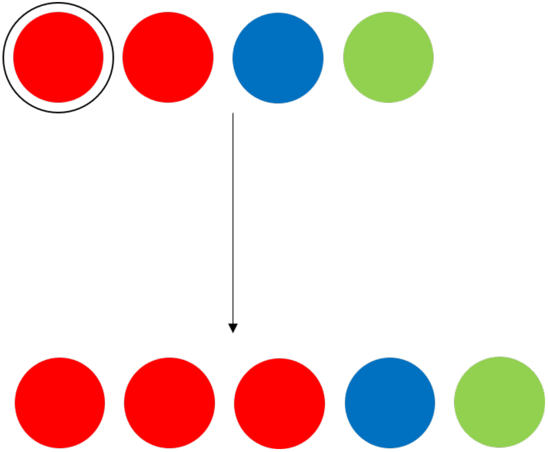
Pólya urns with growing initial compositions

Christopher Dean

UK Easter Probability Meeting 2023

March 29, 2023

Introduction to Pólya urns

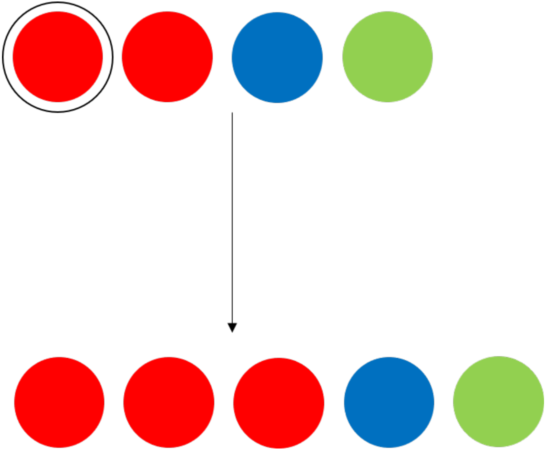


$$\boldsymbol{U}(0) = (2, 1, 1)$$

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{U}(1) = (3, 1, 1)$$

Introduction to Pólya urns



$\boldsymbol{U}(0) = (2, 1, 1)$

$$R = \begin{bmatrix} \color{red}{1} & \color{blue}{0} & \color{green}{0} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\boldsymbol{U}(1) = (3, 1, 1)$

Introduction to Pólya urns

A d-colour generalized Pólya urn process $(\boldsymbol{U}(n))_{n\geq 0}$ is a Markov process which depends on two parameters:

- The initial composition $\boldsymbol{U}(0) \in \mathbb{Z}^d_{\geq 0}$.
- The non-negative integer-valued replacement matrix R.

The process evolves from step n to $n+1$ as follows:

1. Select a ball u.a.r. from the urn.
2. If a ball of colour i is chosen, place the ball back into the urn along with R_{ij} balls of colour j.

Asymptotic behaviour: Two canonical cases

A typical question when studying Pólya urns is how the urn behaves as the number of draws n tends to infinity.

- What does the colour composition $\boldsymbol{U}(n)/\sum_{i=1}^{d} U(n)_i$ converge to?
- What are the size, scale, and shape of the urns fluctuations around its asymptotic colour composition?

Asymptotic behaviour: Two canonical cases

A typical question when studying Pólya urns is how the urn behaves as the number of draws n tends to infinity.

- What does the colour composition $\boldsymbol{U}(n)/\sum_{i=1}^{d} U(n)_i$ converge to?
- What are the size, scale, and shape of the urns fluctuations around its asymptotic colour composition?

This question can be answered for most replacement matrices by solving the following two canonical cases:

Identity replacement matrix: Take $R = SI$, $S \in \mathbb{Z}_{\geq 1}$.

Irreducible replacement matrix: For all $1 \leq i, j \leq d$, if $\boldsymbol{U}(0) = \boldsymbol{e}_i$ there exists $n(i,j) \geq 0$, such that $U(n(i,j))_j > 0$ has positive probability, e.g.

$$R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Growing initial composition urns

Let $(\boldsymbol{U}_n)_{n\geq 1}$ be a sequence of urns with identical replacement matrix, and set $N(n) := \sum_{i=1}^{d} U_n(0)_i$ to be the number of initial balls in the urn.

Growing initial composition urns

Let $(\boldsymbol{U}_n)_{n \geq 1}$ be a sequence of urns with identical replacement matrix, and set $N(n) := \sum_{i=1}^{d} U_n(0)_i$ to be the number of initial balls in the urn.

Recently, (Dimitris & Dimitra, 2019, Borovkov, 2019) categorised the asymptotic behaviour of $\boldsymbol{U}_n(n)$ when the replacement matrix was the identity as $n, N \to \infty$.

Growing initial composition urns

Let $(\boldsymbol{U}_n)_{n\geq 1}$ be a sequence of urns with identical replacement matrix, and set $N(n) := \sum_{i=1}^{d} U_n(0)_i$ to be the number of initial balls in the urn.

Recently, (Dimitris & Dimitra, 2019, Borovkov, 2019) categorised the asymptotic behaviour of $\boldsymbol{U}_n(n)$ when the replacement matrix was the identity as $n, N \to \infty$.

Three regimes appear:

- **Initial Ball Dominant** - $n = o(N)$.
- **Transitional Regime** - $n/N \to 1$ (w.l.o.g.).
- **Time Step Dominant** - $N = o(n)$.

Asymptotic behaviour: Irreducible replacement matrix

(Athreya & Karlin, 1968) Let $\boldsymbol{U}$ have an irreducible replacement matrix R. Let λ_1 denote the Perron-Frobenius eigenvalue of R. Then, a.s.

$$\lim_{n\to\infty} \frac{\boldsymbol{U}(n)}{\sum_{i=1}^{d} U(n)_i} = \boldsymbol{v_1},$$

where $\boldsymbol{v}_1$ is the left eigenvector of λ_1.

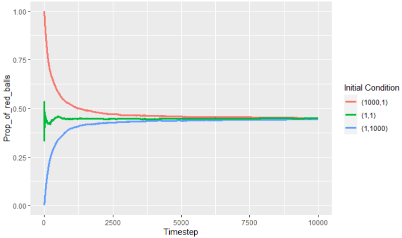


Figure: Simulations of a two colour urn with $R = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$ and $\boldsymbol{v_1} \approx (0.45, 0.55)$.

Asymptotic behaviour: Irreducible replacement matrix

(Janson, 2004) Assume R is diagonalizable with real eigenvalues. Let λ_1 be the Perron-Frobenius eigenvalue and λ_2 the second largest eigenvalue.

Fluctuations of the irreducible urn around $\boldsymbol{v}_1$ as $n \to \infty$:

- Small urns: If $\lambda_2 < \lambda_1/2$, then the fluctuations converge to an Ornstein-Uhlenbeck process of size $n^{1/2}$ and scale nt.
- Critical urns: If $\lambda_2 = \lambda_1/2$, then the fluctuations converge to a Brownian motion of size $n^{1/2} \log(n)^{1/2}$ and scale n^t.
- Large urns: If $\lambda_2 > \lambda_1/2$, then the fluctuations converge to a non-Gaussian random variable of size n^{λ_2/λ_1} and scale n^t that depends on $\boldsymbol{U}(0)$.

Growing initial composition urns

Let $(\boldsymbol{U}_n)_{n\geq 1}$ be a sequence of urns with identical irreducible replacement matrix, and set $N(n) := \sum_{i=1}^{d} U_n(0)_i$ to be the number of initial balls in the urn.

Growing initial composition urns

Let $(\boldsymbol{U}_n)_{n\geq 1}$ be a sequence of urns with identical irreducible replacement matrix, and set $N(n) := \sum_{i=1}^{d} U_n(0)_i$ to be the number of initial balls in the urn.

We assume that:

- **Fixed initial colour composition** - There is a vector $\boldsymbol{\mu}$ such that

$$\boldsymbol{\mu}_n := N^{-1}\boldsymbol{U}_n(0) = \boldsymbol{\mu}, \;\; n \geq 1.$$

- **Balanced replacement matrix** - We add $S \geq 1$ balls to the urn at each time step a.s.

Growing initial composition urns

Let $(\boldsymbol{U}_n)_{n\geq 1}$ be a sequence of urns with identical irreducible replacement matrix, and set $N(n) := \sum_{i=1}^{d} U_n(0)_i$ to be the number of initial balls in the urn.

We assume that:

- **Fixed initial colour composition** - There is a vector $\boldsymbol{\mu}$ such that

$$\boldsymbol{\mu}_n := N^{-1}\boldsymbol{U}_n(0) = \boldsymbol{\mu}, \ \ n \geq 1.$$

- **Balanced replacement matrix** - We add $S \geq 1$ balls to the urn at each time step a.s.

Three regimes appear:

- **Initial Ball Dominant** - $n = o(N)$.
- **Transitional Regime** - $n/N \to 1$ (w.l.o.g.).
- **Time Step Dominant** - $N = o(n)$.

Sub-urn representation

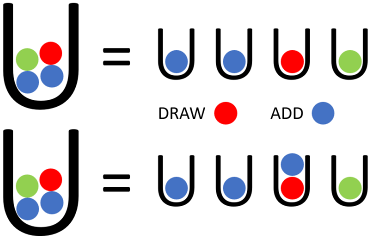
Let $\boldsymbol{U}$ be an urn with N initial balls. We have

$$\boldsymbol{U}(n) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{d} \sum_{j=1}^{U(0)_i} \boldsymbol{U}_{ij}(D_{ij}(n)), \quad n \geq 0.$$

- $\boldsymbol{U}_{ij}$ - Pólya urns with initial conditions $\boldsymbol{e}_i$ and identical replacement matrix to $\boldsymbol{U}$.
- $D_{ij}(n)$ - The number of times the urn $\boldsymbol{U}_{ij}$ has been drawn from by time step n of $\boldsymbol{U}$.

$$\boldsymbol{U}(0) = ({\color{blue}2}, {\color{red}1}, {\color{green}1})$$

$$R = \begin{pmatrix} {\color{blue}0} & {\color{blue}0} & {\color{blue}1} \\ {\color{red}1} & {\color{red}0} & {\color{red}0} \\ {\color{green}0} & {\color{green}1} & {\color{green}0} \end{pmatrix}$$



Sub-urn representation

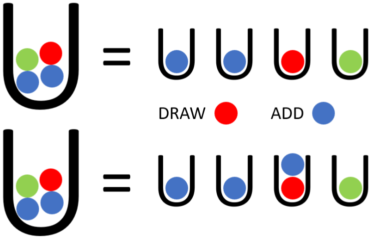
Let $\boldsymbol{U}$ be an urn with N initial balls. We have

$$\boldsymbol{U}(n) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{d} \sum_{j=1}^{U(0)_i} \boldsymbol{U}_{ij}(D_{ij}(n)), \quad n \geq 0.$$

- $\boldsymbol{U}_{ij}$ - Pólya urns with initial conditions $\boldsymbol{e}_i$ and identical replacement matrix to $\boldsymbol{U}$.
- $D_{ij}(n)$ - The number of times the urn $\boldsymbol{U}_{ij}$ has been drawn from by time step n of $\boldsymbol{U}$.

$$\boldsymbol{U}(0) = ({\color{blue}2}, {\color{red}1}, {\color{green}1})$$

$$R = \begin{pmatrix} {\color{blue}0} & {\color{blue}0} & {\color{blue}1} \\ {\color{red}1} & {\color{red}0} & {\color{red}0} \\ {\color{green}0} & {\color{green}1} & {\color{green}0} \end{pmatrix}$$



If the urn is balanced, conditionally on the $D_{ij}(n)$, the $\boldsymbol{U}_{ij}(D_{ij}(n))$ are independent.

Sub-urn representation

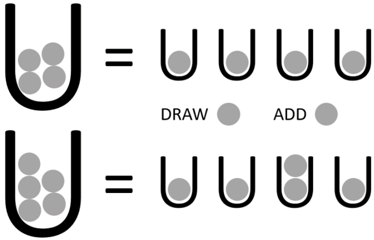
Let $\boldsymbol{U}$ be an urn with N initial balls. We have

$$\boldsymbol{U}(n) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{d} \sum_{j=1}^{U(0)_i} \boldsymbol{U}_{ij}(D_{ij}(n)), \quad n \geq 0.$$

- $\boldsymbol{U}_{ij}$ - Pólya urns with initial conditions $\boldsymbol{e}_i$ and identical replacement matrix to $\boldsymbol{U}$.
- $D_{ij}(n)$ - The number of times the urn $\boldsymbol{U}_{ij}$ has been drawn from by time step n of $\boldsymbol{U}$.

$$\boldsymbol{U}(0) = (2, 1, 1)$$

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$



If the urn is balanced, conditionally on the $D_{ij}(n)$, the $\boldsymbol{U}_{ij}(D_{ij}(n))$ are independent.

Sub-urn representation

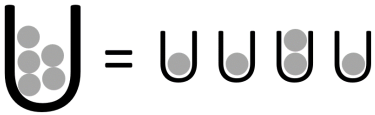
Let $\boldsymbol{U}$ be an urn with N initial balls. We have

$$\boldsymbol{U}(n) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{d} \sum_{j=1}^{U(0)_i} \boldsymbol{U}_{ij}(D_{ij}(n)), \quad n \geq 0.$$

- $\boldsymbol{U}_{ij}$ - Pólya urns with initial conditions $\boldsymbol{e}_i$ and identical replacement matrix to $\boldsymbol{U}$.
- $D_{ij}(n)$ - The number of times the urn $\boldsymbol{U}_{ij}$ has been drawn from by time step n of $\boldsymbol{U}$.

$$\boldsymbol{U}(0) = (2, 1, 1)$$

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$



If the urn is balanced, conditionally on the $D_{ij}(n)$, the $\boldsymbol{U}_{ij}(D_{ij}(n))$ are independent.

Sub-urn representation

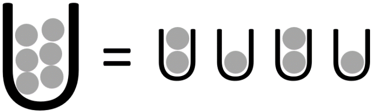
Let $\boldsymbol{U}$ be an urn with N initial balls. We have

$$\boldsymbol{U}(n) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{d} \sum_{j=1}^{U(0)_i} \boldsymbol{U}_{ij}(D_{ij}(n)), \quad n \geq 0.$$

- $\boldsymbol{U}_{ij}$ - Pólya urns with initial conditions $\boldsymbol{e}_i$ and identical replacement matrix to $\boldsymbol{U}$.
- $D_{ij}(n)$ - The number of times the urn $\boldsymbol{U}_{ij}$ has been drawn from by time step n of $\boldsymbol{U}$.

$\boldsymbol{U}(0) = (2, 1, 1)$

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$



If the urn is balanced, conditionally on the $D_{ij}(n)$, the $\boldsymbol{U}_{ij}(D_{ij}(n))$ are independent.

Sub-urn representation

Let $\boldsymbol{U}$ be an urn with N initial balls. We have

$$\boldsymbol{U}(n) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{d} \sum_{j=1}^{U(0)_i} \boldsymbol{U}_{ij}(D_{ij}(n)), \quad n \geq 0.$$

- $\boldsymbol{U}_{ij}$ - Pólya urns with initial conditions $\boldsymbol{e}_i$ and identical replacement matrix to $\boldsymbol{U}$.
- $D_{ij}(n)$ - The number of times the urn $\boldsymbol{U}_{ij}$ has been drawn from by time step n of $\boldsymbol{U}$.

$$\boldsymbol{U}(0) = (2, 1, 1)$$

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

If the urn is balanced, conditionally on the $D_{ij}(n)$, the $\boldsymbol{U}_{ij}(D_{ij}(n))$ are independent.

Sub-urn representation

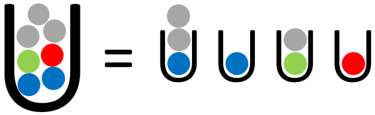
Let $\boldsymbol{U}$ be an urn with N initial balls. We have

$$\boldsymbol{U}(n) \stackrel{\mathrm{d}}{=} \sum_{i=1}^{d} \sum_{j=1}^{U(0)_i} \boldsymbol{U}_{ij}(D_{ij}(n)), \quad n \geq 0.$$

- $\boldsymbol{U}_{ij}$ - Pólya urns with initial conditions $\boldsymbol{e}_i$ and identical replacement matrix to $\boldsymbol{U}$.
- $D_{ij}(n)$ - The number of times the urn $\boldsymbol{U}_{ij}$ has been drawn from by time step n of $\boldsymbol{U}$.

$$\boldsymbol{U}(0) = (2, 1, 1)$$

$$R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$



If the urn is balanced, conditionally on the $D_{ij}(n)$, the $\boldsymbol{U}_{ij}(D_{ij}(n))$ are independent.

Initial ball dominant regime $n = o(N)$

Let the sequence $(\boldsymbol{U}_n)_{n\geq 1}$ have a balanced irreducible replacement matrix and initial colour composition $\boldsymbol{\mu}$. Let $\ell_1(n,t) = \log(1 + Snt/N)$.

Initial Ball Dominant: As $n \to \infty$

$$n^{-1/2}(\boldsymbol{U}_n(\lfloor nt \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_1(n,t)}\boldsymbol{\mu}) \xrightarrow{\mathrm{d}} \boldsymbol{W}_1(t) \text{ in } D[0,\infty),$$

$$N^{-1}\boldsymbol{U}_n(\lfloor nt \rfloor) \xrightarrow{\mathrm{p}} \boldsymbol{\mu} \text{ in } D[0,\infty).$$

- **Asymptotic colour composition $\boldsymbol{\mu}$** - Initial colour composition dominates the asymptotic colour composition.
- **Brownian fluctuations $\boldsymbol{W}_1$** - The draws of the urn are close to a random walk with jump probabilities given by $\boldsymbol{\mu}$ and jumps given by R.

Transitional regime $n \sim N$

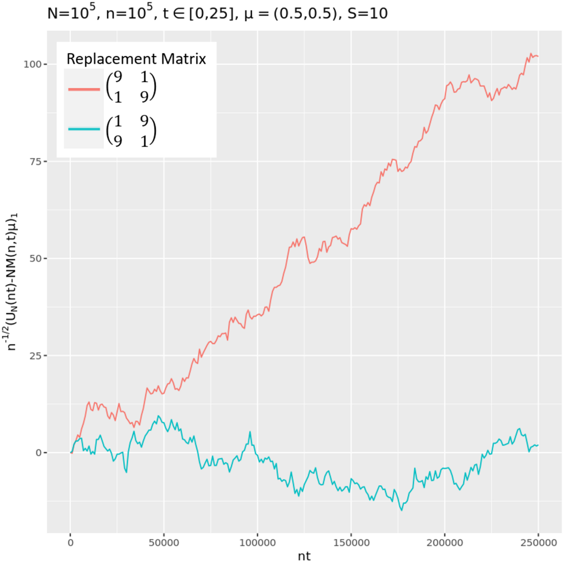
Let the sequence $(\boldsymbol{U}_n)_{n \geq 1}$ have a balanced irreducible replacement matrix and initial colour composition $\boldsymbol{\mu}$. Let $\ell_1(n,t) = \log(1 + Snt/N)$.

Transitional regime: As $n \to \infty$

$$n^{-1/2}(\boldsymbol{U}_n(\lfloor nt \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_1(n,t)}\boldsymbol{\mu}) \xrightarrow{\mathrm{d}} \boldsymbol{W_2}(t) \text{ in } D[0,\infty),$$
$$N^{-1}\boldsymbol{U}_n(\lfloor nt \rfloor) \xrightarrow{\mathrm{p}} \mathrm{e}^{R'S^{-1}\log(1+St)}\boldsymbol{\mu} \text{ in } D[0,\infty).$$

- **Asymptotic colour composition** $\mathrm{e}^{R'S^{-1}\log(1+St)}\boldsymbol{\mu}$ **(normalized)** - Can be seen as the "expected" composition of an urn with initial colour composition $\boldsymbol{\mu}$ number of initial balls 1 and number of draws t.
- **Gaussian fluctuations $\boldsymbol{W_2}$** - The shape of the fluctuations depend on the fluctuations (variance) of the urn with a finite number of draws.

Transitional regime $n \sim N$



Time step dominant regime $N = o(n)$

Time Step Dominant: Assume R is diagonalizable with real eigenvalues. S is the Perron-Frobenius eigenvalue, and let λ_2 be the second largest eigenvalue. Then, as $n \to \infty$

$$\frac{\boldsymbol{U}_n(\lfloor N(n/N)^t \rfloor)}{SN(n/N)^t} \xrightarrow{\mathrm{p}} \boldsymbol{v}_1 \text{ in } D(0, \infty).$$

Time step dominant regime $N = o(n)$

Time Step Dominant: Assume R is diagonalizable with real eigenvalues. S is the Perron-Frobenius eigenvalue, and let λ_2 be the second largest eigenvalue. Then, as $n \to \infty$

$$\frac{\boldsymbol{U}_n(\lfloor N(n/N)^t \rfloor)}{SN(n/N)^t} \xrightarrow{\mathrm{p}} \boldsymbol{v}_1 \text{ in } D(0,\infty).$$

Small Urns: Let $\ell_1(n,t) = \log(1 + Snt/N)$. If $\lambda_2 < S/2$, then as $n \to \infty$

$$n^{-1/2}(\boldsymbol{U}_n(\lfloor nt \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_1(n,t)}\boldsymbol{\mu}) \xrightarrow{\mathrm{d}} t^{1/2}\boldsymbol{W}_s(\log(t)) \text{ in } D(0,\infty).$$

Time step dominant regime $N = o(n)$

Time Step Dominant: Assume R is diagonalizable with real eigenvalues. S is the Perron-Frobenius eigenvalue, and let λ_2 be the second largest eigenvalue. Then, as $n \to \infty$

$$\frac{\boldsymbol{U}_n(\lfloor N(n/N)^t \rfloor)}{SN(n/N)^t} \xrightarrow{\mathrm{p}} \boldsymbol{v}_1 \text{ in } D(0, \infty).$$

Small Urns: Let $\ell_1(n, t) = \log(1 + Snt/N)$. If $\lambda_2 < S/2$, then as $n \to \infty$

$$n^{-1/2}(\boldsymbol{U}_n(\lfloor nt \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_1(n,t)}\boldsymbol{\mu}) \xrightarrow{\mathrm{d}} t^{1/2}\boldsymbol{W}_s(\log(t)) \text{ in } D(0, \infty).$$

Critical Urns: Let $\ell_2(n, t) = \log(1 + S(n/N)^t)$. If $\lambda_2 = S/2$, then as $n \to \infty$

$$\frac{\boldsymbol{U}_n(\lfloor N(n/N)^t \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_2(n,t)}\boldsymbol{\mu}}{N^{1/2}(n/N)^{t/2}\log(n/N)^{1/2}} \xrightarrow{\mathrm{d}} \boldsymbol{W}_c(t) \text{ in } D(0, \infty).$$

Time step dominant regime $N = o(n)$

Time Step Dominant: Assume R is diagonalizable with real eigenvalues. S is the Perron-Frobenius eigenvalue, and let λ_2 be the second largest eigenvalue. Then, as $n \to \infty$

$$\frac{\boldsymbol{U}_n(\lfloor N(n/N)^t \rfloor)}{SN(n/N)^t} \xrightarrow{\mathrm{p}} \boldsymbol{v}_1 \text{ in } D(0,\infty).$$

Small Urns: Let $\ell_1(n,t) = \log(1 + Snt/N)$. If $\lambda_2 < S/2$, then as $n \to \infty$

$$n^{-1/2}(\boldsymbol{U}_n(\lfloor nt \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_1(n,t)}\boldsymbol{\mu}) \xrightarrow{\mathrm{d}} t^{1/2}\boldsymbol{W}_s(\log(t)) \text{ in } D(0,\infty).$$

Critical Urns: Let $\ell_2(n,t) = \log(1 + S(n/N)^t)$. If $\lambda_2 = S/2$, then as $n \to \infty$

$$\frac{\boldsymbol{U}_n(\lfloor N(n/N)^t \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_2(n,t)}\boldsymbol{\mu}}{N^{1/2}(n/N)^{t/2}\log(n/N)^{1/2}} \xrightarrow{\mathrm{d}} \boldsymbol{W}_c(t) \text{ in } D(0,\infty).$$

Large Urns: If $\lambda_2 > S/2$, then as $n \to \infty$

$$\frac{\boldsymbol{U}_n(\lfloor N(n/N)^t \rfloor) - N\mathrm{e}^{R'S^{-1}\ell_2(n,t)}\boldsymbol{\mu}}{N^{1/2}(n/N)^{\lambda_2 t/S}} \xrightarrow{\mathrm{d}} \boldsymbol{V}_\ell \text{ in } D(0,\infty).$$

Thank you!