Multivariate discrete distributions via sums and shares

Journal Item

Multivariate discrete distributions via sums and shares

M.C. Jones[a], Éric Marchand[b]

[a]School of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom
[b]Département de mathématiques, Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1

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 which we present two contrasting applications. 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. Analogies and interactions with related continuous distributions are to the fore throughout.

Keywords: Liouville distribution, Multinomial mixture, Poisson mixture, Pólya distribution, Schur constant distribution.
2010 MSC: Primary 62E15, Secondary 62H05

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 $M_1, \ldots, M_d \in \mathbb{N}_0$. For convenience, write $\mathcal{M}_q = (M_1, \ldots, M_q)$, $q = d - 1$ 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:

i) let the sum T have a distribution with probability mass function (p.m.f.) $p_T(t)$, $t \in \mathbb{N}_0$;

ii) conditionally on $T = t$, share t out between values for $\mathcal{M}_d$, i.e., let $\mathcal{M}_{d-1}|T = t$ have a distribution with p.m.f. $b_{[t]}(m_1, \ldots, m_{d-1})$ on the discrete simplex defined by $\mathcal{M}_{d-1} \in \{0, \ldots, t\}^{d-1}$ such that $M_1 + \cdots + M_{d-1} \in \{0, \ldots, t\}$.

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

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

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

*M.C. Jones. Email address: `m.c.jones@open.ac.uk`

p.m.f. $b_{[t]}(m_1, \ldots, m_{d-1})$ 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 $r_1 = \lambda u_1, \ldots, r_d = \lambda u_d$, where λ is the parameter of the Poisson distribution and $u_1, \ldots, u_d$ 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 $U_1, \ldots, U_d \in (0, 1)$ such that $U_1 + \cdots + U_d = 1$. The resulting joint distributions are considered in general terms in Section 2.

We then specialize again by making the natural choices of Λ following a gamma distribution (so that T is negative binomial) and of $U_1, \ldots, U_d$ 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 [16] 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. Inferential issues are considered briefly in Section 3.2 before we present two rather different illustrative applications of this model in Sections 3.3 and 3.1.

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. [8], 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 $b_{[t]}(m_1, \ldots, m_{d-1})$ to be the p.m.f. of a *multinomial mixture* distribution, viz.

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

where

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

Next we take $p_T(t)$ to be the p.m.f. of a *Poisson mixture* distribution, viz.

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

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

$$p(m_1, \ldots, m_d) = \int \cdots \int_{0<u_1+\cdots+u_{d-1}<1} \frac{(m_1 + \cdots + m_d)!}{m_1! \cdots m_d!} u_1^{m_1} \cdots u_{d-1}^{m_{d-1}} (1 - u_1 - \cdots - u_{d-1})^{m_d} h(u_1, \ldots, u_{d-1}) du_1 \ldots du_{d-1}$$
$$\times \int_0^\infty \frac{e^{-\lambda} \lambda^{m_1+\cdots+m_d}}{(m_1 + \cdots + m_d)!} \ell(\lambda) d\lambda, \quad (2)$$

for all $m_1, \ldots, m_d \in \mathbb{N}_0^d$. Alternatively, one can think of (2) as the result of mixing independent Poisson distributions with parameters $r_1 = \lambda u_1, \ldots, r_d = \lambda u_d$ over the distribution of $R_1 = \Lambda U_1, \ldots, R_d = \Lambda U_d$, where $(R_1, \ldots, R_d)$ follow the continuous analog of (1) in which $\Lambda = R_1 + \cdots + R_d$ plays the role of T and $(R_1, \ldots, R_{d-1})$ plays the role of $\mathcal{M}_{d-1}$, viz.

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

for all $r_1, \ldots, r_d > 0$.

The marginal distributions for $M_1, \ldots, M_d$ are Poisson mixture distributions but the marginal distributions of $R_1, \ldots, R_d$, and hence of $M_1, \ldots, M_d$, are not tractable in general. The moments of $M_1, \ldots, M_d$ are readily available in terms of those of Λ and $U_1, \ldots, U_d$, however. In particular, $\mathrm{E}(M_i) = \mathrm{E}(\Lambda)\mathrm{E}(U_i)$ and

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

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

$$\mathrm{cov}(M_i, M_j) = \mathrm{cov}(R_i, R_j).$$

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

$$\mathrm{cov}(M_i, M_j) = \mathrm{var}(\Lambda)\mathrm{E}(U_i U_j) + \{\mathrm{E}(\Lambda)\}^2 \mathrm{cov}(U_i, U_j).$$

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

$$\mathrm{cov}(M_1, M_2) = \mathrm{var}(\Lambda)\mathrm{E}\{U(1-U)\} - \{\mathrm{E}(\Lambda)\}^2 \mathrm{var}(U),$$

where $U \equiv U_1 \sim H$ on the interval $(0,1)$. Clearly, this covariance is always negative for degenerate Λ (i.e., T is Poisson distributed) and non-degenerate U. In contrast, 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

$$U \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_d), \quad \Lambda \sim \mathrm{Gamma}(a, b),$$

$\alpha_1, \ldots, \alpha_d, a, b > 0$; 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

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

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

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

When $d = 2$, this is the Bailey distribution of Laurent [16] and the distribution in (4) thus represents its multivariate extension.

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

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

for all $r_1, \ldots, r_d > 0$. Of course, $R_1, \ldots, R_d | \Lambda = \lambda \sim \lambda \times \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_d)$ and $\Lambda = R_1 + \cdots + R_d \sim \mathrm{Gamma}(a, b)$. 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. [7] when $d = 2$ and by Peña & Gupta [21] when $d = 3$. It is a particular continuous multivariate Liouville distribution [13].

Moments are readily available. Inserting the moments of the gamma and Dirichlet distributions into the formulas in Section 2, we find that, for $i \in \{1, \ldots, d\}$,

$$\mathrm{E}(M_i) = \frac{a\,\alpha_i}{b\,\alpha_\bullet}, \quad \mathrm{var}(M_i) = \frac{a\alpha_i}{b^2\,\alpha_\bullet^2(1+\alpha_\bullet)} [\alpha_\bullet\{a + 1 + (1+\alpha_\bullet)b\} + (\alpha_\bullet - a)\alpha_i] \tag{6}$$

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

$$\mathrm{cov}(M_i, M_j) = \frac{a(\alpha_\bullet - a)\alpha_i \alpha_j}{b^2\,\alpha_\bullet^2(1+\alpha_\bullet)}. \tag{7}$$

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 random variables distributed as $\mathrm{Gamma}(\alpha_1, b), \ldots, \mathrm{Gamma}(\alpha_d, b)$, and hence (4) to the distribution of independent random variables $\mathrm{NegativeBinomial}(\alpha_1, \theta), \ldots, \mathrm{NegativeBinomial}(\alpha_d, \theta)$.

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

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

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. Namely, for all $i \in \{1, \ldots, d\}$,

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

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

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

Here, ${}_2F_1$ is the Gauss hypergeometric function. A closely related expression was given when $d = 2$ in [16].

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 [10, 11, 25]. For $(m_1, m_2) \in \mathbb{N}_0^2$, define

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

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

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

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 $\ln[\alpha_\bullet(a+1)/\{a(\alpha_\bullet + 1)\}]$ and tends to zero as $t \to \infty$.

3.1. Special cases

3.1.1. Special Case 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. [8] with negative binomial T, which has

$$p(m_1, \ldots, m_d) = \frac{(a)_{m_1+\cdots+m_d}}{(d)_{m_1+\cdots+m_d}} \, \theta^a \, (1-\theta)^{m_1+\cdots+m_d} \tag{8}$$

for all $m_1, \ldots, m_d \in \mathbb{N}_0^d$. Notice that (8) depends on $m_1, \ldots, m_d$ only through the sum $m_1 + \cdots + m_d$. Similarly, the underlying distribution of $(R_1, \ldots, R_d)$ has density

$$f(r_1, \ldots, r_d) \; = \; \frac{b^a \Gamma(d)}{\Gamma(a)} (r_1 + \cdots + r_d)^{a-d} \, e^{-b(r_1+\cdots+r_d)} \tag{9}$$

for all $r_1, \ldots, r_d > 0$, which depends on $r_1, \ldots, r_d$ only through $r_1 + \cdots + r_d$. Immediately, if $a = d$, (8) reduces to the distribution of d independent geometric(θ) random variables and (9) to the distribution of d independent exponential(b) random variables, respectively. Distribution (9) underlies the 'gamma-simplex copula' of McNeil & Nešlehová [19] in a sense to be explained in a more general discussion of distributions with uniform shares in Section 4.1 to follow.

3.1.2. Special Case 2: $\alpha_1, \ldots, \alpha_d \to \infty$ such that $\alpha_i/(\alpha_1 + \cdots + \alpha_d) \to \phi_i$ for all $i \in \{1, \ldots, d\}$.

The Dirichlet mixing distribution becomes degenerate at values $0\phi_1, \ldots, \phi_d \in (0, 1)$ such that $\phi_1 + \cdots + \phi_d = 1$, so the Dirichlet-multinomial distribution reduces to the multinomial distribution and we have

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

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

$$\operatorname{corr}(M_i, M_j) = \sqrt{\frac{\phi_i}{\phi_i + b} \frac{\phi_j}{\phi_j + b}} > 0,$$

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 $d = 2$, we have, for all $m_1, m_2 \in \mathbb{N}_0^2$,

$$p(m_1, m_2) = \frac{(a)_{m_1+m_2}}{m_1!\, m_2!} \theta^a q_1^{m_1} q_2^{m_2}, \tag{10}$$

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

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

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

3.2. Inference

Let $(m_{11}, \ldots, m_{d1}), \ldots, (m_{1n}, \ldots, m_{dn})$ be a sample of independent observations taken from the distribution with density (4); also let $t_j = m_{1j} + \cdots + m_{dj}$, with $j \in \{1, \ldots, n\}$, be the sample totals.

Likelihood inference for the parameters a, θ associated with the distribution of Λ and for the parameters $\alpha_1, \ldots, \alpha_d$ 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 $\alpha_1, \ldots, \alpha_d$.

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. [14]. 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 $\alpha_1, \ldots, \alpha_d$ 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

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

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 the αs being positive.

A full exploration of the many inferential issues, both frequentist and Bayesian, for the models of this paper in their full breadth and generality is deferred to further work. In the next two sections, we present two, rather different, applications of model (4) when $d = 2$.

3.3. *Application: Data on shunter accidents*

We consider a historical bivariate data set due to Adelstein ([2], Table 11A) (also Arbous & Kerrich [5], Table I) consisting of (m_1, m_2) pairs denoting the numbers of accidents for the periods 1937–1942 and 1943–1947, respectively, for each of $n = 122$ (human) shunters on South African Railways. See also [9] and [22] and the references therein. Adelstein [2] showed that a Poisson distribution was an inadequate model for the total number of accidents per shunter, while a negative binomial distribution was adequate. (Dependent) Poisson distributions are, perhaps arguably in the first period, adequate models for the marginals, though negative binomial distributions might be preferred for them, too. Fitting a model like ours with a negative binomial total but less directly specified marginals therefore seems like an interesting exercise to perform. Specifically, we fit model (4) which, in addition to a negative binomial sum, employs a bivariate Pólya sharing distribution.

ML parameter estimates for model (4) were evaluated numerically yielding

$$\hat{a} = 3.419, \quad \hat{\theta} = 0.604, \qquad \hat{\alpha}_1 = 38.4, \quad \hat{\alpha}_2 = 50.0.$$

The above estimates for a and θ were obtained by focussing on solving the score equations, while a plot of the likelihood on a fine grid with $\alpha_1 \le 50, \alpha_2 \le 50$ yielded the estimates for α_1 and α_2. The precise values of the α_is have a negligible effect on the fitted model when they are large, and relate to Special Case 2 in Section 3.1, as discussed in the following paragraph. Our model fits the data well (as do a number of other models; see Table 9 of Sellers et al. [22]): see Table 1 for a comparison of observed and expected numbers of accidents; a chi-squared test based on an admittedly arbitrary combination of cells with small expected frequencies led to a test statistic smaller than the degrees-of-freedom of the asymptotic chi-squared null distribution. The empirical correlation is 0.29; the estimated correlation in our fitted model is 0.23.

So, how is the negative binomial distribution of the sum shared out by our model? Well, both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are large. For large α_1, α_2, p.m.f. (4) with $d = 2$ is approximated by (10) with $\phi_1 = \alpha_1/\alpha_\bullet$ and $\phi_2 = \alpha_2/\alpha_\bullet$, i.e.,

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

Table 1: Observed (O) and expected (under our fitted model; E) numbers of accidents among $n = 122$ shunters.

$m_1 \downarrow$ $m_2 \rightarrow$		0	1	2	3	4	5	6	7+	Total
0	O	21	18	8	2	1	0	0	0	50
	E	21.71	16.65	8.32	3.43	1.27	0.44	0.14	0.07	52.03
1	O	13	14	10	1	4	1	0	0	43
	E	12.78	12.52	7.59	3.66	1.55	0.60	0.22	0.11	39.03
2	O	4	5	4	2	1	0	1	0	17
	E	4.93	5.86	4.16	2.30	1.09	0.46	0.18	0.11	19.09
3	O	2	1	3	2	0	1	0	0	9
	E	1.58	2.20	1.78	1.11	0.58	0.27	0.12	0.08	7.72
4	O	0	0	1	1	0	0	0	0	2
	E	0.45	0.72	0.66	0.45	0.26	0.13	0.06	0.04	2.77
5	O	0	0	0	0	0	0	0	0	0
	E	0.12	0.22	0.22	0.17	0.10	0.06	0.03	0.02	0.94
6	O	0	0	0	0	0	0	0	0	0
	E	0.03	0.06	0.07	0.06	0.04	0.02	0.01	0.01	0.30
7+	O	0	1	0	0	0	0	0	0	1
	E	0.01	0.02	0.03	0.03	0.02	0.01	0.00	0.00	0.12
Total	O	40	39	26	8	6	2	1	0	122
	E	41.61	38.25	22.83	11.21	4.91	1.99	0.76	0.44	

This is the basic bivariate negative binomial distribution produced by mixing independent Poissons with the same mean over a gamma distribution for that mean, which was utilized by Arbous & Kerrich [5] for these data. It is still the case that $T \sim \text{NegativeBinomial}(a, \theta)$, but we also have here that $M_1 \sim \text{NegativeBinomial}(a, \theta(\alpha_1+\alpha_2)/(\alpha_1+\theta\alpha_2))$ and $M_2 \sim \text{NegativeBinomial}\,(a, \theta(\alpha_1+\alpha_2)/(\alpha_2+\theta\alpha_1))$; inter alia, $\mathrm{E}(M_1) = \alpha_1\,\mathrm{E}(T)/(\alpha_1+\alpha_2)$ and $\mathrm{E}(M_2) = \alpha_2\,\mathrm{E}(T)/(\alpha_1+\alpha_2)$. Our model is broadly similar, but not the same; either model provides an adequate description of the data. Given the adequacy of these models based on basic distributional ingredients, it is unclear what value is added by other, more exotic, distributions that have also been fitted to these data.

3.4. Application: Karl Broman's socks

Sum and shares distributions offer potentially useful benchmarks in eliciting prior distributions for discrete-valued parameters, such as population sizes. Here is an illustration. Bååth [6] addressed the following problem as tweeted by Karl Broman:

> Given a sample of eleven orphan (that is, unpaired) socks, how many socks does Karl Broman have in total?

A Bayesian model, describing uncertainty with regards to the total number of socks, as well as the proportions of socks in pairs and orphan socks, is put forth by Bååth, and followed by the implementation of an approximate Bayesian computation algorithm to obtain posterior distributions, such as for the total number of socks. We expand here on the prior construction, which can naturally be linked to p.m.f. (4), and how related properties can enhance the prior construction and aid in analytical posterior computations, not requiring simulation.

Suppose that, initially, there are $T = M_1 + M_2$ pairs of socks and that M_2 socks are now missing. In other words, $2M_1 + M_2$ remain, M_2 of which are orphans and $2M_1$ of which are paired. As in [6], suppose that a prior distribution for (M_1, M_2) is elicited starting with a Negative Binomial T, $M_1, M_2|T = t, p$ distributed as $\text{Multinomial}(t, p, 1-p)$, and p as $\text{Beta}(\alpha_1, \alpha_2)$. This is exactly the distribution in Section 3 when $d = 2$, with p.m.f. given in (4). Posterior distributions are of interest for (M_1, M_2), as well as for the total number of socks $S = 2M_1 + M_2$.

The distributional properties of (4) aid here in prior elicitation for this problem, as well as other problems where the unknowns $(M_1, \ldots, M_d)$ are population sizes modelled as sums and shares. As an illustration, the choices $a = \mu^2/(\sigma^2 - \mu)$, $\theta = \mu/\sigma^2$ are necessary for the prior mean μ and standard deviation σ to be matched by the prior on T. The choice of (α_1, α_2) may be elicited in a familiar fashion with a prior mean and variance for the Beta distributed p. Furthermore, the plausibility of an assigned choice for (α_1, α_2) can be further evaluated by the corresponding covariance or correlation, which are available from (6) and (7), assuming, of course, that one can elicit such quantities.

Turning now to the posterior evaluation, consider n draws from the S socks and denote as X the number of paired socks among those drawn. With the prior p.m.f. for (M_1, M_2) explicitly given by (4), one requires the probabilities $\Pr(X = x|M_1 = k, M_2 = \ell)$ to infer the posterior distribution for (M_1, M_2). Focussing on the case $x = 0$, that is, for n draws producing n orphan socks, a combinatorial argument yields

$$\Pr(X = 0 \mid M_1 = k, M_2 = \ell) = \sum_{j=(n-k)\vee 0}^{n\wedge\ell} \binom{\ell}{j}\binom{k}{n-j} 2^{n-j} \bigg/ \binom{2k+\ell}{n},$$

for $\ell + k \geq n$, and 0 otherwise. We thus obtain for $m_1 + m_2 \geq n$, that $\Pr(M_1 = m_1, M_2 = m_2|X = 0)$ is given by

$$\begin{aligned}\Pr(M_1 = m_1, M_2 = m_2|X = 0) &= \Pr(X = 0|M_1 = m_1, M_2 = m_2)\, p(m_1, m_2)/\Pr(X = 0)\\ &= \frac{1}{K}\frac{(a)_t}{(\alpha_1+\alpha_2)_t}\frac{(\alpha_1)_{m_1}(\alpha_2)_{m_2}}{(t+m_1-n+1)_n}\theta^a(1-\theta)^t \sum_{j=(n-m_1)\vee 0}^{n\wedge m_2} \frac{\binom{n}{j}\, 2^{n-j}}{(m_2-j)!(m_1-n+j)!},\end{aligned}$$

with $K = \Pr(X = 0)$ and $t = m_1 + m_2$. We do not have a closed-form formula for K, but we can approximate it rather nicely, by summing the above probabilities over the pairs (m_1, m_2) such that $m_1 + m_2 \geq n$. Finally, the p.m.f. for the total number of socks is obtained as

$$\Pr(S = y) = \sum_{0\leq k\leq\lfloor(y-1)/2\rfloor} \Pr(M_1 = k, M_2 = y - 2k)$$

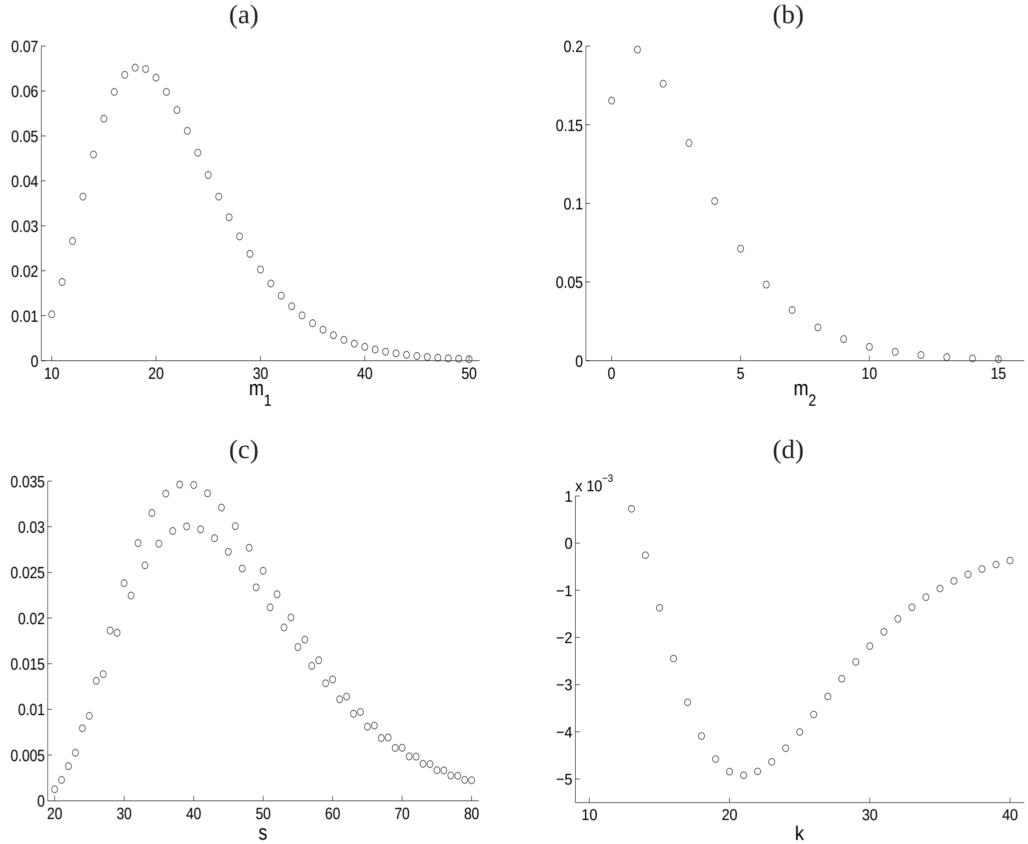

Figure 1: Posterior p.m.f.s for: (a) M_1; (b) M_2; and (c) S. Panel (d) is a plot of $\Pr(S = 2k + 1|X = 0) - \Pr(S = 2k|X = 0)$ as a function of k.

for all $y \in \{n, n + 1, \ldots\}$.

The above development capitalizes on an explicit form for the prior p.m.f. paired with a closed form for the likelihood function at $X = 0$. Furthermore, as motivated above, the prior is nicely elicited as a sum and shares distribution.

Now, Bååth [6] used a Negative Binomial prior for the total number of socks $2(M_1 + M_2)$ — which does not take account of the parity of this number — while we assume a Negative Binomial prior for the total number of pairs of socks $T = M_1 + M_2$. Matching the moments used by Bååth, consider a mean $\mu = 15$ and standard deviation of $\sigma = 7.5$ for T yielding $a = 60/11 \approx 5.455$ and $\theta = 4/15 \approx 0.267$. Furthermore, consider the choice of a Beta(15, 2) prior for the proportion of socks that are pairs, i.e., $\alpha_1 = 15, \alpha_2 = 2$. In terms of prior elicitation, we point out that this combination of $(a, \theta, \alpha_1, \alpha_2)$ implies a correlation of 0.2895.

For $n = 11$, Figure 1 gives the corresponding posterior p.m.f.s for: (a) the number of pairs of paired socks (M_1); (b) the number of orphan socks (M_2); and (c) the total number of socks ($S = 2M_1 + M_2$). Figure 1(d) is related to Figure 1(c). The observed $X = 0$ has pushed the posterior distributions away from 0. For instance, the prior expectations are $E(M_1) = 225/17 \approx 13.235$ and $E(M_2) = 30/17 \approx 1.765$, while the posterior expectations are $E(M_1|X = 0) \approx 21.3$ and $E(M_2|X = 0) \approx 2.82$. There is an unusual oscillation in the posterior p.m.f. of S with its parity determined by that of M_2, and a quite subtle behavior of $\Pr(M_1 \text{ is even}|M_2 = k, X = 0)$ as a function of k. See Figure 1(d) which represents the differences $\Pr(S = 2k + 1|X = 0) - \Pr(S = 2k|X = 0)$ as a function of k.

As it happens, the true collection of socks that Karl Broman had comprised $M_1 = 21$ pairs of socks, $M_2 = 3$ orphan socks, and hence $S = 2M_1 + M_2 = 45$ socks in total. The estimates given by the posterior means of M_1 and M_2, as well as the estimate $E(S|X = 0) \simeq 45.42$, are thus remarkably accurate.

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

$$p(m_1, \ldots, m_d) = \frac{(m_1 + \cdots + m_d)!}{m_1! \cdots m_d!} \frac{\prod_{i=1}^d (\alpha_i)_{m_i}}{(\alpha_\bullet)_{m_1+\cdots+m_d}} \; p_T(m_1 + \cdots + m_d). \tag{11}$$

Note that (11) can be written as

$$p(m_1, \ldots, m_d) = \frac{\prod_{i=1}^d (\alpha_i)_{m_i}}{m_1! \cdots m_d!} \mathcal{F}(m_1 + \cdots + m_d). \tag{12}$$

where $\mathcal{F}(t) = t!p_T(t)/(\alpha_\bullet)_t$. The form of (12) is a discrete analog of the continuous multivariate Liouville distribution [13], which has p.d.f.

$$f(r_1, \ldots, r_d) = \left(\prod_{i=1}^d r_i^{\alpha_i - 1} \right) \mathbb{F}(r_1 + \cdots + r_d), \tag{13}$$

for all $r_1, \ldots, r_d > 0$, or 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 $(R_1, \ldots, R_d)$ associated with (11) — the special case of (3) with Dirichlet h — has the form

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

which is indeed a continuous Liouville distribution of form (13). In the discrete case, our preferred formulation is (11) rather than (12) 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 [17] for the distribution having p.m.f. of form

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

Rearranging appropriately, this can be written

$$p(m_1, \ldots, m_d) = \frac{(m_1 + \cdots + m_d)!}{m_1! \cdots m_d!} \left\{ \prod_{i=1}^d \left(\frac{\theta_i}{\theta_1 + \cdots + \theta_d} \right)^{m_i} \right\} p_T(m_1 + \cdots + m_d). \tag{14}$$

This is none other than distribution (1) with general distribution for T and *multinomial*, Multinomial(t, $\theta_1/\sum_{i=1}^d \theta_i, \ldots, \theta_{d-1}/\sum_{i=1}^d \theta_i$), conditional distribution for $\mathcal{M}_{d-1}|T = t$. It follows that the new multivariate discrete Liouville distribution at (11) is a mixture over $\Theta_1/\sum_{j=1}^d \Theta_j, \ldots, \Theta_{d-1}/\sum_{j=1}^d \Theta_j \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_d)$ of Lingappaiah's multivariate discrete Liouville distribution at (14), and hence is more general than it.

4.1. *Special case: Schur constant distributions*

The special cases of our multivariate discrete Liouville distribution (11) with $\alpha_1 = \cdots = \alpha_d = 1$ are the Schur constant distributions of Castañer et al. [8]. 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 analog with the continuous case. All but two of the observations in this subsection can also be found in [8]; the two new observations are both in the fourth paragraph of this subsection. Castañer et al. provide various applications in insurance for a related counting process.

From (11), the p.m.f. in this case has the simple form

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

this reduces in the bivariate case (see also Ait Aoudia & Marchand [3]) 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 $k \in \mathbb{N}_0$ 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

$$p_S(m_i) = (d-1)\sum_{t=m_i}^{\infty} \frac{(t-m_i+1)_{d-2}}{(t+1)_{d-1}} p_T(t) \tag{15}$$

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

$$p_T(t) = \binom{t+d-1}{d-1} \sum_{j=0}^{d-1} \binom{d-1}{j} (-1)^j p(t+j). \tag{16}$$

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

$$p(m_1, \ldots, m_d) = \sum_{j=0}^{d-1} \binom{d-1}{j} (-1)^j p_S(m_1 + \cdots + m_d + j) \tag{17}$$

and the survival function simply as

$$\Pr(M_1 \geq m_1, \ldots, M_d \geq m_d) = \overline{P}_S(m_1 + \cdots + m_d), \tag{18}$$

where $\overline{P}_S$ is the survival function associated with p_S. While p_T can be specified arbitrarily on $\mathbb{N}_0$, (16) and (17) show that p_S has to be a discrete $(d-1)$-monotone distribution on $\mathbb{N}_0$. Independence in Schur constant distributions corresponds to p_S being geometric. An interesting example of (17) occurs for p_S a Poisson(α) p.m.f., where $\alpha \in (0, 1]$. 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. [4].

The most convenient form for multivariate moments of Schur constant distributions — which appears not to be in [8] — is

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

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

$$\mathrm{cov}(M_i, M_j) = \frac{1}{d^2(d+1)}\left[d\,\mathrm{var}(T) - \{\mathrm{E}(T)\}^2 - d\mathrm{E}(T)\right] = \frac{1}{2}\left[\mathrm{var}(M_i) - \{\mathrm{E}(M_i)\}^2 - \mathrm{E}(M_i)\right].$$

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

$$\mathrm{var}(T) > \mathrm{E}(T)\,\{\mathrm{E}(T)/d + 1\},$$

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

$$\mathrm{var}(M_i) \geq \frac{1}{3}\,\mathrm{E}(M_i)\{\mathrm{E}(M_i) + 1\}.$$

This inequality is the discrete analog of the result $\mathrm{var}(X) \geq \{\mathrm{E}(X)\}^2/3$ for X following a unimodal continuous distribution with mode at 0 given by Johnson & Rogers [15]: it is essentially the case $a = 0$, $\alpha = 1$ of Theorem 3.1 of Abouammoh, Ali & Mashhour [1]. 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 analogs of the continuous distributions underlying Archimedean copulas; see, e.g., Nelsen [20]. Those distributions also have survival functions of the form (18) which, along with their densities, are constant on planes of the form $r_1 + \cdots + r_d = k > 0$ and have equal continuous $(d-1)$-monotone marginal distributions on $\mathbb{R}^+$, the analog of relationship (15) being the so-called Williamson transform; see especially McNeil & Nešlehová [18]. Independence corresponds to exponential marginals.

4.2. Special case: The multivariate generalized Waring distribution

The multivariate generalized Waring distribution [23, 24] 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 [16]. Two contrasting applications of the latter model were investigated.

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 $M_i|R_1, \ldots R_d$, with $i \in \{1, \ldots, d\}$, as independently distributed Poisson(R_i). Thus, as is well illustrated by the identity $\mathrm{cov}(M_i, M_j) = \mathrm{cov}(R_i, R_j)$ in Section 2, the dependence structure of the discrete M_is is induced by that of the continuously distributed R_is, and vice versa.

Finally, for other choices of the distribution of T, continuous multivariate Liouville distributions emerge for the distribution of the R_is, as well as discrete analogs for the distribution of the M_is. Moreover, the latter include the important special case of Schur constant distributions [8] which are expanded upon in Section 4.1. Consideration of the correlations in such distributions led us to a pre-existing but not well known 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.

Acknowledgments

We are grateful to Pankaj Bhagwat for numerical evaluations and for pointing out a computational mistake in an earlier version of Section 3.3. We also thank the Editor-in-Chief, an Associate Editor and two referees for their positive treatment of this article.

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:

$$\Pr(M_i = m_i) = \sum_{j=m_i}^{\infty} (-1)^{j-m_i} \binom{j}{m_i} \mathrm{E}\left\{\binom{M_i}{j}\right\}.$$

This gives

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

using a standard transformation formula for the hypergeometric function.

Appendix B: Product binomial moments for Schur constant distributions

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

Using, for example, (3.3) of Gould [12], it is the case that

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

that for all $i \in \{2, \ldots, d-2\}$,

$$\sum_{m_i=k_i}^{t-m_{i+1}-\cdots-m_{d-1}-k_1-\cdots-k_{i-1}-k_d}\binom{m_i}{k_i}\binom{t-m_i-\cdots-m_{d-1}+i-1}{k_1+\cdots+k_{i-1}+k_d+i-1} = \binom{t-m_{i+1}-\cdots-m_{d-1}+i}{k_1+\cdots+k_i+k_d+i},$$

and that

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

It then follows that

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

as required.

References

[1] A. Abouammoh, A. Ali, A. Mashhour, On characterizations and variance bounds of discrete α-unimodality, Statist. Pap. 35 (1994) 151–161.
[2] A. Adelstein, Accident proneness: A criticism of the concept based upon an analysis of shunters' accidents (with discussion), J. R. Stat. Soc. Ser. A 115 (1952) 354–410.
[3] D. Ait Aoudia, E. Marchand, On a simple construction of a bivariate probability function with a common marginal, Amer. Statist. 68 (2014) 170–173.
[4] D. Ait Aoudia, E. Marchand, F. Perron, Counts of Bernoulli success strings in a multivariate framework, Statist. Probab. Lett. 119 (2016) 1–10.
[5] A. Arbous, J. Kerrich, Accident statistics and the concept of accident-proneness, Biometrics 7 (1951) 340–432.
[6] R. Bååth, Tiny data, approximate Bayesian computation and the socks of Karl Broman, www.sumsar.net/blog/2014/10/tiny-data-and-the-socks-of-karl-broman (2014).
[7] S. Bhattacharya, B. Pradhan, D. Kundu, Analysis of hybrid censored competing risks data, Statist. 48 (2014) 1138–1154.
[8] A. Castañer, M. Claramunt, C. Lefèvre, S. Loisel, Discrete Schur-constant models, J. Multivariate Anal. 140 (2015) 343–362.
[9] F. Famoye, P. Consul, Bivariate generalized Poisson distribution with applications, Metrika 42 (1995) 127–138.
[10] L. Goodman, How to ransack social mobility tables and other kinds of cross-classification tables, Amer. J. Sociol. 75 (1969) 1–39.
[11] L. Goodman, 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 (1985) 10–69.
[12] H. Gould, Combinatorial Identities, Morgantown Printing and Binding Co, Morgantown, VA, 1972.
[13] R. Gupta, D. Richards, Multivariate Liouville distributions, J. Multivariate Anal. 23 (1987) 233–256.
[14] N.L. Johnson, A. Kemp, S. Kotz, Univariate Discrete Distributions, 3rd ed., Wiley, Hoboken, NJ, 2005.
[15] N.L. Johnson, C. Rogers, The moment problem for unimodal distributions, Ann. Math. Statist. 22 (1951) 433–439.
[16] S. Laurent, Some Poisson mixtures distributions with a hyperscale parameter, Braz. J. Probab. Statist. 26 (2012) 265–278.
[17] G. Lingappaiah, Discrete generalized Liouville-type distribution and related multivariate distributions, Trab. Estadíst. Invest. Operativa 35 (1984) 319–330.

[18] A.J. McNeil, J. Nešlehová, Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions, Ann. Statist. 37 (2009) 3059–3097.
[19] A.J. McNeil, J. Nešlehová, From Archimedean to Liouville copulas, J. Multivariate Anal. 101 (2010) 1772–1790.
[20] R.B. Nelsen, An Introduction to Copulas, 2nd ed., Springer, New York, 2006.
[21] E. Peña, A. Gupta, Bayes estimation for the Marshall–Olkin exponential distribution, J. R. Stat. Soc. Ser. B 52 (1990) 379–389.
[22] K. Sellers, D. Morris, N. Balakrishnan, Bivariate Conway–Maxwell–Poisson distribution: Formulation, properties, and inference, J. Multivariate Anal. 150 (2016) 152–168.
[23] E. Xekalaki, The bivariate generalized Waring distribution and its application to accident theory, J. R. Stat. Soc. Ser. A 147 (1984) 488–498.
[24] E. Xekalaki, The multivariate generalized Waring distribution, Commun. Statist. Theor. Meth. 15 (1986) 1047–1064.
[25] G. Yule, M.J. Kendall, An Introduction to the Theory of Statistics, Hafner, New York, 1950.