a and θ is the standard one of ML estimation of the parameters of the negative binomial distribution when both are unknown; for early references, see Section 5.8.3 of Johnson et al. (2005). The cross-term in the 2×2 Fisher information submatrix is $-1/\theta$, meaning that the asymptotic correlation between the ML estimates of a and θ is positive.

The problem of estimating $\alpha_1, ..., \alpha_d$ is one of ML estimation of the parameters of the Pólya distribution based on independent data with different, known, values of t. The score equations reduce to

$$\sum_{j=1}^{n} \{\psi(\alpha_i + m_{ij}) - \psi(\alpha_i)\} = \sum_{j=1}^{n} \{\psi(\alpha_\bullet + t_j) - \psi(\alpha_\bullet)\}, \quad i = 1, ..., d,$$

where ψ is the digamma function (which is an increasing function for positive values of its argument). The $d \times d$ submatrix of the observed information matrix has all its off-diagonal elements the same and equal to

$$\sum_{j=1}^{n} \{\psi'(\alpha_\bullet + t_j) - \psi'(\alpha_\bullet)\} < 0.$$

It follows that the corresponding block of the asymptotic correlation matrix is of equicorrelation type, the correlations between ML estimates of α's being positive.

4. Super Case: the Multivariate Discrete Liouville Distribution

Super cases of the distribution on which we focussed in Sections 3 and 4 abound, of course, by making different choices of distributions for T and $\mathcal{M}_{d-1}|T = t$ in (1). A super case of particular interest might be that in which the sharing distribution remains the Dirichlet-multinomial as in Section 3 but a general distribution is allowed for T (rather than the negative binomial). This results in

$$p(m_1, ..., m_d) = \frac{(m_1 + \cdots + m_d)!}{m_1! \cdots m_d!} \frac{\prod_{i=1}^{d} (\alpha_i)_{m_i}}{(\alpha_\bullet)_{m_1+\cdots+m_d}} \, p_T(m_1 + \cdots + m_d). \tag{10}$$

Note that (10) can be written as

$$p(m_1, ..., m_d) = \frac{\prod_{i=1}^{d} (\alpha_i)_{m_i}}{m_1! \cdots m_d!} \mathcal{F}(m_1 + \cdots + m_d). \tag{11}$$

where $\mathcal{F}(t) = t! p_T(t)/(\alpha_\bullet)_t$. The form of (11) is a discrete analogue of the continuous multivariate Liouville distribution (Gupta & Richards, 1987) which has p.d.f.

$$f(r_1, ..., r_d) = \left(\prod_{i=1}^{d} r_i^{\alpha_i - 1} \right) \mathbb{F}(r_1 + \cdots + r_d), \tag{12}$$

$r_i > 0$, $i = 1, ..., d$, for suitable $\mathbb{F}$. More strikingly, perhaps, if T has a Poisson mixture distribution with general (not necessarily gamma) mixing density ℓ, then the distribution of $(R_1, ..., R_d)$ associated with (10) — the special case of (3) with Dirichlet h — has the form

$$f(r_1, ..., r_d) = \frac{\Gamma(\alpha_\bullet)}{\prod_{i=1}^{d} \Gamma(\alpha_i)} \frac{\prod_{i=1}^{d} r_i^{\alpha_i - 1}}{(r_1 + \cdots + r_d)^{\alpha_\bullet - 1}} \ell(r_1 + \cdots + r_d)$$

which is indeed a continuous Liouville distribution of form (12). In the discrete case, our preferred formulation is (10) rather than (11) because the role of p_T in the former is much clearer than that of $\mathcal{F}$ in the latter.

In the literature, the name multivariate discrete Liouville distribution was used by Lingappaiah (1984) for the distribution having p.m.f. of form

$$p(m_1, ..., m_d) = \frac{\prod_{i=1}^{d} \theta_i^{m_i}}{m_1! \cdots m_d!} \mathcal{G}(m_1 + \cdots + m_d).$$

Rearranging appropriately, this can be written

$$p(m_1, ..., m_d) = \frac{(m_1 + \cdots + m_d)!}{m_1! \cdots m_d!} \left\{ \prod_{i=1}^{d} \left(\frac{\theta_i}{\theta_1 + \cdots + \theta_d} \right)^{m_i} \right\} p_T(m_1 + \cdots + m_d). \tag{13}$$

This is none other than distribution (1) with general distribution for T and *multinomial*, Multinomial$(t, \theta_1/\sum_{i=1}^{d} \theta_i, ..., \theta_{d-1}/\sum_{i=1}^{d} \theta_i)$, conditional distribution for $\mathcal{M}_{d-1}|T = t$. It follows that the new multivariate discrete Liouville distribution at (10) is a mixture over $\Theta_1/\sum_{j=1}^{d} \Theta_j, ..., \Theta_{d-1}/\sum_{j=1}^{d} \Theta_j \sim \text{Dir}(\alpha_1, ..., \alpha_d)$ of Lingappaiah's multivariate discrete Liouville distribution at (13) (and hence is more general than it).

4.1 Special Case: Schur Constant Distributions

The special cases of our multivariate discrete Liouville distribution (10) with $\alpha_1 = \cdots = \alpha_d = 1$ are the Schur constant distributions of Castañer et al. (2015). (A special case of this observation was made in Section 3.1 above.) This multivariate discrete distribution is of considerable independent interest. As well as listing a few of the particular properties of this distribution, we would like to stress its analogue with the continuous case. All but two of the observations in this subsection can also be found in Castañer et al. (2015); the two new observations are both in the fourth paragraph of this subsection. Castañer et al. (2015) provide various applications in insurance for a related counting process.

From (10), the p.m.f. in this case has the simple form

$$p(m_1, ..., m_d) = p_T(m_1 + \cdots + m_d) \,/\, \binom{m_1 + \cdots + m_d + d - 1}{d - 1};$$

this reduces in the bivariate case (for which, see also Ait Aoudia & Marchand, 2014) to $p(m_1, m_2) = p_T(m_1 + m_2)/(m_1 + m_2)$. Notice that the p.m.f. is constant on $m_1 + \cdots + m_d = k$ for each constant value of $k = 0, 1, ...$, hence the name given to the distribution. This is a consequence of the sharing distribution being uniform on the discrete simplex.

For Schur constant distributions, univariate marginal distributions are all the same and can be written in terms of p_T as

$$p_S(m_i) = (d - 1) \sum_{t=m_i}^{\infty} \frac{(t - m_i + 1)_{d-2}}{(t + 1)_{d-1}} p_T(t) \tag{14}$$

(multivariate marginals can be written in rather similar fashion); conversely,

$$p_T(t) = \binom{t + d - 1}{d - 1} \sum_{j=0}^{d-1} \binom{d - 1}{j} (-1)^j p(t + j). \tag{15}$$

In fact, the joint p.m.f. can be written in terms of p_S as

$$p(m_1, ..., m_d) = \sum_{j=0}^{d-1} \binom{d - 1}{j} (-1)^j p_S(m_1 + \cdots + m_d + j) \tag{16}$$

and the survival function simply as

$$\Pr(M_1 \geq m_1, ..., M_d \geq m_d) = \overline{P}_S(m_1 + \cdots + m_d) \tag{17}$$

where $\overline{P}_S$ is the survival function associated with p_S. While p_T can be specified arbitrarily on $\mathbb{N}_0$, (15) and (16) show that p_S has to be a discrete $(d - 1)$-monotone distribution on $\mathbb{N}_0$ (Chee & Wang, 2016). Independence in Schur constant distributions corresponds to p_S being geometric. An interesting example of (16) occurs for p_S a Poisson(α) p.m.f., where $0 < \alpha \leq 1$. Such an example, as well as Dirichlet Poisson mixtures, arises as the distribution of counts of Bernoulli success strings in recent work of Ait Aoudia et al. (2016).

The most convenient form for multivariate moments of Schur constant distributions — which appears not to be in Castañer et al. (2015) — is

$$E\left\{\prod_{i=1}^{d}\binom{M_i}{k_i}\right\} = E\left\{\binom{T}{k_1+\cdots+k_d}\right\} \Big/ \binom{k_1+\cdots+k_d+d-1}{d-1}$$

where $T \sim p_T$. See Appendix B for a proof of this result. Inter alia, for any $1 \le i \ne j \le d$,

$$\begin{aligned}\operatorname{Cov}(M_i, M_j) &= \frac{1}{d^2(d+1)}\left[dV(T) - \{E(T)\}^2 - dE(T)\right] \\ &= \frac{1}{2}\left[V(M_i) - \{E(M_i)\}^2 - E(M_i)\right].\end{aligned}$$

It follows that $\operatorname{Corr}(M_i, M_j) < 1/2$. Pairs of random variables are positively correlated if

$$V(T) > E(T)\left\{\frac{E(T)}{d} + 1\right\},$$

a requirement becoming closer and closer to overdispersion of the distribution of T as d increases. Also, $\operatorname{Corr}(M_i, M_j) \ge -1$ implies that, for M_i following a distribution on $\mathbb{N}_0$ with decreasing p.m.f.,

$$V(M_i) \ge \frac{1}{3}E(M_i)\{E(M_i)+1\}.$$

This apparently new observation is the discrete analogue of the result $V(X) \ge \frac{1}{3}\{E(X)\}^2$ for X following a unimodal continuous distribution with mode at 0 given by Johnson & Rogers (1951). In particular, a univariate distribution with decreasing p.m.f. on $\mathbb{N}_0$ is guaranteed to be overdispersed if its mean is greater than 2.

Schur constant discrete distributions are direct analogues of the continuous distributions underlying Archimedean copulas (e.g. Nelsen, 2006). Those distributions also have survival functions of the form (17) which, along with their densities, are constant on planes of the form $r_1+\cdots+r_d = k > 0$ and have equal continous $(d-1)$-monotone marginal distributions on $\mathbb{R}^+$, the analogue of relationship (14) being the so-called Williamson transform (see especially McNeil & Nešlehová, 2009). Independence corresponds to exponential marginals.

4.2 Special Case: The Multivariate Generalized Waring Distribution

The multivariate generalized Waring distribution (Xekalaki, 1984, 1986) is the multivariate discrete Liouville distribution of this section with general Dirichlet-multinomial sharing distribution and beta-negative binomial (or generalized Waring) distribution for the sum T. We will say no more about this distribution here partly because the distribution is well known and partly because it is a case where the distributions of T and $\mathcal{M}_{d-1}|T = t$ have a parameter in common (one of the parameters of the beta mixing distribution is $\alpha_1 + \cdots + \alpha_d$) so that there is not the inferentially desirable separation between sum and share parameters in this case.

5. Concluding Remarks

The findings of this article expand on a sum and share decomposition to model $d-$variate discrete distributions and more specifically multivariate count data that fall into d distinct categories. From a simple Poisson mixture model for the total T with mixing density ℓ and a sharing distribution mechanism $T = M_1 + \cdots + M_d$ with $M_1, \ldots, M_{d-1}|T, U_1, \ldots, U_{d-1}$ multinomially distributed with $U_1, \ldots, U_{d-1} \sim h$, a rich ensemble of properties, examples and relationships arises. As a main example, in further studying the case of a negative binomial sum and Pólya shares, we obtained a seemingly new model as the joint distribution of $(M_1, \ldots, M_d)$, previously arising in the bivariate case as a Bayesian predictive distribution (Laurent, 2012).

We have addressed the equivalent scheme for generating the distributions above consisting in decomposing $\lambda = \sum_i R_i$ with $R_i = \lambda U_i$ and the U_i as above. This yields $M_i|R_1, \ldots R_d$, $i = 1, ..., d$, as independently distributed Poisson(R_i). Thus, as is well illustrated by the identity $\text{Cov}(M_i, M_j) = \text{Cov}(R_i, R_j)$ (Section 2), the dependence structure of the discrete M_i's is induced by that of the continuously distributed R_i's, and vice versa.

Finally, for other choices of the distribution of T, continuous multivariate Liouville distributions emerge for the distribution of the R_i's, as well as discrete analogues for the distribution of the M_i's. Moreover, the latter include the important special case of Schur constant distributions (Castaner et al. 2015) which are expanded upon in Section 4.1. Consideration of the correlations in such distributions led us to a variance-mean inequality for univariate discrete distributions with decreasing probability mass functions (the distributions' univariate marginals).

In summary, we feel that our findings provide considerable insight and appealing analytics for generating and understanding multivariate discrete distributions via sum and share decompositions.

Appendix A: Marginal P.M.F. Associated with (4)

One can recover the marginal p.m.f.'s through their relationship with the binomial moments:

$$\mathbb{P}(M_i = m_i) = \sum_{j=m_i}^{\infty} (-1)^{j-m_i} \binom{j}{m_i} \mathbb{E}\binom{M_i}{j}.$$

This gives

$$\begin{aligned}
\mathbb{P}(M_i = m_i) &= \sum_{j=m_i}^{\infty} (-1)^{j-m_i} \binom{j}{m_i} \frac{(a)_j\,(\alpha_i)_j}{(\alpha_\bullet)_j\, j!\, b^j} \\
&= \frac{1}{m_i!\, b^{m_i}} \sum_{k=0}^{\infty} \frac{(a+m_i)_k\,(\alpha_i+m_i)_k}{(\alpha_\bullet+m_i)_k\, k!} \left(-\frac{1}{b}\right)^k \\
&= \frac{(a)_{m_i}\,(\alpha_i)_{m_i}}{(\alpha_\bullet)_{m_i}\, m_i!\, b^{m_i}}\, {}_2F_1\left(a+m_i, \alpha_i+m_i; \alpha_\bullet+m_i; -\frac{1}{b}\right) \\
&= \frac{b^a}{(1+b)^{a+m_i}} \frac{(a)_{m_i}\,(\alpha_i)_{m_i}}{(\alpha_\bullet)_{m_i}\, m_i!\, b^{m_i}}\, {}_2F_1\left(a+m_i, \alpha_\bullet-\alpha_i; \alpha_\bullet+m_i; \frac{1}{1+b}\right),
\end{aligned}$$

using a standard transformation formula for the hypergeometric function.

Appendix B: Product Binomial Moments for Schur Constant Distributions

$$\begin{aligned}
E\left\{\prod_{i=1}^{d}\binom{M_i}{k_i}\right\} &= \sum_{m_1=k_1}^{\infty}\cdots\sum_{m_d=k_d}^{\infty}\binom{m_1}{k_1}\cdots\binom{m_d}{k_d}\frac{p_T(m_1+\cdots+m_d)}{\binom{m_1+\cdots+m_d+d-1}{d-1}} \\
&= \sum_{m_1=k_1}^{\infty}\cdots\sum_{m_{d-1}=k_{d-1}}^{\infty}\binom{m_1}{k_1}\cdots\binom{m_{d-1}}{k_{d-1}} \\
&\qquad \times \sum_{t=m_1+\cdots+m_{d-1}+k_d}^{\infty}\binom{t-m_1-\cdots-m_{d-1}}{k_d}\frac{p_T(t)}{\binom{t+d-1}{d-1}}.
\end{aligned}$$

Using, for example, (3.3) of Gould (1972), it is the case that

$$\sum_{m_1=k_1}^{t-m_2-\cdots-m_{d-1}-k_d}\binom{m_1}{k_1}\binom{t-m_2-\cdots-m_{d-1}-m_1}{k_d} = \binom{t-m_2-\cdots-m_{d-1}+1}{k_1+k_d+1},$$

$$\begin{aligned}
&\sum_{m_i=k_i}^{t-m_{i+1}-\cdots-m_{d-1}-k_1-\cdots-k_{i-1}-k_d}\binom{m_i}{k_i}\binom{t-m_i-\cdots-m_{d-1}+i-1}{k_1+\cdots+k_{i-1}+k_d+i-1} \\
&\qquad = \binom{t-m_{i+1}-\cdots-m_{d-1}+i}{k_1+\cdots+k_i+k_d+i}, \qquad 2\le i\le d-2,
\end{aligned}$$

and

$$\sum_{m_{d-1}=k_{d-1}}^{t-k_1-\cdots-k_{d-2}-k_d}\binom{m_{d-1}}{k_{d-1}}\binom{t-m_{d-1}+d-2}{k_1+\cdots+k_{d-2}+k_d+d-2} = \binom{t+d-1}{k_1+\cdots+k_d+d-1}.$$

It then follows that

$$
\begin{aligned}
E\left\{\prod_{i=1}^{d}\binom{M_i}{k_i}\right\} &= \sum_{t=k_1+\cdots+k_d}^{\infty}\binom{t+d-1}{k_1+\cdots+k_d+d-1}\frac{p_T(t)}{\binom{t+d-1}{d-1}} \\
&= \frac{1}{\binom{k_1+\cdots+k_d+d-1}{d-1}}\sum_{t=k_1+\cdots+k_d}^{\infty}\binom{t}{k_1+\cdots+k_d}p_T(t)
\end{aligned}
$$

as required.

References

Ait Aoudia, D. & Marchand, É. (2014) On a simple construction of a bivariate probability function with a common marginal. *Amer. Statist.*, 68, 170–173.

Ait Aoudia, D., Marchand, É. & Perron, F. (2016) Counts of Bernoulli success strings in a multivariate framework. *Statist. Probab. Lett.,* 119, 1–10.

Bhattacharya, S., Pradhan, B. & Kundu, D. (2014) Analysis of hybrid censored competing risks data. *Statist.*, 48, 1138–1154.

Castañer, A., Claramunt, M.M., Lefèvre, C. & Loisel, S. (2015) Discrete Schur-constant models. *J. Multivariate Anal.*, 140, 343–362.

Chee, C.S. & Wang, Y. (2016) Nonparametric estimation of species richness using discrete k-monotone distributions. *Comput. Statist. Data Anal.,* 93, 107–118.

Goodman, L.A. (1969) How to ransack social mobility tables and other kinds of cross-classification tables. *Amer. J. Sociol.,* 75, 1–39.

Goodman, L.A. (1985) The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetric models for contingency tables with or without missing entries. *Ann. Statist.*, 13, 10–69.

Gould, H.W. (1972) *Combinatorial Identities.* Morgantown Printing and Binding Co.

Gupta, R.D. & Richards, D.S.P. (1987) Multivariate Liouville distributions. *J. Multivariate Anal.*, 23, 233–256.

Johnson, N.L., Kemp, A.W & Kotz, S. (2005) *Univariate Discrete Distributions*, Third Edition. Wiley.

Johnson, N.L. & Rogers, C.A. (1951) The moment problem for unimodal distributions. *Ann. Math. Statist.*, 22, 433–439.

Laurent, S. (2012) Some Poisson mixtures distributions with a hyperscale parameter. *Braz. J. Probab. Statist.*, 26, 265–278.

Lingappaiah, G.S. (1984) Discrete generalized Liouville-type distribution and related multivariate distributions. *Trab. Estadíst. Invest. Operativa*, 35, 319–330.

McNeil, A.J. & Nešlehová, J. (2009) Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Ann. Statist.,* 37, 3059–3097.

McNeil, A.J. & Nešlehová, J. (2010) From Archimedean to Liouville copulas. *J. Multivariate Anal.*, 101, 1772–1790.

Nelsen, R.B. (2006) *An Introduction to Copulas*, Second Edition. Springer.

Peña, E.A. & Gupta, A.K. (1990) Bayes estimation for the Marshall-Olkin exponential distribution. *J. Roy. Statist. Soc. Ser. B*, 52, 379–389.

Xekalaki, E. (1984) The bivariate generalized Waring distribution and its application to accident theory. *J. Roy. Statist. Soc. Ser. A*, 147, 488–498.

Xekalaki, E. (1986) The multivariate generalized Waring distribution. *Comm. Statist. Theor. Meth.*, 15, 1047–1064.

Yule, G.U. & Kendall, M.G. (1950) *An Introduction to the Theory of Statistics*. Hafner.