EXCHANGEABLE PAIRS OF BERNOULLI RANDOM VARIABLES, KRAWTCHOUCK POLYNOMIALS, AND EHRENFEST URNS

Persi Diaconis and Robert Griffiths *
Stanford University and Oxford University

January 30, 2011

Summary

This paper explores characterizations of Bivariate Binomial distributions of the Lancaster form with Krawtchouk polynomial eigenfunctions which are equivalent to a characterization by Eagleson (1969). Generalized Ehrenfest urns with Binomial stationary distributions have transition functions which are equivalent to the Lancaster distributions. There is a partial extension to d-colour ball Ehrenfest urns.

Keywords: Bivariate Binomial distributions; Correlation sequences; Ehrenfest Urns; Krawtchouk polynomials; Lancaster distributions.

Running Head: Correlation sequences, Krawtchouk Polynomials, Ehrenfest Urns

1. Introduction

A classical problem, which has become known as the *Lancaster problem* (Lancaster, 1969; Koudou, 1996) is to characterize bivariate distributions with given marginals and orthogonal polynomial eigenfunctions. Let (X, Y) have a bivariate distribution with marginal distributions $f(x), f(y)$ and let $\{P_n(x)\}_{n\in\mathbb{N}}$ be a complete orthogonal polynomial set on $f(x)$, such that for $m, n \in \mathbb{N}$

$$\mathbb{E}\Big[P_m(X)P_n(X)\Big] = \delta_{mn}h_n^{-1}.$$

*Corresponding author: Professor R. C. Griffiths, Department of Statistics, University of Oxford, 1 South Parks Rd, Oxford, OX1 3TG, UK; Ph +44 1865 281237; Fax +44 1865 272595; email griff@stats.ox.ac.uk

Exchangeable Lancaster bivariate distributions have the form

$$f(x)f(y)\Big\{1+\sum_{n=1}^{\infty}\rho_n h_n P_n(x)P_n(y)\Big\}, \tag{1}$$

where $\{\rho_n\}_{n\in\mathbb{N}}$ is a sequence of constants with $\rho_0 = 1$. The expansion holds in mean square, assuming that $\sum_{n=1}^{\infty}\rho_n^2 < \infty$. The sequence is a correlation sequence between the orthogonal polynomial sets $\{P_m(X)\}_{m\in\mathbb{N}}$ and $\{P_n(Y)\}_{n\in\mathbb{N}}$ because

$$\mathbb{E}\Big[P_m(X)P_n(Y)\Big] = \delta_{mn}h_n^{-1}\rho_n.$$

A characterization of which sequences are correlation sequences amounts to finding necessary and sufficient conditions such that (1) is non-negative and thus a bivariate distribution. The set of correlation sequences $\mathcal{S}$ is a convex set so that if $\{\rho_n\}_{n\in\mathbb{N}}, \{\omega_n\}_{n\in\mathbb{N}} \in \mathcal{S}$ then for $\lambda \in [0,1]$ $\lambda\{\rho_n\}_{n\in\mathbb{N}} + (1-\lambda)\{\omega_n\}_{n\in\mathbb{N}} \in \mathcal{S}$. There is also closure under direct products with $\{\rho_n\omega_n\}_{n\in\mathbb{N}} \in \mathcal{S}$. Since $\mathcal{S}$ is a convex set it is characterized by its extreme points. A more general problem is to characterize correlation sequences when the marginal distributions and orthogonal polynomial sets are not identical.

The bivariate distribution leads to transition functions for a reversible Markov chain with stationary distribution f of

$$f(y;x) = f(y)\Big\{1+\sum_{n=1}^{\infty}\rho_n h_n P_n(x)P_n(y)\Big\}. \tag{2}$$

For such Markov chains $\{P_n\}_{n\in\mathbb{N}}$ are the right eigenfunctions of the transition distribution f with eigenvalues $\{\rho_n\}_{n\in\mathbb{N}}$ such that

$$\mathbb{E}\Big[P_n(Y)\mid X\Big] = \rho_n P_n(X).$$

We call these Markov chains with orthogonal polynomial eigenfunctions. Conversely if a reversible Markov chain has stationary distribution $f(x)$ and transition distributions $f(y;x)$, then $f(y;x)f(x)$ is an exchangeable bivariate distribution with marginals $f(x), f(y)$. The k-step transition functions from the single step transition function (2) are

$$f^{(k)}(y;x) = f(y)\Big\{1+\sum_{n=1}^{\infty}\rho_n^k h_n P_n(x)P_n(y)\Big\}. \tag{3}$$

From the representation (3), the L^2 or chi-square distance to stationarity can be bounded via

$$\chi^2(k) = \sum \frac{\big(f^{(k)}(y;x) - f(y)\big)^2}{f(y)} = \sum_{n=1}^{\infty} \rho_n^{2k} h_n P_n^2(x).$$

Now, analytic estimates using knowledge of ρ_n and $P_n(x)$ are used to bound the right-hand side, determining how large k must be to make it suitably small. For many examples, see Diaconis, Khare and Saloff-Coste (2008).

If $\{X_k\}_{k\in\mathbb{N}}$ is a reversible Markov chain with transition functions (2) started in stationarity, then the joint distribution of $(X, Y) = (X_0, X_k)$ is exchangeable with marginal distributions $f(x), f(y)$. Embedding the process into continuous time by taking $\{Z(t) = X_{N(t)}\}_{t\geq 0}$ with $\{N(t)\}_{t\geq 0}$ a Poisson process of rate λ independent of $\{X_k\}_{k\in\mathbb{N}}$, then $\{Z(t)\}_{t\geq 0}$ has transition functions

$$\begin{aligned}&\sum_{k=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!} f^{(k)}(y;x)\\ &\quad = f(y)\Big\{1 + \sum_{n=1}^{\infty} e^{-(1-\rho_n)\lambda t} h_n P_n(x) P_n(y)\Big\}. \qquad (4)\end{aligned}$$

Bochner (1954) studied the general form of the eigenvalues of a continuous time Markov process with given stationary distribution and eigenfunctions. Eigenvalues have the form in (4) or can be obtained as a limit

$$\lim_{\lambda\to\infty} e^{-\big(1-\rho_n(\lambda)\big)\lambda t} = e^{-c_n t}$$

from sequences depending on λ such that $\rho_n(\lambda) \to 1$ and $\big(1 - \rho_n(\lambda)\big)\lambda \to c_n$. Characterizations of $\{\rho_n\}_{n\in\mathbb{N}}$ are known for most distributions in the Meixner class. For the Normal distribution the sequence is a moment sequence of a random variable on $[-1, 1]$ and for the Gamma, Poisson and Negative Binomial distributions the sequence is a moment sequence of a random variable on $[0, 1]$. Characterizations for distributions which do not have an infinite support interval are more difficult, though characterizations for Beta distributions and Binomial distributions are known. A good general reference is Koudou (1996). Griffiths (2009) obtains characterizations of reversible continuous time stochastic processes with polynomial eigenfunctions related to the bivariate distribution characterizations. In this paper we study bivariate distributions when the marginals have Binomial distributions $b(x;N,p)$ and $b(y;N,p)$, where

$$b(x;N,p) = \binom{N}{x} p^x (1-p)^{N-x},\ x = 0,1,\dots,N.$$

The Lancaster distributions are then, for $x,y = 0,1,\dots,N$,

$$p(x;y) = b(x;N,p)b(y;N,p)\Big\{1 + \sum_{n=1}^{N} \rho_n h_n Q_n(x) Q_n(y)\Big\}, \qquad (5)$$

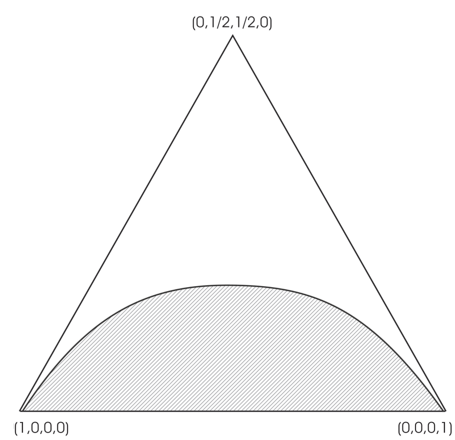
where $\{Q_n(z)\}_{n=0}^N$ are the Krawtchouk polynomials scaled so that $Q_n(0) = 1,\ n = 0,1,\dots,N$. We also study reversible Markov chains and processes which have Binomial stationary distributions and Kratchouk polynomial eigenfunctions.

Examples of Bivariate Binomial distributions which have orthogonal polynomial eigenfunctions are the following.

Aitken and Gonin's 2×2 contingency table marginals

Let $\big\{p_{\xi\eta}\big\}_{\xi,\eta\in\{0,1\}}$ be an exchangeable probability distribution and suppose that N independent observations form a contingency table from this underlying distribution. Aitken and Gonin (1935) show that (X,Y), the marginal row and column counts of the number of entries 1,1, has a joint distribution (5) with $\rho_n = \tau^n$, where $\tau =$ correlation(ξ,η).

Figure 1: The set of exchangeable binary joint distributions



A general bivariate binary distribution $(p_{00}, p_{01}, p_{10}, p_{11})$ is exchangeable if and only if $p_{01} = p_{10}$. These can be pictured as the convex set of Figure 1. The three extreme points $(1, 0, 0, 0), (0, \frac{1}{2}, \frac{1}{2}, 0), (0, 0, 0, 1)$ correspond to drawing from an urn without replacement. For urns containing two balls {0,0} or {0,1} or {1,1} respectively, the quadratic curve $((1-p)^2, p(1-p), p(1-p), p^2)$ represents the independent Bernoulli measures. The shaded regions below the curve represents mixtures of Bernoulli measures. These can be characterized as measures with non-negative correlations (Suppes and Zanotti, 1980,1981). They apply this to show the impossibility of hidden variables in quantum mechanics. Further details on the geometry of exchangeability and connections to deFinetti's theorem are in Diaconis (1977).

Random elements in common

Let $X = U + V$, $Y = V + W$, where U, V, W are independent Binomial random variables with respective parameters $N - M, p$; M, p; $N - M, p$, $M < N$. Then (X, Y) has a joint distribution (5) with $\rho_n = M_{[n]}/N_{[n]}$, where $a_{[n]} = a(a-1)\cdots(a-n+1)$. This example is related to a Markov chain in Diaconis, Khare and Saloff-Coste (2008). Exchangeability of the marginal distributions can be relaxed in both the last two examples to have non-identical Binomial distributions.

Cumulative Bernoulli trials

Hoare and Rahman (1983) consider a Markov chain with a fixed total number of N balls which are either red or blank coloured. If X is the number of red balls then in the first stage of a transition, independently each red ball remains red with probability α, or changes to blank with probability $1-\alpha$. Let K be the number of red balls after this first stage. In the second stage the $N - K$ blank balls independently change to red with probability β or remain blank with probability $1 - \beta$. The conditional *pgf* (probability generating function) for a transition from X to Y is

$$\mathbb{E}_K\Big[s^K(1-\beta+\beta s)^{N-K}\Big]$$

where K has a Binomial (X, α) distribution. Let X_t be the number of red balls after t transitions. $\{X_t\}_{t\in\mathbb{N}}$ is a reversible Markov Chain with a Binomial (N, γ) stationary distribution, where $\gamma = \beta/(1-\alpha+\alpha\beta)$. In a stationary process (X_t, X_{t+1}) has a joint distribution (5) with $\rho_n = (1 - \beta/\gamma)^n$.

An Ehrenfest Urn

An urn has N balls coloured red or blue. Transitions in a Markov chain are made by selecting a ball at random and changing its colour. X_t is the number of red balls after t transitions. $\{X_t\}_{t\in\mathbb{N}}$ is a reversible Markov Chain with a Binomial $(N, \frac{1}{2})$ stationary distribution.

An Ehrenfest (p) Urn

Let $q \leq p$, $p+q=1$. In a transition choose a ball at random and change the colour of a ball drawn according to the transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ q/p & 1-q/p \end{bmatrix} \tag{6}$$

That is, if a blue ball is selected it is changed to red with probability 1, whereas a red ball is changed to blue with probability q/p. $\{X_t\}_{t\in\mathbb{N}}$ is a reversible Markov Chain with a Binomial (N,p) stationary distribution.

This process is an example of applying the Metropolis algorithm to the original Ehrenfest urn. For discussion and a complete analysis, see Bassetti and Diaconis (2006).

A Generalized Ehrenfest (p) Urn

At a single transition choose $z \in \{0,1,\dots,N\}$ balls at random without replacement according to a probability distribution $\{a(z)\}_{z=0}^N$ and independently change the colours of the balls drawn according to the transition matrix (6). $\{X_t\}_{t\in\mathbb{N}}$ is a reversible Markov Chain with a Binomial (N,p) stationary distribution.

This generalization will be used extensively in what follows.

Eagleson (1969) characterizes correlation sequences of bivariate Binomial distributions (X,Y) with distribution (5) as having a representation

$$\rho_n = \sum_{z=0}^{N} a(z)Q_n(z), \tag{7}$$

where $\{a(z)\}_{z=0}^N$ is an arbitary probability distribution. Extreme point correlation sequences are the actual polynomials themselves, $\{\rho_n = Q_n(z)\}_{n=0}^N$ obtained for fixed z when $a(w) = \delta_{wz}$, $w=0,\dots,N$. The characterization holds when $p \geq q(=1-p)$. (If $p<q$ the roles of $x, N-x$; and p,q are

interchanged.) If $a(z) = b(z; N, p)$ then $\rho_0 = 1$, $\rho_n = 0$, $n = 1, \ldots, N$ and X, Y are independent. If $a(z) = \delta_{z0}$, then $\rho_n = 1$, $n = 0, \ldots, N$ and $X = Y$.

Eagleson's work is connected to Bochner's characterization of the Fourier transform on a group in terms of a positive definiteness condition. In a series of papers, Bochner extended this to characterize the transform of a probability using eigenfunctions. Let K be a self-adjoint operator on a compact space with eigenfunctions $f_j(x)$ (so $Kf_j(x) = \lambda_j f_j(x)$). If $P(x)$ is a probability measure, define $\hat{P}(j) = \int f_j dP$. Roughly put, for the cases he considered, Bochner shows that a sequence t_j can be represented as $t_j = \hat{P}(j)$ if and only if for all sequences $\{a_j\}$ with $\sum a_j P_j(x) \geq 0$ for all x, it holds that $\sum a_j t_j P_j(x) \geq 0$ for all x. Eagleson treated the case of general finite state space and applied his results to the Krawtchouck polynomials. There have been some extensions; Vere-Jones (1971) shows that the Krawtchouck polynomials are the spherical functions for the hyperoctahedral group and thus relates Eagleson's theorem to Bochner's theorem for groups. For extensions using the language of hypergroups, see the excellent surveys of Bakry and Huet (2008) and their references.

In this paper we prove the equivalence of three characterizations of bivariate Binomial random variables (X, Y) which have distributions of the form (5) with Krawtchouk polynomial eigenfunctions. The first is the Eagleson (1969) characterization that correlation sequences have the representation (7). The second characterization is that $(X, Y) = \sum_{i=1}^N (\xi_i, \eta_i)$, where $\{(\xi_i, \eta_i)\}_{i=1}^N$ are Bernoulli (p) pairs, independent within the marginal sequences, and with a random distribution of correlations in the two-point set $\{-q/p, 1\}$. The third characterization is that the conditional distributon of Y given X can be identified with transition functions in the generalized Ehrenfest urn model. The distribution $\{a(z)\}_{z=0}^N$ in Eagleson's characterization (7) is the same as the distribution of the number of balls drawn in the urn model. There is also a characterization when the distributions have Binomial (N, p_1) and Binomial (N, p_2) marginals; and another with Binomial (N, p) and Binomial (M, p) marginals.

Section 5 compares the Lancaster distributions with the Frechét bounds. Section 6 treats Markov chains with binomial stationary distributions; passing to various limits gives results for Gaussian processes and $M/M/\infty$ queues. In Section 7 we consider d-dimensional generalized Ehrenfest urn processes, extending the 2-dimensional case. The eigenvector, eigenvalue structure is generally not clear in these higher dimensional processes.

2. The Krawtchouk Polynomials

The Krawtchouk polynomials $\big\{K_n(x;N,p)\big\}_{n=0}^N$ are orthogonal on the Binomial (N,p) distribution. A comprehensive treatment is in Ishmail (2005). MacWilliams and Sloane (1997) list many properties and identities with applications in coding theory. The polynomials have a generating function

$$G(z;x) = \sum_{n=0}^N K_n(x;N,p)\frac{z^n}{n!} = (1+qz)^x(1-pz)^{N-x}. \tag{8}$$

The scaling is such that

$$K_n(0;N,p) = n!\binom{N}{n}(-p)^n,\ \ \mathbb{E}\big[K_n(X;N,p)^2\big] = n!^2\binom{N}{n}(pq)^n.$$

Defining $Q_n(x) = K_n(x;N,p)/K_n(0;N,p)$, so that $Q_n(0) = 1$,

$$h_n = \mathbb{E}\big[Q_n(X)^2\big]^{-1} = \binom{N}{n}\big(p/q\big)^n \tag{9}$$

and

$$\begin{aligned} Q_n(x) &= \sum_{\nu=0}^N \big(-q/p\big)^\nu \frac{\binom{x}{\nu}\binom{N-x}{n-\nu}}{\binom{N}{n}} && (10)\\ &= \sum_{\nu=0}^N \big(-q/p\big)^\nu \frac{\binom{n}{\nu}\binom{N-n}{x-\nu}}{\binom{N}{x}}. && (11)\end{aligned}$$

The generating function for $\{Q_n(x)\}_{n=0}^N$ is

$$\sum_{n=0}^N \binom{N}{n} Q_n(x)z^n = \big(1-(q/p)z\big)^x\big(1+z\big)^{N-x}. \tag{12}$$

There is the self-dual relationship that $Q_n(x) = Q_x(n)$, seen directly from the hypergeometric identity of terms in (10) and (11). We use the notation $Q_n(x;N,p)$ and $h_n(N,p)$ when it is important to distinguish parameters, otherwise we suppress N,p.

If X is Binomial (N,p) then $X = \sum_{i=1}^N \xi_i$ where $\big\{\xi_i\big\}_{i=1}^N$ is a sequence of independent Bernoulli (p) trials. Orthogonal polynomials on the Bernoulli distribution taking values 0 or 1 with probabilities q,p are $\{1, K_1(\xi;1,p)\}$, where $K_1(\xi;1,p) = \xi - p$, with $\xi = 0,1$. Denote the first polynomial in this special case of the Binomial with $N = 1$ by $u(\xi) = \xi - p$. The generating function for the Krawtchouk polynomials (8) can be expressed as

$$G(z;X) = \prod_{j=1}^{N} \big(1 + u(\xi_j)z\big), \tag{13}$$

because if X of the entries in $\{\xi_i\}_{i=1}^N$ are 0 and $N - X$ are 1, then there are X terms in the product of $1+u(0)z = 1-pz$ and $N-X$ of $1+u(1)z = 1+qz$, showing the identity with (8). Let S_N be the symmetric group on $\{1,\dots,N\}$, then from the generating function (13)

$$K_n(X;N,p) = n! \sum_{\sigma\in S_N} u(\xi_{\sigma(1)})\cdots u(\xi_{\sigma(n)}), \tag{14}$$

and similarly

$$Q_n(X) = \binom{N}{n}^{-1} \sum_{\sigma\in S_N} v(\xi_{\sigma(1)})\cdots v(\xi_{\sigma(n)}), \tag{15}$$

where $v(\xi) = (p-\xi)/p$. The sums are the nth symmetric products in the independent, identically distributed, sequences of random variables $\{u(\xi_j)\}_{j=1}^N$ and $\{v(\xi_j)\}_{j=1}^N$. This is an interesting important representation for the Krawtchouk polynomials which is not well known. It is applied in the next section.

3. Eagleson's characterization of Bivariate Binomial distributions

Eagleson (1969) shows that his characterization (7) of $\{\rho_n\}_{n=0}^N$ is equivalent to the non-negativity of the triple product sum

$$K(x,y,z) = \sum_{n=0}^{N} h_n Q_n(x)Q_n(y)Q_n(z) \geq 0,\ x,y,z = 0,1,\dots N, \tag{16}$$

where h_n is defined in (9). Following Eagleson's introduction such triple product sums have been widely used for other families, (Bakry and Huet, 2008). The property (16) actually follows from the representation (15).

Using duality

$$\begin{aligned} K(x,y,z) &= \sum_{n=0}^{N} h_n Q_n(x)Q_n(y)Q_n(z) \\ &= q^{-N}\sum_{n=0}^{N}\binom{N}{n}p^n q^{N-n}Q_x(n)Q_y(n)Q_z(n) \\ &= q^{-N}\mathbb{E}\big[Q_x(W_N)Q_y(W_N)Q_z(W_N)\big] \qquad (17) \end{aligned}$$

where W_N is Binomial (N,p). The positivity of (17) follows using (15) because of independence of variables and that $\mathbb{E}\big[v(\xi)^r\big] \geq 0$ for $r = 1,2,3$. When $r = 3$,

$$\mathbb{E}\big[v(\xi)^3\big] = p^{-3}[qp^3 + p(-q)^3] = p^{-2}q(p^2 - q^2) \geq 0.$$

The non-negativity of $K(x,y,z)$ implies that

$$p_{xy}(z) = \binom{N}{z}p^z(1-p)^{N-z}K(x,y,z), z = 0,1,\dots,N$$

is a probability distribution for fixed x,y. The duplication formula

$$Q_n(x)Q_n(y) = \sum_{z=0}^{N} Q_n(z)p_{xy}(z) \qquad (18)$$

holds because of the way $K(x,y,z)$ is defined in (16). Using duality, the duplication formula (18) shows that the product of two different orthogonal polynomials can be expressed as a linear combination of polynomials in the same family with positive coefficients.

4. Characterizations of Bivariate Binomial distributions

We build a set of characterizations from the very simplest bivariate distribution when $N = 1$ and P is a 2×2 transition matrix with stationary distribution $(p_0 = q, p_1 = p)$. Then

$$\begin{aligned} p_{xy} &= p_y\Big\{1 + \rho_1 h_1 Q_1(x)Q_1(y)\Big\} \\ &= p_y\Big\{1 + \rho_1(x-p)(y-p)/(pq)\Big\},\ x,y = 0,1. \qquad (19) \end{aligned}$$

ρ_1 is the usual correlation coefficient in the bivariate distribution $p_x p_{xy}$. Clearly with $p \geq q$, $p_{xy} \geq 0$ for $x,y = 0,1$ if and only if $p_{01} \geq 0$ and

$p_{00} \geq 0$. That is $-q/p \leq \rho_1 \leq 1$. The extreme points of the possible correlation coefficient are $-q/p, 1$. In Eagleson's characterization $\rho_1 = a(0) + a(1)Q_1(1) = a(0) - a(1)q/p$, so $a(1) = p(1-\rho_1)$ and $a(0) = 1 - a(1)$.

Theorem 1. *Three equivalent characterizations of bivariate Binomial (N,p) random variables (X,Y) with Krawtchouk polynomial eigenfunctions are the following.*

(i) The distribution of (X,Y) has a Lancaster expansion

$$p(x;y) = b(x;N,p)b(y;N,p)\Big\{1 + \sum_{n=1}^{N} \rho_n h_n Q_n(x)Q_n(y)\Big\}, \tag{20}$$

for $x,y = 0,1,\dots,N$. The correlation sequence $\{\rho_n\}_{n=0}^N$ has a representation

$$\rho_n = \sum_{z=0}^{N} a(z)Q_n(z), \tag{21}$$

where $\{a(z)\}_{z=0}^N$ is a probability distribution (Eagleson, 1969).

(ii) There is a representation

$$(X,Y) = \sum_{i=1}^{N} (\xi_i, \eta_i), \tag{22}$$

where $\{(\xi_i,\eta_i)\}_{i=1}^N$ are Bernoulli (p) pairs of random variables, independent within marginal sequences $\{\xi_i\}_{i=1}^N$ and $\{\eta_i\}_{i=1}^N$, with random correlations $\{\tau_i\}_{i=1}^N$ such that random variables within and between the pairs $\{(\xi_i,\eta_i)\}_{i=1}^N$ are conditionally independent given $\{\tau_i\}_{i=1}^N$. The distribution of $\{\tau_i\}_{i=1}^N$ can be always be taken to be exchangeable with correlations in the two-point set $\{-q/p, 1\}$.

(iii) (X,Y) is distributed as (X_t, X_{t+1}) for all $t \in \mathbb{N}$ in a stationary generalized Ehrenfest (p) urn process, where X_0 has a Binomial (N,p) distribution.

The distribution $\{a(z)\}_{z=0}^N$ has an intrepretation in the three characterizations of:
(i)

$$\begin{aligned} \rho_n &= \sum_{z=0}^{N} a(z)Q_n(z),\ n = 0,1,\dots,N & (23)\\ a(z) &= P(Y = z \mid X = 0),\ z = 0,1,\dots,N. & (24) \end{aligned}$$

(ii) $\{a(z)\}_{z=0}^{N}$ *is the distribution of the number of correlation coefficients* $\{\tau_i\}_{i=1}^{N}$ *in the two-point set* $\{-q/p, 1\}$ *equal to* $-q/p$*;*

(iii) $\{a(z)\}_{z=0}^{N}$ *is the distribution of the number of balls drawn in each transition in a stationary generalized Ehrenfest* (p) *urn process.*

Proof.
(i) This is Eagleson's (1969) characterization.
(ii) Take $\{\tau_i\}_{i=1}^{N}$ to be an exchangeable sequence, because if not it could be replaced by $\{\tau_{\theta(i)}\}_{i=1}^{N}$, where θ is a random permutation on S_N. This leaves the distribution of (X, Y) invariant. Each pair (ξ_i, η_i) has the property $\mathbb{E}\big[\eta_i - p|\xi_i\big] = \tau_i(\xi_i - p)$, and therefore

$$\begin{aligned}
&\mathbb{E}\Big[\binom{N}{n}Q_n(Y) \mid X\Big] &(25)\\
&= \mathbb{E}\Big[\mathbb{E}\Big(\sum_{\sigma\in S_N} v(\eta_{\sigma(1)})\cdots v(\eta_{\sigma(n)}) \mid \{\xi_j\},\{\tau_j\}\Big) \mid X\Big]\\
&= \mathbb{E}\Big[\sum_{\sigma\in S_N} \tau_{\sigma(1)}\ldots\tau_{\sigma(n)} v(\xi_{\sigma(1)})\cdots v(\xi_{\sigma(n)}) \mid X\Big]\\
&= \mathbb{E}\Big[\tau_1\cdots\tau_n\Big]\mathbb{E}\Big[\sum_{\sigma\in S_N} v(\xi_{\sigma(1)})\cdots v(\xi_{\sigma(n)}) \mid X\Big]\\
&= \mathbb{E}\Big[\tau_1\cdots\tau_n\Big]\binom{N}{n}Q_n(X), &(26)
\end{aligned}$$

showing that

$$\rho_n = \mathbb{E}\big[\tau_1\cdots\tau_n\big] \tag{27}$$

and

$$\mathbb{E}\big[Q_m(X)Q_n(Y)\big] = \delta_{mn}h_n\rho_n,\ m,n = 0,1,\ldots,N.$$

In general $\tau_i \in [-q/p, 1],\ i = 1,\ldots,N$. There is a decomposition into extreme points which are endpoints of the interval

$$\tau_i = \lambda_i \times (-q/p) + (1-\lambda_i)\times 1,\ \lambda_i \in [0,1],\ i = 1,\ldots,N.$$

The correlation coefficients can always be taken to be in the two-point set $\{-q/p, 1\}$. Define

$$\tau_i^\star = \begin{cases} -q/p & \text{with probability } \lambda_i\\ 1 & \text{with probability } 1-\lambda_i. \end{cases} \tag{28}$$

Then the distribution of (X,Y) is invariant under the change in correlations because

$$\rho_n = \mathbb{E}\Big[\tau_1 \cdots \tau_n\Big] = \mathbb{E}\Big[\tau_1^\star \cdots \tau_n^\star\Big].$$

To complete the proof in *(ii)* we show that the extreme points in the representation (27) are $\{Q_n(z)\}_{n=0}^N$, $z = 0,1,\dots,N$, the same as in *(i)*. Choose a sequence of correlations $\{\tau_j^z\}_{j=1}^N$ to have a uniform permutation distribution on a sequence with z entries $-q/p$, and $N-z$ entries 1. The number of entries in a collection of n correlations equal to $-q/p$ has a hypergeometric distribution so

$$\begin{aligned} \rho_n &= \mathbb{E}\Big[\tau_1^z \cdots \tau_n^z\Big] \\ &= \sum_{\nu=0}^N \big(-q/p\big)^\nu \frac{\binom{z}{\nu}\binom{N-z}{n-\nu}}{\binom{N}{n}} \\ &= Q_n(z). \end{aligned} \tag{29}$$

The condition that $K(x,y,z) \geq 0$ is not explicitly used in the proof of *(ii)*. Effectively it is replaced by the fact that the correlations are mixtures of extreme point Bernoulli correlations $-q/p$ and 1.

(iii) In the generalized Ehrenfest (p) urn, label the balls $1,\dots,N$. Before a given transition let ξ_i be the indicator function as to whether the ith ball is red. Let η_i be a similar indicator function for red balls after a transition is made. Balls change colour according to the transition matrix P given by (6). Under P the regression of η_i on ξ_i is

$$\mathbb{E}\big[\eta_i \mid \xi_i\big] = P(\eta_i = 1 \mid \xi_i) = p - (q/p)(\xi_i - p),$$

in agreement with the second column in P when $\xi_i = 0$ or 1. If z balls are chosen to change colour then it follows that the transition distribution is identical to the distribution described by *(ii)* with an extreme point correlation sequence (29). □

The examples of bivariate Binomial distributions in the Introduction have correlation sequences and mixing distributions shown in Table 1. Mixing distributions are straightforward to calculate by taking the joint *pgf* of (X,Y), $\mathbb{E}\big[s^X t^Y\big]$, setting $s=0$, then rescaling $\mathbb{E}\big[t^Y I\{X=0\}\big]$ to be a *pgf*.

Table 1: Correlation sequences and mixing distributions.

Model	ρ_n	$a(z)$
Contingency table	$\text{corr}(X,Y)^n$	$b(z;N,p_{01}/q)$
Random elements in common	$M_{[n]}/N_{[n]}$	$b(z;N-M,p)$
Cumulative Binomial trials	$(1-\beta/\gamma)^n$	$b(z;N,\beta)$
Ehrenfest (p) urn	$Q_n(1)=-(n/N-p)/p$	δ_{z1}

In the next two theorems we relax the assumption of exchangeability of (X,Y) by considering different trial parameters N,M in Theorem 2, and different Bernoulli probabilities p_1,p_2 in Theorem 3. The non-negativity of three triple product sums with non-identical Krawtchouk polynomials (33), (45) and (48) is of independent interest.

Theorem 2. *Two equivalent characterizations of bivariate random variables (X,Y) with Binomial (N,p) and Binomial (M,p) ($M\le N$) marginal distributions and Krawtchouk polynomial eigenfunctions are the following.*

(i) The distribution of (X,Y) has a Lancaster expansion

$$b(x;N,p)b(y;M,p)\Big\{1+\sum_{n=1}^{M}\rho_n\sqrt{h_n(N,p)h_n(M,p)}Q_n(x;N,p)Q_n(y;M,p)\Big\}. \tag{30}$$

for $x=0,1,\dots,N$, $y=0,1,\dots,M$. The correlation sequence $\{\rho_n\}_{n=0}^M$ has a representation

$$\rho_n=\sqrt{M_{[n]}/N_{[n]}}\sum_{z=0}^{M}a(z)Q_n(z;M,p) \tag{31}$$

where $\{a(z)\}_{z=0}^M$ is a probability distribution.

(ii) There is a representation

$$X=\sum_{i=1}^{N}\xi_i,\ Y=\sum_{i=1}^{M}\eta_i, \tag{32}$$

where $\{(\xi_i,\eta_i)\}_{i=1}^M$ are Bernoulli (p) pairs of random variables, independent within marginal sequences $\{\xi_i\}_{i=1}^M$ and $\{\eta_i\}_{i=1}^M$, with random correlations $\{\tau_i\}_{i=1}^M$ such that random variables within and between the pairs $\{(\xi_i,\eta_i)\}_{i=1}^M$ are conditionally independent given $\{\tau_i\}_{i=1}^M$. The distribution of $\{\tau_i\}_{i=1}^M$ can be always be taken to be exchangeable with correlations in the two-point set

$\{-q/p, 1\}$. $\{\xi_i\}_{i=M+1}^N$ *is a Bernoulli* (p) *sequence of random variables independent of* $\{(\xi_i, \eta_i)\}_{i=1}^M$.

Proof. *(i)* A triple product sum is again important in the proof.

$$\begin{aligned} & K(x,y,z;N,p,M,p) \\ = & \sum_{n=0}^{M} h_n(M,p)Q_n(x;N,p)Q_n(y;M,p)Q_n(z;M,p) \qquad (33) \\ = & q^{-M}\sum_{n=0}^{M} b(n;M,p)Q_x(n;N,p)Q_y(n;M,p)Q_z(n;M,p) \\ = & q^{-M}\mathbb{E}\Big[(Q_x(W;N,p)Q_y(W;M,p)Q_z(W;M,p)\Big], \qquad (34) \end{aligned}$$

for $x = 0, 1, \dots, N$, $y, z = 0, 1, \dots, M$, where W has a Binomial (M, p) distribution. In the symmetric function representation for the Krawtchouk polynomials choose $W = \sum_{i=1}^M \xi_i$ and $\xi_j = 0$ for $j = M + 1, \dots, N$. $Q_x(W; N, p)$ is then proportional to the xth symmetric product in the N variables $\{v(\xi_i)\}_{i=0}^N$ and $Q_y(W; M, p), Q_z(W; M, p)$ are symmetric products in the first M variables. Then $K(x, y, z; N, p, M, p) \geq 0$ from a similar argument to that used in (17) and by noting $v(0) = 1 > 0$. Mixing $K(x, y, z; N, p, M, p)$ over z with a distribution $\{a(z)\}_{z=0}^M$ completes the sufficiency of *(ii)* and the necessity follows by identifying $\{a(z)\}_{z=0}^M$ as the distribution of Y given $X = 0$.

To show the equivalence of *(ii)* and *(i)* note that with $X_M = \sum_{i=1}^M \xi_i$,

$$\mathbb{E}\Big[\sqrt{h_n(N,p)}Q_n(X;N,p) \mid X_M\Big] = \sqrt{M_{[n]}/N_{[n]}}\sqrt{h_n(M,p)}Q_n(X_M;M,p),$$

so

$$\begin{aligned} \rho_n &= \mathbb{E}\Big[\sqrt{h_n(N,p)h_n(M,p)}Q_n(X;M,p)Q_n(Y;M,p)\Big] \\ &= \sqrt{M_{[n]}/N_{[n]}}\mathbb{E}\Big(\sqrt{h_n(N,p)h_n(M,p)}Q_n(X_M;M,p)Q_n(Y;N,p)\Big) \\ &= \sqrt{M_{[n]}/N_{[n]}}\sum_{z=0}^{M} a(z)Q_n(z;M,p). \end{aligned}$$

□

In a Bernoulli pair (ξ, η) with non-identical marginal distributions the joint distribution of the pair is

$$f_{\xi,\eta} = p_{1\xi}p_{2\eta}\Big\{1 + \tau\frac{\xi - p_1}{\sqrt{p_1q_1}} \times \frac{\eta - p_2}{\sqrt{p_2q_2}}\Big\},\ \xi, \eta \in 0, 1 \qquad (35)$$

with $p_{1\xi}, p_{2\eta}$ marginal Bernoulli p_1, p_2 distributions, and τ the correlation coefficient. Arrange indexing so that without loss of generality

$$p_1 \geq q_1,\ p_2 \geq q_2, \text{ and } p_1 \geq p_2.$$

Then $f_{\xi,\eta} \geq 0,\ \xi, \eta \in \{0, 1\}$ if and only if

$$-\sqrt{\frac{q_1 q_2}{p_1 p_2}} \leq \tau \leq \sqrt{\frac{q_1 p_2}{p_1 q_2}}. \tag{36}$$

Theorem 3. *Three equivalent characterizations of random variables (X, Y) with Binomial (N, p_1) and Binomial (N, p_2) marginal distributions ($p_1 \geq q_1,\ p_2 \geq q_2$ and $p_1 \geq p_2$) and Krawtchouk polynomial eigenfunctions are the following.*

(i) The distribution of (X, Y) has a Lancaster expansion

$$p(x; y) = b(x; N, p_1) b(y; N, p_2) \times \Big\{1 + \sum_{n=1}^{N} \rho_n \sqrt{h_n(N, p_1) h_n(N, p_2)} Q_n(x; N, p_1)) Q_n(y; N, p_2)\Big\}, \tag{37}$$

for $x, y = 0, 1, \ldots, N$. The correlation sequence $\{\rho_n\}_{n=0}^N$ has a representation

$$\rho_n = \Big(\frac{p_2 q_1}{q_2 p_1}\Big)^{n/2} \sum_{z=0}^{N} a(z) Q_n(z; N, p_2), \tag{38}$$

where $\{a(z)\}_{z=0}^N$ is a probability distribution.

(ii) There is a representation

$$(X, Y) = \sum_{i=1}^{N} (\xi_i, \eta_i), \tag{39}$$

where $\{(\xi_i, \eta_i)\}_{i=1}^N$ are Bernoulli (p_1, p_2) pairs of random variables, independent within marginal sequences $\{\xi_i\}_{i=1}^N$ and $\{\eta_i\}_{i=1}^N$, with random correlations $\{\tau_i\}_{i=1}^N$ such that random variables within and between the pairs $\{(\xi_i, \eta_i)\}_{i=1}^N$ are conditionally independent given $\{\tau_i\}_{i=1}^N$. The distribution of

$\{\tau_i\}_{i=1}^N$ can be always be taken to be exchangeable with correlations in the two-point set

$$\Big\{-\sqrt{\frac{q_1q_2}{p_1p_2}},\sqrt{\frac{q_1p_2}{p_1q_2}}\Big\}. \tag{40}$$

(iii) Y given $X = x$ is distributed as the transition function in a generalized Ehrenfest (p_1, p_2) urn with blue and red balls. z balls are chosen and colour changes are made according to a transition matrix P_1, and in the remaining $N - z$ balls colour changes are made according to a transition matrix P_2.

$$P_1 = \begin{pmatrix} 0 & 1 \\ \frac{q_2}{p_1} & 1-\frac{q_2}{p_1} \end{pmatrix} \quad P_2 = \begin{pmatrix} 1 & 0 \\ 1-\frac{p_2}{p_1} & \frac{p_2}{p_1} \end{pmatrix} \tag{41}$$

The distribution $\{a(z)\}_{z=0}^N$ has an interpretation in the three characterizations of:
(i)

$$\rho_n = \Big(\frac{p_2q_1}{q_2p_1}\Big)^{n/2}\sum_{z=0}^N a(z)Q_n(z;N,p_2),\ n=0,\dots,N, \tag{42}$$

$$a(z) = P(Y=z \mid X=0),\ z=0,1,\dots,N; \tag{43}$$

(ii) $\{a(z)\}_{z=0}^N$ is the distribution of the number of correlation coefficients $\{\tau_i\}_{i=0}^N$ in the two-point set (40) equal to the first co-ordinate.

(iii) $\{a(z)\}_{z=0}^N$ is the distribution of the number of balls drawn in each transition in a stationary generalized Ehrenfest (p_1, p_2) urn process.

Proof. We omit most of the proof since it is similar to the proof of Theorems 1 and 2. The extreme point sequences are shown to be

$$\rho_n = \Big(\frac{p_2q_1}{q_2p_1}\Big)^{n/2}Q_n(z;N,p_2), \tag{44}$$

by a similar conditional expectation as (26). Take independent Bernoulli pairs $\{(\xi_j,\eta_j)\}_{j=1}^N$ such that z pairs have correlation $-\sqrt{q_1q_2/p_1p_2}$ and $N-z$ have correlation $\sqrt{q_1p_2/p_1q_2}$. $Q_n(z;N,p_2)$ is an average of the n-th products of N variables with z values $-q_2/p_2$ and $N-z$ values 1, and

$$\sqrt{\frac{p_2q_1}{q_2p_1}}\Big\{-\frac{q_2}{p_2},1\Big\} = \Big\{-\sqrt{\frac{q_1q_2}{p_1p_2}},\sqrt{\frac{q_1p_2}{p_1q_2}}\Big\},$$

giving the first factor in (44). Although it is not needed for the proof, the form of the extreme point correlation sequences implies a non-negative triple product sum

$$\begin{aligned} &K(x,y,z;N,p_1,N,p_2)\\ &\quad=\sum_{n=0}^{N} h_n(p_2)Q_n(x;N,p_1)Q_n(y;N,p_2)Q_n(z;N,p_2)\geq 0. \end{aligned}\tag{45}$$

There is another direct proof of (45) using a similar formula to (17) and the symmetric function representation of the polynomials. Calculations needed are to show that if ξ is a Bernoulli (p_2) random variable then

$$\begin{aligned} \mathbb{E}\big[v(\xi;p_1)\big] &= \frac{p_1-p_2}{p_1}\geq 0, \text{ and}\\ \mathbb{E}\big[v(\xi;p_1)v(\xi;p_2)^2\big] &= \frac{q_2}{p_1p_2}\Big[p_1p_2-q_1q_1\Big]\geq 0. \end{aligned}\tag{46}$$

We remark that if X is Binomial (N,p_1) and Y is Binomial (M,p_2) then if (X,Y) has Krawtchouk polynomial eigenfunction

$$\begin{aligned} \rho_n &= \sqrt{\frac{h_n(M,p_2)}{h_n(N,p_1)}}\mathbb{E}\Big[Q_n(Y;M,p_2)\mid X=0\Big]\\ &= \Big(\frac{M_{[n]}}{N_{[n]}}\Big)^{1/2}\Big(\frac{p_2q_1}{p_1q_2}\Big)^{n/2}\mathbb{E}\Big[Q_n(Z;M,p_2)\Big], \end{aligned}\tag{47}$$

for a random variable Z on $0,1,\dots,M$. Under the conditions $p_1\geq q_1$, $p_2\geq q_2$ and $p_1\geq p_2$, $M\leq N$ as in Theorems 2 and 3,

$$\begin{aligned} &K(x,y,z;N,p_1,M,p_2)\\ &=\sum_{n=0}^{M} h_n(N,p_2)Q_n(x;N,p_1)Q_n(y;M,p_2)Q_n(z;M,p_2)\geq 0 \end{aligned}\tag{48}$$

because of the calculations (46). Therefore a characterization of correlation sequences of bivariate Binomial distributions with the conditions on p_1,p_2, M,N is (47). This is not completely general however, because of the assumption that $M\leq N$.

Theorems 1 and 3 characterize distributions for which there is a finite exchangeability of Bernoulli pairs $\{(\xi_i,\eta_i)\}_{i=1}^N$. If there is an infinite exchangeable sequence of Bernoulli pairs the distribution structure is much easier. Then $\{\tau_i\}_{i=1}^{\infty}$ is an infinite exchangeable sequence with entries in (40) and there exists a de Finetti representation with a measure μ on $[0,1]$ such that

$$\begin{aligned}\rho_n &= \mathbb{E}\big[\tau_1 \cdots \tau_n\big] \\ &= \int_{[0,1]} \Bigg[-\sqrt{\frac{q_1 q_2}{p_1 p_2}}\lambda + \sqrt{\frac{q_1 p_2}{p_1 q_2}}(1-\lambda) \Bigg]^n d\mu(\lambda). \qquad (49)\end{aligned}$$

That is, with infinite exchangeability of Bernoulli pairs the distribution of (X,Y) is the distribution of marginal $1,1$ counts in a contingency table of N observations where the probability of an observation falling in cell (ξ,η) is $f_{\xi,\eta}$ in (35) where τ has a random distribution in a continuous interval with endpoints in (40). If τ is fixed then (37) is the joint distribution of counts $1,1$ in a usual contingency table. □

5. Fréchet bounds

A general bivariate distribution function $F(x,y)$ with marginal distribution functions $G(x)$, $H(y)$ is bounded by Frechét's distribution functions, (Fréchet, 1951)

$$F_*(x,y) = \max\Big\{G(x) + H(y) - 1, 0\Big\} \qquad (50)$$

$$F^*(x,y) = \min\Big\{G(x), H(y)\Big\}, \qquad (51)$$

so that

$$F_*(x,y) \le F(x,y) \le F^*(x,y),\ x,y \in \mathbb{R}. \qquad (52)$$

The covariance of (X,Y) can be expressed as

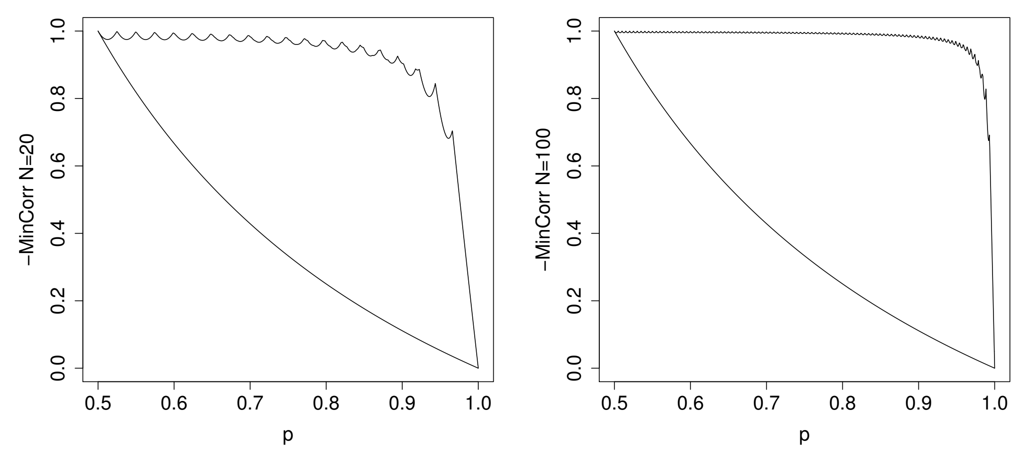
$$\mathrm{Cov}(X,Y) = \int\int \Big[F(x,y) - G(x)H(y)\Big] dx dy, \qquad (53)$$

provided the covariance is finite (Fréchet, 1951; Hoeffding, 1940). Frechét's bounds (51) and the identity (53) provide bounds for the correlation coefficient for a general (X,Y) of

$$\rho_{F_*} \le \rho_F \le \rho_{F^*}. \qquad (54)$$

In the Binomial distribution $P(X \le k) = 1 - I_p(k+1; N-k)$, where $I_p(a,b)$ is the incomplete Beta function. Much smaller correlations are achievable with the minimal bound than in distributions (5) which have a minimal possible correlation of $-q/p$ when $p \geq q$. We plot examples of the negative of the minimal correlation with $p \geq 1/2$ in Figure 2 together with a curve of q/p. The minimal correlation as a function of p is symmetric about $p = 1/2$, so a graph for $p \leq 1/2$ would be a reflection about $p = 1/2$ of the graph for $p \geq 1/2$. The non-monotonicity in p is of interest. Griffiths, Milne and Wood (1979) plot the minimal Frechét correlation in bivariate Poisson distributions with repect to the mean λ which has a similar property. Presumably the non-monotonicity occurs because of discreteness as it does for confidence intervals for p from the Binomial distribution, Brown, Cai and DasGupta (2001).

Figure 2: Minimal Correlation plots in a Bivariate Binomal $(p, -\rho_{F_*})$
$N = 20, 100$



6. Continuous time Markov processes with Binomial stationary distributions

A reversible Markov process $\{X(t)\}_{t\geq 0}$ with Binomial stationary distribution has Krawtchouk polynomial eigenfunctions if and only if the transition functions, for $y = 0, 1, \ldots, N$ have the form

$$f(x, y; t) = b(y; N, p)\Big\{1 + \sum_{n=1}^{N} e^{-d_n t} h_n Q_n(x) Q_n(y)\Big\} \tag{55}$$

with

$$d_n = \sum_{z=0}^{N} b(z)\Big(1 - Q_n(z)\Big) \tag{56}$$

where $b(z) \geq 0, z = 0, 1, \ldots, N$. We have not seen this result previously, but it is similar to a characterization in Bochner (1954) and follows the construction (4) in the Introduction. $\{X(t)\}_{t\geq 0}$ can be constructed from a subordinated Ehrenfest urn process. Let $\{X_k^\circ\}_{k\in\mathbb{N}}$ be a generalized Ehrenfest urn model with a mixing distribution for z, the number of balls chosen to change, of $b(z)/\sum_{j=1}^N b(j)$ and $\{N(t)\}_{t\geq 0}$ an independent Poisson process of rate $\lambda = \sum_{j=1}^N b(j)$. Then

$$\{X(t)\}_{t\geq 0} =^{\mathcal{D}} \{X_{N(t)}^\circ\}_{t\geq 0}$$

has transition functions (55).

Limit distributions

Normal limit

Let $X^{(N)}$ be Binomial (N, p). Then $(X^{(N)} - Np)/\sqrt{Npq}$ converges in distribution to a $N(0,1)$ distribution. Let $x^{(N)} = v^{(N)}\sqrt{Npq} + Np$ be a sequence such that $v^{(N)} \to v \in (-\infty, \infty)$ as $N \to \infty$. It is a standard result that

$$h_n^{1/2} Q_n(x^{(N)}) \to \frac{(-1)^n}{\sqrt{n!}} H_n(v)$$

where $\{H_n(v)\}_{n\in\mathbb{N}}$ are the Hermite-Chebycheff polynomials with generating function

$$\sum_{n=0}^\infty H_n(v)\frac{\zeta^n}{n!} = e^{\zeta v - \frac{1}{2}\zeta^2}.$$

In the generalized Ehrenfest urn models the extreme processes have transition functions

$$P(X_k = y \mid X_0 = x) = \binom{N}{y} p^y q^{N-y}\Big\{1 + \sum_{n=1}^N Q_n^k(z) h_n Q_n(x) Q_n(y)\Big\}.$$

Let $x^{(N)} = u^{(N)}\sqrt{Npq} + Np$, $y^{(N)} = v^{(N)}\sqrt{Npq} + Np$, $k^{(N)} = [Nt]$, so that as $N \to \infty$, $x^{(N)} \to x$, $y^{(N)} \to y$. Since

$$Q_n^{[Nt]}(z) \sim \Big(1 - \frac{qnz}{pN}\Big)^{[Nt]} \to e^{-(q/p)nzt},$$

then

$$\{\xi_N(t)\}_{t\geq 0} = \Big\{\frac{Y([Nt]) - Np}{\sqrt{Npq}}\Big\}_{t\geq 0}$$

converges in distribution to the Ornstein-Uhlenbeck process $\{\xi(t)\}_{t\geq 0}$ with transition density

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}\Big\{1+\sum_{n=1}^{\infty}e^{-(q/p)nzt}(n!)^{-1}H_n(u)H_n(v)\Big\},$$

where $\xi(0)=u,\xi(t)=v$.

Poisson limit

Let $X^{(N)}$ be Binomial (N,p). Take $N\to\infty$, with $q^{(N)}\to 0$, $Nq^{(N)}\to\lambda$, $x_N\to x\in[0,\infty)$. We suppose that $p\geq q$, so we need to take $q\to 0$. Then

$$\binom{N}{n}Q_n(N-x^{(N)};N,1-\lambda/N)\to\frac{(-\lambda)^n}{n!}C_n(x;\lambda),$$

where $\{C_n(x;\lambda)\}_{n\in\mathbb{N}}$ are the Poisson-Charlier orthogonal polynomials with generating function

$$\sum_{n=0}^{\infty}C_n(x;\lambda)\frac{z^n}{n!}=\big(1-z/\lambda\big)^x e^z.$$

To obtain a limit in the generalized Ehrenfest urn model extreme points measure time in units of N^2. Let $X'_k=N-X_k$, $k^{(N)}=[N^2t]$. Then

$$Q_n^{[N^2t]}(z)\sim\Big(1-\frac{n\lambda z}{N^2}\Big)^{[N^2t]}\to e^{-n\lambda zt}.$$

The transition distribution of $X'_{[N^2t]}$ converges to

$$e^{-\lambda}\frac{\lambda^v}{v!}\left\{1+\sum_{n\geq 1}e^{-n\lambda zt}\frac{\lambda^n}{n!}C_n(u;\lambda)C_n(v;\lambda)\right\},\ v=0,1,\ldots.$$

The limit transition functions are those of a $M/M/\infty$ Queue, with arrival and service rates $\lambda,1$ and a time change $t\to tz$. See, for example, Schoutens (2000). Griffiths (2009) shows that every reversible Markov process with stationary Poisson distribution and Poisson-Charlier polynomial eigenfunctions has the distribution of a subordinated $M/M/\infty$ queue.

Eigenvalue characterizations of Markov processes

Eagleson's characterization of eigenvalues of bivariate Binomial distributions with Krawtchouk polynomial eigenfunctions is that they are of the form (7). The eigenvalues $\{c_n(t)\}_{n=0}^N$ of the transition functions in a reversible continuous time Markov chain with Binomial stationary distribution and Krawtchouk polynomial eigenfunctions have a Bochner style characterization as being of the form (56), which can be written as

$$c_n(t) = \exp\Big\{-b(0)nt - t\sum_{z=1}^{N} b(z)\big(1-Q_n(z)\big)\Big\}, \tag{57}$$

where $\{b_n\}_{n=0}^N$ are non-negative constants. Sarmanov and Bratoeva (1967) prove that $\{\rho_n\}_{n\in\mathbb{N}}$ is a correlation sequence in a bivariate distribution (1) with standard normal marginal distributions and $\{P_n\}$ the orthonormal Hermite-Chebycheff polynomials if and only if

$$\rho_n = \mathbb{E}\big[Z^n\big], \tag{58}$$

where Z is a random variable on $[-1,1]$. A characterization of the eigenvalues $\{c_n(t)\}_{n\in\mathbb{N}}$ of a reversible Markov process with Normal stationary distribution and Hermite-Chebycheff eigenfunctions is that

$$c_n(t) = \exp\Big\{-t\int_{-1}^{1}\frac{1-\rho^n}{1-\rho}G(d\rho)\Big\}, \tag{59}$$

where G is a finite measure on $[-1,1]$. If G has an atom of a at 1, then the point at 1 contributes an in the integral. The proof again follows Bochner (1954). Griffiths (2009) shows that when G has support only on $[0,1]$ then such reversible Markov process with stationary Normal distribution and Hermite-Chebycheff polynomial eigenfunctions have the distribution of a subordinated Ornstein-Uhlenbeck process. Our interest is in a limit construction from the Binomial sequences (7), (57), to the Normal sequences (58), (59). We take $p = q = 1/2$. The construction is straightforward by noting that as $N \to \infty$ with $p = q = 1/2$, $-1 \le \phi \le 1$,

$$Q_n\Big(\Big[N\frac{1-\phi}{2}\Big], N, \frac{1}{2}\Big) \to \phi^n.$$

Sequences $\{a^N(x)\}$ and $\{b^N(z)\}$ can now be chosen, depending on N, so that (7) and (57) converge to (58) and (59). The Poisson case is quite similar, with distributions on $[0,1]$ instead of $[-1,1]$. As $q \to 0$, $N \to \infty$ and $Nq \to \lambda$, with $0 \le \phi \le 1$

$$Q_n([N(1-\phi)], N, p) \to \phi^n.$$

The construction is then similar to the Normal case.

7. An Ehrenfest Urn with N balls of d colours

An analogy of the 2-colour ball generalized Ehrenfest urn scheme to $d \ge 2$ colours is to recolour balls in a transition according to permutation matrices. In the 2-colour ball urn, changing colours of z balls in a transition amounts to a transposition of colours for z balls and an identity permutation of colours on the remaining $N - z$ balls. Denote the $d!$ 0-1 permutation matrices by $\{P_i\}_{i=1}^{d!}$ and let the permutations represented by the matrices be $\{\sigma_i\}_{i=1}^{d!}$, so that in the ith permutation matrix $P_{ikl} = 1$ is equivalent to $\sigma_i(k) = l$. Choose $\sigma_{d!}$ as the identity permutation. The extreme points of $d \times d$ transition matrices which are doubly stochastic are the $d!$ 0-1 permutation matrices. In the urn model transitions are made by choosing balls in groups $j = (j_1, \dots, j_{d!})$, where $|j| = N$ (with notation $|j| = \sum_i j_i$), then changing their colours within groups in a deterministic way according to the permutations so that a ball of colour c in group i is mapped to colour $\sigma_i(c)$. The mapping in the last group is the identity. A general transition distribution is found by mixing over the group numbers j with probability distribution $\{a(j)\}_{\{j:|j|=N\}}$. Let Z_{ic} be the number of balls of colour c drawn in group i. The *pgf* of the transition distribution is

$$\sum_{|j|=N} a(j) \prod_{i=1}^{d!} \mathbb{E}\Big[\prod_{c=1}^{d} s_{\sigma_i(c)}^{Z_{ic}} \mid j, \{x_c\}\Big], \tag{60}$$

where if there are totals of $\{x_c\}_{c=1}^{d}$ numbers of balls of different colours in the urn before change, then $\{Z_{ic}\}$ has distribution

$$\frac{\prod_{c=1}^{d} \binom{x_c}{z_{1c}, \dots, z_{d!c}}}{\binom{N}{j_1, \dots, j_{d!}}}, \tag{61}$$

with $z_{\cdot c} = x_c$ and $z_{i\cdot} = j_i$, $c = 1, \dots, d$, $i = 1, \dots, d!$. After a transition the number of balls of colour c is $Y_c = \sum_{i=1}^{d!} Z_{i\sigma_i^{-1}(c)}$. The chain is not necessarily reversible. If balls are labelled and a single ball followed through transitions is clear that it will have a uniform stationary distribution on the d colours. It is thus intuitive that the stationary distribution of ball colours in the urn is Multinomial. This is seen formally by showing that the expected value of the *pgf* (60) when $\{x_c\}_{c=1}^d$ has a Multinomial distribution is again Multinomial. If $\{x_c\}_{c=1}^d$ has a Multinomial $(N, d^{-1}, \dots, d^{-1})$ distribution then $\{Z_{ic}\}$ given j has an independent Multinomial distribution within groups

$$\prod_{i=1}^{d!} \binom{j_i}{z_{i1}, \dots, z_{id}} d^{-j_i}, \tag{62}$$

and therefore the unconditional *pgf* (60) is

$$\sum_{|j|=N} a(j) \prod_{i=1}^{d!} \Big[\sum_{c=1}^{d} d^{-1} s_{\sigma_i(c)} \Big]^{j_i} = \Big[\sum_{c=1}^{d} d^{-1} s_c \Big]^N, \tag{63}$$

since for each i, $\sum_{c=1}^d s_{\sigma_i(c)} = \sum_{c=1}^d s_c$. This shows that there is a Multinomial stationary distribution. The joint *pgf* of (X, Y) is similarly

$$\begin{aligned} G(t,s) &= \sum_{\{j:|j|=N\}} a(j) \prod_{i=1}^{d!} \Big[\sum_{c=1}^{d} d^{-1} t_c s_{\sigma_i(c)} \Big]^{j_i} \\ &= \sum_{\{j:|j|=N\}} a(j) \prod_{i=1}^{d!} \Big[\sum_{c=1}^{d} d^{-1} t_{\sigma_i^{-1}(c)} s_c \Big]^{j_i}. \end{aligned} \tag{64}$$

The form of (64) implies that

$$\mathbb{E}\Big[\prod_{c=1}^{d} Y_c^{r_c} \mid X \Big]$$

is a polynomial of degree $|r|$ in X, and similarly for the conditional moments of X given Y. This implies that the eigenfunctions of (X, Y) are orthogonal polynomials in X and Y, however their form and the corresponding eigenvectors are generally not clear.

In the special case where permutations $\{\sigma_i\}_{i=1}^{d!}$ are chosen independently with probabilities $\{\alpha_i\}_{i=1}^{d!}$, and $\{a(j)\}_{\{j:|j|=N\}}$ has a Multinomial (N, α) distribution then (64) becomes

$$\Big[\sum_{c=1}^{d} \sum_{i=1}^{d!} \alpha_i d^{-1} t_c s_{\sigma_i(c)} \Big]^N = \Big[\sum_{c=1}^{d} \sum_{k=1}^{d} d^{-1} P_{ck} t_c s_k \Big]^N, \tag{65}$$

where

$$P = \sum_{k=1}^{d!} \alpha_k P_k, \tag{66}$$

the extreme point representation of a doubly stochastic transition matrix. We know that the distribution (65) has eigenfunctions which are orthogonal polynomials on the Multinomial $(N, d^{-1}, \dots, d^{-1})$ distribution (Griffiths, 1971; Zhou and Lange, 2009). If $\{\alpha_k\}_{k=1}^{d!}$ has a random distribution it is still not clear what the eigenvector, eigenvalue structure is.

The eigenstructure for a collection of generalized Ehrenfest urns has recently been calculated by Mizukawa (2010). He works things out in three scenarios. In the first, a ball is selected uniformly at random and placed into a uniformly chosen different urn. In the second and third, the urns are arranged cyclically. A randomly chosen ball is dropped into the following urn or into an adjacent urn. In each case, there is a group-theoretic symmetry so that the machinery of Gelfand pairs can be used. One interesting feature is the appearance of complex reflection groups.

The next theorem connects d-colour ball generalized Ehrenfest urns with bivariate Multinomial random variables (X, Y) which are the sum of pairs of indicator random variables in N trials, each with a random bivariate distribution. This structure can be regarded as an extension of the 2-colour case where the Bernoulli random variables in the sum have random correlations, because in a 2×2 distribution knowing the correlation and fixed marginal distributions is equivalent to knowing the bivariate distribution. The extreme points have a *pgf* of the form (64) when $\{a(j)\}_{\{j:|j|=N\}}$ puts weight on a single j.

Theorem 4. *Let (X, Y) have a bivariate d-dimensional Multinomial distribution such that*

$$P(X = x) = \binom{N}{x} \frac{1}{d^N},\ x \in \mathbb{N}^d, |x| = N,$$

and similarly for Y. Two equivalent classes of bivariate distributions with these marginal distributions are the following.

(i)

$$(X, Y) = \sum_{i=1}^{N} (\xi_i, \eta_i),$$

where $\{(\xi_i, \eta_i)\}_{i=1}^{N}$ are d-dimensional pairs of random vectors, independent within marginal sequences $\{\xi_i\}_{i=1}^{N}$ and $\{\eta_i\}_{i=1}^{N}$ which are uniformly dis-

tributed on the unit vectors $\{e_k\}_{k=1}^d$. *The pairs have random bivariate distributions expressed as matrices* $\{Q_i\}_{i=1}^N$, *such that they are conditionally independent given* $\{Q_i\}_{i=1}^N$. *The distribution of* $\{Q_i\}_{i=1}^N$ *can always be taken to be exchangeable in the set of permutation distributions* $\{d^{-1}P_j\}_{j=1}^{d!}$.
(ii) (X,Y) *is distributed as* (X_t, X_{t+1}) *for all* $t \in \mathbb{N}$ *in a stationary* d*-dimensional generalized Ehrenfest urn process with* pgf *(64).*

Proof.
(ii) $\Longrightarrow$ (i). If (X,Y) has a *pgf* (64) then take the random matrices $\{Q_i\}_{i=1}^N$ to have a permutation distribution on j_i copies of $d^{-1}P_i$, $i = 1,\dots,d!$ where j has the distribution $\{a(j)\}_{\{j:|j|=N\}}$.
(i) $\Longrightarrow$ (ii). The *pgf* of (X,Y) in (i) is

$$\mathbb{E}\Big[\prod_{i=1}^N t'Q_i s\Big], \tag{67}$$

where t and s are d-dimensional and prime denotes transpose. Expectation is over the random matrices. There is a decomposition

$$Q_i = \sum_{k=1}^{d!} b_{ik} d^{-1} P_k,$$

where B is a non-negative random $N \times d!$ matrix with row sums unity. The *pgf* (67) is therefore

$$\mathbb{E}\Big[\prod_{i=1}^N \sum_{k=1}^{d!} b_{ik} d^{-1} t' P_k s\Big]. \tag{68}$$

Let $\{V_i\}_{i=1}^N$ be random variables, conditionally independent given B such that for $i = 1,\dots N$

$$P(V_i = k \mid B) = b_{ik},\ k = 1,\dots,d!.$$

Let M_k be the number of entries in $\{V_i\}_{i=1}^N$ equal to k and $M = (M_k)$. Then we can identify the distribution

$$a(j) = P(M = j),$$

and (68) is equal to (64). □

If there is an infinite exchangeable sequence $\{(\xi_j,\eta_j)\}_{j=1}^\infty$ with fixed marginal distributions within each pair, then there is a de Finetti representation such that the counts (N_{ij}) in cells (i,j) are distributed as in a contingency table with random probabilities of falling into cells. That is, there is a random $d \times d$ doubly stochastic matrix Φ with

$$P\big((n_{ij})\big) = \binom{N}{(n_{ij})} \mathbb{E}\Big[\prod_{i,j\in[d]} \Phi_{ij}^{n_{ij}}\Big] d^{-N}.$$

The joint *pgf* of (X, Y) is then

$$\mathbb{E}\Big[\sum_{i=1}^{d}\sum_{j=1}^{d} d^{-1}\Phi_{ij} s_i t_j\Big]^N. \tag{69}$$

As we expect a characterization of the distribution of (X, Y) by (69) under infinite exchangeability of $\{(\xi_j, \eta_j)\}_{j=1}^{\infty}$ is much easier than the characterization in Theorem 4 under finite exchangeability.

We briefly consider exchangeability of (X, Y) with joint *pgf* (64), corresponding to a reversible urn process. Label the $d!$ permutations so that the last L, (say) $\pi_1, \ldots, \pi_L$ are idempotent and the 1st $2M$ are arranged as

$$\sigma_1, \sigma_1^{-1}, \ldots, \sigma_M, \sigma_M^{-1}, \pi_1, \ldots, \pi_L,$$

where M, L are determined constants, with $2M + L = d!$. Then indexing the ordered sequence $j = (j_1, \ldots, j_{d!})$ by permutations instead of $1, \ldots, d!$ as

$$j = (j_{\sigma_1}, j_{\sigma_1^{-1}}, \ldots, j_{\sigma_M}, j_{\sigma_M^{-1}}, j_{\pi_1}, \ldots, j_{\pi_L}),$$

the distribution of (X, Y) is exchangeable if and only if $\{a(j)\}_{\{j:|j|=N\}}$ only depends on unordered pairs and the idempotent permutations

$$\{j_{\sigma_1}, j_{\sigma_1^{-1}}\}, \ldots, \{j_{\sigma_M}, j_{\sigma_M^{-1}}\}, j_{\pi_1}, \ldots, j_{\pi_L}.$$

This is seen by writing the *pgf* $G(t, s)$ (64) as

$$\sum_{|j|=N} a(j) \prod_{i=1}^{M} \Big[\sum_{c=1}^{d} d^{-1} t_c s_{\sigma_i(c)}\Big]^{j_{\sigma_i}} \Big[\sum_{c=1}^{d} d^{-1} t_{\sigma_i(c)} s_c\Big]^{j_{\sigma_i^{-1}}} \prod_{i=1}^{L} \Big[\sum_{c=1}^{d} d^{-1} t_c s_{\pi_i(c)}\Big]^{j_{\pi_i}}, \tag{70}$$

and noting that $G(t, s) = G(s, t)$ if and only if $\{(j_{\sigma_i}, j_{\sigma_i^{-1}})\}_{j=1}^{M}$ is exchangeably distributed within pairs. Also in the idempotent permutation terms

$$\sum_{c=1}^{d} t_c s_{\pi_i(c)} = \sum_{c=1}^{d} t_{\pi_i(c)} s_c.$$

The easiest case is when $\{a(j)\}_{\{j:|j|=N\}}$ just has weight on idempotent permutations.

As an example, if $d=3$ and the $3!=6$ permutations are ordered as

$$(231, 312), 321, 213, 132, 123$$

then $a(j_1,\dots,j_6)$ has to be exchangeable in j_1, j_2.

7.1. Joint Multinomial distributions and Markov chains

Let (X, Y) be two R and C dimensional random vectors with a *pgf* of

$$\Big(\sum_{i=1}^{R}\sum_{j=1}^{C} p_{ij}^{R,C} s_i t_j\Big)^N, \tag{71}$$

which is the *pgf* of the joint marginal counts in a $R \times S$ contingency table with N observations, where the probability of an observation falling in cell (i,j) is $p_{ij}^{R,C}$. The distribution of (X, Y) has a Lancaster expansion in orthogonal polynomials on the Multinomial distribution (Griffiths, 1971; Zhou and Lange, 2009) that we now describe briefly.

Let $\{u_j^{(l)}\}_{l=0}^{d-1}$ be a complete set of orthonormal functions on a probability distribution $\{p_j\}_{j=1}^{d}$ with $u_j^{(0)}=1$ for all j such that for $k,l=0,1,\dots,d-1$

$$\sum_{j=1}^{d} u_j^{(k)} u_j^{(l)} p_j = \delta_{kl}.$$

Define a collection of orthogonal polynomials on the Multinomial distribution

$$m(x, \{p_j\}) = \binom{N}{x}\prod_{j=1}^{d} p_j^{x_j}$$

$\big\{Q_n(X;\{u^{(l)}\})\big\}$, with $n=(n_1,\dots n_{d-1})$, and $|n|\le N$, as the coefficients of $w_1^{n_1}\cdots w_{d-1}^{n_{d-1}}$ in the generating function

$$G(x,w;\{u^{(l)}\}) = \prod_{j=1}^{d}\Big(1+\sum_{l=1}^{d-1} w_l u_j^{(l)}\Big)^{x_j}. \tag{72}$$

It is straightfoward to show, by using the generating function, that

$$\mathbb{E}\Big[Q_n(X;\{u^{(l)}\})Q_{n'}(X;\{u^{(l)}\})\Big] = \delta_{nn'}\binom{N}{|n|}\binom{|n|}{n}. \tag{73}$$

The transform of $Q_n(X;\{u^{(l)}\})$ is equal to

$$\mathbb{E}\Big[\prod_{j=1}^{d}\phi_j^{X_j}Q_n(X;\{u^{(l)}\})\Big]$$
$$= \binom{N}{|n|}\binom{|n|}{n}T_0(\phi)^{N-|n|}T_1(\phi)^{n_1}\cdots T_{d-1}(\phi)^{n_{d-1}}, \qquad (74)$$

where

$$T_i(\phi) = \sum_{j=1}^{d} p_j\phi_j u_j^{(i)},\ i = 0,\dots,d-1.$$

Let $Z_1,\dots,Z_N$ be independent identically distributed random variables such that

$$P(Z = k) = p_k,\ k = 1,\dots,d.$$

Then with

$$X_i = |\{Z_k : Z_k = i, k = 1,\dots,d\}|,$$
$$G(X,w;\{u^{(l)}\}) = \prod_{k=1}^{N}\Big(1+\sum_{l=1}^{d-1}w_l u_{Z_k}^{(l)}\Big). \qquad (75)$$

We also have from (75) that

$$Q_n(X;\{u^{(l)}\}) = \sum_{\{A_l\}}\prod_{k_1\in A_1}u_{Z_{k_1}}^{(1)}\cdots\prod_{k_{d-1}\in A_{d-1}}u_{Z_{k_{d-1}}}^{(d-1)}, \qquad (76)$$

where summation is over all partitions of subsets of $\{1,\dots,N\}$, $\{A_l\}$ such that $|A_l| = n_l$, $l = 1,\dots,d$. This is an analogue of the symmetric function representation (15) for the Krawtchouck polynomials. $Q_n(x;\{u^{(l)}\})$ is a polynomial of degree n in $(S_1(x),\dots,S_{d-1}(x))$, with

$$S_k(x) = \sum_{j=1}^{d} u_j^{(k)}x_j,\ k = 1,\dots,d-1$$

whose only term of maximal degree $|n|$ is $\prod_1^{d-1}S_k^{n_k}$. Another form is

$$S_k(X) = \sum_{j=1}^{N} u_{Z_j}^{(k)},\ k = 1,\dots,d-1.$$

Let

$$p_{ij}^{R,C} = p_{i\cdot}p_{\cdot j}\Big\{1+\sum_{l=1}^{\min(R,C)-1}\rho_l u_i^{(l)}v_j^{(l)}\Big\},\ i = 1,\dots,R,\ j = 1,\dots,C$$

be a Lancaster expansion in orthonormal functions $\{u^{(l)}\}$, $\{v^{(l)}\}$ on the respective marginal distributions $\{p_{i\cdot}\}$, $\{p_{\cdot j}\}$. Then the joint distribution of (X,Y) has a Lancaster expansion

$$\begin{aligned} P(x,y) &= m(x,\{p_{i\cdot}\})m(y,\{p_{\cdot j}\}) \\ &\quad \times \Big\{1 + \sum_{\{n:1\leq |n|\leq N\}} \gamma_n Q_n^\circ(x;\{u^{(l)}\})Q_n^\circ(y;\{v^{(l)}\})\Big\}, \end{aligned} \tag{77}$$

where $\gamma_n = \prod_{k=1}^{d-1}\rho_k^{n_k}$, and $\{Q_n^\circ\}$ are the polynomials scaled to be orthonormal, that is

$$Q_n^\circ = \frac{Q_n}{\sqrt{\binom{N}{|n|}\binom{|n|}{n}}}.$$

7.2. d-urn Ehrenfest urn models

Mizukawa (2010) finds the eigenfunctions for several multivariate Ehrenfest models. In his models there are d urns and N balls. For a correspondence with our model, numbers of balls in the d urns are thought of as numbers of d distinctly coloured balls. Three different transition schemes are considered where a ball is selected at random, then in the different models the ball is moved into: *(a)* a randomly chosen different urn; *(b)* the next urn right (mod d); or *(c)* one of the two adjacent urns (mod d) with equal probability.

An extended model where transitions of a single ball from colour i to colour j are given in a $d\times d$ transition probability matrix P with stationary distribution $\{p_i\}$ is now considered. Denoting x as the d-dimensional colour configuration in the urns before a transition and Y after a transition, Y has a conditional *pgf*

$$\sum_{i,j=1}^{d} \frac{x_i}{N} p_{ij} t_j \prod_{l=1}^{d} t_l^{x_l-\delta_{il}}. \tag{78}$$

If X has a Multinomial $(N,\{p_i\})$ distribution then the joint *pgf* of (X,Y), from (78), is

$$\Big(\sum_{i,j=1}^{d} s_i p_i p_{ij} t_j\Big)\Big(\sum_{k=1}^{d} p_k s_k t_k\Big)^{N-1}. \tag{79}$$

The d-dimensional chain is seen to have a Multinomial stationary distribution from (79) by setting $s_i = 1$, $i = 1,\dots,d$ producing an unconditional Multinomial *pgf* for Y of $\Big(\sum_{k=1}^{d} p_k t_k\Big)^N$. Transition matrices in Mizukawa's schemes are:

(a) $P = \big((1-\delta_{ij})/(d-1)\big)$, with a representation (66) of $P = N_d^{-1}\sum_{k=1}^{N_d} P_k'$, where $\{P_k'\}$ is the set of $N_d = d!\sum_{r=0}^{d}(-1)^r\frac{1}{r!}$ permutation matrices with diagonal entries zero.
(b) $P = P^+$, a permutation matrix of a complete directed cycle through $1,\dots,d$ where $p^+_{i\ i+1(\mathrm{mod}\ j)} = 1$ and $p^+_{ij} = 0$ otherwise.
(c) $P = (P^+ + P^-)/2$, where P^+ is the matrix in *(b)*, and P^- is a similar permutation matrix of a directed cycle in the opposite direction.
The transition matrices in *(a)* and *(c)* are reversible, being symmetric, but the transition matrix in *(b)* is not reversible.

The joint *pgf* of (X,Y), (79), is

$$\Big(\sum_{k=1}^{d!} a_k \sum_{c=1}^{d} d^{-1} t_{\sigma_k^{-1}(c)} s_c\Big)\Big(\sum_{l=1}^{d} d^{-1} t_l s_l\Big)^{N-1}, \tag{80}$$

which is of the form (64) where the distribution $\{a(j)\}_{\{j:|j|=N\}}$ has support only on

$$\{j : j_k = 1, j_{d!} = N-1, j_i = 0, i \neq k, d!; k = 1,\dots,d!-1\}$$
$$\cup\,\{j : j_{d!} = N, j_i = 0, i \neq d!\}.$$

Let P have a Lancaster expansion

$$p_{ij} = p_j\Big\{1 + \sum_{k=1}^{d-1}\rho_k u_i^{(k)} v_j^{(k)}\Big\},\ i,j = 1,\dots,d \tag{81}$$

where $\{u^{(k)}\}$ and $\{v^{(k)}\}$ are orthonormal function sets on $\{p_i\}$. Then working with the generating functions of the orthogonal polynomials on X and Y,

$$\mathbb{E}\Big[G(X,\{u^{(l)}\},w)G(Y,\{v^{(l)}\},z)\Big] = \Big(1 + \sum_{k=1}^{d-1}\rho_k w_k z_k\Big)\Big(1 + \sum_{k=1}^{d-1} w_k z_k\Big)^{N-1} \tag{82}$$

The eigenvalues of (X,Y) are the coefficients of $\prod_{k=1}^{d-1}\big(w_k z_k\big)^{n_k}$ in (82) divided by the normalizing constant (73), $\binom{N}{|n|}\binom{|n|}{n}$;

$$1,\ \gamma_n = \sum_{k=1}^{d-1}\rho_k\frac{n_k}{N},$$

with corresponding orthonormal functions in a Lancaster expansion (89) of

$$Q_n^\circ(x;\{u^{(l)}\}),\ Q_n^\circ(y;\{v^{(l)}\}).$$

If instead of one ball being chosen initially, s balls are chosen and independently reallocated to urns according to a transition probability matrix P, then the eigenvalues are

$$\binom{N}{s}^{-1} \sum_{j} \binom{N-|n|}{s-|j|} \prod_{l=1}^{d-1} \binom{n_l}{j_l} \rho_l^{j_l},$$

where summation is over $d-1$ dimensional vectors $\{j : 0 \leq j_k \leq n_k, k = 1, \ldots, d-1\}$. The corresponding orthonormal functions in the Lancaster expansion are the same as with one ball being chosen.

The Lancaster expansions (81) for Mizukawa's transition matrices P have the following correlation sequences and orthogonal functions.
(a) $\rho_k = -1/(d-1)$, $k = 1, \ldots, d-1$; $\{u^{(k)}\}_{k=0}^{d-1}$ is any complete orthonormal set of functions with $u_j^{(0)} = 1$, $j = 1, \ldots, d$ and $\{v^{(k)}\}_{k=0}^{d-1}$ is the same set of functions.
(b) $\rho_k = 1$, $k = 1, \ldots, d-1$; $\{u^{(k)}\}_{k=0}^{d-1}$ is any complete orthonormal set of functions with $u_j^{(0)} = 1$, $j = 1, \ldots, d$ and $v_j^{(k)} = u_{j-1 (\text{mod } d)}^{(k)}$, $k = 1, \ldots, d-1$, $j = 1, \ldots, d$.
(c) The two orthonormal sets of functions are identical and equal to

$$\left\{\sqrt{2}\cos\left(2\pi k(j-1)/d\right)\right\}_{k=1}^{d_c} \bigcup \left\{\sqrt{2}\sin\left(2\pi k(j-1)/d\right)\right\}_{k=1}^{d_s}, \tag{83}$$

$j = 1, \ldots, d$, where $d_c = d_s = (d-1)/2$ if d is odd, and $d_c = d/2$, $d_s = d/2-1$ if d is even. The correlation sequence is $\big\{\cos(2\pi k/d)\big\}_{k \leq d/2}$, repeated with the two cos and sin sets of functions in (83) for $k = 1, \ldots, d_c$ and $k = 1, \ldots, d_s$. The classical eigenfunction complex variable expansion

$$p_{ij} = d^{-1} \sum_{k=0}^{d-1} \cos(2\pi k/d) e^{2\pi \iota k(i-1)/d} e^{-2\pi k \iota (j-1)/d}, \tag{84}$$

where $\iota = \sqrt{-1}$, gives rise to the Lancaster expansion by taking real parts of the right side and identifying identical circular functions for $k \leq d/2$ and $k \geq d/2$.

Extending the definition of the orthogonal polynomials to a biorthogonal system is useful in expressing one-step transition function expansions as eigenfunction expansions. The k-step transition functions then have the same form with the eigenvalues raised to the k-th power. In a Lancaster expansion this does not hold if the transition functions are not reversible.

Let $\{\alpha^{(l)}\}_{l=0}^{d-1}$ and $\{\beta^{(l)}\}_{l=0}^{d-1}$ be biorthogonal functions, which are possibly complex, on $\{p_i\}$ with $\alpha^{(0)} = \beta^{(0)} \equiv 1$ such that

$$\sum_{i=1}^{d} p_i \alpha_i^{(k)} \beta_i^{(l)} = \delta_{kl}. \tag{85}$$

Define

$$\big\{Q_n(X;\{\alpha^{(l)}\})\big\},\ \big\{Q_n(X;\{\beta^{(l)}\})\big\} \tag{86}$$

in a similar way to previously by generating functions (72) with $\{u^{(l)}\}$ replaced respectively by $\{\alpha^{(l)}\}$ and $\{\beta^{(l)}\}$. Then (86) are biorthogonal polynomial systems on the Multinomial $(N,\,\{p_i\})$ distribution.

For example consider N particles in states $1,\dots,d$ making independent transitions according to a transition matrix P with stationary distribution $\{p_i\}_{i=1}^{d}$ such that the eigenfunction expansion of P is

$$p_{ij} = p_j\Big\{1 + \sum_{k=1}^{d-1} \rho_k \alpha_i^{(k)} \beta_j^{(k)}\Big\},\ i,j = 1,\dots,d. \tag{87}$$

P is not necessarily reversible and the eigenvalues and eigenfunctions are possibly complex. The left eigenvectors of P are $\{p_i\beta_i^{(k)}\}_{k=0}^{d-1}$; the right eigenvectors are $\{\alpha_j^{(k)}\}_{k=0}^{d-1}$ and the eigenvalues are $\{\rho_k\}_{k=0}^{d-1}$, with $\rho_0 = 1$.

Let X_i be the number of particles in state i, $i = 1,\dots,d$. X has a Multinomial $(N,\,\{p_i\})$ stationary distribution and the *pgf* for transitions from x to Y is

$$\prod_{i=1}^{d}\Big(\sum_{j=1}^{d} p_{ij} t_j\Big)^{x_i}. \tag{88}$$

In a similar way to the contingency table example the eigenfunction expansion for the transition functions from x to y is

$$m(y,\{p_j\})\Big\{1 + \sum_{\{n:1\leq|n|\leq N\}} \gamma_n Q_n^{\circ}(x;\{\alpha^{(l)}\}) Q_n^{\circ}(y;\{\beta^{(l)}\})\Big\}, \tag{89}$$

where

$$\gamma_n = \prod_{j=1}^{d-1} \rho_j^{n_j}. \tag{90}$$

In the k-step transition functions γ_n is replaced by γ_n^k.

Mizukawa's examples all have transition matrices for a single ball change which are circulant matrices. Let P be a general $d\times d$ circulant transition matrix with first row $\{q_j\}_{j=1}^d$ and the other rows rotated successively from the first so that the ith row is $\{q_{(j-i)(\mathrm{mod}\ d)+1}\}_{j=1}^d$. P is doubly stochastic, with a uniform stationary distribution on $1,\dots,d$. An eigenfunction expansion of P is

$$p_{ij} = d^{-1}\sum_{k=0}^{d-1}\tau_k e^{2\pi\iota k(i-1)/d}e^{-2\pi\iota k(j-1)/d},\ i,j = 1,\dots,d, \tag{91}$$

where

$$\tau_k = \sum_{r=0}^{d-1} q_{r+1}e^{2\pi\iota rk/d}. \tag{92}$$

P is reversible if and only if it is symmetric, because it has a uniform stationary distribution, then $q_j = q_{-(j-1)(\mathrm{mod}\ d)+1}$, and

$$\tau_k = \begin{cases} q_1 + (-1)^k q_{\lfloor \frac{d}{2}\rfloor+1} + 2\sum_{r=1}^{\lfloor\frac{d}{2}\rfloor-1} q_{r+1}\cos(2\pi kr/d) & \text{if } d \text{ is even}\\ q_1 + 2\sum_{r=1}^{\lfloor\frac{d-1}{2}\rfloor} q_{r+1}\cos(2\pi kr/d) & \text{if } d \text{ is odd.}\end{cases}$$

A Lancaster expansion can be constructed with orthogonal functions (83) and correlation sequence $\{\tau_l\}_{k\le d/2}$, repeated in the two sets of functions.

We now consider a biorthogonal expansion for the transition functions of the urn configuration, rather than the Lancaster expansion. Construct the functions (89) by taking the biorthogonal systems

$$\alpha_i^{(l)} = e^{2\pi\iota l(i-1)/d}\ \ \beta_j^{(l)} = e^{-2\pi\iota l(j-1)/d}, l = 0,\dots,d-1,\ i,j = 1,\dots,d.$$

It is straightforward, but detailed, to verify that the *pgf* of (89) when

$$\gamma_n = \sum_{l=0}^{d-1}\tau^l n_l/N \tag{93}$$

is the correct *pgf* (78). The eigenfunction expansion for the transition functions of $Y \mid x$ is therefore (89) with eigenvalues (93). The eigenfunctions have a particularly nice form, expressed as monomial symmetric polynomials in the d-th roots of unity, as recognized by Mizukawa. If the balls are ordered so that $Z_j = k$ if ball j is in urn k, then from (76),

$$Q_n(X;\{u^{(l)}\}) = \sum_{\{A_l\}}\prod_{k_1\in A_1} e^{(2\pi\iota 1/d)(Z_{k_1}-1)}\cdots\prod_{k_{d-1}\in A_{d-1}} e^{(2\pi\iota(d-1)/d)(Z_{k_{d-1}}-1)}, \tag{94}$$

where summation is over all partitions of subsets of $\{1,\dots,N\}$, $\{A_l\}$ such that $|A_l| = n_l$, $l = 1,\dots,d$. Let $n_0 = N - |n|$, then $(n_j)_{j=0}^{d-1}$ is an ordered partition of N. Regard $\lambda = (Z_k - 1)_{k=1}^N$ as a partition $\big(0^{x_1}1^{x_2}\cdots(d-1)^{x_d}\big)$. Let $\xi = e^{2\pi\iota/d}$ be the d-th root of unity, and $\Xi = \big((1)^{n_0}(\xi)^{n_1}\cdots(\xi^{d-1})^{n_{d-1}}\big)$. Then directly from (94),

$$Q_n(X;\{\alpha^{(l)}\}) = m_\lambda(\Xi), \tag{95}$$

where $m_\lambda(\Xi)$ is a monomial symmetric polynomial.

References

Aitken, A. C. and Gonin, H. T. (1935) On fourfold sampling with and without replacement. *Proc. Roy. Soc. Edinb.* **55** 114–125.

Bakry, D., Huet, N. (2006) The hypergroup property and representation of Markov Kernels. *Séminare de Probabilitiés XLI, Lecture notes in Mathematics*, Vol 1934, 295–347, Springer.

Bassetti, F., Diaconis, P. (2006) Examples comparing importance sampling and the Metropolis algorithm. *Illinois J. Math.* **50** 67–91.

Bochner, S. (1954) Positive zonal functions on spheres. *Proc. Nat. Acad. Sci. USA* **40** 1141–1147.

Brown, L. D., Cai, T. T. and DasGupta, A. (2001) Interval estimation for a binomial proportion. *Statist. Sci.* **16** 101–133.

Diaconis, P. (1977) Finite forms of de Finetti's theorem on exchangeability. *Synthese* **36** 271–281.

Diaconis, P. Khare, K. and Saloff-Coste, L. (2008) Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.*, **23** 151–200.

Eagleson, G. K. (1969) A characterization theorem for positive definite sequences on the Krawtchouk polynomials. *Austral. J. Statist.*, **11** 29–38.

Fréchet, M. (1951) Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon,* Sect. A, Sér. 3, **14** 53–77.

Hoeffding, W. (1940) Masstabinvariant Korrelationstheorie. *Inst. Angew. Math. Univ. Berlin* **5** 181–233.

Gasper, G. (1972) Banach algebras for Jacobi series and positivity of a kernel. *Ann. of Math., 2nd Ser.*, **95** 261–280.

Griffiths, R. C. (1971) Orthogonal polynomials on the Multinomial distribution. *Austral. J. Statist.* **13** 27–35. Corrigenda (1972) *Austral. J. Statist.* **14** 270.

Griffiths, B. [R. C.] (2009) Stochastic processes with orthogonal polynomial eigenfunctions. *J. Comput. Appl. Math.* **23** 739–744.

Griffiths, R. C., Milne, R. K. and Wood, R. (1979) Aspects of correlation in bivariate Poisson distributions and processes. *Austral. J. Statist.*, **21** 238–255.

Hoare, M. R. and Rahman, M. (1983) Cumulative Bernoulli trials and Krawtchouk processes. *Stochastic Process. Appl.* **16** 113–139.

Ishmail, M. E. H. (2005) *Classical and Quantum Orthogonal Polynomials in one variable*, Volume 98 of *Encyclopedia of Mathematics and its Applications.* Cambridge: Cambridge University Press.

Koudou, A. E. (1996) Probabilités de Lancaster, *Exposition. Math.* **14(3)** 247–275.

Lancaster H. O. (1969) *The chi-squared distribution*, John Wiley & Sons.

MacWilliams, F. J. and Sloane, N. J. (1997) *The theory of error-correcting codes.* North-Holland Pub. Co., Amsterdam and New York.

Mizukawa, H. (2010) Finite Gelfand pair approaches for Ehrenfest diffusion model. arXiv:1009.1205v1 [math.CO] 7 Sep 2010

Sarmanov, O. V. and Bratoeva Z. N. (1967) Probabilistic properties of bilinear expansions of Hermite polynomials. *Theor. Probability Appl.* **12** 470–481.

Schoutens, W. (2000) *Stochastic processes and orthogonal polynomials.* Lecture notes in mathematics **146** Springer-Verlag.

Suppes, P. and Zanotti, M. (1980) A new proof of the impossibility of hidden variables using the principles of exchangeability and the identity of conditional distribution. *Studies in the Foundations of Quantum Mechanics* Phil. Sci. Assoc., East Lansing, Michigan.

Suppes, P. and Zanotti, M. (1981) When are probabilistic explanations possible? *Synthese* **48** 191–199.

Vere-Jones, D. (1971) Finite bivariate distributions and semigroups of non-negative matrices. *Q. J. Math.* **22** 247–270.

Zhou, H. and Lange, K. (2009) Composition Markov chains of Multinomial type. *Adv. Appl. Probab.*, **41** 270–291.

Version 3.1, Compiled January 30, 2011.