On uniform and α-monotone discrete distributions

Authors: M.C. Jones
– Department of Mathematics and Statistics, The Open University,
U.K. (m.c.jones@open.ac.uk)

Abstract:

• In this partly expository article, I am concerned with some simple yet fundamental aspects of discrete distributions that are either uniform or have α-monotone probability mass functions. In the univariate case, building on work of F.W. Steutel published in 1988, I look at Khintchine's theorem for discrete monotone distributions in terms of mixtures of discrete uniform distributions, along with similar results for discrete α-monotone distributions. In the multivariate case, I develop a new general family of multivariate discrete distributions with uniform marginal distributions associated with copulas and consider families of multivariate discrete distributions with α-monotone marginals associated with these.

Key-Words:

• *Khintchine's theorem; multivariate geometric distribution; multivariate discrete uniform distribution, multivariate Poisson distribution.*

AMS Subject Classification:

• Primary 62E10; Secondary 62H05.

1.   INTRODUCTION

In this partly expository article, I am concerned with some simple yet fundamental aspects of distributions on $\mathbb{N}_0 \equiv 0, 1, \ldots$, whose probability mass functions (p.m.f.'s) p are uniform or more generally monotone nonincreasing or even more generally α-monotone (see below), together with certain extensions of these distributions to $\mathbb{N}_0^d \equiv \mathbb{N}_0 \times \cdots \times \mathbb{N}_0$, especially $\mathbb{N}_0^2$, and subsets thereof. As a prime example of a univariate distribution with a non-uniform monotone nonincreasing p.m.f. — a 'monotone p.m.f.' for short — think of the geometric distribution; the Poisson distribution turns out to be an example of an α-monotone distribution.

The main topics to be considered in this article, by section, are:

§2   Khintchine's theorem for monotone distributions on $\mathbb{N}_0$, re-interpreted in terms of mixtures of discrete uniform distributions, and a consequent variance inequality for univariate discrete monotone distributions;

§3   a general family of multivariate discrete distributions with uniform marginal distributions associated in an attractive yet novel way with copulas;

§4   univariate α-monotone distributions on $\mathbb{N}_0$ which, for $0 < \alpha < 1$, are a 'stronger' subset of monotone distributions, and which are of interest for $\alpha > 1$ also, when they can be non-monotone and include many familiar distributions. Originally introduced by Steutel (1988), I pursue further interpretation and properties;

§5   families of multivariate discrete distributions with α-monotone marginals associated with the distributions of Sections 3 and 4. Their correlation structures are explicit and relatively straightforward.

Potential Bayesian applications of Khintchine's theorem for discrete distributions (§2) are to the provision of monotone prior distributions for discrete-valued parameters and of nonparametric priors for α-monotone discrete distributions (similar to e.g. Brunner & Lo, 1989, in the continuous case). Families of multivariate discrete distributions with separation between marginal and dependence parameters (§3 and especially §5) can, as in the continuous case, form good test-beds for simulation studies; in particular, as a referee suggests, the opportunity arises to simulate correlated discrete variables with a given correlation matrix and univariate margins. Distributions with monotone and especially α-monotone marginals can be used as models for appropriate data too, of course. I look briefly at alternative multivariate geometric and Poisson distributions to those in e.g. Davy & Rayner (1996) and Bermúdez & Karlis (2011), respectively, while alternatives to existing multivariate binomial (e.g. Westfall & Young, 1989) and multivariate negative binomial (e.g. Shi & Valdez, 2014) distributions are also readily available but not developed explicitly.

All mathematical manipulations made in this article have the major benefit of being simple and direct. As I go along, it will often be useful to point out analogies and connections with results for continuous data which have uniform or α-monotone probability density functions (p.d.f.'s) f on $\mathbb{R}^+$, and their multivariate extensions.

---

2. DISCRETE KHINTCHINE'S THEOREM

Let f be a monotone p.d.f. on $\mathbb{R}^+$. Then, the renowned Khintchine's Theorem (Khintchine, 1938, Feller, 1971) says that $X \sim f$ can be written as a uniform scale mixture, either as $X = UY$, where U and Y are independent, $U \sim \text{Uniform}(0,1)$ and $Y \sim G$ for some cumulative distribution function (c.d.f.) G on $\mathbb{R}^+$, or equivalently as $X|Y = y \sim \text{Uniform}(0, y)$, $Y \sim G$. If f is differentiable, then typically G has a p.d.f. g such that $g(x) = -xf'(x)$. (The distribution of Y is not absolutely continuous if f has support $(0, b)$ say, when $b < \infty$ and $f(b) > 0$; see Section 4.)

Implicit in Steutel's (1988) paper on "discrete α-monotonicity" — of which, more in Section 4 — is a corresponding result to Khintchine's theorem in the discrete case. (See also the earlier work of Medgyessy, 1972.) It is framed in terms of binomial thinning, as first proposed by Steutel and van Harn (1979). For values of $\theta \in [0, 1]$, the random variable $N_{m,\theta}$ is the binomially thinned version of the count $m \in \mathbb{N}_0$ if

$$N_{m,\theta} \equiv \theta \circ m \equiv \sum_{j=1}^{m} B_j$$

where the sum is understood to be zero if $m = 0$. Here, $B_1, ..., B_m$ are independent Bernoulli(θ) random variables. (Note that if $\theta = 1$, $N_{m,\theta} = m$ and if $\theta = 0$, $N_{m,\theta} = 0$.) A useful equivalent way of expressing $N_{m,\theta} = \theta \circ m$ is as

$$N_{m,\theta} = \theta \circ m \sim \text{Binomial}(m, \theta)$$

where Binomial$(0, \theta)$ is understood to be the degenerate distribution at zero.

The above is binomial thinning for fixed θ and m, extensions to which are to mix over distributions for their random variable versions, Θ and/or M. So, consider the distribution of $N = \Theta \circ M \sim p$ on $\mathbb{N}_0$ where $\Theta \sim h$ on $(0, 1)$, independently of $M \sim q$ on $\mathbb{N}_0$. This distribution can be expressed as

$$N|M = m \sim \text{BinMix}(m), \qquad M \sim q,$$

with the binomial mixture distribution 'BinMix' defined as follows: $N_m \equiv \Theta \circ m \sim$ BinMix(m) if

(2.1) $$N_m|\Theta = \theta \sim \text{Binomial}(m, \theta), \qquad \Theta \sim h.$$

Steutel's (1988) observation is that taking $\Theta \sim \text{Uniform}(0,1)$ is equivalent to p being a monotone p.m.f. on $\mathbb{N}_0$. I now note that in that case, where $h(\theta) = I(0 < \theta < 1)$ and $I(\cdot)$ denotes the indicator function,

$$N_m = \Theta \circ m \sim \text{Uniform}\{0, ..., m\},$$

that is, the binomial mixture distribution reduces to the uniform distribution on $\{0, ..., m\}$. To see this, note that, for each $x \in \{0, ..., m\}$,

$$\int_0^1 \binom{m}{x} \theta^x (1-\theta)^{m-x} \, d\theta = \binom{m}{x} B(x+1, m-x+1) = \frac{1}{m+1}$$

(here, $B(\cdot, \cdot)$ is the beta function). This is, of course, a very special case of the beta-binomial distribution (see Johnson, Kemp and Kotz, 2005, Section 6.9.2).

The discrete analogue of Khintchine's theorem can therefore be given most simply — and not unexpectedly given its continuous analogue — as a discrete uniform mixture, as in Result 2.1.

Result 2.1 *A p.m.f. p on $\mathbb{N}_0$ is monotone if and only if $N \sim p$ can be written as*

$$N|M = m \sim \text{Uniform}\{0, ..., m\}, \qquad M \sim q,$$

where q is any p.m.f. on $\mathbb{N}_0$. In fact, the p.m.f.s p and q are related by

$$(2.2) \qquad p(n) = \sum_{m=n}^{\infty} \frac{q(m)}{m+1}, \qquad q(m) = (m+1)\{p(m) - p(m+1)\}.$$

Also, the corresponding c.d.f.s P and Q are related by

$$Q(n) = P(n) - (n+1)p(n+1).$$

Example 2.1

(a) Let $N \sim \text{Geometric}(p)$, $0 < p < 1$, which has strictly decreasing p.m.f. In this case,
$$q(m) = (m+1)\, p^2 (1-p)^m,$$
that is, $M \sim \text{NegativeBinomial}(2, p)$, which is the distribution of the sum of two independent Geometric(p) random variables.

(b) Let $N \sim \text{Poisson}(\mu)$ with $0 < \mu \leq 1$. Then, p is monotone on $\mathbb{N}_0$, and Result 2.1 applies with

$$q(m) = (m+1-\mu)\, p(m).$$

One of a number of ways of interpreting q is that it is the distribution of $M_0 + B$ where $B \sim \text{Bernoulli}(\mu)$, independent of $M_0 \sim \text{Poisson}(\mu)$.

(c) Now let $M \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Then, N has the strictly decreasing p.m.f.

$$p(n) = \frac{e^{-\lambda}}{\lambda} \sum_{j=n+1}^{\infty} \frac{\lambda^j}{j!} = \frac{1}{\lambda} \Gamma(\lambda; n+1)$$

where $\Gamma(\cdot; \cdot)$ is the incomplete gamma function ratio. From (2.3) below, $\mathbb{E}(N) = \lambda/2$ and $\mathbb{V}(N) = \lambda(6 + \lambda)/12$, so p is overdispersed as well as decreasing.

(d) The distribution of part (c) is a special case of taking $q(m) = (m+1)\, r(m+1)/\mu_r$ where r is an arbitrary p.m.f. on $\mathbb{N}_0$ with finite mean μ_r. Then, $p(n) = \overline{R}(n)/\mu_r$ where $R(n) = P(R > n)$ and $R \sim r$, so p is clearly monotone.

(e) There is no distribution satisfying $p = q$. If there were, p must satisfy $p(m+1)/p(m) = m/(m+1)$, $m = 0, 1, ...$, and this was shown by Leo Katz in the 1940s not to correspond to a valid distribution (see Johnson et al., 2005, Section 2.3.1).

Either directly or as a consequence of more general results for mixed binomial thinning, it is easy to show that

$$(2.3) \qquad \mathbb{E}(N) = \mathbb{E}(M)/2, \qquad \mathbb{V}(N) = \left[4\mathbb{V}(M) + 2\mathbb{E}(M) + \{\mathbb{E}(M)\}^2\right]/12.$$

Since $\mathbb{V}(M) \geq 0$ and $\mathbb{E}(M) = 2\mathbb{E}(N)$, the following variance-mean inequality arises.

Result 2.2 *If N follows a monotone p.m.f. on $\mathbb{N}_0$, then*

$$\mathbb{V}(N) \geq \mathbb{E}(N)\{1 + \mathbb{E}(N)\}/3,$$

and any monotone distribution is overdispersed if $\mathbb{E}(N) > 2$.

This inequality and observation arose in Jones and Marchand (2019) from a different perspective. The inequality is the discrete analogue of the inequality $\mathbb{V}(X) \geq \{\mathbb{E}(X)\}^2/3$ of Johnson and Rogers (1951) in the continuous monotone case.

3.   MULTIVARIATE DISCRETE UNIFORM DISTRIBUTIONS

Write c and C for the p.d.f. and cumulative distribution function (c.d.f.) of an absolutely continuous copula on $(0, 1)^d$ (e.g. Nelsen, 2006, Joe, 1997, 2014). This section and the next can be seen as an investigation of a role for such multivariate continuous uniform distributions in providing the dependence properties of certain multivariate discrete distributions, starting in this section with

multivariate discrete distributions with discrete uniform marginal distributions, referred to from here on as multivariate discrete uniform distributions. Note that this is quite different from the use of a copula in conjunction with the discontinuous c.d.f.'s and quantile functions of discrete marginals, a common practice but with a number of "dangers and limitations", as discussed by Genest and Nešlehová (2007). That said, a multivariate discrete uniform distribution does *not* fulfil the same role for multivariate discrete distributions as a copula does for multivariate continuous distributions because univariate discrete c.d.f.'s, when considered as functions of their random variable, are not distributed as discrete uniforms i.e., if X has distribution F, and F is discrete, then $F(X)$ is not uniform. In contrast, $F(X)$ is (continuous) uniform when F is continuous.

The fact that a binomial distribution mixed over a continuous uniform distribution for its probability parameter is itself a discrete uniform distribution suggests that a multivariate discrete uniform distribution can be defined as the distribution of $(N_1, ..., N_d)$ on $\{0, ..., m_1\} \times \cdots \times \{0, ..., m_d\}$ such that

$$N_i | \Theta_i = \theta_i \sim \text{Binomial}(m_i, \theta_i) \quad \text{independently for} \quad i = 1, ..., d,$$
$$\Theta^{(d)} \equiv \{\Theta_1, ..., \Theta_d\} \sim c(\theta_1, ..., \theta_d).$$

The joint p.m.f. of $(N_1, ..., N_d)$ is

$$(3.1) \quad p_U(n_1, ..., n_d \mid m_1, ..., m_d) = \left\{ \prod_{i=1}^{d} \binom{m_i}{n_i} \right\}$$
$$\times \int_0^1 \cdots \int_0^1 \left\{ \prod_{i=1}^{d} \theta_i^{n_i} (1 - \theta_i)^{m_i - n_i} \right\} c(\theta_1, ..., \theta_d) \, d\theta_1 \, ... \, d\theta_d.$$

Its univariate marginal distributions are discrete uniform by construction because those of the copula are continuous uniform.

Moments of this construction are readily available and, in particular, correlations are determined by those of the copula as follows. Since $\text{Cov}(N_i, N_j | \Theta^{(d)} = \theta^{(d)}) = 0$, it is the case that

$$\text{Cov}(N_i, N_j) = \text{Cov}\{\mathbb{E}(N_i | \Theta^{(d)} = \theta^{(d)}), \mathbb{E}(N_j | \Theta^{(d)} = \theta^{(d)})\}$$
$$(3.2) \qquad\qquad = m_i m_j \text{Cov}(\Theta_i, \Theta_j).$$

Also, since $\mathbb{V}(N_i) = m_i(m_i + 2)/12$, $\mathbb{V}(N_j) = m_j(m_j + 2)/12$, it is the case that

$$\text{Corr}(N_i, N_j) = \frac{m_i m_j \text{Corr}(\Theta_i, \Theta_j)/12}{\sqrt{m_i(m_i + 2)m_j(m_j + 2)}/12}$$
$$(3.3) \qquad\qquad = \sqrt{\frac{m_i}{m_i + 2}} \sqrt{\frac{m_j}{m_j + 2}} \text{Corr}(\Theta_i, \Theta_j).$$

So, while the correlation of N_i and N_j has the same sign as that of Θ_i and Θ_j, it reduces to one-third that of the copula in the binary case, and increases, tending

to a factor of 1, as the marginal supports grow larger. Note that $\mathrm{Corr}(\Theta_i, \Theta_j)$ is Spearman's rho.

The existence of this simple relationship between discrete and continuous uniform correlations is a reason for preferring the current construction to discretisations of the copula, although the two can be very similar, as the following simple example shows.

Example 3.1

Consider the bivariate Farlie–Gumbel–Morgenstern (FGM) copula given by

$$C(u,v) = uv\{1 + \phi(1-u)(1-v)\}, \qquad c(u,v) = 1 + \phi(1-2u)(1-2v),$$

on $0 < u, v < 1$ with $-1 \leq \phi \leq 1$. Entering this into (3.1) when $d = 2$ gives

$$p_{FGM}(n_1, n_2) = \frac{1}{(m_1 + 1)(m_2 + 1)} \left\{ 1 + \phi \frac{(2n_1 - m_1)(2n_2 - m_2)}{(m_1 + 2)(m_2 + 2)} \right\};$$

its correlation, from (3.3) and e.g. Example 2.4 of Joe (1997), is

$$\sqrt{\frac{m_1}{m_1 + 2}} \sqrt{\frac{m_2}{m_2 + 2}} \frac{\phi}{3}.$$

A natural discretisation of any C in the bivariate case is

$$
\begin{aligned}
p'(n_1, n_2) = \ &C\left(\frac{n_1 + 1}{m_1 + 1}, \frac{n_2 + 1}{m_2 + 1}\right) + C\left(\frac{n_1}{m_1 + 1}, \frac{n_2}{m_2 + 1}\right) \\
&- C\left(\frac{n_1 + 1}{m_1 + 1}, \frac{n_2}{m_2 + 1}\right) - C\left(\frac{n_1}{m_1 + 1}, \frac{n_2 + 1}{m_2 + 1}\right)
\end{aligned}
$$

which turns out in the FGM case to equate to

$$(3.4) \qquad p'_{FGM}(n_1, n_2) = \frac{1}{(m_1 + 1)(m_2 + 1)} \left\{ 1 + \phi \frac{(2n_1 - m_1)(2n_2 - m_2)}{(m_1 + 1)(m_2 + 1)} \right\};$$

this differs just a little from p_{FGM}. The correlation associated with this model, calculated directly from (3.4), is similar to that of p_{FGM}, but a little larger; it is

$$\sqrt{\frac{m_1(m_1 + 2)}{(m_1 + 1)^2}} \sqrt{\frac{m_2(m_2 + 2)}{(m_2 + 1)^2}} \frac{\phi}{3}.$$

Formula (3.1) is a particular way of constructing multivariate distributions with uniform univariate marginals. If a multivariate discrete uniform distribution is specified by other means, there is not necessarily a copula leading to it via construction (3.1). Even when there is, as with copula discretisation, there is not

generally a unique copula leading to that distribution. The following simple, if extreme, example makes this clear.

Example 3.2

Let $d = 2$ and $m_1 = m_2 = 1$. In this case, the elements of the joint p.m.f. of (N_1, N_2) depend only on $p_U(0,0) \leq 1/2$, since $p_U(0,1) = \{1 - 2p_U(0,0)\}/2$, $p_U(1,0) = \{1 - 2p_U(0,0)\}/2$ and $p_U(1,1) = p_U(0,0)$. Write $\mathbb{E}_C$ for expectation under the copula. Then, from (3.1), we have

$$
\begin{aligned}
p_U(0,0) &= \mathbb{E}_C\{(1 - \Theta_1)(1 - \Theta_2)\} = \mathbb{E}_C(\Theta_1\Theta_2), \\
p_U(0,1) &= \mathbb{E}_C\{(1 - \Theta_1)\Theta_2\} = \tfrac{1}{2} - \mathbb{E}_C(\Theta_1\Theta_2), \\
p_U(1,0) &= \mathbb{E}_C\{\Theta_1(1 - \Theta_2)\} = \tfrac{1}{2} - \mathbb{E}_C(\Theta_1\Theta_2), \\
p_U(1,1) &= \mathbb{E}_C(\Theta_1\Theta_2).
\end{aligned}
$$

Therefore, any copula with $\mathbb{E}_C(\Theta_1\Theta_2) = p_U(0,0)$ will give rise to this bivariate binary uniform distribution. (In fact, the uniform marginals of the copula are not required for this argument: the copula can be replaced by any distribution on $(0,1) \times (0,1)$ with marginal means equal to $1/2$ and $\mathbb{E}(\Theta_1\Theta_2) = p_U(0,0)$.) However, the product moment requirement translates to $\mathrm{Corr}(\Theta_1, \Theta_2) = 12p_U(0,0) - 3$, which restricts the existence of such a mixing distribution to when $1/6 \leq p_U(0,0) \leq 1/3$.

4.   DISCRETE α-MONOTONICITY

I now return to the univariate domain. To set the scene, I first describe the situation in the continuous case. There, α-monotonicity was introduced by Olshen and Savage (1970) (see also Dharmadhikari and Joag-Dev, 1988, and Bertin, Cuculescu and Theodorescu, 1997): the distribution of a continuous random variable X is said to be α-monotone if and only if the distribution of X^α is monotone, $\alpha > 0$. Then, X can be written in the form $X = A_\alpha Y$ say, where $A_\alpha \sim$ Beta$(\alpha, 1)$, independently of $Y \sim g$ on $\mathbb{R}^+$, in a similar manner to Khintchine's theorem; equivalently, $X = U^{1/\alpha}Y$ where $U \sim$ Uniform$(0,1)$. Clearly $\alpha = 1$ corresponds to ordinary monotonicity. By construction, if a distribution is α_0-monotone say, then is it also α-monotone for all $\alpha > \alpha_0$. In particular, α-monotone distributions with $\alpha < 1$ are also ordinary monotone.

Providing an alternative view of an equivalent formulation of Abouammoh (1987/1988), Steutel (1988) first put forward discrete α-monotonicity in the following manner: for $\alpha > 0$, $N \sim p$ is discrete α-monotone if $N = A_\alpha \circ M_\alpha = U^{1/\alpha} \circ M_\alpha$, where $A_\alpha \sim$ Beta$(\alpha, 1)$, $U \sim$ Uniform$(0,1)$ and either of these is independent of $M_\alpha \sim q_\alpha$ on $\mathbb{N}_0$. The distribution of N can now be recognized, from Section 2, as being that of

$$(4.1) \qquad N|M_\alpha = m_\alpha \sim \text{BetaBinomial}(m_\alpha, \alpha, 1), \qquad M_\alpha \sim q_\alpha,$$

where the BetaBinomial($m_\alpha, \alpha, 1$) distribution has p.m.f.

$$(4.2) \qquad p_{BB1}(x) = \frac{\alpha\, m_\alpha!\, \Gamma(x + \alpha)}{x!\, \Gamma(m_\alpha + \alpha + 1)}$$

for $x \in \{0, ..., m_\alpha\}$. This is because now $h(\theta) = \alpha\theta^{\alpha-1}I(0 < \theta < 1)$ in (2.1) so that the binomial mixture distribution becomes

$$\alpha \int_0^1 \binom{m_\alpha}{x} \theta^{x+\alpha-1}(1-\theta)^{m_\alpha-x}\, d\theta = \alpha \binom{m_\alpha}{x} B(x + \alpha, m_\alpha - x + 1) = p_{BB1}(x).$$

(4.1) and (4.2) lead directly to confirmation of Steutel's (1988) formula

$$p(n) = \alpha \frac{\Gamma(n + \alpha)}{n!} \sum_{m=n}^{\infty} \frac{m!\, q_\alpha(m)}{\Gamma(m + \alpha + 1)}.$$

Steutel then observes that

$$(4.3) \qquad (n + \alpha)p(n) - (n + 1)p(n + 1) = \alpha q_\alpha(n)$$

from which it can be concluded that discrete α-monotonicity corresponds to p having the simple property that

$$(n + \alpha)p(n) \geq (n + 1)p(n + 1).$$

Here, the inequality is strict except when $q_\alpha(n) = 0$. The corresponding c.d.f.s P and Q_α are related by

$$\alpha Q_\alpha(n) = \alpha P(n) - (n + 1)p(n + 1),$$

which can be readily checked to give rise to (4.3). Comments above on continuous α-monotonicities for various values of α continue to hold in the discrete case.

It can be added that (4.3) can also be written

$$(4.4) \qquad q(n) = (1 - \alpha)p(n) + \alpha\, q_\alpha(n)$$

where $q = q_1$ is as at (2.2) in Result 2.1. To corroborate and interpret (4.4) in the case that $0 < \alpha \leq 1$, an alternative way of expressing α-monotonicity arises from writing $A_\alpha = UV$ where $U \sim \text{Uniform}(0, 1)$ independently of some appropriate V; this is possible when $0 < \alpha \leq 1$ because then $\text{Beta}(\alpha, 1)$ is monotone (non-increasing). Moreover, $\text{Beta}(\alpha, 1)$ is then a distribution on a finite interval with non-zero density at its upper endpoint. As signposted at the start of Section 2, the density of V is not $-xf'(x)$ if f has support $(0, b)$ and $f(b) > 0$; in fact,

$$V \sim \begin{cases} Y & \text{with probability } 1 - \alpha, \\ b & \text{with probability } \alpha, \end{cases}$$

where $Y \sim -xf'(x)/\{1 - f(b)\}$ on $(0, b)$. When $b = 1$ and $h(x) = \alpha x^{\alpha-1}$ so that $h(1) = \alpha$, it turns out that $-xh'(x)/\{1 - h(1)\} = h(x)$. In the case of discrete

α-monotonicity with $0 < \alpha \leq 1$, it follows that $N = A_\alpha \circ M = (UV) \circ M = U \circ (V \circ M)$ so that $N = U \circ N_0$ where $U \sim \text{Uniform}(0,1)$ and

$$N_0 \sim \begin{cases} N & \text{with probability } 1 - \alpha, \\ M & \text{with probability } \alpha, \end{cases}$$

which is immediately seen to be equivalent to (4.4).

By any of a number of routes, it can be shown that, for α-monotone distributions for any $\alpha > 0$,

$$\mathbb{E}(N) = \frac{\alpha \mathbb{E}(M_\alpha)}{\alpha + 1}, \quad \mathbb{V}(N) = \frac{\alpha \left[(\alpha + 1)^2 \mathbb{V}(M_\alpha) + (\alpha + 1)\mathbb{E}(M_\alpha) + \{\mathbb{E}(M_\alpha)\}^2 \right]}{(\alpha + 1)^2(\alpha + 2)}.$$

Since $\mathbb{V}(M_\alpha) \geq 0$ and $\mathbb{E}(M_\alpha) = (\alpha + 1)\mathbb{E}(N)/\alpha$, the following variance-mean inequality ensues.

Result 4.1 *If N follows an α-monotone p.m.f. on $\mathbb{N}_0$ for all $\alpha \geq \alpha_{min}$ say, then*

$$\mathbb{V}(N) \geq \frac{\mathbb{E}(N)\{\alpha_{min} + \mathbb{E}(N)\}}{\alpha_{min}(\alpha_{min} + 2)} \geq \frac{\mathbb{E}(N)\{\alpha + \mathbb{E}(N)\}}{\alpha(\alpha + 2)}.$$

The 'outside' inequality is essentially Theorem 3.1 of Abouammoh, Ali and Mashhour (1994) with $a = 0$ and Corollary 5.3.21 of Bertin et al. (1997). An α-monotone distribution is thereby guaranteed to be overdispersed if $\mathbb{E}(N) > \alpha_{min}(\alpha_{min} + 1)$. Of course, the outside inequality in Result 4.1 reduces to Result 2.2 when $\alpha = 1$.

Example 4.1

(a) $N \sim \text{Geometric}(p)$ is α-monotone for $\alpha \geq 1 - p \equiv \alpha_{min}$. Using (4.3), the corresponding p.m.f. of M_α is

$$q_\alpha(m) = \{(m + 1)p - (1 - \alpha)\}p(1 - p)^m/\alpha.$$

As noted in Example 2.1(a), $M_1 \sim \text{NegativeBinomial}(2, p)$ while it can now also be observed that M_{1-p} has the distribution of $M_1 + 1$. The dispersion inequality for α-monotone distributions confirms the overdispersion of the geometric distribution for all $0 < p < 1$.

(b)   Let $N \sim \text{Poisson}(\mu)$ with $0 < \mu \leq \alpha$. Then, the Poisson p.m.f. p is α-monotone on $\mathbb{N}_0$, and formula (4.3) applies to give

$$q_\alpha(m) = (m + \alpha - \mu)\, p(m)/\alpha.$$

Now, q_α is the distribution of $M_0 + B$ where $B \sim \text{Bernoulli}(\mu/\alpha)$, independent of $M_0 \sim \text{Poisson}(\mu)$. In particular, q_μ is the length-biased form of the Poisson distribution which is, in fact, the distribution of $M_0 + 1$. The dispersion inequality is, of course, not satisfied for any $\mu > 0$.

(c) Both of the above examples together with binomial and negative binomial distributions are covered by the Katz family, for which

$$(1 + n)\, p(n + 1) = (a + bn)\, p(n);$$

see Section 2.3.1 of Johnson et al. (2005). In general, $a > 0$ and $b < 1$, but α-monotonicity restricts the range of a to $0 < a \leq \alpha$. For any Katz distribution,

$$q_\alpha(m) = \{(\alpha - a) + (1 - b)m\}p(m)$$

reducing to $q_a(m) = (1 - b)mp(m)/a$ when $\alpha = a$. Let $K_{a,b}$ be a random variable following the Katz distribution with parameters a and b. Then, the latter length-biased distribution is also the distribution of $K_{a+b,b} + 1$. Since $\mathbb{E}(K_{a,b}) = a/(1-b)$ and $\mathbb{V}(K_{a,b}) = a/(1-b)^2$, the dispersion inequality yields overdispersion if $(a+1)(1-b) < 1$ while a Katz distribution is actually overdispersed for $0 < b < 1$. The general results reduce to those of part (a) when $a = b = 1 - p$ and part (b) when $a = \mu, b = 0$. They give results for the Binomial(k, p) distribution when $a = kp/(1 - p), b = -p/(1 - p)$ and to the NegativeBinomial(k, p) distribution when $a = k(1 - p), b = (1 - p)$.

5. MULTIVARIATE DISCRETE DISTRIBUTIONS WITH α-MONOTONE UNIVARIATE MARGINALS

Combining Sections 2 and 3 further, it is natural to develop discrete distributions on $\mathbb{N}_0^d$ with monotone univariate marginals as the distribution of $N^{(d)} \equiv (N_1, ..., N_d)$ where

$$
\begin{aligned}
N_i | M_i = m_i, \Theta_i = \theta_i &\sim \text{Binomial}(m_i, \theta_i) \quad \text{independently for} \quad i = 1, ..., d, \\
M^{(d)} \equiv \{M_1, ..., M_d\} &\sim q(m_1, ..., m_d), \\
\Theta^{(d)} \equiv \{\Theta_1, ..., \Theta_d\} &\sim c(\theta_1, ..., \theta_d),
\end{aligned}
$$

where q is now an arbitrary p.m.f. on $\mathbb{N}_0^d$ and $M^{(d)}$ is independent of $\Theta^{(d)}$. This is, of course, equivalent to mixing the multivariate discrete uniform distribution of Section 3 over q:

$$N^{(d)}|M^{(d)} = \{m_1, ..., m_d\} \sim p_U(n_1, ..., n_d|m_1, ..., m_d), \qquad M^{(d)} \sim q(m_1, ..., m_d).$$

To additionally fold in the work of Section 4, to provide multivariate discrete distributions with α-monotone marginal distributions (more properly $\alpha^{(d)}$-monotone marginal distributions where $\alpha^{(d)} \equiv \{\alpha_1, ..., \alpha_d\}$), the key is to replace $\Theta^{(d)}$ by $\Theta_\alpha^{(d)} \equiv \{\Theta_1^{1/\alpha_1}, ..., \Theta_d^{1/\alpha_d}\}$. Let the resulting random variable be $N_\alpha^{(d)}$. The joint p.m.f. of $N_\alpha^{(d)}$ is

$$p_D(n_1,...,n_d;\alpha_1,...\alpha_d) = \sum_{m_1=n_1}^{\infty} \cdots \sum_{m_d=n_d}^{\infty} q(m_1,...,m_d) \left\{ \prod_{i=1}^{d} \binom{m_i}{n_i} \right\}$$

$$(5.1) \qquad \times \int_0^1 \cdots \int_0^1 \left\{ \prod_{i=1}^{d} \theta_i^{n_i/\alpha_i}(1-\theta_i^{1/\alpha_i})^{m_i-n_i} \right\} c(\theta_1,...,\theta_d)\, d\theta_1 ... d\theta_d.$$

Its univariate marginal distributions have the α_1-monotone, α_2-monotone, ..., α_d-monotone p.m.f.'s of Section 4 by construction. The form of (5.1) involves d infinite sums and integrals but, as will be seen below, certain special cases simplify considerably. Moments remain readily available and correlations are as follows. Using (2.3) and (3.2),

$$\mathrm{Cov}(N_i,N_j) = \mathbb{E}(M_iM_j)\,\mathrm{Cov}(\Theta_i^{1/\alpha_i},\Theta_j^{1/\alpha_j}) + \frac{\alpha_i}{\alpha_i+1}\frac{\alpha_j}{\alpha_j+1}\,\mathrm{Cov}(M_i,M_j)$$

so that
(5.2)
$$\mathrm{Corr}(N_i,N_j) = \frac{\mathbb{E}(M_iM_j)\,\mathrm{Corr}(\Theta_i^{1/\alpha_i},\Theta_j^{1/\alpha_j})+\sqrt{\alpha_i(\alpha_i+2)\alpha_j(\alpha_j+2)}\,\mathrm{Cov}(M_i,M_j)}{\sqrt{[(\alpha_i+1)^2\mathbb{V}(M_i)+(\alpha_i+1)\mathbb{E}(M_i)+\{\mathbb{E}(M_i)\}^2][(\alpha_j+1)^2\mathbb{V}(M_j)+(\alpha_j+1)\mathbb{E}(M_j)+\{\mathbb{E}(M_j)\}^2]}}.$$

In the following two subsections, I will take a brief look at two major particular cases of this in terms of the form of distribution for M. These distributions and their properties are analogues of those given in Section 3 of Bryson and Johnson (1982) in the continuous case when $d = 2$. They are theoretically interesting but for the most part may prove to have limited practical applicability.

5.1. When $M_1,...,M_d$ are mutually independent

Let $M_i \sim q_i$, independently for $i = 1,...,d$. This allows the dependence structure of p_D to depend only on that of C ameliorated by the value of $\alpha^{(d)}$. The joint p.d.f. of $N_\alpha^{(d)}$ is given by the obvious small change to (5.1). The correlation of N_i and N_j, given by (5.2), reduces to
(5.3)
$$\mathrm{Corr}(N_i,N_j) = \sqrt{\frac{\mathbb{E}(M_i)}{(\alpha_i+1)^2\mathbb{D}(M_i)+\mathbb{E}(M_i)+\alpha_i+1}} \sqrt{\frac{\mathbb{E}(M_j)}{(\alpha_j+1)^2\mathbb{D}(M_j)+\mathbb{E}(M_j)+\alpha_j+1}}\,\mathrm{Corr}(\Theta_i^{1/\alpha_i},\Theta_j^{1/\alpha_j}).$$

where $\mathbb{D}(M) = \mathbb{V}(M)/\mathbb{E}(M)$ is the index of dispersion of M. Again, this has the same sign as the correlation associated with the copula and is always a reduction of the absolute value of the correlation compared with that of the copula, sometimes considerably so.

Example 5.1

This example concerns a family of multivariate distributions with geometric marginal distributions. Following Example 2.1(a), let $q_i(m) = (m+1)\,p_i^2(1-p_i)^m$ with $\mathbb{E}(M_i) = 2(1-p_i)/p_i$ and $\mathbb{V}(M_i) = 2(1-p_i)/p_i^2$, $i = 1, ..., d$. The corresponding multivariate geometric distribution arises by taking $\alpha_1 = \cdots = \alpha_d = 1$. Reduction of (5.1) in this case requires simplification of terms of the form $\sum_{m=n}^{\infty}(m+1)p^2(1-p)^m \binom{m}{n}\theta^n(1-\theta)^{m-n}$ which is achieved by noting that, with $0 < \psi \equiv (1-p)(1-\theta) < 1$,

$$\sum_{m=n}^{\infty}(m+1)\binom{m}{n}\psi^{m-n} = (n+1)\sum_{m=n}^{\infty}\binom{m+1}{n+1}\psi^{m-n}$$

$$= (n+1)\sum_{j=0}^{\infty}\binom{n+j+1}{j}\psi^j = \frac{n+1}{(1-\psi)^{n+2}}.$$

This results in the joint p.m.f.

$$p_G(n_1,\dots,n_d;p_1,...,p_d) = \prod_{i=1}^{d}(n_i+1)p_i^2(1-p_i)^{n_i}$$

$$\times \int_0^1 \cdots \int_0^1 \left[\prod_{i=1}^{d}\frac{\theta_i^{n_i}}{\{1-(1-p_i)(1-\theta_i)\}^{n_i+2}}\right] c(\theta_1,...,\theta_d)\,d\theta_1\,...\,d\theta_d$$

with correlations

$$\mathrm{Corr}(N_i,N_j) = \frac{1}{3}\sqrt{(1-p_i)(1-p_j)}\,\mathrm{Corr}(\Theta_i,\Theta_j).$$

The correlations associated with this family of multivariate geometric distributions are therefore limited to the range $-1/3 < \mathrm{Corr}(N_i,N_j) < 1/3$, although the range of correlations decreases as the p_i's increase.

Example 5.2

In a similar manner to Example 5.1, this example concerns a family of multivariate distributions with Poisson marginals. It arises by taking $q_i(m) = \mu_i^{m-1}e^{-\mu_i}/(m-1)!$, $m = 1, 2, ...$, and $\alpha_j = \mu_j$, $j = 1, ..., d$ (cf. Example 4.1(b)). In this case, simplification of (5.1) requires simplification of sums of the form $\sum_{m=n}^{\infty}e^{-\mu}\mu^{m-1}\binom{m}{n}\theta^{n/\mu}(1-\theta^{1/\mu})^{m-n}/(m-1)!$. Now, with $\Omega \equiv \mu(1-\theta^{1/\mu}) > 0$,

$$\sum_{m=n}^{\infty}m\frac{\Omega^{m-n}}{(m-n)!} = \sum_{m=n}^{\infty}(m-n)\frac{\Omega^{m-n}}{(m-n)!} + n\sum_{m=n}^{\infty}\frac{\Omega^{m-n}}{(m-n)!} = (\Omega+n)e^{\Omega}.$$

The corresponding joint p.m.f. is

$$p_P(n_1,\dots,n_d;\mu_1,...,\mu_d) = \prod_{i=1}^{d}\frac{\mu^{n_i}}{n_i!}$$

$$\times \int_0^1 \cdots \int_0^1 \left\{\prod_{i=1}^{d}\theta_i^{n_i/\mu_i}\left(1-\theta_i^{1/\mu_i}+\frac{n_i}{\mu_i}\right)e^{-\mu_i\theta_i^{1/\mu_i}}\right\} c(\theta_1,...,\theta_d)\,d\theta_1\,...\,d\theta_d.$$

Since $\mathbb{E}(M_i) = \mu_i + 1, \mathbb{V}(M_i) = \mu_i$, $i = 1, ..., d$, the correlations associated with these distributions are

$$\text{Corr}(N_i, N_j) = \frac{1}{\sqrt{(\mu_i + 2)(\mu_j + 2)}} \, \text{Corr}(\Theta_i^{1/\mu_i}, \Theta_j^{1/\mu_j})$$

so that $-1/2 < \text{Corr}(N_i, N_j) < 1/2$. In this case, the range of correlations decreases as the mean parameters increase.

5.2. When $M_1, ..., M_d$ are Equal or Most Strongly Dependent

---

Let $M_1 = \cdots = M_d = M$ say, $i = 1, ..., d$, with $M \sim q_0$. This particular comonotonicity also allows the dependence structure of p_D to depend on that of C, but with an opportunity for higher correlations. Let $n_{\max} = \max(n_1, ..., n_d)$. The joint p.d.f. of $N_\alpha^{(d)}$ is given by

$$p_D(n_1, ..., n_d; \alpha_1, ..., \alpha_d) = \sum_{m=n_{\alpha,\max}}^{\infty} q_0(m) \left\{ \prod_{i=1}^{d} \binom{m}{n_i} \right\}$$

$$\times \int_0^1 \cdots \int_0^1 \left\{ \prod_{i=1}^{d} \theta_i^{n_i/\alpha_i} (1 - \theta_i^{1/\alpha_i})^{m-n_i} \right\} c(\theta_1, ..., \theta_d) \, d\theta_1 ... d\theta_d.$$

Its correlations are, from (5.2),

$$(5.4) \quad \rho_{ij} \equiv \text{Corr}(N_i, N_j) = \frac{\{\mathbb{D}(M) + \mathbb{E}(M)\} \text{Corr}(\Theta_i^{1/\alpha_i}, \Theta_j^{1/\alpha_j}) + \sqrt{\alpha_i(\alpha_i+2)\alpha_j(\alpha_j+2)} \, \mathbb{D}(M)}{\sqrt{[(\alpha_i+1)^2 \mathbb{D}(M) + \mathbb{E}(M) + \alpha_i + 1][(\alpha_j+1)^2 \mathbb{D}(M) + \mathbb{E}(M) + \alpha_j + 1]}},$$

which are all equal if $\alpha_1 = \cdots = \alpha_d$. If r_{ij} denotes the correlation at (5.3) when both M_i and M_j have the distribution of M, then

$$\rho_{ij} = r_{ij} + \frac{\mathbb{D}(M)\left\{ \text{Corr}(\Theta_i^{1/\alpha_i}, \Theta_j^{1/\alpha_j}) + \sqrt{\alpha_i(\alpha_i+2)\alpha_j(\alpha_j+2)} \right\}}{\sqrt{[(\alpha_i+1)^2 \mathbb{D}(M) + \mathbb{E}(M) + \alpha_i + 1][(\alpha_j+1)^2 \mathbb{D}(M) + \mathbb{E}(M) + \alpha_j + 1]}}$$

which is typically greater than r_{ij}, certainly whenever $\alpha_i(\alpha_i + 2)\alpha_j(\alpha_j + 2) > 1$.

Example 5.3

While in Sections 3 and 5.1 the independence copula with density $c(\theta_1, ..., \theta_d) = \prod_{i=1}^{d} I(0 < \theta_i < 1)$ results in distributions with independent marginals, this is not the case here because of the commonality of M. In fact, using the independence copula, the joint p.m.f. of $N_\alpha^{(d)}$ depends only on $n_{\max}$ and is given by

$$p_I(n_1, ..., n_d; \alpha_1, ..., \alpha_d) = \sum_{m=n_{\max}}^{\infty} q_0(m)(m!)^d \prod_{i=1}^{d} \frac{\alpha_i \, \Gamma(n_i + \alpha_i)}{n_i! \, \Gamma(m + 1 + \alpha_i)},$$

reducing to

$$p_I(n_1, ..., n_d; 1, ..., 1) = \sum_{m=n_{1,\max}}^{\infty} \frac{q_0(m)}{(m+1)^d}.$$

The corresponding correlations are, in general,

$$\mathrm{Corr}(N_i, N_j) = \sqrt{\frac{\alpha_i(\alpha_i+2)}{(\alpha_i+1)^2 \mathbb{D}(M) + \mathbb{E}(M) + \alpha_i + 1}} \sqrt{\frac{\alpha_j(\alpha_j+2)}{(\alpha_j+1)^2 \mathbb{D}(M) + \mathbb{E}(M) + \alpha_j + 1}} \, \mathbb{D}(M),$$

which are all positive. When $\alpha_1 = \cdots = \alpha_d = 1$,

$$0 < \mathrm{Corr}(N_i, N_j) = \frac{3\mathbb{D}(M)}{4\mathbb{D}(M) + \mathbb{E}(M) + 2} < \frac{3}{4}.$$

Example 5.4

For a general copula, let us contrast the correlation structure associated with the specific multivariate geometric and Poisson distributions of Examples 5.1 and 5.2 when $M_1, ..., M_d$ are independent with the corresponding distributions when $M_1 = \cdots = M_d = M$.

(a) Let $\alpha_1 = \cdots = \alpha_d = 1$ and $M \sim \mathrm{NegativeBinomial}(2, p)$. Then, the corresponding family of multivariate distributions with $\mathrm{Geometric}(p)$ marginals has correlations

$$\mathrm{Corr}(N_i, N_j) = \frac{1}{2} + \frac{(3 - 2p) \, \mathrm{Corr}(\Theta_i, \Theta_j)}{6}.$$

In this case, $0 < \mathrm{Corr}(N_i, N_j) < 1$, contrasting with a range of $(-1/3, 1/3)$ in Example 5.1. In fact, these correlations are always greater than those when $p_i = p_j = p$ in the independent M's case because $\alpha(\alpha+2) = 3 > 1$. In the case of the independence copula as in Example 5.3, $\mathrm{Corr}(N_i, N_j) = 1/2$.

(b) Let $\alpha_1 = \cdots = \alpha_d = \mu$ and $M = M_1 + 1$ where $M_1 \sim \mathrm{Poisson}(\mu)$, as in Example 5.2. Then, the corresponding family of multivariate Poisson distributions has correlations

$$\mathrm{Corr}(N_i, N_j) = \left(\frac{\mu}{\mu+1}\right)^2 + \frac{(\mu^2 + 3\mu + 1) \, \mathrm{Corr}(\Theta_i^{1/\mu}, \Theta_j^{1/\mu})}{(\mu+1)^2 (\mu+2)}.$$

It is certainly the case that $-1/2 < \mathrm{Corr}(N_i, N_j) < 1$ (contrasting with $(-1/2, 1/2)$ in Example 5.2) although slightly more negative correlation is possible for certain very small μ. The correlation is greater than that when $\mu_i = \mu_j$ in Example 5.2 whenever $\mathrm{Corr}(\Theta_i^{1/\mu}, \Theta_j^{1/\mu}) > -\mu(\mu + 2)$. In the case of the independence copula, $0 < \mathrm{Corr}(N_i, N_j) = \mu^2/(\mu + 1)^2 < 1$.

Finally, if $M_1, ..., M_d$ are not the same, then the strongest dependence is comonotonicity or the Fréchet upper bound. The expression for p_D does not simplify but the pair $\{N_i, N_j\}$ can be more highly correlated in comparison to Section 5.1.

ACKNOWLEDGMENTS

REFERENCES

[1]   ABOUAMMOH, A.M. (1987/1988). On discrete α-unimodality, *Statistica Neerlandica*, **41**, 239–244. Correction, **42**, 141.

[2]   ABOUAMMOH, A.M., A.M. ALI AND A.F. MASHHOUR (1994). On characterizations and variance bounds of discrete α-unimodality, *Statistical Papers*, **35**, 151–161.

[3]   BERMÚDEZ, L. AND D. KARLIS (2011). Bayesian multivariate Poisson models for insurance ratemaking, *Insurance: Mathematics and Economics*, **48**, 226–236.

[4]   BERTIN, E.M.J., I. CUCULESCU AND R. THEODORESCU (1997). *Unimodality of Probability Measures*, Kluwer, Dordrecht.

[5]   BRUNNER, L.J. AND A.Y. LO (1989). Bayes methods for a symmetric unimodal density and its mode, *Annals of Statistics*, **17**, 1550–1566.

[6]   BRYSON, M.C. AND M.E. JOHNSON (1982). Constructing and simulating multivariate distributions using Khintchine's theorem, *Journal of Statistical Computation and Simulation*, **16**, 129–137.

[7]   DAVY, P.J. AND J.C.W. RAYNER (1996). Multivariate geometric distributions, *Communications in Statistics – Theory and Methods*, **25**, 2971–2987.

[8]   DHARMADHIKARI, S. AND K. JOAG-DEV (1988). *Unimodality, Convexity, and Applications*, Academic Press, Boston, MA.

[9]   FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol 2, Wiley, New York.

[10] GENEST, C. AND J. NEŠLEHOVÁ (2007). A primer on copulas for count data, *Astin Bulletin*, **37**, 475–515.

[11] JOE, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

[12] JOE, H. (2014). *Dependence Modeling with Copulas*, Chapman & Hall/CRC Press, Boca Raton, FL.

[13] JOHNSON, N.L., A.W. KEMP & S. KOTZ (2005). *Univariate Discrete Distributions* (3rd ed.), Wiley, Hoboken, NJ.

[14] JOHNSON, N.L. AND C.A. ROGERS (1951), The moment problem for unimodal distributions, *Annals of Mathematical Statistics*, **22**, 433–439.

[15] JONES, M.C. AND É. MARCHAND (2019). Multivariate discrete distributions via sums and shares, *Journal of Multivariate Analysis*, **171**, 83–93.

[16] KHINTCHINE, A.Y. (1938). On unimodal distributions, *Izvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki*, **2**, 1–7.

[17] MEDGYESSY, P. (1972). On the unimodality of discrete distributions, *Periodica Mathematica Hungarica*, **2**, 245–257.

[18] NELSEN, R.B. (2006), *An Introduction to Copulas* (2nd ed.), Springer, New York.

[19] OLSHEN, R.A. AND L.J. SAVAGE (1970). A generalized unimodality, *Journal of Applied Probability*, **7**, 21–34.

[20] SHI, P. AND E.A. VALDEZ (2011). Multivariate negative binomial models for insurance claim counts, *Insurance: Mathematics and Economics*, **55**, 18–29.

[21] STEUTEL, F.W. (1988). Note on discrete α-unimodality, *Statistica Neerlandica*, **42**, 137–140.

[22] STEUTEL, F.W. AND K. VAN HARN (1979). Discrete analogues of self-decomposability and stability, *Annals of Probability*, **7**, 893–899.

[23] WESTFALL, P.H. AND S.S. YOUNG (1989). P value adjustments for multiple tests in multivariate binomial models, *Journal of the American Statistical Association*, **84**, 780–786.