PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 30, Fasc. 1 (2010), pp. 103–120

THE SUPERPOSITION MARKOV CHAIN: FINITE OCCUPANCY WITH COUPLING, AND THE ASYMPTOTICS THEREOF

BY

BERND GÜNTHER (MÜHLHEIM)

Abstract. Picking up a lottery example by Markov we analyze a variant of the finite occupancy problem that assumes complete symmetry among the target cells but drops the customary assumptions about independence. Assuming that the distribution of the number of non-empty cells approaches a non-trivial asymptotic limit for large state space sizes the limit time evolution is studied.

2000 AMS Mathematics Subject Classification: 60C05, 60F05, 60J10, 94A29.

Key words and phrases: Random allocation, weak limit theorems, finite Markov chains, Stirling polynomials.

1. INTRODUCTION

The classical occupancy problem in its most simple form distributes a finite number of particles among a finite number of equivalent cells and studies the ensuing patterns. A vast number of generalizations has been developed weakening the assumptions of equivalence or of finiteness or similar; see [5] and [8] for recent overviews. However, the assumption of independence is almost always adhered to. There are a few exceptions: Markoff [7], Section 21, counts the numbers that have shown up at least once in so many drawings of a lottery and provides some special results. In [4], Chapter VII, allocation by complexes is investigated, but the authors assume that the size of their complexes is asymptotically small (of order $o\left(\sqrt[4]{n}\right)$) compared to the state space size n, which is an obstacle to practical applications. In our paper, rather than enforcing a particular asymptotic behaviour by whatever kind of assumptions it is our philosophy to consider prescribed asymptotics as initial conditions and investigate their time evolution, and to expect the limit to reflect as many properties of the finite case as possible.

As an example for the kind of application we have in mind we present the following problem from coding theory: Given a database containing a large number of records, certain properties of the records shall be encoded in bitstrings, one for each record, to facilitate searches [9], [2]. Each property is encoded in a randomly chosen bit pattern, and each record is assigned the superposition of the bit patterns for all properties it satisfies. An error may occur if one of the codes is covered by other code patterns, and this irregularity must be minimized. Best results are achieved if the number of 1-bits in each code pattern is *fixed*.

In our exposition we will adhere to complete symmetry of the cells but will dispense with their independence; we are interested in the number of non-empty cells. Using the image of the urn model we are going to study the following game: in each round a fixed number of equivalent balls is distributed among a finite number of equivalent urns, each urn receiving at most one ball. In the following rounds, the same number of balls is distributed without knowledge of the previous distributions. Each urn may hold as many balls as you like, but has only two externally observable states: empty or non-empty.

This scheme may be generalized by choosing the number of balls at random, using the same random variable in each round. Thus we include the classical, binomial occupancy scheme, but not its multinomial extensions.

Let n denote the number of urns. By an *isotropic allocation scheme* we mean a random variable X assuming subsets of $\{1, \ldots, n\}$ as values, whose distribution is invariant under any permutation of $\{1, \ldots, n\}$. Thus the distribution of X is completely determined by that of $\#X$, the number of elements of X. From independent repetitions $X_1, X_2, \ldots$ of X we can construct a Markov chain $X^{(1)}, X^{(2)}, \ldots$ by $X^{(m)} := X_1 \cup \ldots \cup X_m$, our superposition Markov chain. Urn number k is considered non-empty at Markov time m if $k \in X^{(m)}$. Observe that the assumption of independence of the urns would immediately enforce binomial distribution.

We will start our paper with a preparatory section on isotropic allocation schemes, then our Markov chain and its combinatorial aspect will be described. In studying its asymptotic aspects for large numbers of urns n it will be assumed that the expected number of balls $E\#X = n(1-q)$ allocated in one round scales *proportionally* with n, and we will study the asymptotic distribution of the normalized random variables

$$\tilde{X}_m = \frac{1}{q^m \sqrt{n}} [\#X^{(m)} - n(1-q^m)].$$

It will be shown that, if $\tilde{X}_1$ is asymptotically distributed like a random variable Y satisfying some mild regularity conditions, then the whole process $\tilde{X}_1, \tilde{X}_2, \ldots$ is asymptotically distributed like $Y_1 + Z_1, \ldots, \sum_{k=1}^{m} (Y_k + Z_k), \ldots$, where Y_k are independent versions of Y and Z_k are independent $\mathcal{N}\left(0, \sigma_k^2\right)$-distributed random variables with variance $\sigma_k^2 = (1/q - 1)(1/q^{k-1} - 1)$, a non-stationary Markov chain with independent increments. This raises the question of asymptotic independence of the urns. In our concluding section we will show that significant coupling is discernible in the asymptotic limit of all cases except the binomial case, that is already independent in finite context.

Special consideration will be given to the classical binomial case and to the fixed weight case, where the generating random variable $\#X$ is constant. It will be seen that the former is not a permissible approximation to the latter, they are asymptotically distinct.

2. ISOTROPIC ALLOCATION SCHEMES

For any isotropic allocation scheme X with values in $\{1,\dots,n\}$ and any subset $A \subseteq \{1,\dots,n\}$ the probabilities

$$p_k := P(X = A), \tag{2.1}$$

$$F_k := P(X \subseteq A), \tag{2.2}$$

$$G_k := P\,(X \supseteq A) \tag{2.3}$$

depend only on the size $k := \#A$. They satisfy the conditions

$$\sum_{k=0}^{n} \binom{n}{k} p_k = 1, \quad F_k = \sum_{j=0}^{k} \binom{k}{j} p_j \quad \text{and} \quad G_k = \sum_{j=k}^{n} \binom{n-k}{j-k} p_j$$

and are most conveniently handled via generating functions:

$$f(t) := \sum_{k=0}^{n} \binom{n}{k} p_k t^k, \tag{2.4}$$

$$F(t) := \sum_{k=0}^{n} \binom{n}{k} F_k t^k, \tag{2.5}$$

$$G(t) := \sum_{k=0}^{n} \binom{n}{k} G_{n-k} t^k, \tag{2.6}$$

$$F(t) = (1+t)^n f\left(\frac{t}{1+t}\right), \tag{2.7}$$

$$G(t) = (-1)^n F(-1-t), \tag{2.8}$$

$$G(t) = t^n f\left(\frac{1+t}{t}\right). \tag{2.9}$$

This allows easy conversion between the three sets of parameters p_k, F_k and G_k, for instance we have

$$p_k = \sum_{j=0}^{k} (-1)^{j+k} \binom{k}{j} F_j.$$

Moments $\mu_r = E\big((\#X)^r\big)$ and central moments $\mu_r^* = E\big((\#X - \mu_1)^r\big)$ are given by the formulas

$$\mu_r = \sum_{k=0}^{r} k! \binom{n}{k} \mathfrak{S}_r^{(k)} G_k, \tag{2.10}$$

$$\mu_r^* = \sum_{k=0}^{r}\sum_{j=0}^{k}(-1)^k\binom{r}{k}\mathfrak{S}_k^{(j)}(n)_j\,(n-\mu_1)^{r-k}\,F_{n-j}, \tag{2.11}$$

$$F_{n-r} = \frac{1}{(n)_r}\sum_{k=0}^{r}\sum_{j=0}^{k}(-1)^j\binom{k}{j}S_r^{(k)}\,(n-\mu_1)^{k-j}\,\mu_j^*. \tag{2.12}$$

Here $\mathfrak{S}_r^{(k)}$ denote the Stirling numbers of the second kind, $S_r^{(k)}$ Stirling numbers of the first kind and $(n)_r$ the falling factorial power $(n)_r = n(n-1)\dots(n-r+1)$. We notice in particular that

$$\mu_1 = n\,(1-F_{n-1}) \quad \text{and} \quad \mu_2^* = nF_{n-1} - n^2F_{n-1}^2 + n(n-1)F_{n-2}.$$

The most important examples for isotropic distributions will be:

1. The binomial distribution $p_k = (1-q)^k q^k$, $F_k = q^k$ for $0 \leqslant q \leqslant 1$.
2. The fixed weight distribution with weight w,

$$p_k = 0 \text{ for } k \neq w \quad \text{and} \quad p_w = \binom{n}{w}^{-1},$$

$$F_k = 0 \text{ for } k < w \quad \text{and} \quad F_k = \binom{k}{w}\binom{n}{w}^{-1} \text{ for } k \geqslant w,$$

where w is an integer, $0 \leqslant w \leqslant n$. We observe that

$$F_k = \binom{n}{k}^{-1}\binom{nF_{n-1}}{n-k} \quad \text{and} \quad G_k = \binom{n}{k}^{-1}\binom{nG_1}{k} \text{ for } k \geqslant w.$$

3. Two weight distributions with parameters $w_1 < \vartheta < w_2$, where $w_i \in \mathbb{N}_0$, $\vartheta \in \mathbb{R}$, and

$$p_k = \begin{cases} (w_2-\vartheta)\,(w_2-w_1)^{-1}\binom{n}{w_1}^{-1} & \text{for } k = w_1, \\ (\vartheta-w_1)\,(w_2-w_1)^{-1}\binom{n}{w_2}^{-1} & \text{for } k = w_2, \\ 0 & \text{elsewhere.} \end{cases} \tag{2.13}$$

Evidently, this is a convex combination of two fixed weight distributions and the parameters are chosen such that the expectation is ϑ.

The isotropic distributions constitute a convex space whose extreme points are the fixed weight distributions. The isotropic distributions with given expectation ϑ constitute a convex space whose extreme points are the two weight distributions for parameters $w_1 < \vartheta < w_2$ plus, if ϑ is an integer, the fixed weight distribution with weight ϑ. Therefore fixed weight and two weight distributions will frequently occur as solutions of optimization problems. For instance, it is an easy application of Jensen's inequality to show

$$F_k \geqslant \binom{n}{k}^{-1}\binom{nF_{n-1}}{n-k} \qquad \text{for } k \geqslant n\,(1 - F_{n-1}) - 1,$$

where nF_{n-1} is not required to be an integer, and

$$G_k \geqslant \binom{n}{k}^{-1}\binom{nG_1}{k} \qquad \text{for } k \leqslant nG_1 + 1.$$

As long as the right-hand sides of these two inequalities are positive, they coincide with the parameters F_k, G_k, respectively, in case of fixed weight distributions. This means that fixed weight distributions minimize the parameters F_k and G_k for any given expectation. It now follows from (2.10) that fixed weight distributions minimize all moments.

3. THE MARKOV CHAIN

Given two independent isotropic allocation schemes X and X' we can construct a third one by superposition $X'' := X \cup X'$. For any subset $A \subseteq \{0, \ldots, n\}$ with $a = \#A$ we obtain $F''_a = P\,(X \cup X' \subseteq A) = P\,(X \subseteq A)\,P\,(X' \subseteq A) = F_a F'_a$. Moreover,

$$\begin{aligned} P(X'' = A) &= \sum_{B \subseteq A} P(X' = B, A \setminus B \subseteq X \subseteq A) \\ &= \sum_{B \subseteq A} \left\{ \sum_{j=0}^{b} \binom{b}{j} p_{a-j} \right\} P(X' = B), \end{aligned} \tag{3.1}$$

where $b = \#B$, as can be seen by setting $j = \#X \cap B$. Thus passage from state $X' = B$ to state $X'' = A$ is described by a $2^n \times 2^n$-dimensional transition matrix $\mathbf{P} = (P_{AB})$,

$$P_{AB} = \begin{cases} \sum_{j=0}^{b} \binom{b}{j} p_{a-j} & \text{if } A \supseteq B, \\ 0 & \text{elsewhere.} \end{cases} \tag{3.2}$$

Notice that $\mathbf{P}$ is a triangular matrix containing our friends, the parameters $F_a = P_{AA}$ as diagonal elements, and therefore as eigenvalues. We also recognize the first column as $G_a = P_{A\emptyset}$.

Diagonalization is easy: we define a (non-orthogonal) transformation matrix $\mathbf{T} = (T_{AB})$ by

$$T_{AB} := \begin{cases} (-1)^{a-b} & \text{for } A \supseteq B, \\ 0 & \text{elsewhere,} \end{cases} \tag{3.3}$$

$$T^{-1}_{AB} = \begin{cases} 1 & \text{for } A \supseteq B, \\ 0 & \text{elsewhere.} \end{cases} \tag{3.4}$$

Then a short computation shows that $\mathbf{TPT}^{-1} = \mathbf{F}$ equals the diagonal matrix $\mathbf{F} = (F_{AB})$ with $F_{AA} = F_a$.

Using a sequence of independent repetitions $X_1, X_2, \ldots$ of an isotropic allocation scheme X we set up a Markov chain $X^{(1)}, X^{(2)}, \ldots$ with stationary transition probabilities P_{AB} by $X^{(m)} = X_1 \cup \ldots \cup X_m$. Observe that $X^{(m)}$ is an isotropic allocation scheme with parameters F^m_k. In particular, setting[1] $q := F_{n-1}$, $0 \leqslant q \leqslant 1$, the expectation is given by $EX^{(m)} = n(1 - F^m_{n-1}) = n\,(1 - q^m)$ and the variance by $\mu_2^{*(m)} = nq^m - n^2q^{2m} + n(n-1)F^m_{n-2}$.

EXAMPLE 3.1. The classical case of independent urns, where the generating random variable X is of binomial distribution with parameter $1 - q$. Then $F^m_k = q^{mk}$ and therefore $X^{(m)}$ is binomially distributed with parameter $1 - q^m$. The variance is $\sigma^2 = nq^m\,(1 - q^m) = n(1-q)q^m \sum_{j=0}^{m-1} q^j$.

EXAMPLE 3.2. The fixed weight case: Suppose X has fixed weight distribution with weight w and set $q := 1 - w/n$. At Markov time $m > 1$ our distribution is no longer of fixed weight but is determined by the parameters

$$F^m_k = \binom{k}{w}^m \binom{n}{w}^{-m},$$

in particular

$$F^m_{n-1} = q^m \quad \text{and} \quad F^m_{n-2} = \left(\frac{nq(nq-1)}{n(n-1)}\right)^m,$$

and therefore $EX^{(m)} = n\,(1 - q^m)$, and

$$\sigma^2 = nq^m\,(1 - q^m) - n(n-1)q^m \left[q^m - \left(\frac{q - 1/n}{1 - 1/n}\right)^m\right].$$

This case is significantly more complicated than the binomial case and its investigation is one of the main objectives of this paper. Expanding the variance in a power series over $1/n$ one obtains

$$\sigma^2 = n[q^m\,(1 - q^m) - m(1-q)q^{2m-1}] + \frac{m(m-1)}{2}(1-q)^2q^{2m-2} \pm \ldots,$$

[1] In our paper, the letter q will always denote $q = 1 - n^{-1}EX$, n will be reserved for the size of our state space, and m for the Markov time.

neglecting terms of higher order. In practical applications q will be close to 1, and then the leading term

$$n[q^m(1-q^m)-m(1-q)q^{2m-1}]=n(1-q)^2q^m\sum_{j=0}^{m-2}(j+1)q^j$$

is of *quadratic* order in $1-q$, and hence much smaller than the binomial value, which is of *linear* order. It should already be clear right now that the binomial distribution cannot be used as an approximation for the fixed weight case.

4. THE COMBINATORIAL ASPECT

The probability $p_k^{(m)}=P(X^{(m)}=A)$, where $\#A=k$, can be computed by the following symbolic theorem:

THEOREM 4.1. *The probability generating function*

$$f(t)=\sum_{k=0}^{n}\binom{n}{k}p_k^{(m)}t^k$$

at Markov time m is generated by the function

$$\tilde{f}(t;x_1,\dots,x_m)=\{1+t[\prod_{j=1}^{m}(1+x_j)-1]\}^n. \tag{4.1}$$

To apply this theorem one must expand $\tilde{f}\in\mathbb{R}[x_1,\dots,x_m]$ as polynomial in the variables x_j:

$$\tilde{f}(t;x_1,\dots,x_m)=\sum_{k,a_1,\dots,a_m}\binom{n}{k}B^{(k)}_{a_1,\dots,a_m}t^kx_1^{a_1}\dots x_m^{a_m}, \tag{4.2}$$

and then make the formal replacement $x_j^r\to p_r$ to obtain the customary generating function

$$f(t)=\sum_{k,a_1,\dots,a_m}\binom{n}{k}B^{(k)}_{a_1,\dots,a_m}t^kp_{a_1}\dots p_{a_m}, \tag{4.3}$$

thus expressing the probabilities $p_k^{(m)}$ at Markov time m in terms of those at Markov time 1. For example, in the binomial situation we have to perform the formal replacement $x_j^r\to q^n(1/q-1)^r$ that amounts essentially to a variable substitution and leads to

$$f(t)=q^{mn}\tilde{f}\left(t;\frac{1}{q}-1,\dots,\frac{1}{q}-1\right)=\{q^m+t\,[1-q^m]\}^n,$$

the familiar generating function for binomial probabilities with parameter $1-q^m$. We will see below that the numbers $B^{(k)}_{a_1,\dots,a_m}$ are quite familiar combinatorial objects.

Proof. Observing that the summand -1 cancels the constant term of the product $\prod_{j=1}^{m}(1+x_j)$ in (4.1) we see that the coefficients $B^{(k)}_{a_1,\dots,a_m}$ are non-negative and we can therefore *define* numbers

$$p'_k := \sum_{a_1,\dots,a_m} B^{(k)}_{a_1,\dots,a_m} p_{a_1}\dots p_{a_m} \geqslant 0, \text{ i.e. } f(t) = \sum_{k=0}^{n}\binom{n}{k}p'_k t^k.$$

Then $\sum_{k=0}^{n}\binom{n}{k}p'_k$ is obtained by the symbolic replacement from

$$\tilde{f}\,(1; x_1,\dots,x_m) = \Big\{\prod_{j=1}^{m}(1+x_j)\Big\}^n = \prod_{j=1}^{m}(1+x_j)^n.$$

Replacing $x_j^{a_j}$ by p_{a_j} turns the factor

$$(1+x_j)^n = \sum_{k=0}^{n}\binom{n}{k}x_j^k$$

into

$$\sum_{k=0}^{n}\binom{n}{k}p_k = 1,$$

and therefore

$$\sum_{k=0}^{n}\binom{n}{k}p'_k = 1.$$

Hence the numbers p'_k define isotropic allocation probabilities and in order to show that they are the correct ones it suffices to compare the parameters

$$F'_k = \sum_{j=0}^{k}\binom{k}{j}p'_j$$

to F^m_k. But

$$\sum_{k=0}^{n}\binom{n}{k}F'_k t^k = (1+t)^n f\left(\frac{t}{1+t}\right)$$

is obtained by the symbolic replacement from

$$(1+t)^n\tilde{f}\left(\frac{t}{1+t}; x_1,\dots,x_m\right)$$
$$= \Big\{1+t\prod_{j=1}^{m}(1+x_j)\Big\}^n = \sum_{k=0}^{n}\binom{n}{k}t^k\prod_{j=1}^{m}(1+x_j)^k.$$

The factor $(1+x_j)^k$ is turned into

$$\sum_{i=0}^{k}\binom{k}{i}p_i = F_k,$$

therefore

$$\sum_{k=0}^{n}\binom{n}{k}F'_k t^k=\sum_{k=0}^{n}\binom{n}{k}t^k F_k^m$$

and *a fortiori* $F'_k = F_k^m$. ■

The numbers $B^{(k)}_{a_1,\ldots,a_m}$ are determined by

$$\big[\prod_{j=1}^{m}(1+x_j)-1\big]^k=\sum_{a_1,\ldots,a_k}B^{(k)}_{a_1,\ldots,a_m}x_1^{a_1}\ldots x_m^{a_m},\tag{4.4}$$

and this provides clues for their evaluation and combinatorial interpretation. The first of the following equations holds by a direct application of the binomial theorem, the second one can be shown by induction on m:

$$B^{(k)}_{a_1,\ldots,a_m}=\sum_{j=0}^{k}(-1)^{j-k}\binom{k}{j}\prod_{i=1}^{m}\binom{j}{a_i},\tag{4.5}$$

$$B^{(k)}_{a_1,\ldots,a_m}=\sum_{0=j_0\leqslant j_1\leqslant\cdots\leqslant j_m=k}\prod_{i=1}^{m}\binom{a_i}{j_i-j_{i-1}}\binom{j_i}{a_i}.\tag{4.6}$$

To reveal the combinatorial meaning we follow MacMahon [6] (Vol. I, Section I, Chapter II.18) and consider

$$\sum_{a_1,\ldots,a_m}B^{(k)}_{a_1,\ldots,a_m}(tx_1)^{a_1}\ldots(tx_m)^{a_m}=\big[\prod_{j=1}^{m}(1+tx_j)-1\big]^k,\tag{4.7}$$

that is

$$\sum_{a_1,\ldots,a_m}B^{(k)}_{a_1,\ldots,a_m}(tx_1)^{a_1}\ldots(tx_m)^{a_m}=\big[\sum_{j=1}^{m}t^j\sigma_j\,(x_1,\ldots,x_m)\big]^k,\tag{4.8}$$

where σ_j are the elementary symmetric polynomials. Now suppose we are given a totality of $a_1+\ldots+a_m$ objects of m different categories, a_j objects of type j. For short, one speaks of objects of specification $a_1,\ldots,a_m$. By comparing coefficients with those of MacMahon it follows that $B^{(k)}_{a_1,\ldots,a_m}$ is the number of ways objects of specification $a_1,\ldots,a_m$ can be distributed among k *distinct* parcels such that no parcel remains empty and no parcel receives more than one object of a kind.

5. THE LIMIT PROCESS MODEL

In order to give meaning to asymptotic limits we consider a family of isotropic allocation schemes X_n, one for each state space size n or at least for n ranging over a cofinal subsequence of integers. When there is a risk of ambiguity, the letter n will be affixed as subscript to the parameter under consideration; so for instance

$\mu_{nr}^{*(m)}$ will denote the r-th central moment at Markov time m in a state space of size n. The expectation will scale proportionally to n, that means, if at state space size n we have $E\#X_n = n\,(1-q_n)$, we assume

$$\lim_{n\to\infty} q_n = q \quad \text{with } 0<q<1. \tag{5.1}$$

We will study the weak limit (i.e. limit in distribution) of the normalized random variable

$$\frac{1}{q_n\sqrt{n}}[X_n - n\,(1-q_n)],$$

and we will assume from now on that the limit distribution Y has finite moments $E|Y|^r < \infty$ and that its characteristic function

$$\chi(t) = Ee^{itY} = \sum_{r=0}^{\infty} \frac{i^r \mu_r}{r!} t^r$$

is analytic in a neighborhood of 0.

Our main tool will be Stirling polynomials $\mathcal{S}_k(x)$ (see [10], [3], [1]), defined by the generating function

$$\sum_{k=0}^{\infty} \frac{1}{k!} \mathcal{S}_k(x) t^k = \left(\frac{t}{1-e^{-t}}\right)^{x+1}. \tag{5.2}$$

$\mathcal{S}_k(x)$ is a polynomial of degree k with leading coefficient 2^{-k}, and the Stirling numbers are given by

$$\mathfrak{S}_n^{(m)} = (-1)^{n-m}\binom{n}{m}\mathcal{S}_{n-m}(-m-1), \tag{5.3}$$

$$S_n^{(m)} = (-1)^{n-m}\binom{n-1}{m-1}\mathcal{S}_{n-m}(n-1). \tag{5.4}$$

Unfortunately, authors disagree about normalization conventions for Stirling polynomials; we prefer to follow Roman [10] (Chapter 4, Section 4.8). Jordan [3] uses

$$\psi_k(x) = \frac{\mathcal{S}_{k+1}(x)}{(k+1)!(x+1)},$$

whereas Graham et al. [1] consider the "polynomials"

$$\sigma_k(x) = \frac{\mathcal{S}_k(x-1)}{k!x}.$$

The devil played tricks and placed a misprint on Jordan's formula in [3], p. 225, where the leading coefficient of $\psi_m(x)$ should correctly be given as

$$1/(m+1)!2^{m+1}.$$

Thus armed, we can prove:

THEOREM 5.1. *Let us consider two independent isotropic allocation schemes X and X' in $\{1, \ldots, n\}$ with expectations $n(1-q)$ and $n(1-q')$, respectively. Then $X'' = X \cup X'$ is an isotropic allocation scheme with expectation $n(1-qq')$. If asymptotic approximations Y for $(q\sqrt{n})^{-1}\,[\#X - n(1-q)]$ and Y' for $(q'\sqrt{n})^{-1}\,[\#X' - n(1-q')]$ are given, then the independent sum $Y + Y' + Z$ is an asymptotic approximation for $(qq'\sqrt{n})^{-1}\,[\#X'' - n(1-qq')]$, where Z is an $\mathcal{N}(0,\sigma^2)$-distributed random variable with $\sigma^2 = (1/q'-1)\,(1/q-1)$.*

Proof. We intend to express the central moments of $X \cup X'$ in terms of those of X and X' and start by rewriting equation (2.12) with the aid of Stirling polynomials:

$$(n)_j F_{n-j} = n^j q^j \sum_{i,\ell} P_{i\ell}(j)\mu_i^* n^{-i-\ell}, \tag{5.5}$$

$$(n)_j F'_{n-j} = n^j q'^j \sum_{i,\ell} P'_{i\ell}(j)\mu_i'^* n^{-i-\ell}, \tag{5.6}$$

$$P_{i\ell}(x) := (-1)^{i+\ell}\binom{x-\ell}{i}\binom{x-1}{\ell}\mathcal{S}_\ell(x-1)q^{-\ell-i}, \tag{5.7}$$

$$P'_{i\ell}(x) := (-1)^{i+\ell}\binom{x-\ell}{i}\binom{x-1}{\ell}\mathcal{S}_\ell(x-1)q'^{-\ell-i}. \tag{5.8}$$

Notice that $P_{i\ell}(x)$ is a polynomial in x of degree $2\ell + i$ with leading coefficient $(-1)^{i+\ell}/(i!\ell!2^\ell q^{i+\ell})$ and $P'_{i\ell}(x)$ is a polynomial in x of degree $2\ell + i$ with leading coefficient $(-1)^{i+\ell}/(i!\ell!2^\ell q'^{i+\ell})$. Consequently,

$$(n)_j (n)_j F''_{n-j} = n^{2j}\left(qq'\right)^j \sum_{i_1,i_2,\ell} P''_{i_1 i_2 \ell}(j)\mu_{i_1}^* \mu_{i_2}'^* n^{-i_1-i_2-\ell}, \tag{5.9}$$

$$P''_{i_1 i_2 \ell}(x) = \sum_{\ell_1+\ell_2=\ell} P_{i_1\ell_1}(x)P'_{i_2\ell_2}(x), \tag{5.10}$$

where $P''_{i_1 i_2 \ell}(x)$ is a polynomial of degree $2\ell + i_1 + i_2$ with leading coefficient

$$\frac{(-1)^{i_1+i_2+\ell}}{i_1!i_2!\ell!2^\ell q^{i_1}q'^{i_2}}\left(\frac{1}{q}+\frac{1}{q'}\right)^\ell. \tag{5.11}$$

We have to cancel one of the factors $(n)_j$:

$$[(n)_j]^{-1} = \sum_{u=j-1}^{\infty} \mathfrak{S}_u^{(j-1)} n^{-u-1} \tag{5.12}$$

$$= n^{-j}\sum_{u=0}^{\infty}(-1)^u\binom{u+j-1}{u}\mathcal{S}_u(-j)n^{-u} \tag{5.13}$$

$$= n^{-j}\sum_u Q_u(j)n^{-u}, \tag{5.14}$$

where

$$Q_u(x) := (-1)^u \binom{x+u-1}{u} \mathcal{S}_u(-x) \tag{5.15}$$

is a polynomial of degree $2u$ with leading coefficient $1/(u!2^u)$.

Now, by equations (5.9) and (5.14),

$$(n)_j F''_{n-j} = \left(nqq'\right)^j \sum_{i_1,i_2,v} T_{i_1 i_2 v}(j) \mu^*_{i_1} \mu'^*_{i_2} n^{-i_1-i_2-v}, \tag{5.16}$$

$$T_{i_1 i_2 v}(x) := \sum_{\ell+u=v} P''_{i_1 i_2 \ell}(x) Q_u(x), \tag{5.17}$$

where $T_{i_1 i_2 v}(x)$ is a polynomial of degree $2v + i_1 + i_2$ with leading coefficient

$$\frac{(-1)^{i_1+i_2+v}}{v! i_1! i_2! 2^v q^{i_1} q'^{i_2}} \left(\frac{1}{q} + \frac{1}{q'} - 1\right)^v. \tag{5.18}$$

Then we obtain consecutively

$$\mu''^*_r = \sum_{k=0}^{r} \sum_{j=0}^{k} (-1)^k \binom{r}{k} \mathfrak{S}^{(j)}_k \left(nqq'\right)^{r-k} (n)_j F''_{n-j}, \tag{5.19}$$

$$\begin{aligned}\mu''^*_r = \left(nqq'\right)^r \sum_{j,k,i_1,i_2,v} (-1)^j \binom{r}{k} \binom{k}{j} \mathcal{S}_{k-j}(-j-1) \left(qq'\right)^{-k+j} \\ \times T_{i_1 i_2 v}(j) \mu^*_{i_1} \mu'^*_{i_2} n^{-i_1-i_2-v-k+j},\end{aligned} \tag{5.20}$$

$$\mu''^*_r = \left(nqq'\right)^r \sum_{i_1,i_2,w} \left[\sum_k (-1)^k \binom{r}{k} R_{i_1,i_2,w}(k)\right] \mu^*_{i_1} \mu'^*_{i_2} n^{-w}, \tag{5.21}$$

$$\mu''^*_r = (-1)^r \left(nqq'\right)^r \sum_{i_1,i_2,w} \left[\Delta^r_k R_{i_1,i_2,w}(k)\right] \mu^*_{i_1} \mu'^*_{i_2} n^{-w}, \tag{5.22}$$

where

$$\begin{aligned} R_{i_1,i_2,w}(x) = \\ = \sum_{j=0}^{w-i_1-i_2} \frac{(-1)^j}{j!} (x)_j \mathcal{S}_j(-x+j-1) T_{i_1,i_2,w-i_1-i_2-j}(x-j) \left(qq'\right)^{-j}.\end{aligned} \tag{5.23}$$

Here Δ_x is the difference operator applied to argument x, $\Delta_x f(x) = f(x+1) - f(x)$. We observe that $\Delta^r_x f(x) = \sum_{k=0}^r (-1)^{r+k} f(x+k)$. $R_{i_1,i_2,w}(x)$ is a polynomial of degree $2w - i_1 - i_2$ with leading coefficient

$$\frac{(-1)^{i_1+i_2}(1-q)^{w-i_1-i_2}(1-q')^{w-i_1-i_2}}{(w-i_1-i_2)! i_1! i_2! 2^{w-i_1-i_2} q^{w-i_2} q'^{w-i_1}}. \tag{5.24}$$

For $r > 2w - i_1 - i_2$ the polynomial $\Delta_x^r R_{i_1,i_2,w}(x)$ is zero, and for $r \leqslant 2w - i_1 - i_2$ it is a polynomial of degree $2w - i_1 - i_2 - r$ with leading coefficient

$$\frac{(-1)^{i_1+i_2}(2w-i_1-i_2)_r(1-q)^{w-i_1-i_2}(1-q')^{w-i_1-i_2}}{(w-i_1-i_2)!i_1!i_2!2^{w-i_1-i_2}q^{w-i_2}q'^{w-i_1}}. \tag{5.25}$$

Therefore

$$\frac{\mu_r''^*}{(qq')^r\sqrt{n^r}} = (-1)^r \sum_{i_1,i_2,w} [\Delta_k^r R_{i_1,i_2,w}(k)] \frac{\mu_{i_1}^*}{\sqrt{n^{i_1}}} \frac{\mu_{i_2}'^*}{\sqrt{n^{i_2}}} n^{-w+(i_1+i_2+r)/2}. \tag{5.26}$$

By assumption of weak convergence of the random variables

$$\frac{1}{q\sqrt{n}}\left[\#X - n(1-q)\right] \quad \text{and} \quad \frac{1}{q'\sqrt{n}}\left[\#X' - n(1-q')\right]$$

the central moments of Y and Y' are given by

$$\mu_{\infty,i_1}^* = \lim_{n\to\infty} \frac{\mu_{n,i_1}^*}{q^{i_1}\sqrt{n^{i_1}}} \quad \text{and} \quad \mu_{\infty,i_2}'^* = \lim_{n\to\infty} \frac{\mu_{n,i_2}'^*}{q'^{i_2}\sqrt{n^{i_2}}}.$$

In equation (5.26) all summands except those with $w = (i_1 + i_2 + r)/2$ vanish for $n \to \infty$. Hence there must be a limit

$$\mu_{\infty,r}''^* = \lim_{n\to\infty} \frac{\mu_{n,r}''^{(*)}}{(qq')^r\sqrt{n^r}}$$

satisfying the equation

$$\frac{\mu_{\infty,r}''^*}{r!} = \tag{5.27}$$
$$= \sum_{\substack{i_1,i_2 \\ i_1+i_2\leqslant r \\ i_1+i_2\equiv r (\mathrm{mod}\ 2)}} \frac{1}{\big((r-i_1-i_2)/2\big)!} \left(\frac{(1-q)(1-q')}{2qq'}\right)^{(r-i_1-i_2)/2} \frac{\mu_{\infty,i_1}^*}{i_1!} \frac{\mu_{\infty,i_2}'^*}{i_2!},$$

which is a convolution equation for the Taylor coefficients of the characteristic functions:

$$\chi''(t) = \chi(t)\chi'(t)\exp\left(-\frac{1}{2}\left(\frac{1}{q}-1\right)\left(\frac{1}{q'}-1\right)t^2\right). \tag{5.28}$$

This immediately implies our theorem. ∎

THEOREM 5.2. *If the normalized variable* $(q\sqrt{n})^{-1}[\#X^{(1)} - n(1-q)]$ *taken at Markov time* $m = 1$ *is for large state space size* n *asymptotically distributed like a random variable* Y, *then the whole Markov chain*

$$\frac{1}{q\sqrt{n}}[\#X^{(1)} - n(1-q)], \ldots, \frac{1}{q^m\sqrt{n}}[\#X^{(m)} - n\,(1-q^m)], \ldots$$

is asymptotically distributed like $Y_1, \ldots, \sum_{k=1}^{m}(Y_k + Z_k), \ldots,$ *where* Y_k *are copies of* Y *and* Z_k *are* $\mathcal{N}\big(0, \sigma_k^2\big)$*-distributed random variables with variance* $\sigma_k^2 = (1/q - 1)\big(1/q^{k-1} - 1\big)$, *all independent.*

Notice that the limit process is a non-stationary Markov chain with independent increments $Y_m + Z_m$.

Proof. We have to show that the joint distribution of the finite sections $\widetilde{\#X}^{(1)}, \ldots, \widetilde{\#X}^{(m)}$ converges weakly to that of $Y_1 + Z_1, \ldots, \sum_{k=1}^{m}(Y_k + Z_k)$, where

$$\widetilde{\#X}_k = \frac{1}{q^k\sqrt{n}}[\#X^{(k)} - n(1-q^k)].$$

By the Markov property it suffices to check the distribution for $m = 1$, which holds by assumption, and to check the transition probabilities. For the latter we restrict our probability space to the condition $\widetilde{\#X}^{(m)} = t$, i.e. to $\#X^{(m)} = n\,(1-q^m) + tq^m\sqrt{n}$. This is equivalent to a fixed weight distribution with parameter $q_n = q^m(1 - t/\sqrt{n}) \to q^m$. The limit distribution is independent of t. An application of Theorem 5.1 now concludes the proof. ■

EXAMPLE 5.1. Let us see how the classical binomial case fits. In this context $(q\sqrt{n})^{-1}[\#X^{(1)} - n(1-q)]$ is asymptotically $\mathcal{N}\big(0, \sigma^2\big)$-distributed with $\sigma^2 = 1/q - 1$. Then $Y_k + Z_k$ is normally distributed with variance $(1/q - 1)(q^{k-1})^{-1}$ and Theorem 5.2 states that $(q^m\sqrt{n})^{-1}[\#X^{(m)} - n\,(1-q^m)]$ is asymptotically normal with variance $q^{-m} - 1$, as one might have guessed!

EXAMPLE 5.2. The fixed weight case with weight w. Let us set $q = 1 - w/n$. Here $\#X$ is constant, and therefore $Y \equiv 0$. From Theorem 5.2 it follows that $(q^m\sqrt{n})^{-1}[\#X^{(m)} - n\,(1-q^m)]$ is asymptotically normal with variance

$$\sum_{k=1}^{m}(1/q - 1)\,(1/q^{k-1} - 1) = 1/q^m - 1 - m\,(1/q - 1)\,.$$

We emphasize again that the variance is *much* smaller than in the binomial situation with the same parameter q.

EXAMPLE 5.3. The case above requires rational q. For irrational q we consider fixed weight distributions with weight w such that $n(1-q) - 1 < w < n(1-q)$. The asymptotics coincide with those of Example 5.3.

EXAMPLE 5.4. Of course, not all limit distributions are normal, since Theorem 5.2 leaves us almost complete freedom of construction. Consider for instance $0 < q < 1$ and $\beta \in \mathbb{R}$ and assume

$$\sqrt{n} > \frac{|\beta|}{\min(q, 1-q)}.$$

Set $q_1 := q + \beta/\sqrt{n}$, $q_2 := q - \beta/\sqrt{n}$ and let $w_1, w_2 \in \{0, \dots, n\}$ be the integers closest to nq_1, nq_2, that means $|w_i - nq_i| < 1$. Now apply the two weight allocation schemes with weights w_1, w_2, selecting weight w_i with probability

$$p_i := \frac{1}{2}\binom{n}{w_i}^{-1}.$$

The expectation is nq, and the normalized variable assumes each of the values $(w_i - nq)/\sqrt{n}$ with probability $\frac{1}{2}$, where

$$\left|\frac{w_1 - nq}{\sqrt{n}} - \beta\right| < \frac{1}{\sqrt{n}} \quad \text{and} \quad \left|\frac{w_2 - nq}{\sqrt{n}} + \beta\right| < \frac{1}{\sqrt{n}}.$$

This converges in distribution to the random variable that assumes the two values $\pm\beta$ with probability $\frac{1}{2}$ each. By Theorem 5.2, $(q^m\sqrt{n})^{-1}[\#X^{(m)} - n\,(1-q^m)]$ is asymptotically distributed like the independent sum of a binomial random variable distributed on the lattice $\beta\mathbb{Z}$ and a normally distributed variable with variance $1/q^m - 1 - m\,(1/q - 1)$.

6. COUPLING

Since our limit process has independent increments in contrast to the finite case, we might ask what happens to the coupling of our cells in the asymptotic limit. As it turns out, it is just as strong as in the finite setting.

For any subset $A \subseteq \{1, \dots, n\}$, $a := \#A$, the restriction $X_{|A} := A \cap X$ of our isotropic allocation scheme to A is an isotropic allocation scheme on A, whose distribution depends only on the number a. The original scheme can be reconstructed by the formula

$$\#X = \#X_{|A} + \#X_{|\complement A}, \tag{6.1}$$

and we say that our scheme is *independent* if for all A the decomposition (6.1) is an independent sum. Otherwise we speak of *coupling*.

In finite context the question is not very exciting, because it is easy to see that the binomial case is the only independent one. We will now investigate the limit case, and have to work out the asymptotic distribution of $X_{|A}$.

LEMMA 6.1. $\#X_{|A}$ *has expectation* $a(1-q)$, *and if the random variable* $(q\sqrt{n})^{-1}[\#X - n(1-q)]$ *is asymptotically distributed like a random variable* Y, *if the size* $a = \#A$ *scales proportionally with* n, $a = \vartheta n$ *with* $0 < \vartheta < 1$, *then the restricted random variable* $(q\sqrt{a})^{-1}[\#X_{|A} - a(1-q)]$ *is asymptotically distributed like* $Y_\vartheta = \sqrt{\vartheta}\,(Y + Z)$, *where* Z *is independent normally distributed with variance* $(1/\vartheta - 1)\,(1/q - 1)$ *and mean* 0.

Proof. For any $B \subseteq A$ we have $\tilde{P}(X_{|A} \subseteq B) = P(X \subseteq B \cup \complement A)$, and therefore $X_{|A}$ is characterized by the parameters $F'_b = F_{b+n-a}$. In particular, the expectation is $a(1 - F'_{a-1}) = a(1 - F_{n-1}) = a(1-q)$.

Now let $X'' \subseteq \{0, \dots, n\}$ be the fixed weight allocation scheme with weight $n - a$ and consider the independent union $X' = X \cup X''$. Then for any subset $B \subseteq \{0, \dots, n\}$ and $b = \#B \geqslant n - a$:

$$\begin{aligned}
P(X' = B) &= \sum_{\#C = n-a} P(X \cup C = B) P(X'' = C) \\
&= \binom{n}{n-a}^{-1} \sum_{\#C=n-a} P(X \cup C = B), \\
P(\#X' = b) &= \binom{n}{n-a}^{-1} \sum_{\#C=n-a} P\big(\#(X \cup C) = b\big) \\
&= \binom{n}{n-a}^{-1} \sum_{\#C=n-a} P\big(\#(X_{|\complement C}) + n - a = b\big) \\
&= P(\#X_{|A} + n - a = b).
\end{aligned}$$

Now our lemma follows from Theorem 5.1. ∎

Asymptotically the decomposition (6.1) is replaced by a decomposition of the limits:

$$Y = \sqrt{\vartheta} Y_\vartheta + \sqrt{1-\vartheta} Y_{1-\vartheta}, \tag{6.2}$$

where Y_ϑ and $Y_{1-\vartheta}$ are determined by Lemma 6.1. The factor $\sqrt{\vartheta}$ serves to adjust the $(\sqrt{a})^{-1}$-normalization of the restriction to the whole state space. We say that our isotropic allocation scheme is *asymptotically independent* if the decomposition (6.2) is an independent sum for all $0 < \vartheta < 1$. This is the case if and only if the characteristic functions are related by

$$\chi(t) = \chi_\vartheta(\sqrt{\vartheta} t) \chi_{1-\vartheta}(\sqrt{1-\vartheta} t).$$

By Lemma 6.1 the characteristic functions of the restrictions are given by

$$\chi_\vartheta\left(\frac{t}{\sqrt{\vartheta}}\right) = \chi\,(t) \exp\left[-\frac{1}{2}\left(\frac{1}{\vartheta} - 1\right)\left(\frac{1}{q} - 1\right) t^2\right].$$

Therefore asymptotic independence is equivalent to

$$\chi(t) = \chi(\vartheta t)\chi\big((1-\vartheta)t\big)\exp\left[-\vartheta(1-\vartheta)\left(\frac{1}{q}-1\right)t^2\right]$$

for all $0 < \vartheta < 1$. Since $\chi(0) = 1$ and since $\chi(t)$ is assumed analytic in a neighborhood of 0, we may pass to logarithms. Hence asymptotic independence is equivalent to $\ln\chi(t) = \ln\chi(\vartheta t) + \ln\chi\big((1-\vartheta)t\big) - \vartheta(1-\vartheta)(1/q-1)t^2$. Comparing power series coefficients this can happen only if $\ln\chi(t) = -(1/2)(1/q-1)t^2$. Thus we have proved:

THEOREM 6.1. *An isotropic allocation scheme X with expectation $n(1-q)$ is asymptotically independent if and only if $(q\sqrt{n})^{-1}[\#X - n(1-q)]$ is asymptotically normal with mean 0 and variance $1/q - 1$.*

We emphasize that asymptotic normality alone is *not* sufficient for independence, in addition the variance must have a very specific value. This value happens to be the asymptotic limit of a binomially distributed random variable for parameter q, so we end up with the limit of a case that is already independent in finite context. In contrast, the asymptotic limit of a fixed weight distribution does *not* qualify, neither does its evolution at later Markov times.

In many cases there is linear correlation. Recall that the linear correlation coefficient c of two random variables X and Y is defined by

$$\begin{aligned} c &= \frac{E\big((X-\bar{X})(Y-\bar{Y})\big)}{\sqrt{E(X-\bar{X})^2E(Y-\bar{Y})^2}} \\ &= \frac{E(X+Y-\bar{X}-\bar{Y})^2 - E(X-\bar{X})^2 - E(Y-\bar{Y})^2}{2\sqrt{E(X-\bar{X})^2E(Y-\bar{Y})^2}}. \end{aligned}$$

We are going to compute the correlation coefficient c_ϑ of the random variables $\sqrt{\vartheta}Y_\vartheta$ and $\sqrt{1-\vartheta}Y_{1-\vartheta}$ from (6.2). If the variance of the original variable is (asymptotically) denoted by σ^2, then it follows from the above that $\sqrt{\vartheta}Y_\vartheta$ has variance $\sigma_\vartheta^2 = \vartheta^2\sigma^2 + \vartheta(1-\vartheta)(1/q-1)$, and $\sqrt{1-\vartheta}Y_{1-\vartheta}$ has variance $\sigma_{1-\vartheta}^2 = (1-\vartheta)^2\sigma^2 + \vartheta(1-\vartheta)(1/q-1)$. Therefore

$$c_\vartheta = \frac{\sigma^2 - (1/q-1)}{\sqrt{\sigma^2 + (1/\vartheta - 1)(1/q-1)}\sqrt{\sigma^2 + \big(1/(1-\vartheta)-1\big)(1/q-1)}}.$$

The absolute value of c_ϑ is maximal for $\vartheta = \frac{1}{2}$ with

$$c_{1/2} = \frac{\sigma^2 - (1/q-1)}{\sigma^2 + (1/q-1)}.$$

In general, we expect a negative correlation, because an increase of your own share of balls can be achieved by stealing out of your opponents urns. But positive correlations can also occur:

EXAMPLE 6.1. Example 5.4 in Section 5 has $\sigma^2 = \beta^2$, and therefore $c_\vartheta > 0$ if $\beta > \sqrt{1/q - 1}$. Here the dominating influence on the correlation is the total number of balls.

REFERENCES

[1] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics*: *A Foundation for Computer Science*, 2nd edition, Addison-Wesley Publishing Group, Amsterdam 1994.

[2] B. Günther, *On the probability distribution of superimposed random codes*, IEEE Trans. Inform. Theory 54 (7) (2008), pp. 3206–3210. DOI 10.1109/TIT.2008.924658.

[3] C. Jordan, *Calculus of Finite Differences*, AMS Chelsea, 1965. Reprint 1979.

[4] V. F. Kolchin, B. A. Sevast'yanov and V. P. Christyakov, *Random Allocations*, Scripta Series in Mathematics, Wiley, 1978.

[5] S. Kotz and N. Balakrishnan, *Advances in urn models during the past two decades*, in: *Advances in Combinatorial Methods and Applications to Probability and Statistics*, N. Balakrishnan (Ed.), Birkhäuser, 1997, pp. 203–257.

[6] P. A. MacMahon, *Combinatory Analysis*, Phoenix Editions, Dover 2004. Reprint Vol. I (1915), Vol. II (1916) and Intro (1920).

[7] A. A. Markoff, *Wahrscheinlichkeitsrechnung*, Teubner, 1912.

[8] S. M. Mirakhmedov, *Asymptotic normality associated with generalized occupancy problems*, Statist. Probab. Lett. 77 (15) (2007), pp. 1549–1558.

[9] C. S. Roberts, *Partial-match retrieval via the method of superimposed codes*, Proceedings of the IEEE 67 (12) (1979), pp. 1624–1642.

[10] S. Roman, *The Umbral Calculus*, Dover 2005.

DB-Systel GmbH
Helpertseestraße 21
63165 Mühlheim, Germany
E-mail: dr.bernd.guenther@t-online.de

Received on 19.9.2008;
revised version on 20.6.2009