ON THE JOINT BEHAVIOR OF TYPES OF COUPONS IN GENERALIZED COUPON COLLECTION

HOSAM M. MAHMOUD,* *The George Washington University*

ROBERT T. SMYTHE,** *Oregon State University*

Abstract

The 'coupon collection problem' refers to a class of occupancy problems in which j identical items are distributed, independently and at random, to n cells, with no restrictions on multiple occupancy. Identifying the cells as coupons, a coupon is 'collected' if the cell is occupied by one or more of the distributed items; thus, some coupons may never be collected, whereas others may be collected once or twice or more. We call the number of coupons collected exactly r times *coupons of type r*. The coupon collection model we consider is general, in that a random number of purchases occurs at each stage of collecting a large number of coupons; the sample sizes at each stage are independent and identically distributed according to a *sampling distribution*. The joint behavior of the various types is an intricate problem. In fact, there is a variety of joint central limit theorems (and other limit laws) that arise according to the interrelation between the mean, variance, and range of the sampling distribution, and of course the phase (how far we are in the collection processes). According to an appropriate combination of the mean of the sampling distribution and the number of available coupons, the phase is sublinear, linear, or superlinear. In the sublinear phase, the normalization that produces a Gaussian limit law for uncollected coupons can be used to obtain a multivariate central limit law for at most two other types—depending on the rates of growth of the mean and variance of the sampling distribution, we may have a joint central limit theorem between types 0 and 1, or between types 0, 1, and 2. In the linear phase we have a multivariate central limit theorem among the types 0, 1, . . . , k for any fixed k.

Keywords: Urn model; random structure; stochastic process; multivariate martingale; multivariate central limit theorem; sampling; coupon collection; phase; phase transition

2010 Mathematics Subject Classification: Primary 60C05; 60F05; 05A05
Secondary 60G42; 60G48; 60E05

1. Coupon collection

The most basic form of the 'coupon collection problem' concerns the distribution of j items, independently and at random, to n cells (thought of as coupons), where a coupon is 'collected' if the cell is occupied. Questions of interest then include the following. (i) What is the distribution of the number of uncollected coupons (or the number collected exactly once, etc.)? (ii) What is the expected number of items that must be placed in order that all cells be occupied (all coupons collected)?

There are many variations of this problem, going back at least to de Moivre (1718) and Laplace (1774). The problem gained popularity in the 1930s, when the Dixie Cup Company

sold ice cream cups with a cardboard cover that had hidden on the underside a coupon (carrying likeable items such as cute animals, movie stars, and Major League baseball players). This marketing strategy is meant to encourage fans of such items to complete sets of their favorites, and, thus, increase the sales. Many companies followed suit and there has been a myriad of such schemes; many are now obsolete.

A generalized form of the classical coupon collector's problem assumes that the consumer purchases a random number, $S \geq 1$, of items each time and that the promoting company guarantees that the S associated coupons are distinct. The collector obtains S coupons at each stage, of which some or all may already be in his/her possession. The company promises that all sold collections of size S are equally likely. Ideally, when one such collection is sold, it is immediately replaced in the market to maintain the uniformity of all subsets of any feasible size. In the Dixie Cup scheme $S \equiv 1$. Kobza *et al.* (2007) and Stadje (1990) provided surveys.

While collecting coupons, some may never be obtained, others may be collected once or twice or more. We call the number of coupons collected exactly r times *coupons of type r*. We investigate in this paper the joint behavior of coupons of different types across the phases of collection.

2. Setup as an urn scheme

Consider the following setup for coupon collection with random sampling. At the start we have n coupons to be collected, where n may be large. Let S_j be an independent, identically distributed sequence of random variables, all distributed like a generic random variable $S = S(n) \in \{1, 2, \ldots, s_n\}$, with $s_n \leq n$. For technical reasons that will become evident later on, we keep the range of S small, relative to n. Specifically, we work with $s_n = o(\sqrt{n})$. So, the mean $\mu_S(n)$ and standard deviation $\sigma_S(n)$ of $S(n)$ are also $o(\sqrt{n})$. At the jth stage the collector purchases a random number, S_j, of coupons.

Let $X_{j,r}^{(n)}$ be the number of coupons that have been collected exactly r times after j samples have been purchased. For fixed $k \geq 0$, let

$$X_j^{(n)} = \begin{pmatrix} X_{j,0}^{(n)} \\ X_{j,1}^{(n)} \\ \vdots \\ X_{j,k}^{(n)} \end{pmatrix}.$$

Represent the coupons as balls of $k + 2$ different colors (labeled $0, 1, \ldots, k + 1$) in an urn. Coupons that have been collected r times are balls of color r for $r = 0, \ldots, k$. Color $k + 1$ is special—it represents all the balls that have been drawn more than k times. There is no need to study the number of balls of color $k + 1$, as it is determined by the number of balls of all the other colors. Specifically,

$$X_{j,k+1}^{(n)} = n - \sum_{r=0}^{k} X_{j,r}^{(n)}.$$

3. Organization

The rest of the paper is organized in sections as follows. We introduce the notation in Section 4, and state the results in Section 5. In Section 6 we formulate the basic stochastic

recurrence, which gives a matrix recurrence for the mean and covariance. We present an exact solution to the mean recurrence in Subsection 6.1, and we present exact and asymptotic solutions to the covariance recurrence in Subsection 6.2. The multivariate martingale underlying the process is derived in Section 7, and the analysis in the sublinear and linear phases is given in Sections 8 and 9, respectively. We conclude in Section 10 with some illustrating examples.

4. Notation and setup

The notation $\mathrm{Bin}(n, p)$ stands for a binomial random variable on n trials with rate of success p per trial, and $\mathcal{N}_k(\mathbf{0}, \mathbf{\Sigma})$ stands for a multivariate normal vector of k components with mean $\mathbf{0}$ (of k components) and $k \times k$ covariance matrix $\mathbf{\Sigma}$. In this investigation some of the multivariate normal distributions refer to what some books call singular multivariate normal distributions, where $\mathbf{\Sigma}$ is a singular matrix, but a number of linear combinations together define a proper multivariate normal distribution of lower dimension. Let $\mathrm{Hypergeo}(n, m, w)$ be a hypergeometric random variable that represents the number of white balls in a sample of size m balls taken without replacement at random (all subsets of size m being equally likely) from an urn containing a total of n white and red balls, of which w are white. The mean and variance for this standard distribution are well known; see Stuart and Ord (1987, Article 5.14).

The transpose of a matrix $\mathbf{W}$ is denoted by $\overline{\mathbf{W}}$. We will use the matrix norm $\|\cdot\|$, defined as the square root of the sum of the squares of the matrix components, or, equivalently, the square root of the sum of the squares of its eigenvalues. We use the notation $\boldsymbol{o}(b_n)$ and $\boldsymbol{O}(b_n)$ for matrices in which each component is respectively $o(b_n)$ and $O(b_n)$ in the usual scalar sense. The probabilistic versions $\boldsymbol{o}_{\mathcal{L}_1}(b_n)$ and $\boldsymbol{o}_{\mathrm{P}}(b_n)$ will stand for a sequence of random matrices where each component is $o(b_n)$ respectively in the $\mathcal{L}_1$ norm and in probability. We will use the notation '$\xrightarrow{\mathrm{D}}$' to denote convergence in distribution and '$\xrightarrow{\mathrm{P}}$' to denote convergence in probability. In the sequel all matrix convergence, be it deterministic or probabilistic (in $\mathcal{L}_1$ and in probability), is considered componentwise. We let $\mathcal{F}_j$ be the sigma-field generated by the first j draws. Note that the sequence $\{\mathcal{F}_j\}_{j=0}^{\infty}$ of sigma-fields is increasing. Thus, it can be the filtration of a martingale sequence. Unless otherwise stated, all asymptotic equivalents and bounds are taken as $n \to \infty$.

The following special matrices will be used:

$$
\mathbf{B} := \begin{pmatrix}
0 & 0 & 0 & 0 & \dots & 0 & 0 \\
1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix},
$$

$$
\mathbf{M} := \frac{\mu_S(n)}{n}\mathbf{B} + \left(1 - \frac{\mu_S(n)}{n}\right)\mathbf{I} = \mathbf{I} + \frac{\mu_S(n)}{n}(\mathbf{B} - \mathbf{I}),
$$

$$
\mathbf{F} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad
\mathbf{G} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix};
$$

the matrices $\mathbf{B}$ and $\mathbf{M}$ are of dimension $(k+1) \times (k+1)$, and $\mathbf{F}$ and $\mathbf{G}$ are $(k+1)$-component vectors. Note that high powers of $\mathbf{B}$ vanish. Specifically, $\mathbf{B}^i$ is the zero matrix for $i \geq k+1$.

For $0 \leq i \leq k$, $\boldsymbol{B}^i$ is a matrix of 0s, except for the entries on the forward minor diagonal starting at position $(i, 0)$ and ending at position $(k, k - i)$, which are all 1.

In what follows we will use the functions

$$g_n = \frac{(\sigma_S^2(n) - \mu_S(n))n + \mu_S^2(n)}{n^2(n-1)}, \qquad h_n = \frac{\mu_S(n)n - \sigma_S^2(n) - \mu_S^2(n)}{n(n-1)},$$

and the functions $f_{\ell,m}(y)$, which are the coefficients of linearity in the covariance between coupons of types ℓ and m in the usual coupon collection ($S \equiv 1$). They appear in the work of Kolchin *et al.* (1978, p. 38). The first few are

$[f_{\ell,m}]_{0 \leq \ell, m \leq 2}$

$$= e^{-2\lambda} \begin{pmatrix} e^\lambda - 1 - \lambda & -\lambda^2 & -\dfrac{\lambda^3}{2} + \dfrac{\lambda^2}{2} \\[2mm] -\lambda^2 & \lambda e^\lambda - \lambda + \lambda^2 - \lambda^3 & -\dfrac{\lambda^4}{2} + \lambda^3 - \lambda^2 \\[2mm] -\dfrac{\lambda^3}{2} + \dfrac{\lambda^2}{2} & -\dfrac{\lambda^4}{2} + \lambda^3 - \lambda^2 & e^\lambda(\lambda^2 + 1 - \lambda) - \dfrac{\lambda^5}{4} + \dfrac{3\lambda^4}{4} - \lambda^3 - 1 \end{pmatrix}.$$

We denote the matrix $[f_{\ell,m}]_{0 \leq \ell, m \leq k}$ by $\boldsymbol{J}_{k+1}$. We also need the matrix

$$\boldsymbol{J}'_{k+1} = e^{-2\lambda}\left[\frac{\lambda^{\ell+m-1}}{\ell!\, m!}(\lambda - \ell)(\lambda - m)\right]_{0 \leq \ell, m \leq k}.$$

The vector

$$\overline{\boldsymbol{v}}_{k+1} = \left(1, \lambda, \frac{\lambda^2}{2}, \ldots, \frac{\lambda^k}{k!}\right)$$

appears as the coefficient of linearity of $\mathrm{E}[X_j^{(n)}]$ in the linear phase.

We let $\mathbf{1}_{\mathcal{E}}$ denote the indicator of event $\mathcal{E}$, that is, a function that assumes the value 1 if $\mathcal{E}$ occurs and the value 0 otherwise. We will use the backward operator $\nabla a_i = a_i - a_{i-1}$.

5. The results

Central to all the analysis is a careful handling of the covariance structure of the process. We present an exact formula for the covariance in Proposition 1. We do not deal with the case where $0 = \liminf_{n \to \infty} \sigma_S^2(n) < \limsup_{n \to \infty} \sigma_S^2(n)$. Up to $O(1)$ draws nothing much of interest happens. We investigate the joint behavior of balls after j draws for j in two phases.

(a) The growing sublinear phase, where $j = j_n$ grows to ∞ with n, but $j_n = o(n/\mu_S(n))$.

(b) The linear phase,

$$\mu_S(n) j_n \sim \lambda_n n$$

for some positive λ_n that is bounded away from 0 and ∞. That is, for positive constants Q_1 and Q_2 and all $n \geq 1$,

$$0 < Q_1 \leq \lambda_n \leq Q_2 < \infty.$$

Theorem 1. *Consider coupon collection with a sampling distribution having range in $\{1, 2, \ldots, s_n\}$, where $s_n = o(\sqrt{n})$, with mean $\mu_S(n)$ and variance $\sigma_S^2(n)$. Let $X_{j,r}^{(n)}$ be the number*

of balls of type r, $0 \le r \le k$, and let $X_j^{(n)}$ be the vector with components $X_{j,r}^{(n)}$, $r = 0, \ldots, k$. Assume that we are in the growing sublinear phase, where $j_n \to \infty$ and $j_n = o(n/\mu_S(n))$. Let

$$\alpha_n := \frac{\mu_S^2(n) j_n^2/2n}{\mu_S^2(n) j_n^2/2n + \sigma_S^2(n) j_n}.$$

Suppose that we are in the upper sublinear phase,

$$s_n^2 = o\left(\frac{\mu_S^2(n) j_n^2}{2n} + \sigma_S^2(n) j_n\right).$$

The following statements hold.

(a) *If $\alpha_n \to \alpha \in (0, 1]$, for $k = 2$,*

$$\left(X_{j_n}^{(n)} - \left(1 - \frac{\mu_S(n)}{n}\right)^{j_n}\left(\begin{matrix} n \\ \mu_S(n) j_n \\ \frac{\mu_S^2(n) j_n^2}{2n} \end{matrix}\right)\right)\left(\sqrt{\frac{\mu_S^2(n) j_n^2}{2n} + \sigma_S^2(n) j_n}\right)^{-1}$$

$$\xrightarrow{D} \mathcal{N}_3\left(0, \begin{pmatrix} 1 & -(\alpha+1) & \alpha \\ -(\alpha+1) & 3\alpha+1 & -2\alpha \\ \alpha & -2\alpha & \alpha \end{pmatrix}\right),$$

subject to the additional condition that $\mu_S^2(n) j_n^2/2n + \sigma_S^2(n) j_n$ grows to ∞.

(b) *If $\alpha_n \to 0$, for $k = 1$,*

$$\frac{X_{j_n}^{(n)} - (1 - \mu_S(n)/n)^{j_n}\left(\begin{matrix} n \\ \mu_S(n) j_n \end{matrix}\right)}{\sigma_S(n)\sqrt{j_n}} \xrightarrow{D} \mathcal{N}_2\left(0, \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right),$$

subject to the additional condition that $1/\sigma_S^2(n) = o(j_n)$.

Remark. In Theorem 1(b) the covariance matrix is singular. In this case each of $X_{j_n,0}^{(n)}$ and $X_{j_n,1}^{(n)}$ (shifted by its mean and scaled by $\sigma_S(n)\sqrt{j_n}$) satisfies a univariate central limit theorem with variances and covariances coinciding with those in the limiting bivariate normal distribution. The same can be said about $X_{j_n,0}^{(n)}$, $X_{j_n,1}^{(n)}$, and $X_{j_n,2}^{(n)}$ in the $\alpha = 1$ case of Theorem 1(a): each (appropriately normalized) satisfies a univariate central limit theorem with variances and covariances coinciding with those in the limiting trivariate normal distribution.

Theorem 2. *Consider coupon collection with a sampling distribution having range in $\{1, 2, \ldots, s_n\}$, where $s_n = o(\sqrt{n})$, with mean $\mu_S(n)$ and variance $\sigma_S^2(n)$. Let $X_{j,r}^{(n)}$ be the number of balls of type r, $0 \le r \le k$, and let $X_j^{(n)}$ be the vector with components $X_{j,r}^{(n)}$, $r = 0, \ldots, k$. Assume that we are in the linear phase, where $j_n \sim \lambda_n n/\mu_S(n)$ for $0 < Q_1 \le \lambda_n \le Q_2 < \infty$. Let*

$$\beta_n := \frac{\mu_S(n)}{\mu_S(n) + \sigma_S^2(n)}.$$

If $\lambda_n \to \lambda > 0$, the following statements hold.

(a) *If $\beta_n \to \beta \in (0, 1]$,*

$$\frac{X_{j_n}^{(n)} - n\mathrm{e}^{-\lambda}\boldsymbol{v}_{k+1}}{\sqrt{n}} \xrightarrow{\mathrm{D}} \mathcal{N}_{k+1}\left(\boldsymbol{0},\ \boldsymbol{J}_{k+1} + \frac{1-\beta}{\beta}\boldsymbol{J}'_{k+1}\right).$$

(b) *If $\beta_n \to 0$, and λ is not a positive integer $r \leq k$,*

$$\frac{X_{j_n}^{(n)} - n\mathrm{e}^{-\lambda}\boldsymbol{v}_{k+1}}{\sigma_S(n)\sqrt{n/\mu_S(n)}} \xrightarrow{\mathrm{D}} \mathcal{N}_{k+1}(\boldsymbol{0},\ \boldsymbol{J}'_{k+1}).$$

If $\lambda = r \leq k$ then $(X_{j_n,r}^{(n)} - \lambda^r \mathrm{e}^{-\lambda} n/r!)/\sqrt{n}$ is asymptotically normal with variance given by the rth diagonal entry of $\boldsymbol{J}_{k+1}$.

The alert reader must have noticed that the case of superlinear growth (when $\mu_S(n)\, j_n$ grows faster than n) has not been considered in this work. For the case of $X_{j,0}^{(n)}$ (uncollected coupons), some asymptotic results appear in Smythe (2011). It turns out that in the superlinear phase, the asymptotic variance of each $X_{j,r}^{(n)}$, $r = 0, 1, \ldots, k$, is of a different order, so there is no multivariate central limit theorem of the type considered in the present paper. Some results in this case will appear elsewhere.

6. Stochastic recurrence and moments

When the collector purchases a sample of size S_j coupons, any coupon appearing in that sample is acquired one more time. That is, if the corresponding ball in the urn is of color r (for $r = 0, \ldots, k$), its color is upgraded to $r+1$; the number of balls of color $r+1$ goes up by 1, and the number of balls of color r goes down by 1. Let $H_{j,r}^{(n)}$ be the number of balls of color r in the jth sample (which is of size S_j). So, $H_{j,r}^{(n)}$ has the distribution of the Hypergeo$(n, S_j, X_{j-1,r}^{(n)})$ random variable. We have the recurrence system

$$X_{j,r}^{(n)} = X_{j-1,r}^{(n)} + H_{j,r-1}^{(n)} - H_{j,r}^{(n)} \quad \text{for } r = 0, \ldots, k, \tag{1}$$

interpreting $H_{j,-1}^{(n)} \equiv 0$. From this stochastic recurrence we can find moments. We illustrate this only on the first two moments. As we will see, it is quite tedious to obtain the second moment; exact higher moments would be a real challenge to find by such direct methods, and we later find their asymptotics by alternative means.

6.1. The mean

The counts $X_{j,r}^{(n)}$ (interpreting $X_{j,-1}^{(n)} \equiv 0$) have averages

$$\begin{aligned}
\mathrm{E}[X_{j,r}^{(n)}] &= \mathrm{E}[X_{j-1,r}^{(n)}] + \frac{1}{n}\mathrm{E}[X_{j-1,r-1}^{(n)}S_j] - \frac{1}{n}\mathrm{E}[X_{j-1,r}^{(n)}S_j] \\
&= \mathrm{E}[X_{j-1,r}^{(n)}] + \frac{1}{n}\mathrm{E}[X_{j-1,r-1}^{(n)}]\,\mathrm{E}[S_j] - \frac{1}{n}\mathrm{E}[X_{j-1,r}^{(n)}]\,\mathrm{E}[S_j] \\
&= \left(1 - \frac{\mu_S(n)}{n}\right)\mathrm{E}[X_{j-1,r}^{(n)}] + \frac{\mu_S(n)}{n}\mathrm{E}[X_{j-1,r-1}^{(n)}].
\end{aligned}$$

This can be represented in matrix form:

$$
\mathrm{E}[X_j^{(n)}] =
\begin{pmatrix}
1 - \dfrac{\mu_S(n)}{n} & 0 & 0 & 0 & \cdots & 0 & 0 \\[2mm]
\dfrac{\mu_S(n)}{n} & 1 - \dfrac{\mu_S(n)}{n} & 0 & 0 & \cdots & 0 & 0 \\[2mm]
0 & \dfrac{\mu_S(n)}{n} & 1 - \dfrac{\mu_S(n)}{n} & 0 & \cdots & 0 & 0 \\[2mm]
\vdots & \vdots & \ddots & & & \vdots & \vdots \\[2mm]
0 & 0 & 0 & 0 & \cdots & \dfrac{\mu_S(n)}{n} & 1 - \dfrac{\mu_S(n)}{n}
\end{pmatrix}
\mathrm{E}[X_{j-1}^{(n)}]
$$

$$
=
\begin{pmatrix}
1 - \dfrac{\mu_S(n)}{n} & 0 & 0 & 0 & \cdots & 0 & 0 \\[2mm]
\dfrac{\mu_S(n)}{n} & 1 - \dfrac{\mu_S(n)}{n} & 0 & 0 & \cdots & 0 & 0 \\[2mm]
0 & \dfrac{\mu_S(n)}{n} & 1 - \dfrac{\mu_S(n)}{n} & 0 & \cdots & 0 & 0 \\[2mm]
\vdots & \vdots & & \ddots & & \vdots & \vdots \\[2mm]
0 & 0 & 0 & 0 & \cdots & \dfrac{\mu_S(n)}{n} & 1 - \dfrac{\mu_S(n)}{n}
\end{pmatrix}^{j}
\begin{pmatrix} n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.
$$

The rows and columns of these $(k+1) \times (k+1)$ matrices are indexed by the coupon types $0, 1, \ldots, k$. A matrix decomposition will help us asymptotically simplify this matrix expression. We can write the average vector as

$$
\mathrm{E}[X_j^{(n)}] = \left(\frac{\mu_S(n)}{n} \boldsymbol{B} + \left(1 - \frac{\mu_S(n)}{n} \right) \boldsymbol{I} \right)^j
\begin{pmatrix} n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}
= \sum_{i=0}^{k} \frac{\mu_S^i(n)}{n^i} \left(1 - \frac{\mu_S(n)}{n} \right)^{j-i} \binom{j}{i} \boldsymbol{B}^i
\begin{pmatrix} n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}
$$

for $j \geq k$. Note that we stopped the sum at k, because all higher powers of $\boldsymbol{B}$ are identically $\boldsymbol{0}$. Thus,

$$
\mathrm{E}[X_j^{(n)}] =
\begin{pmatrix}
n \left(1 - \dfrac{\mu_S(n)}{n} \right)^j \\[3mm]
j \mu_S(n) \left(1 - \dfrac{\mu_S(n)}{n} \right)^{j-1} \\[3mm]
\dfrac{j(j-1)\mu_S^2(n)}{2n} \left(1 - \dfrac{\mu_S(n)}{n} \right)^{j-2} \\[3mm]
\vdots \\[3mm]
\dfrac{\mu_S^k(n)}{n^{k-1}} \binom{j}{k} \left(1 - \dfrac{\mu_S(n)}{n} \right)^{j-k}
\end{pmatrix}.
\tag{2}
$$

6.2. The covariance structure

Toward the covariance calculation, let us take the conditional expectation of the cross products of the counts, i.e.

$$
\mathrm{E}[X_{j,\ell}^{(n)} X_{j,m}^{(n)} \mid \mathcal{F}_{j-1}] = (X_{j-1,\ell}^{(n)} + H_{j,\ell-1}^{(n)} - H_{j,\ell}^{(n)})(X_{j-1,m}^{(n)} + H_{j,m-1}^{(n)} - H_{j,m}^{(n)}),
$$

yielding the expectation

$$
\begin{aligned}
A_j^{(n)}(\ell, m) &:= \mathrm{E}[X_{j,\ell}^{(n)} X_{j,m}^{(n)}] \\
&= \mathrm{E}[X_{j-1,\ell}^{(n)} X_{j-1,m}^{(n)}] \\
&\quad + (\mathrm{E}[X_{j-1,\ell}^{(n)} X_{j-1,m-1}^{(n)}] - 2\,\mathrm{E}[X_{j-1,\ell}^{(n)} X_{j-1,m}^{(n)}] + \mathrm{E}[X_{j-1,\ell-1}^{(n)} X_{j-1,m}^{(n)}]) \frac{\mathrm{E}[S_j]}{n} \\
&\quad + \mathrm{E}[H_{j,\ell-1}^{(n)} X_{j,m-1}^{(n)} - H_{j,\ell-1}^{(n)} X_{j,m}^{(n)} - H_{j,\ell}^{(n)} X_{j,m-1}^{(n)} + H_{j,\ell}^{(n)} X_{j,m}^{(n)}].
\end{aligned}
$$

We use the known expressions for the mean and covariance of the multihypergeometric distribution (see Stuart and Ord (1987, Article 5.14)). Let us now construct the matrix $\mathrm{E}[X_j^{(n)} \overline{X}_j^{(n)}] =:$ $A_j^{(n)} = [A_j^{(n)}(\ell, m)]_{0 \le \ell, m, \le k}$. After some rather lengthy algebraic operations we reach the recurrence

$$
A_j^{(n)} = A_{j-1}^{(n)} + \frac{\mu_S(n)}{n}(B A_{j-1}^{(n)} + A_{j-1}^{(n)} \overline{B}) + B A_{j-1}^{(n)} \overline{B} + h_n \tilde{K}_j^{(n)},
$$

where $\tilde{K}_j^{(n)}$ is the diagonal matrix $\mathrm{diag}(\mathrm{E}[X_{j-1,0}^{(n)}], \dots, \mathrm{E}[X_{j-1,k}^{(n)}])$. The recurrence in this form is not easy to iterate. Nonetheless, a reorganization in the form

$$
A_j^{(n)} = g_n G A_{j-1}^{(n)} \overline{G} + C A_{j-1}^{(n)} \overline{C} + h_n K_j^{(n)},
$$

where $G = B - I$, $C = I + \mu_S(n) G / n$, and $K_j^{(n)} = (B - I) \tilde{K}_j^{(n)} (\overline{B} - I)$, helps us iterate the recurrence.

A few steps of iteration reveal a pattern, which can then be proved by induction. The recurrence has the exact solution

$$
\begin{aligned}
\mathrm{E}[X_j^{(n)} \overline{X}_j^{(n)}] &= h_n \sum_{i=0}^{j-1} \sum_{r=0}^{i} g_n^{i-r} \binom{i}{r} C^r G^{i-r} K_{j-i}^{(n)} \overline{G}^{i-r} \overline{C}^r \\
&\quad + \sum_{r=0}^{j} g_n^{j-r} \binom{j}{r} C^r G^{j-r} A_0^{(n)} \overline{G}^{j-r} \overline{C}^r.
\end{aligned}
$$

It can be shown (by induction for example) that

$$
C^p := \frac{1}{n^p} \left[\binom{p}{\ell - m} \mu_S(n)^{\ell - m} (n - \mu_S(n))^{p - \ell + m} \right]_{0 \le \ell, m \le k}
$$

and

$$
G^q := \left[(-1)^{\ell - m + q} \binom{q}{\ell - m} \right]_{0 \le \ell, m \le k};
$$

interpret $\binom{y}{x}$ as 0 whenever x is negative or $y < x$. Multiplying out, we obtain

$$
\begin{aligned}
&[C^r G^{j-r} A_0^{(n)} \overline{G}^{j-r} \overline{C}^r]_{\ell, m} \\
&= \frac{1}{n^{2r-2}} \left(\sum_{p=0}^{\ell} (-1)^p \binom{r}{\ell - p} \binom{j-r}{p} \mu_S^{\ell - p}(n)(n - \mu_S(n))^{r - \ell + p} \right) \\
&\quad \times \left(\sum_{q=0}^{m} (-1)^q \binom{r}{m - q} \binom{j-r}{q} \mu_S^{m - q}(n)(n - \mu_S(n))^{r - m + q} \right).
\end{aligned}
$$

These calculations are facilitated by a symbolic algebra system such as MATHEMATICA® or MAPLE®. Similarly, we have

$$[\boldsymbol{C}^r \boldsymbol{G}^{i-r} \boldsymbol{K}_{j-i} \overline{\boldsymbol{G}}^{i-r} \overline{\boldsymbol{C}}^r]_{\ell,m}$$

$$= \frac{1}{n^{2r}} \sum_{v=0}^{k} \left(\sum_{p=0}^{k} (-1)^{p-v} \binom{r}{\ell-p} \binom{i-r+1}{p-v} \mu_S(n)^{\ell-p} (n - \mu_S(n))^{r-\ell+p} \right)$$

$$\times \frac{\mu_S^v(n)}{n^{v-1}} \binom{j}{v} \left(1 - \frac{\mu_S(n)}{n} \right)^{j-i-1-v}$$

$$\times \left(\sum_{q=0}^{k} (-1)^{q-v} \binom{r}{m-q} \binom{i-r+1}{q-v} \mu_S(n)^{m-q} (n - \mu_S(n))^{r-m+q} \right).$$

Putting these calculations together, we obtain an exact expression for the covariance.

Proposition 1. *We have*

$$\mathrm{cov}[X_{j,\ell}^{(n)}, X_{j,m}^{(n)}] = \frac{\mu_S(n)n - \sigma_S^2(n) - \mu_S^2(n)}{n(n-1)}$$

$$\times \left[\sum_{i=0}^{j-1} \sum_{r=0}^{i} \left(\frac{(\sigma_S^2(n) - \mu_S(n))n + \mu_S^2(n)}{n^2(n-1)} \right)^{i-r} \binom{i}{r} \frac{1}{n^{2r-2}} \right.$$

$$\times \sum_{v=0}^{k} \left(\sum_{p=0}^{k} (-1)^{p-v} \binom{r}{\ell-p} \binom{i-r+1}{p-v} \mu_S^{\ell-p}(n) \right.$$

$$\times (n - \mu_S(n))^{r-\ell+p} \Big)$$

$$\times \frac{\mu_S^v(n)}{n^{v-1}} \binom{j}{v} \left(1 - \frac{\mu_S(n)}{n} \right)^{j-i-1-v}$$

$$\times \left(\sum_{q=0}^{k} (-1)^{q-v} \binom{r}{m-q} \binom{i-r+1}{q-v} \mu_S^{m-q}(n) \right.$$

$$\times (n - \mu_S(n))^{r-m+q} \Big) \bigg]$$

$$+ n^2 \left[\sum_{r=0}^{j} \left(\frac{(\sigma_S^2(n) - \mu_S(n))n + \mu_S^2(n)}{n^2(n-1)} \right)^{j-r} \binom{j}{r} \right.$$

$$\times \left(\sum_{p=0}^{\ell} (-1)^p \binom{r}{\ell-p} \binom{j-r}{p} \mu_S^{\ell-p}(n)(n - \mu_S(n))^{r-\ell+p} \right)$$

$$\times \left(\sum_{q=0}^{m} (-1)^q \binom{r}{m-q} \binom{j-r}{q} \mu_S^{m-q}(n)(n - \mu_S(n))^{r-m+q} \right) \bigg]$$

$$- \frac{\mu_S^{\ell+m}(n)}{n^{\ell+m-2}} \binom{j}{\ell} \binom{j}{m} \left(1 - \frac{\mu_S(n)}{n} \right)^{2j-\ell-m}.$$

7. The underlying multivariate martingale

Condition recurrence (1) on $\mathcal{F}_{j-1}$ to obtain

$$\mathrm{E}[X_{j,r}^{(n)} \mid \mathcal{F}_{j-1}] = \left(1 - \frac{\mu_S(n)}{n}\right) X_{j-1,r}^{(n)} + \frac{\mu_S(n)}{n} X_{j-1,r-1}^{(n)}.$$

Putting the recurrences for different colors together in one matrix form, we have

$$\mathrm{E}[\boldsymbol{X}_j^{(n)} \mid \mathcal{F}_{j-1}] = \left(\frac{\mu_S(n)}{n} \boldsymbol{B} + \left(1 - \frac{\mu_S(n)}{n}\right) \boldsymbol{I}\right) \boldsymbol{X}_{j-1}^{(n)}.$$

Therefore,

$$\mathrm{E}[\boldsymbol{X}_j^{(n)} \mid \mathcal{F}_{j-1}] := \boldsymbol{M} \boldsymbol{X}_{j-1}^{(n)},$$

and $\boldsymbol{Y}_j^{(n)} := \boldsymbol{M}^{-j} \boldsymbol{X}_j^{(n)}$ is a martingale and $\tilde{\boldsymbol{Y}}_j^{(n)} = \boldsymbol{M}^{-j} \boldsymbol{X}_j^{(n)} - \boldsymbol{X}_0^{(n)}$ is a centered martingale.

For suitable scale factors ξ_n for each phase, we will check Lindeberg's conditional condition, that is,

$$U_n := \sum_{i=1}^{j_n} \mathrm{E}\left[\left\|\frac{1}{\xi_n} \nabla \tilde{\boldsymbol{Y}}_i^{(n)}\right\|^2 \mathbf{1}_{\{\|\nabla \tilde{\boldsymbol{Y}}_i^{(n)}/\xi_n\| > \varepsilon\}} \;\middle|\; \mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 0,$$

and the conditional variance condition, that is,

$$\boldsymbol{V}_n := \sum_{i=1}^{j_n} \mathrm{cov}\left[\frac{1}{\xi_n} \nabla \tilde{\boldsymbol{Y}}_i^{(n)} \;\middle|\; \mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} \boldsymbol{\Gamma}$$

for a covariance matrix $\boldsymbol{\Gamma}$.

When both conditions hold, the sum $\sum_{j=1}^{j_n} \xi_n^{-1} \nabla \tilde{\boldsymbol{Y}}_j^{(n)} = \xi_n^{-1}(\boldsymbol{M}^{-j_n} \boldsymbol{X}_{j_n}^{(n)} - \boldsymbol{X}_0^{(n)})$ converges to the multinormally distributed random vector $\mathcal{N}_{k+1}(\boldsymbol{0}, \boldsymbol{\Gamma})$; this follows from an appropriate extension of the univariate martingale central limit theorem in Hall and Heyde (1980, p. 58), via, for example, the Cramér–Wold device. Some of the limiting covariance matrices that appear in this work are singular, with fewer linear combinations that have a proper nonsingular multivariate normal distribution (with a positive definite covariance matrix). The following lemma gives an exact computation of $\boldsymbol{V}_n$, and will be helpful in all the phases via an appropriate asymptotic analysis.

Lemma 1. *We have*

$$\boldsymbol{V}_n = \frac{1}{\xi_n^2} \sum_{i=1}^{j_n} [g_n \boldsymbol{M}^{-i} (\boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{B} \boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} \overline{\boldsymbol{B}} \boldsymbol{B} \boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} \overline{\boldsymbol{B}}) \overline{\boldsymbol{M}}^{-i}$$

$$+ h_n \boldsymbol{M}^{-i} (\boldsymbol{D}_i^{(n)} - \boldsymbol{B} \boldsymbol{D}_i^{(n)} - \boldsymbol{D}_i^{(n)} \overline{\boldsymbol{B}} + \boldsymbol{B} \boldsymbol{D}_i^{(n)} \overline{\boldsymbol{B}}) \overline{\boldsymbol{M}}^{-i}],$$

where $\boldsymbol{D}_i^{(n)}$ is the diagonal matrix $\mathrm{diag}(X_{i-1,0}^{(n)}, X_{i-1,1}^{(n)}, \ldots, X_{i-1,k}^{(n)})$.

Proof. Start with the definition of conditional covariance to obtain

$$\mathrm{cov}[\nabla \tilde{\boldsymbol{Y}}_i^{(n)} \mid \mathcal{F}_{i-1}] = \mathrm{cov}[\nabla (\boldsymbol{M}^{-i} \boldsymbol{X}_i^{(n)} - \boldsymbol{X}_0^{(n)}) \mid \mathcal{F}_{i-1}]$$

$$= \mathrm{E}[(\boldsymbol{M}^{-i} \boldsymbol{X}_i^{(n)} - \boldsymbol{M}^{-i+1} \boldsymbol{X}_{i-1}^{(n)})(\overline{\boldsymbol{X}}_i^{(n)} \overline{\boldsymbol{M}}^{-i} - \overline{\boldsymbol{X}}_{i-1}^{(n)} \overline{\boldsymbol{M}}^{-i+1}) \mid \mathcal{F}_{i-1}].$$

Using the stochastic recurrence (1), we expand the products in the covariance and write it as

$$
\begin{aligned}
\mathrm{cov}[&\nabla \tilde{Y}_i^{(n)} \mid \mathscr{F}_{i-1}] \\
&= g_n \boldsymbol{M}^{-i}(\boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{B}\boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} \overline{\boldsymbol{B}} + \boldsymbol{B}\boldsymbol{X}_{i-1}^{(n)} \overline{\boldsymbol{X}}_{i-1}^{(n)} \overline{\boldsymbol{B}})\overline{\boldsymbol{M}}^{-i} \\
&\quad + h_n \boldsymbol{M}^{-i}(\boldsymbol{D}_i^{(n)} - \boldsymbol{B}\boldsymbol{D}_i^{(n)} - \boldsymbol{D}_i^{(n)} \overline{\boldsymbol{B}} + \boldsymbol{B}\boldsymbol{D}_i^{(n)} \overline{\boldsymbol{B}})\overline{\boldsymbol{M}}^{-i}.
\end{aligned}
$$

We now sum these terms to obtain an exact expression for $\boldsymbol{V}_n$.

8. The sublinear phase

We are in the sublinear phase when $j = j_n = o(n/\mu_S(n))$. In this phase we take

$$
\xi_n = \sqrt{\frac{\mu_S^2(n)j_n^2}{2n} + \sigma_S^2(n)j_n}.
$$

Lemma 2. *For $j = o(n/\mu_S(n))$ in the sublinear phase,*

$$
\|\nabla \tilde{Y}_j^{(n)}\| \le 4(k+1)\sqrt{2k+1}\, s_n.
$$

Proof. Set

$$
\overline{\boldsymbol{H}}_j^{(n)} = (H_{j,0}^{(n)}, H_{j,1}^{(n)}, \ldots, H_{j,k}^{(n)}),
$$

so as to write (1) in the form

$$
\boldsymbol{X}_j^{(n)} = \boldsymbol{X}_{j-1}^{(n)} + \boldsymbol{B}\boldsymbol{H}_j^{(n)} - \boldsymbol{H}_j^{(n)}. \tag{3}
$$

We bound each component of $\boldsymbol{X}_{j-1}^{(n)}$ by n, and it follows that, for large n,

$$
\begin{aligned}
\|\nabla \tilde{Y}_j^{(n)}\| &= \|\boldsymbol{Y}_j^{(n)} - \boldsymbol{Y}_{j-1}^{(n)}\| \\
&= \|\boldsymbol{M}^{-j}\boldsymbol{X}_j^{(n)} - \boldsymbol{M}^{-j+1}\boldsymbol{X}_{j-1}^{(n)}\| \\
&= \|\boldsymbol{M}^{-j}\|\,\|\boldsymbol{X}_j^{(n)} - \boldsymbol{M}\boldsymbol{X}_{j-1}^{(n)}\| \\
&= \left\|\left(\boldsymbol{I} + \boldsymbol{O}\!\left(\frac{\mu_S(n)}{n}\right)\right)\right\|\,\|(\boldsymbol{X}_{j-1}^{(n)} + \boldsymbol{B}\boldsymbol{H}_j^{(n)} - \boldsymbol{H}_j^{(n)}) - \boldsymbol{M}\boldsymbol{X}_{j-1}^{(n)}\| \\
&\le 2\sqrt{k+1}\left(\left\|\frac{\mu_S(n)}{n}(\boldsymbol{I} - \boldsymbol{B})\right\|\,\|\boldsymbol{X}_{j-1}^{(n)}\| + \|\boldsymbol{B} - \boldsymbol{I}\|\,\|\boldsymbol{H}_j^{(n)}\|\right) \\
&= 2\sqrt{k+1}\|\boldsymbol{B} - \boldsymbol{I}\|\left(\frac{\mu_S(n)}{n}\|\boldsymbol{X}_{j-1}^{(n)}\| + \|\boldsymbol{H}_j^{(n)}\|\right) \\
&\le 2\sqrt{(k+1)(2k+1)}\left(\frac{\mu_S(n)}{n}(\sqrt{k+1})n + s_n\sqrt{k+1}\right) \\
&\le 2(k+1)\sqrt{2k+1}(\mu_S(n) + s_n) \\
&\le 2(k+1)(\sqrt{2k+1})2s_n.
\end{aligned}
$$

Lemma 3. *If j_n is sublinear but restricted to the upper sublinear phase where*

$$
s_n = o(\xi_n) = o\!\left(\sqrt{\frac{\mu_S^2(n)j_n^2}{2n} + \sigma_S^2(n)j_n}\right),
$$

and if $\xi_n \to \infty$, we then have

$$U_n := \sum_{i=1}^{j_n} \mathrm{E}\left[\left\|\frac{1}{\xi_n}\nabla\tilde{\boldsymbol{Y}}_i^{(n)}\right\|^2 \mathbf{1}_{\{\|\nabla\tilde{\boldsymbol{Y}}_i^{(n)}/\xi_n\|>\varepsilon\}} \;\middle|\; \mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 0.$$

Proof. For any given $\varepsilon > 0$, according to Lemma 2, the sets

$$\left\{\left\|\frac{1}{\sqrt{\mu_S^2(n)j_n^2/2n + \sigma_S^2(n)j_n}}\nabla\tilde{\boldsymbol{Y}}_i^{(n)}\right\| > \varepsilon\right\}$$

are empty for all n greater than some positive integer $n_0(\varepsilon)$. For $n \geq n_0(\varepsilon)$, in view of the restriction of j_n in an upper sublinear phase, we have

$$U_n \leq \sum_{j=1}^{n_0(\varepsilon)} \frac{16(k+1)^2(2k+1)s_n^2}{\mu_S^2(n)j_n^2/2n + \sigma_S^2(n)j_n}$$

$$= 16(k+1)^2(2k+1)n_0(\varepsilon)\frac{s_n^2}{\mu_S^2(n)j_n^2/2n + \sigma_S^2(n)j_n}$$

$$\to 0 \quad \text{as } n \to \infty.$$

It was shown in Mahmoud (2010) that in the sublinear phase most of the draws produce type-0 balls, which are converted into type-1 balls. That is, for any $0 \leq i \leq j_n$,

$$X_{i,0}^{(n)} = n - i\mu_S(n) + o_\mathrm{P}(s_n j_n), \qquad X_{i,1}^{(n)} = i\mu_S(n) + o_\mathrm{P}(s_n j_n),$$

and

$$X_{i,r}^{(n)} = o_\mathrm{P}(s_n j_n) \quad \text{for } 2 \leq r \leq k+1.$$

We can represent this as

$$\boldsymbol{X}_i^{(n)} = n\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix} + i\mu_S(n)\begin{pmatrix}-1\\1\\0\\\vdots\\0\end{pmatrix} + o_\mathrm{P}(s_n j_n),$$

so that $\boldsymbol{X}_i^{(n)} = n\boldsymbol{F} + i\mu_S(n)\boldsymbol{G} + o_\mathrm{P}(s_n j_n)$ and $\boldsymbol{X}_i^{(n)}\overline{\boldsymbol{X}}_i^{(n)} = n^2\boldsymbol{F}\overline{\boldsymbol{F}} + in\mu_S(n)(\boldsymbol{F}\overline{\boldsymbol{G}} + \boldsymbol{G}\overline{\boldsymbol{F}}) + o_\mathrm{P}(ns_n j_n)$. We shall also need the diagonal matrix $\boldsymbol{D} = \mathrm{diag}(-1, 1, 0, \ldots, 0)$. In the entire sublinear phase, $\boldsymbol{M}^{-i} = \boldsymbol{I} + \boldsymbol{O}(\mu_S(n)/n)$. By Lemma 1 we can develop asymptotics:

$$V_n = \frac{g_n}{\xi_n^2}\sum_{i=1}^{j_n}\left(\boldsymbol{I} + \boldsymbol{O}\left(\frac{\mu_S(n)}{n}\right)\right)$$

$$\times [(n^2\boldsymbol{F}\overline{\boldsymbol{F}} + in\mu_S(n)(\boldsymbol{F}\overline{\boldsymbol{G}} + \boldsymbol{G}\overline{\boldsymbol{F}}) + o_\mathrm{P}(ns_n j_n))$$

$$-\boldsymbol{B}(n^2\boldsymbol{F}\overline{\boldsymbol{F}} + in\mu_S(n)(\boldsymbol{F}\overline{\boldsymbol{G}} + \boldsymbol{G}\overline{\boldsymbol{F}}) + o_\mathrm{P}(ns_n j_n))$$

$$-(n^2\boldsymbol{F}\overline{\boldsymbol{F}} + in\mu_S(n)(\boldsymbol{F}\overline{\boldsymbol{G}} + \boldsymbol{G}\overline{\boldsymbol{F}}) + o_\mathrm{P}(ns_n j_n))\overline{\boldsymbol{B}}$$

$$+\boldsymbol{B}(n^2\boldsymbol{F}\overline{\boldsymbol{F}} + in\mu_S(n)(\boldsymbol{F}\overline{\boldsymbol{G}} + \boldsymbol{G}\overline{\boldsymbol{F}}) + o_\mathrm{P}(ns_n j_n))\overline{\boldsymbol{B}}]\left(\boldsymbol{I} + \boldsymbol{O}\left(\frac{\mu_S(n)}{n}\right)\right)$$

$$+ \frac{h_n}{\xi_n} \sum_{i=1}^{j_n} \left(I + O\left(\frac{\mu_S(n)}{n} \right) \right)$$

$$\times [(nF\overline{F} + in\mu_S(n)D + o_P(ns_n j_n))$$
$$- B(nF\overline{F} + in\mu_S(n)D + o_P(ns_n j_n))$$
$$- (nF\overline{F} + in\mu_S(n)D + o_P(ns_n j_n))\overline{B}$$
$$+ B(nF\overline{F} + in\mu_S(n)D + o_P(ns_n j_n))\overline{B}] \left(I + O\left(\frac{\mu_S(n)}{n} \right) \right).$$

Let $R := F\overline{G} + G\overline{F}$. Collecting the like terms, after a lengthy calculation we obtain

$$V_n = \frac{\sigma_S^2(n) j_n}{\xi_n^2} (F\overline{F} - BF\overline{F} - F\overline{F}B + BF\overline{F}B)$$

$$+ \frac{\mu_S(n) j_n^2}{2\xi_n^2 n(n-1)} (((\sigma_S^2(n) - \mu_S(n))n + \mu_S^2(n))(R - BR - R\overline{B} + BR\overline{B})$$

$$+ (\mu_S(n)n - \sigma_S^2(n) - \mu_S^2(n))(D - BD - D\overline{B} + BD\overline{B})) + o_P(1)$$

$$= \frac{\sigma_S^2(n) j_n}{\xi_n^2} \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$+ \frac{\mu_S(n) j_n^2}{2\xi_n^2 n(n-1)} \left(n\sigma_S^2(n) \begin{pmatrix} -2 & 3 & -1 & 0 & \cdots & 0 \\ 3 & -4 & 1 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \right.$$

$$+ (n\mu_S(n) - \mu_S^2(n)) \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ -2 & 4 & -2 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\left. + \sigma_S^2(n) \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \right) + o_P(1).$$

Curiously, this expression can give many different asymptotics according to the interplay between the factors $\mu_S(n)$, $\sigma_S^2(n)$, the range s_n, and the phase of j_n. In all cases $\xi_n \to \infty$ is required for convergence, and we consistently used the range condition $s_n = o(\sqrt{n})$. When the term $\sigma_S^2(n) j_n$ is dominant in ξ_n, i.e. when $\mu_S(n) j_n/\sqrt{n} = o(\sigma_S(n)\sqrt{j_n})$, and, thus, $\alpha_n \to 0$, we obtain the convergence

$$
V_n \xrightarrow{\text{P}}
\begin{pmatrix}
1 & -1 & 0 & 0 & \cdots & 0 \\
-1 & 1 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 0
\end{pmatrix}.
$$

Under this scaling, Theorem 1(b) follows. However, when

$$
\frac{\mu_S(n) j_n/\sqrt{2n}}{\sqrt{\mu_S^2(n) j_n^2/2n + \sigma_S^2(n) j(n)}} \to \alpha \in (0, 1],
$$

we obtain Theorem 1(a).

As noted previously, when $\alpha = 1$, the matrix in part (a) of Theorem 1 is singular. The reason for this singularity can be most easily seen in the standard case of $\mu_S(n) \equiv 1$ and $\sigma_S(n) \equiv 0$ (a special case of $\alpha = 1$). In this case we have, for example (for large n),

$$
X_{j_n,0}^{(n)} \stackrel{\text{D}}{\approx} \mathcal{N}\left(n e^{-j_n/n}, \frac{j_n^2}{2n} \right), \qquad
X_{j_n,1}^{(n)} \stackrel{\text{D}}{\approx} \mathcal{N}\left(j_n e^{-j_n/n}, \frac{2 j_n^2}{n} \right)
$$

(cf. Kolchin *et al.* (1978, p. 38)); here the notation '$\stackrel{\text{D}}{\approx}$' means approximate equality in distribution. The correlation between $X_{j_n,0}^{(n)}$ and $X_{j_n,1}^{(n)}$ is approximately $(-j_n^2/n) e^{-2j_n/n}/(j_n^2/n) \to -1$ as $n \to \infty$, so the bivariate limit distribution is degenerate; the same result holds for types 0 and 2, and for types 1 and 2, and for more general values of $\mu_S(n)$ and $\sigma_S(n)$ that result in $\alpha = 1$. In other words, noting that, for $\alpha = 1$, the rank of the covariance matrix in Theorem 1(a) is 1, there is a 'proper' multivariate central limit theorem for one combination of the three random variables—in fact, $X_{j_n,0}^{(n)}$ is asymptotically normal, and $X_{j_n,1}^{(n)}$ and $X_{j_n,2}^{(n)}$ are both linear combinations of $X_{j_n,0}^{(n)}$.

For $r > 2$, the variables $X_{j_n,r}^{(n)}$ have asymptotically normal distributions in part (a) of Theorem 1, but the variances can be of smaller order than $\mu_n^2 j_n^2/n$, as is the case in standard coupon collection, with $\mu_S(n) \equiv 1$ and $\sigma_S(n) \equiv 0$.

In Theorem 1(b), a univariate central limit theorem holds for $r \geq 2$, but with variances of lower order than $\sigma_n^2 j_n$. As the covariance matrix indicates, $X_{j_n,0}^{(n)}$ and $X_{j_n,1}^{(n)}$ have correlation -1, but the reason in this case is different from that in part (a). In this case, the term involving $\sigma_S^2(n)$ dominates the variances of $X_{j_n,0}^{(n)}$ and $X_{j_n,1}^{(n)}$, and, asymptotically, the randomness in both variables is due to $\{S_i(n)\}$, the random sequence of draws, causing the degeneracy.

9. The linear phase

We are in the linear phase when $j = j_n \sim \lambda_n n/\mu_S(n)$. In this phase we take

$$
\xi_n = \sqrt{n\left(1 + \frac{\sigma_S^2(n)}{\mu_S(n)}\right)}.
$$

We show next that in the linear phase both $X_j^{(n)}$ and $\overline{X}_j^{(n)}$ can be approximated by the leading term of the mean with ignorable errors.

Lemma 4. *For $j \sim \lambda_n n / \mu_S(n)$ and every $r \in \{0, 1, \ldots, k\}$, we have*

$$X_{j,r}^{(n)} = \lambda_n n e^{-\lambda_n} + o_{\mathcal{L}_1}(n).$$

Proof. This proof involves extensive computation, and we only highlight its salient points. Let $j = j_n$ be in the linear phase, i.e. $j \sim \lambda_n n / \mu_S(n)$. The exact covariance, as given in Proposition 1, has three parts: the first part is a double sum on $r = 0, \ldots, i$ and $i = 0, \ldots, j-1$; the middle part is a sum on $r = 0, \ldots, j$; and the third part is the negative of the product of the means of $X_{j,\ell}^{(n)}$ and $X_{j,m}^{(n)}$. By direct inspection of the exact mean (2) we see that in the linear phase

$$\mathrm{E}[X_{j,r}^{(n)}] = \frac{\lambda_n^r n}{r!} e^{-\lambda_n} + O(\mu_S(n)).$$

Therefore, the third part is $-\lambda_n^{2r} n^2 e^{-2\lambda_n} / \ell! \, m! + o(n^2)$. The middle part is also asymptotically $\lambda_n^{2r} n^2 e^{-2\lambda_n} / \ell! \, m! + o(n^2)$. We thus obtain an exact cancellation of the n^2 term between the second and third parts, leaving $o(n^2)$.

In the first part, the term $\mu_S^\nu(n) n^{-\nu+1} \binom{j}{\nu} (1 - \mu_S(n)/n)^{j-i-1-\nu}$ is the mean of the number of balls of type ν after $j - i - 1$ draws. As we vary i and r (each up to the linear phase), this term remains $O(n)$. Combined with all other elements, the first part is $o(n^2)$. The three parts combined give $o(n^2)$ covariance between types ℓ and m (for $0 \le \ell, m \le k$).

Consequently, we have

$$\mathrm{E}\left[\left(X_{j,r}^{(n)} - \frac{\lambda_n^r}{r!} n e^{-\lambda_n}\right)^2\right] = \mathrm{var}[X_{j,r}^{(n)}] + \left(\mathrm{E}[X_{j,r}^{(n)}] - \frac{\lambda_n^r}{r!} n e^{-\lambda_n}\right)^2 = o(n^2).$$

So, by Jensen's inequality,

$$\mathrm{E}\left[\left| X_{j,r}^{(n)} - \frac{\lambda_n^r}{r!} e^{-\lambda_n} n \right|\right] \le \sqrt{\mathrm{E}\left[\left(X_{j,r}^{(n)} - \frac{\lambda_n^r}{r!} e^{-\lambda_n} n\right)^2\right]} = o(n),$$

which implies that

$$X_{j,r}^{(n)} = \frac{\lambda_n^r}{r!} e^{-\lambda_n} n + o_{\mathcal{L}_1}(n).$$

Lemma 5. *For $j \sim \lambda_n n / \mu_S(n)$ in the linear phase,*

$$\|\nabla \tilde{Y}_j^{(n)}\| = 4(k+1)^{5/2} \sqrt{2k+1} \, s_n.$$

Proof. It follows from the stochastic recurrence (3) that

$$
\begin{aligned}
\|\nabla \tilde{Y}_j^{(n)}\| &= \|Y_j^{(n)} - Y_{j-1}^{(n)}\| \\
&= \|M^{-j} X_j^{(n)} - M^{-j+1} X_{j-1}^{(n)}\| \\
&= \|M^{-j}\| \|X_j^{(n)} - M X_{j-1}^{(n)}\| \\
&= \left\| e^{(B-I)\lambda_n} \left(I + O\left(\frac{\mu_S(n)}{n}\right)\right) \right\| \|(X_{j-1}^{(n)} + B H_j^{(n)} - H_j^{(n)}) - M X_{j-1}^{(n)}\| \\
&\le 2 \|e^{(B-I)\lambda_n}\| \left(\left\| \frac{\mu_S(n)}{n}(I - B) \right\| \|X_{j-1}^{(n)}\| + \|B - I\| \|H_j^{(n)}\| \right)
\end{aligned}
$$

$$= 2\mathrm{e}^{-\lambda_n}(k+1)\|\mathrm{e}^{\lambda_n \boldsymbol{B}}\|\|\boldsymbol{B}-\boldsymbol{I}\|\left(\frac{\mu_S(n)}{n}\|\boldsymbol{X}_{j-1}^{(n)}\|+\|\boldsymbol{H}_j^{(n)}\|\right)$$

$$\leq 2\mathrm{e}^{-\lambda_n}(k+1)\|\mathrm{e}^{\lambda_n \boldsymbol{B}}\|\sqrt{2k+1}(s_n\sqrt{k+1}+s_n\sqrt{k+1}).$$

Note that

$$\mathrm{e}^{\lambda_n \boldsymbol{B}} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \lambda_n & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{\lambda_n^k}{k!} & \dfrac{\lambda_n^{k-1}}{(k-1)!} & \dfrac{\lambda_n^{k-2}}{(k-2)!} & \cdots & 1 \end{pmatrix},$$

with norm less than $(k+1)\mathrm{e}^{\lambda_n}$.

Lemma 6. *For* $j \sim \lambda_n n/\mu_S(n)$ *in the linear phase,*

$$U_n := \sum_{i=1}^{j_n} \mathrm{E}\left[\left\|\frac{1}{\xi_n}\nabla\tilde{\boldsymbol{Y}}_i^{(n)}\right\|^2 \mathbf{1}_{\{\|\nabla\tilde{\boldsymbol{Y}}_i^{(n)}/\xi_n\|>\varepsilon\}} \,\bigg|\, \mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 0.$$

Proof. For any given $\varepsilon > 0$, according to Lemma 5 and the constraint $s_n = o(\sqrt{n})$, the sets $\{\|\nabla\tilde{\boldsymbol{Y}}_i^{(n)}/\xi_n\| > \varepsilon\}$ are empty for all n greater than some positive integer $n_0(\varepsilon)$. For $n \geq n_0$, we have

$$U_n \leq \sum_{j=1}^{n_0(\varepsilon)} \frac{16(k+1)^5(2k+1)s_n^2}{n(1+\sigma_S^2(n)/\mu_S(n))} \leq \frac{16(k+1)^5(2k+1)n_0(\varepsilon)s_n^2}{n(1+\sigma_S^2(n)/\mu_S(n))} \to 0 \quad \text{as } n \to \infty.$$

To handle the conditional variance condition, we break up the sum over 1 to $j_n \sim \lambda_n n/\mu_S(n)$ at some point near the beginning of the linear phase. More precisely, choose a small positive $\varepsilon < Q_1$ and break up the sum in V_n into a sum going from 1 to $\lfloor \varepsilon n/\mu_S(n)\rfloor - 1$ and a sum starting at $\lfloor \varepsilon n/\mu_S(n)\rfloor$ and ending at j_n. For large n, we write V_n in the form

$$V_n = \frac{1}{\xi_n^2}\sum_{i=1}^{j_n}[g_n\boldsymbol{M}^{-i}(\boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{B}\boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)}\overline{\boldsymbol{B}} + \boldsymbol{B}\boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)}\overline{\boldsymbol{B}})\overline{\boldsymbol{M}}^{-i}$$

$$+ h_n\boldsymbol{M}^{-i}(\boldsymbol{D}_i^{(n)} - \boldsymbol{B}\boldsymbol{D}_i^{(n)} - \boldsymbol{D}_i^{(n)}\overline{\boldsymbol{B}} + \boldsymbol{B}\boldsymbol{D}_i^{(n)}\overline{\boldsymbol{B}})\overline{\boldsymbol{M}}^{-i}]$$

$$= \left(\frac{1}{\xi_n^2}\sum_{i=1}^{\lfloor \varepsilon n/\mu_S(n)\rfloor-1} + \frac{1}{\xi_n^2}\sum_{i=\lfloor \varepsilon n/\mu_S(n)\rfloor}^{j_n}\right)$$

$$\times [g_n\boldsymbol{M}^{-i}(\boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{B}\boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)} - \boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)}\overline{\boldsymbol{B}} + \boldsymbol{B}\boldsymbol{X}_{i-1}^{(n)}\overline{\boldsymbol{X}}_{i-1}^{(n)}\overline{\boldsymbol{B}})\overline{\boldsymbol{M}}^{-i}$$

$$+ h_n\boldsymbol{M}^{-i}(\boldsymbol{D}_i^{(n)} - \boldsymbol{B}\boldsymbol{D}_i^{(n)} - \boldsymbol{D}_i^{(n)}\overline{\boldsymbol{B}} + \boldsymbol{B}\boldsymbol{D}_i^{(n)}\overline{\boldsymbol{B}})\overline{\boldsymbol{M}}^{-i}]$$

$$=: \boldsymbol{a}_n' + \boldsymbol{a}_n''.$$

According to the restriction $s_n = o(\sqrt{n})$, we obtain

$$\boldsymbol{a}_n' = \frac{g_n}{n(1+\sigma_S^2(n)/\mu_S(n))}\sum_{i=1}^{\lfloor \varepsilon n/\mu_S(n)\rfloor-1}\boldsymbol{O}(n^2) + \frac{h_n}{n(1+\sigma_S^2(n)/\mu_S(n))}\sum_{i=1}^{\lfloor \varepsilon n/\mu_S(n)\rfloor-1}\boldsymbol{O}(n)$$

$$= \boldsymbol{o}(\varepsilon) \quad \text{as } \varepsilon \to 0.$$

Let

$$\overline{N}_i^{(n)} = \mathrm{e}^{-i\mu_S(n)/n}\left(n, i\mu_S(n), \frac{i^2\mu_S^2(n)}{2n}, \ldots, \frac{i^k\mu_S^k(n)}{k!\,n^{k-1}}\right).$$

According to Lemma 4, for $i \leq j$ in the linear phase,

$$X_{i-1}^{(n)} = N_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n),$$

and, subsequently,

$$X_{i-1}^{(n)}\overline{X}_{i-1}^{(n)} = N_i^{(n)}\overline{N}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n^2).$$

Let

$$\boldsymbol{L}_i^{(n)} = \mathrm{e}^{-i\mu_S(n)/n}\mathrm{diag}\left(n, i\mu_S(n), \ldots, \frac{i^k\mu_S^k(n)}{k!\,n^{k-1}}\right),$$

and go further with the computation:

$$
\begin{aligned}
\boldsymbol{a}_n'' &= \frac{g_n}{\xi_n^2}\sum_{i=\lfloor \varepsilon n\rfloor}^{j_n} \mathrm{e}^{-i\mu_S(n)(\boldsymbol{B}-\boldsymbol{I})/n}\left(\boldsymbol{I} + \boldsymbol{O}\left(\frac{\mu_S(n)}{n}\right)\right)\\
&\quad \times [(N_i^{(n)}\overline{N}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n^2)) - \boldsymbol{B}(N_i^{(n)}\overline{N}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n^2))\\
&\quad\quad - (N_i^{(n)}\overline{N}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n^2))\overline{\boldsymbol{B}} + \boldsymbol{B}(N_i^{(n)}\overline{N}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n^2))\overline{\boldsymbol{B}}]\\
&\quad \times \mathrm{e}^{-i\mu_S(n)(\overline{\boldsymbol{B}}-\boldsymbol{I})/n}\left(\boldsymbol{I} + \boldsymbol{O}\left(\frac{\mu_S(n)}{n}\right)\right)\\
&\quad + h_n\sum_{i=\lfloor \varepsilon n\rfloor}^{j_n}\mathrm{e}^{-i\mu_S(n)(\boldsymbol{B}-\boldsymbol{I})/n}\left(\boldsymbol{I} + \boldsymbol{O}\left(\frac{\mu_S(n)}{n}\right)\right)\\
&\quad \times [(\boldsymbol{L}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n)) - \boldsymbol{B}(\boldsymbol{L}_i^{(n)} + \boldsymbol{o}(n))\\
&\quad\quad - (\boldsymbol{L}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n))\overline{\boldsymbol{B}} + \boldsymbol{B}(\boldsymbol{L}_i^{(n)} + \boldsymbol{o}_{\mathcal{L}_1}(n))\overline{\boldsymbol{B}}]\\
&\quad \times \mathrm{e}^{-i\mu_S(n)(\overline{\boldsymbol{B}}-\boldsymbol{I})/n}\left(\boldsymbol{I} + \boldsymbol{O}\left(\frac{\mu_S(n)}{n}\right)\right).
\end{aligned}
$$

To be able to go through this computation, we first simplify the matrix exponentiation, i.e.

$$\mathrm{e}^{-i\mu_S(n)(\boldsymbol{B}-\boldsymbol{I})/n} = \mathrm{e}^{i\mu_S(n)/n}$$

$$
\times \begin{pmatrix}
1 & 0 & 0 & \cdots & 0\\
-\dfrac{i\mu_S(n)}{n} & 1 & 0 & \cdots & 0\\
\dfrac{i^2\mu_S^2(n)}{2n^2} & -\dfrac{i\mu_S(n)}{n} & 1 & \cdots & 0\\
\vdots & \vdots & & \ddots & \vdots\\
\dfrac{(-i)^k\mu_S^k(n)}{k!\,n^k} & \dfrac{(-i)^{k-1}\mu_S^{k-1}(n)}{(k-1)!\,n^{k-1}} & \dfrac{(-i)^{k-2}\mu_S^{k-2}(n)}{(k-2)!\,n^{k-2}} & \cdots & 1
\end{pmatrix},
$$

and $e^{-i\mu_S(n)(\bar{B}-I)/n}$ is, of course, its transpose. Multiplying out, we obtain the second sum

$$a_n'' = \frac{g_n}{n(1+\sigma_S^2(n)/\mu_S(n))} \sum_{i=\lfloor \varepsilon n/\mu_S(n)\rfloor}^{j_n} \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} (n^2 + o(n^2))$$

$$+ \frac{h_n}{n(1+\sigma_S^2(n)/\mu_S(n))} \sum_{i=\lfloor \varepsilon n/\mu_S(n)\rfloor}^{j_n} (e^{i\mu_S(n)/n} + o(1))c_{i,n},$$

where $c_{i,n}$ is an effectively computable matrix, for example, for $0 \le r \le k$, we have

$$c_{i,n}(0, r) = \frac{(-1)^r i^{r-1} \mu_S^{r-1}(n)}{r! \, n^{r-1}} (i\mu_S(n) + rn)$$

and

$$c_{i,n}(1, 1) = \frac{1}{n}(i^2 \mu_S^2(n) + 3in\mu_S(n) + n^2),$$

etc.

When we put everything together, many cancellations take place, and a tremendous amount of calculation is needed. We only hint to how one covariance may be obtained to give an indication of the work involved. Let us take the $(0, 0)$ entry (which is one of the simplest). Mahmoud (2010) gave this calculation in detail for the case of a bounded range, $s_n = O(1)$, in which $\beta_n \to \beta \in (0, 1]$. Let us take here the opposite case, when the variance dominates the mean, i.e. when $\mu_S(n) = o(\sigma_S^2(n))$, a case where $\beta_n \to 0$. We obtain

$$V_n(0, 0) = a_n'(0, 0) + a_n''(0, 0)$$

$$= O(\varepsilon) + \left(\frac{g_n}{n\sigma_S^2(n)/\mu_S(n)} \sum_{i=\lfloor \varepsilon n/\mu_S(n)\rfloor}^{j_n} n^2 \right.$$

$$+ \frac{h_n}{n\sigma_S^2(n)/\mu_S(n)} \sum_{i=\lfloor \varepsilon n/\mu_S(n)\rfloor}^{j_n} n e^{i\mu_S(n)/n} \bigg) (1 + o(1))$$

$$= O(\varepsilon) + \left(\frac{n\mu_S(n)g_n}{\sigma_S^2(n)} \left(j_n - \left\lfloor \frac{\varepsilon n}{\mu_S(n)} \right\rfloor + 1 \right) \right.$$

$$+ \frac{\mu_S(n)h_n}{\sigma_S^2(n)} \left[\left(\frac{e^{(j_n+1)\mu_S(n)/n} - 1}{e^{\mu_S(n)/n} - 1} \right) - \left(\frac{e^{\lfloor \varepsilon n/\mu_S(n)\rfloor} - 1}{e^{\mu_S(n)/n} - 1} \right) \right] \bigg) (1 + o(1)).$$

Let ε approach 0, and write the limit

$$V_n(0, 0) = \left(\frac{n\mu_S(n)g_n}{\sigma_S^2(n)}(j_n + 1) + \frac{\mu_S(n)h_n}{\sigma_S^2(n)} \left(\frac{e^{(j_n+1)\mu_S(n)/n} - 1}{e^{\mu_S(n)/n} - 1} \right) \right) (1 + o(1)).$$

We are in a phase where $j_n \sim \lambda_n n/\mu_S(n)$, and the latter expression is asymptotically

$$V_n(0, 0) \sim \frac{n\mu_S(n)((\sigma_S^2(n) - \mu_S(n))n + \mu_S^2(n))}{n^2(n-1)\sigma_S^2(n)} \left(\frac{\lambda_n n}{\mu_S(n)} \right)$$

$$+ \frac{\mu_S(n)(\mu_S(n)n - \sigma_S^2(n) - \mu_S^2(n))}{n(n-1)\sigma_S^2(n)} \left(\frac{e^{\lambda_n} - 1}{\mu_S(n)/n} \right).$$

We now use the assumptions that the mean and variance are small relative to n, and the dominance of the variance to arrive at

$$V_n(0, 0) \sim \lambda_n + \frac{\mu_S(n)}{\sigma_S^2(n)}(e^{\lambda_n} - 1) = \lambda_n + o(1).$$

So, if $\lambda_n \to \lambda > 0$, we have $V_n(0, 0) \to \lambda$. In a like manner we can obtain the other entries of the limit of V_n, and find that they are all 0, except the entries (ℓ, m) for $0 \le \ell, m \le 1$, which are $(-1)^{\ell+m}\lambda$.

For $\beta > 0$, an application of the martingale central limit theorem gives

$$\frac{1}{\sqrt{n}} \left(M^{-j_n} X_{j_n}^{(n)} - \begin{pmatrix} n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \xrightarrow{\mathrm{D}} e^{\lambda} e^{-\lambda B} \mathcal{N}_{k+1} \left(0, J_{k+1} + \frac{1-\beta}{\beta} J'_{k+1} \right).$$

As we assumed that $\lambda_n \to \lambda > 0$, we have $M^{-j_n} \to e^{\lambda} e^{-\lambda B}$. By an application of the (multivariate) Slutsky theorem, we obtain the statement of Theorem 2(a).

For $\beta = 0$, further cancellations occur, obliterating the terms of order n in $n(1 + \sigma_S^2(n)/\mu_S(n))V_n$, leaving behind terms of the order $n\sigma_S^2(n)/\mu_S(n)$. Calculations (not shown) similar to those highlighted in the case of $\beta > 0$ give

$$\frac{1}{\sigma_S(n)\sqrt{n/\mu_S(n)}} \left(M^{-j_n} X_{j_n}^{(n)} - \begin{pmatrix} n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \xrightarrow{\mathrm{D}} \mathcal{N}_{k+1} \left(0, \lambda \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \right).$$

Again, as we assumed that $\lambda_n \to \lambda > 0$, we have $M^{-j_n} \to e^{\lambda} e^{-\lambda B}$, and an application of the multivariate Slutsky theorem yields

$$\frac{X_{j_n}^{(n)} - ne^{-\lambda} v_{k+1}}{\sigma_S(n)\sqrt{n/\mu_S(n)}} \xrightarrow{\mathrm{D}} \mathcal{N}_{k+1}(0, J'_{k+1}),$$

provided that λ is not a positive integer that is at most k. When λ is an integer that is at most k, the covariance matrix J'_{k+1} is degenerate (if $\lambda = r$, the rth row and column are all 0s) and only the term from J_{k+1} is present in the variance.

10. Illustrating examples

The covariance formula in Proposition 1 is not easy to reduce, however, we can manage to obtain compact forms for small ℓ and m. For example, extracting the $(0, 1)$ entry from this form, we obtain

$$\mathrm{E}[X_{j,0}^{(n)} X_{j,1}^{(n)}] = h_n \sum_{i=0}^{j-1} \sum_{r=0}^{i} g_n^{i-r} \binom{i}{r} \frac{(n - \mu_S(n))^{j-i+2r-2}}{n^{j-i+2r-2}} ((i+1)\mu_S(n) - (i+1-r)n)$$

$$+ n^2 \sum_{r=0}^{j} g_n^{j-r} \binom{j}{r} \frac{(n - \mu_S(n))^{2r-1}(j\mu_S(n) - (j-r)n)}{n^{2r}}.$$

Reducing these sums we obtain

$$\mathrm{E}[X_{j,0}^{(n)}, X_{j,1}^{(n)}] = j(\mu_S(n)n - \sigma_S^2(n) - \mu_S^2(n))\left(\frac{(n-\mu(n))(n-\mu_S^2(n)-1)+\sigma_S^2(n)}{n(n-1)}\right)^{j-1}.$$

Subtracting $\mathrm{E}[X_{j,0}^{(n)}]\,\mathrm{E}[X_{j,1}^{(n)}]$, we obtain the covariance

$$\mathrm{cov}[X_{j,0}^{(n)}X_{j,1}^{(n)}] = j(\mu_S(n)n - \sigma_S^2(n) - \mu_S^2(n))\left(\frac{(n-\mu(n))(n-\mu_S^2(n)-1)+\sigma_S^2(n)}{n(n-1)}\right)^{j-1}$$
$$- jn\mu_S(n)\left(1 - \frac{\mu_S(n)}{n}\right)^{2j-1}.$$

Likewise,

$$\mathrm{var}[X_{j,0}^{(n)}] = n\left[(n-1)\left(\frac{(n-n\mu_S(n))(n-\mu_S(n)-1)+\sigma_S^2(n)}{n(n-1)}\right)^{j-1} + \left(\frac{n-\mu_S(n)}{n}\right)^{j}\right]$$
$$- n^2\left(\frac{n-\mu_S(n)}{n}\right)^{2j}$$

and

$$\mathrm{var}[X_{j,1}^{(n)}] = j\mu_S(n)\left(\frac{n-\mu_S(n)}{n}\right)^{j-1} - j^2\mu_S^2(n)\left(\frac{n-\mu_S(n)}{n}\right)^{2j-2}$$
$$+ j[j(\sigma_S^2(n)+\mu_S^2(n)-\mu_S(n)n)^2 + (n-1)(\sigma_S^2(n)n - \mu_S(n)(n-\mu_S(n)))]$$
$$\times ((n-\mu_S(n))(n-\mu_S(n)-1)+\sigma_S^2(n))^{j-2}\left(\frac{1}{n(n-1)}\right)^{j-1}.$$

10.1. Standard coupon collection

Consider the standard Dixie Cup problem, where $S \equiv 1$. We illustrate the covariance computation at the beginning of this section on coupons of types 0 and 1. Here we have the covariances

$$\mathrm{cov}[X_{j,0}^{(n)}, X_{j,1}^{(n)}] = j(n-1)\left(\frac{n-2}{n}\right)^{j-1} - jn\left(\frac{n-1}{n}\right)^{2j-1} \sim -\mathrm{e}^{-2\lambda}\lambda^2 n,$$

$$\mathrm{var}[X_{j,0}^{(n)}] = n\left[(n-1)\left(\frac{n-2}{n}\right)^{j} + \left(\frac{n-1}{n}\right)^{j}\right] - n^2\left(\frac{n-1}{n}\right)^{2j}$$
$$\sim \mathrm{e}^{-2\lambda}(\lambda\mathrm{e}^{\lambda} - \lambda - 1)n,$$

$$\mathrm{var}[X_{j,1}^{(n)}] = j\left(\frac{n-1}{n}\right)^{j} - j^2\left(\frac{n-1}{n}\right)^{2j-2} + \frac{j(j-1)(n-1)}{n}\left(\frac{n-2}{n}\right)^{j-2}$$
$$\sim \mathrm{e}^{-2\lambda}(\lambda\mathrm{e}^{\lambda} - \lambda + \lambda^2 - \lambda^3)n.$$

Thus, in the linear phase (when $\mu_S(n)j_n = j_n \sim \lambda_n n$ for λ_n convergent to a fixed $\lambda > 0$), the asymptotic covariance matrix of types 0 and 1 is

$$\mathrm{cov}[X_j^{(n)}] = \boldsymbol{J}_2 n + \boldsymbol{o}(n) = \mathrm{e}^{-2\lambda}\begin{pmatrix} \mathrm{e}^{\lambda} - 1 - \lambda & -\lambda^2 \\ -\lambda^2 & \lambda\mathrm{e}^{\lambda} - \lambda + \lambda^2 - \lambda^3 \end{pmatrix} n + \boldsymbol{o}(n),$$

as in Kolchin *et al.* (1978, p. 38).

In this example

$$\sqrt{\frac{\mu_S^2(n) j_n^2}{2n} + \sigma_S^2(n) j_n} = \sqrt{\frac{j_n^2}{2n}},$$

and so $s_n = 1 = o(\sqrt{j_n^2/n})$ is only satisfied in the upper sublinear phase $\sqrt{n} = o(j_n)$. Here $\alpha_n \equiv 1$, and Theorem 1(a) gives

$$\frac{1}{j_n/\sqrt{n}} \left(X_{j_n}^{(n)} - \left(1 - \frac{1}{n}\right)^{j_n} \begin{pmatrix} n \\ j_n \\ \frac{j_n^2}{2n} \end{pmatrix} \right) \xrightarrow{\text{D}} \mathcal{N}_3 \left(\mathbf{0}, \frac{1}{2} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & 1 \\ 1 & -2 & 1 \end{pmatrix} \right),$$

starting at the phase $j_n/\sqrt{n} \to \infty$ and going all the way to the end of the sublinear phase at $j_n = o(n)$. (The limiting covariance matrix is singular, but each of the types 0, 1, and 2 satisfies a univariate central limit theorem.)

In the linear phase, when $j_n \sim \lambda_n n$ (for λ_n convergent to a fixed $\lambda > 0$), we have $\beta_n \equiv 1$ in Theorem 2(a), giving

$$\frac{X_{j_n}^{(n)} - n \mathrm{e}^{-\lambda} \mathbf{v}_{k+1}}{\sqrt{n}} \xrightarrow{\text{D}} \mathcal{N}_{k+1}(\mathbf{0}, \mathbf{J}_{k+1}).$$

The results are not very different if a fixed number, $S \equiv s$, is acquired in each purchase. Essentially, all the results above stay the same, with the rth component of the shift factor (asymptotic mean of the sampling distribution) scaled by s^r for $r = 0, 1, 2$, and the limiting covariance matrix multiplied by s^2.

10.2. An example with a sampling distribution of fixed range

Suppose that the sampling distribution has distribution $1 + \mathrm{Bin}(s, \frac{1}{2})$, with fixed s. Here, $\mu_S(n) = 1 + \frac{1}{2}s$ and $\sigma_S^2(n) = \frac{1}{4}s$. In this example,

$$s_n = s + 1 = o\left(\sqrt{\frac{\mu_S^2(n) j_n^2}{2n} + \sigma_S^2(n) j_n} \right).$$

So, $s_n = o(j_n)$ throughout the entire growing sublinear phase. Subsequently, $\alpha_n \to 0$ in Theorem 1(b), and we obtain the central limit result

$$\frac{1}{\sqrt{j_n}} \left(X_{j_n}^{(n)} - \left(1 - \frac{s+2}{2n}\right)^{j_n} \begin{pmatrix} n \\ \frac{(s+2)j_n}{2} \end{pmatrix} \right) \xrightarrow{\text{D}} \mathcal{N}_2 \left(\mathbf{0}, \frac{s^2}{16} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right),$$

applying in the entire growing sublinear phase, where $j_n \to \infty$ and $j_n = o(n)$. (Again, the matrix is singular.)

In the linear phase, when $j_n \sim 2\lambda_n n/(s + 2)$ for λ_n convergent to a fixed $\lambda > 0$, we obtain $\beta = (2s + 4)/(3s + 4) > 0$ and Theorem 2(a) gives

$$\frac{X_{j_n}^{(n)} - n \mathrm{e}^{-\lambda} \mathbf{v}_{k+1}}{\sqrt{n}} \xrightarrow{\text{D}} \mathcal{N}_{k+1} \left(\mathbf{0}, \mathbf{J}_{k+1} + \frac{s}{2s + 4} \mathbf{J}'_{k+1} \right).$$

10.3. An example with a nearly degenerate sampling distribution on a two-point set

Suppose that the sampling distribution is the two-point distribution

$$P(S_j = k) = \begin{cases} \dfrac{5}{\ln n} & \text{if } k = \lceil \ln n \rceil - 1, \\[2mm] 1 - \dfrac{5}{\ln n} & \text{if } k = \lceil \ln n \rceil, \end{cases}$$

for $n \geq 149$. Here, we have

$$\mu_S(n) = \frac{5}{\ln n}(\lceil \ln n \rceil - 1) + \left(1 - \frac{5}{\ln n}\right)\lceil \ln n \rceil = \lceil \ln n \rceil - \frac{5}{\ln n} \sim \ln n,$$

$$\sigma_S^2(n) = \frac{5}{\ln n}(\lceil \ln n \rceil - 1)^2 + \left(1 - \frac{5}{\ln n}\right)\lceil \ln n \rceil^2 - \mu_S^2(n) = \frac{5}{\ln n} - \frac{25}{\ln^2 n} \to 0.$$

The sampling distribution is concentrated at the point $\lceil \ln n \rceil$; in fact, $S(n) - \lceil \ln n \rceil \xrightarrow{P} 0$. In this example,

$$s_n = \lceil \ln n \rceil = o\left(\sqrt{\frac{j_n^2 \ln^2 n}{2n} + \frac{5 j_n}{\ln n}}\right)$$

for any j_n in the phase $\ln n = o(j_n)$. The term $j_n^2 \ln^2 n / 2n$ does not begin to dominate $5 j_n / \ln n$ until j_n is at least of the order $n / \ln^3 n$, which is very close to the linear phase (which begins at the order $n / \ln n$). So, we have three sublinear phases. In the early sublinear phase, beginning at $\ln n = o(j_n)$ and going up to $o(n / \ln^3 n)$, the term $5 j_n / \ln n$ is dominant. In this phase $\alpha_n \to 0$, and by Theorem 1(b) we obtain (for types 0 and 1)

$$\frac{1}{\sqrt{j_n / \ln n}} \left(X_{j_n}^{(n)} - \left(1 - \frac{\lceil \ln n \rceil}{n}\right)^{j_n}\left(\frac{n}{j_n \ln n}\right)\right) \xrightarrow{D} \mathcal{N}_2\left(\mathbf{0}, 5\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right).$$

Then comes the middle sublinear phase, where $j_n^2 \ln^2 n / 2n \sim q 5 j_n / \ln n$ for some $q > 0$. Theorem 1(a) applies with $\alpha = q/(q+1)$. In the upper sublinear phase, $n / \ln^3 n = o(j_n)$, but j_n remains sublinear, i.e. $j_n = o(n / \ln n)$. Theorem 1(a) applies with $\alpha = 1$. Hence, only in the middle sublinear phase do we obtain a nondegenerate trivariate central limit theorem for this example.

In the linear phase, when $j_n \sim \lambda_n n / \ln n$ (for λ_n convergent to a fixed $\lambda > 0$) and $\beta_n \to 1$, for types $0, 1, \ldots, k$, we obtain

$$\frac{X_{j_n}^{(n)} - n e^{-\lambda} \mathbf{v}_{k+1}}{\sqrt{n}} \xrightarrow{D} \mathcal{N}_{k+1}(\mathbf{0}, \mathbf{J}_{k+1}).$$

10.4. An example with a uniform sampling distribution on a growing range

Consider coupon collection where at each purchase a uniformly distributed number on the set $\{1, 2, \ldots, 12\lceil n^{1/8}\rceil\}$ is acquired. In this example,

$$\mu_S(n) \sim 6n^{1/8} \quad \text{and} \quad \sigma_S^2(n) = 12n^{1/4}.$$

With these rates of growth we have the condition

$$s_n = 12\lceil n^{1/8}\rceil = o\left(\sqrt{\frac{\mu_S^2(n) j_n^2}{2n} + \sigma_S^2(n) j_n}\right),$$

which means that $s_n = o(\sqrt{18 j_n^2/n^{3/4} + 12 j_n n^{1/4}})$ is automatically satisfied in the growing sublinear phase. In the growing sublinear phase, where $j_n = o(n^{7/8})$, $\sigma_S^2(n) j_n$ dominates $\mu_S^2(n) j_n^2/(2n)$ in the upper sublinear phase $n^{1/4} = o(j_n)$. So, Theorem 1(b) applies, giving

$$\frac{1}{n^{1/8}\sqrt{j_n}}\left(X_{j_n}^{(n)} - \left(1 - \frac{12\lceil n^{1/8}\rceil + 1}{2n}\right)^{j_n}\binom{n}{6 j_n n^{1/8}}\right) \xrightarrow{\text{D}} \mathcal{N}_2\left(\mathbf{0}, 12\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right),$$

holding throughout the entire sublinear phase (ending at $o(j_n^{7/8})$).

For Theorem 2, we compute $\mu_S(n)/(\mu_S(n) + \sigma_S^2(n)) \to 0$. In the linear phase, when $j_n \sim \frac{1}{6}\lambda_n n^{7/8}$ (for λ_n convergent to a fixed $\lambda > 0$), by Theorem 2(b), $X_{j_n}^{(n)}$ (properly normalized) has limiting covariance matrix J_{k+1}', provided that λ is not a positive integer, and, for $r = 0, 1, \ldots k$, each type satisfies a univariate central limit theorem.

References

De Moivre, A. (1718). *The Doctrine of Chances*. W. Pearfon, London (reprinted: Taylor & Francis, London, 1967).

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

Kobza, J. E., Jacobson, S. H. and Vaughan, D. E. (2007). A survey of the coupon collector's problem with random sample sizes. *Methodology Comput. Appl. Prob.* **9,** 573–584.

Kolchin, V. F., Sevast'yanov, B. A. and Chistyakov, V. P. (1978). *Random Allocations*. John Wiley, New York.

Laplace, P. S. (1774). Mémoire sur les suites récurro-récurrents et leurs usages dans la théorie des hasards. *Mém. Acad. R. Sci. Paris* **6,** 353–371.

Mahmoud, H. M. (2010). Gaussian phases in generalized coupon collection. *Adv. Appl. Prob.* **42,** 994–1012.

Smythe, R. T. (2011). Generalized coupon collection: the superlinear case. *J. Appl. Prob.* **48,** 189–199.

Stadje, W. (1990). The collector's problem with group drawings. *Adv. Appl. Prob.* **22,** 866–882.

Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, Vol. 1, 5th edn. Oxford University Press, New York.