General Urn Schemes and Branching Markov Chains

Antar Bandyopadhyay
(Joint work with **Debleena Thacker**)

Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, New Delhi and Kolkata
http://www.isid.ac.in/∼antar

Bangalore Probability Seminar
Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, Bangalore
March 20, 2017

Pólya's Urn Scheme with k-Colors

- Suppose we have k colors, indexed by the set $[k] := \{1, 2, \ldots, k\}$.

Pólya's Urn Scheme with k-Colors

- Suppose we have k colors, indexed by the set $[k] := \{1, 2, \ldots, k\}$.
- We start with an initial configuration, say, $U_0 = (U_{0,1}, U_{0,2}, \ldots, U_{0,k})$,

Pólya's Urn Scheme with k-Colors

- Suppose we have k colors, indexed by the set $[k] := \{1, 2, \ldots, k\}$.
- We start with an initial configuration, say, $U_0 = (U_{0,1}, U_{0,2}, \ldots, U_{0,k})$, where $U_{0,j}$ denotes the *number of balls of color j in the urn* (at the beginning).

Pólya's Urn Scheme with k-Colors

- Suppose we have k colors, indexed by the set $[k] := \{1, 2, \ldots, k\}$.
- We start with an initial configuration, say, $U_0 = (U_{0,1}, U_{0,2}, \ldots, U_{0,k})$, where $U_{0,j}$ denotes the *number of balls of color j in the urn* (at the beginning).
- A ball is drawn *(uniformly) at random* from the urn and returned to the urn along with another ball of its color. This process is then repeated.

Pólya's Urn Scheme with k-Colors

- Suppose we have k colors, indexed by the set $[k] := \{1, 2, \ldots, k\}$.
- We start with an initial configuration, say, $U_0 = (U_{0,1}, U_{0,2}, \ldots, U_{0,k})$, where $U_{0,j}$ denotes the *number of balls of color j in the urn* (at the beginning).
- A ball is drawn *(uniformly) at random* from the urn and returned to the urn along with another ball of its color. This process is then repeated.
- Let $U_n := (U_{n,1}, U_{n,2}, \cdots, U_{n,k})$ denote the configuration of the urn after n draws,

Pólya's Urn Scheme with k-Colors

- Suppose we have k colors, indexed by the set $[k] := \{1, 2, \ldots, k\}$.
- We start with an initial configuration, say, $U_0 = (U_{0,1}, U_{0,2}, \ldots, U_{0,k})$, where $U_{0,j}$ denotes the *number of balls of color j in the urn* (at the beginning).
- A ball is drawn *(uniformly) at random* from the urn and returned to the urn along with another ball of its color. This process is then repeated.
- Let $U_n := (U_{n,1}, U_{n,2}, \cdots, U_{n,k})$ denote the configuration of the urn after n draws, where $U_{n,j}$ represent the *number of balls of color j in the urn* after n draws.

Pólya's Urn Scheme with k-Colors

- Suppose we have k colors, indexed by the set $[k] := \{1, 2, \ldots, k\}$.
- We start with an initial configuration, say, $U_0 = (U_{0,1}, U_{0,2}, \ldots, U_{0,k})$, where $U_{0,j}$ denotes the *number of balls of color j in the urn* (at the beginning).
- A ball is drawn *(uniformly) at random* from the urn and returned to the urn along with another ball of its color. This process is then repeated.
- Let $U_n := (U_{n,1}, U_{n,2}, \cdots, U_{n,k})$ denote the configuration of the urn after n draws, where $U_{n,j}$ represent the *number of balls of color j in the urn* after n draws.
- Let Z_n denote the random color of the $(n+1)$-th draw and χ_{n+1} be a (random) row vector with all entries 0 except the Z_n-th entry been 1, then

$$U_{n+1} = U_n + \chi_{n+1}.$$

Pólya's Urn Scheme with a Replacement Matrix

- We can consider more general replacement mechanism encoded as

$$R := \begin{array}{cccccc} & \text{Red} & \text{Green} & \text{Blue} & \cdots & \text{Yellow} \\ \text{Red} & \alpha & \beta & \gamma & \cdots & \eta \\ \text{Green} & a & b & c & \cdots & e \\ \text{Blue} & x & y & z & \cdots & t \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Yellow} & \phi & \chi & \psi & \cdots & \omega \end{array}$$

where $\alpha, \beta, \gamma, \ldots, \eta$; $a, b, c, \ldots, e$; $x, y, z, \ldots, t$; and $\phi, \chi, \psi, \ldots, \omega$; are non-negative integers.

Pólya's Urn Scheme with a Replacement Matrix

- We can consider more general replacement mechanism encoded as

$$R := \begin{array}{c|ccccc} & \text{Red} & \text{Green} & \text{Blue} & \cdots & \text{Yellow} \\ \text{Red} & \alpha & \beta & \gamma & \cdots & \eta \\ \text{Green} & a & b & c & \cdots & e \\ \text{Blue} & x & y & z & \cdots & t \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Yellow} & \phi & \chi & \psi & \cdots & \omega \end{array}$$

  where $\alpha, \beta, \gamma, \ldots, \eta$; $a, b, c, \ldots, e$; $x, y, z, \ldots, t$; and $\phi, \chi, \psi, \ldots, \omega$; are non-negative integers.

- Once color of a chosen ball is noted, say Z_n, then balls are added in the urn according to the Z_n-th row of the matrix R.

Pólya's Urn Scheme with a Replacement Matrix

- We can consider more general replacement mechanism encoded as

$$R := \begin{array}{c|ccccc} & \text{Red} & \text{Green} & \text{Blue} & \cdots & \text{Yellow} \\ \hline \text{Red} & \alpha & \beta & \gamma & \cdots & \eta \\ \text{Green} & a & b & c & \cdots & e \\ \text{Blue} & x & y & z & \cdots & t \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Yellow} & \phi & \chi & \psi & \cdots & \omega \end{array}$$

  where $\alpha, \beta, \gamma, \ldots, \eta$; $a, b, c, \ldots, e$; $x, y, z, \ldots, t$; and $\phi, \chi, \psi, \ldots, \omega$; are non-negative integers.

- Once color of a chosen ball is noted, say Z_n, then balls are added in the urn according to the Z_n-th row of the matrix R.

- With the same notations as earlier, we can then write

$$U_{n+1} = U_n + \chi_{n+1} R.$$

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*
- For such a model it is customary to divide by the total row sum and to use a *stochastic matrix* instead.

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*
- For such a model it is customary to divide by the total row sum and to use a *stochastic matrix* instead.
- In this formalization it is natural to consider U_0 as the initial proportion of colors instead of the number of colors.

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*
- For such a model it is customary to divide by the total row sum and to use a *stochastic matrix* instead.
- In this formalization it is natural to consider U_0 as the initial proportion of colors instead of the number of colors. Note that this results to some loss of generality!

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*
- For such a model it is customary to divide by the total row sum and to use a *stochastic matrix* instead.
- In this formalization it is natural to consider U_0 as the initial proportion of colors instead of the number of colors. Note that this results to some loss of generality! (Fortunately, recoverable in the classical setup.)

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*
- For such a model it is customary to divide by the total row sum and to use a *stochastic matrix* instead.
- In this formalization it is natural to consider U_0 as the initial proportion of colors instead of the number of colors. Note that this results to some loss of generality! (Fortunately, recoverable in the classical setup.)
- From now on for the rest of the talk we will only consider balanced schemes.

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*
- For such a model it is customary to divide by the total row sum and to use a *stochastic matrix* instead.
- In this formalization it is natural to consider U_0 as the initial proportion of colors instead of the number of colors. Note that this results to some loss of generality! (Fortunately, recoverable in the classical setup.)
- From now on for the rest of the talk we will only consider balanced schemes.
- Note that $U_{n,j}$'s now can be fractions and hence is not really the number of balls of color j.

Balanced Urn Scheme

- If we assume that the row sums are all equal then such a replacement matrix is called a *balanced replacement matrix*
- For such a model it is customary to divide by the total row sum and to use a *stochastic matrix* instead.
- In this formalization it is natural to consider U_0 as the initial proportion of colors instead of the number of colors. Note that this results to some loss of generality! (Fortunately, recoverable in the classical setup.)
- From now on for the rest of the talk we will only consider balanced schemes.
- Note that $U_{n,j}$'s now can be fractions and hence is not really the number of balls of color j.
- In fact, if we consider the (row) vector $\frac{U_n}{n+1}$ then it represents the distribution of the colors in the urn after the n draws.

Baclwell and MacQueen Urn

- In 1973 David Blackwell and James B. MacQueen introduced a new urn scheme to construct an earlier discovered *prior distribution* then called the *Ferguson distribution* (which now a days in Bayesian Statistics literature known as the *Dirichlet Process Prior*).

Baclwell and MacQueen Urn

- In 1973 David Blackwell and James B. MacQueen introduced a new urn scheme to construct an earlier discovered *prior distribution* then called the *Ferguson distribution* (which now a days in Bayesian Statistics literature known as the *Dirichlet Process Prior*).
- They consider the same process as that of Pólya's Urn, except have the colors index by a Polish space (possibly uncountable).

Baclwell and MacQueen Urn

- In 1973 David Blackwell and James B. MacQueen introduced a new urn scheme to construct an earlier discovered *prior distribution* then called the *Ferguson distribution* (which now a days in Bayesian Statistics literature known as the *Dirichlet Process Prior*).
- They consider the same process as that of Pólya's Urn, except have the colors index by a Polish space (possibly uncountable).
- The driving equation also remains same, except it takes the form

$$U_{n+1} = U_n + \delta_{Z_n},$$

where δ_z is the *Dirac Measure* at z.

Generalized Urn Schemes with Colors Indexed by a Polish Space

- Let $S \subseteq \mathbb{R}^d$ be a *Polish space*. We endow S with its Borel σ-algebra $\mathcal{S}$.

Generalized Urn Schemes with Colors Indexed by a Polish Space

- Let $S \subseteq \mathbb{R}^d$ be a *Polish space*. We endow S with its Borel σ-algebra $\mathcal{S}$.
- Let $\mathcal{M}(S)$ and $\mathcal{P}(S)$ denote respectively the set of all *finite measures* and the set of all *probability measures* on the measurable space $(S, \mathcal{S})$.

Generalized Urn Schemes with Colors Indexed by a Polish Space

- Let $S \subseteq \mathbb{R}^d$ be a *Polish space*. We endow S with its Borel σ-algebra $\mathcal{S}$.
- Let $\mathcal{M}(S)$ and $\mathcal{P}(S)$ denote respectively the set of all *finite measures* and the set of all *probability measures* on the measurable space $(S, \mathcal{S})$.
- Let $R : S \times \mathcal{S} \to [0, 1]$ be a Markov kernel on S.

Generalized Urn Schemes with Colors Indexed by a Polish Space

- Let $S \subseteq \mathbb{R}^d$ be a *Polish space*. We endow S with its Borel σ-algebra $\mathcal{S}$.
- Let $\mathcal{M}(S)$ and $\mathcal{P}(S)$ denote respectively the set of all *finite measures* and the set of all *probability measures* on the measurable space $(S, \mathcal{S})$.
- Let $R : S \times \mathcal{S} \to [0, 1]$ be a Markov kernel on S.
- By a *configuration* of the urn at time $n \geq 0$, we will consider a finite measure $U_n \in \mathcal{M}(S)$, such that, if Z_n represents the randomly chosen color at the $(n+1)$-th draw then the conditional distribution of Z_n given the "*past*", is given by

$$\mathbf{P}\left(Z_n \in ds \,\middle|\, U_n, U_{n-1}, \cdots, U_0\right) \propto U_n(ds).$$

Generalized Urn Schemes with Colors Indexed by a Polish Space

- Formally, starting with $U_0 \in \mathcal{P}(S)$ we define $(U_n)_{n \geq 0} \subseteq \mathcal{M}(S)$ recursively as follows

$$U_{n+1}(A) = U_n(A) + R(Z_n, A), \qquad A \in \mathcal{S},$$

where,

$$\mathbf{P}\left(Z_n \in ds \,\middle|\, U_n, U_{n-1}, \cdots, U_0\right) = \frac{U_n(ds)}{n+1}.$$

Generalized Urn Schemes with Colors Indexed by a Polish Space

- Formally, starting with $U_0 \in \mathcal{P}(S)$ we define $(U_n)_{n \geq 0} \subseteq \mathcal{M}(S)$ recursively as follows
$$U_{n+1}(A) = U_n(A) + R(Z_n, A), \qquad A \in \mathcal{S},$$
where,
$$\mathbf{P}\left(Z_n \in ds \,\middle|\, U_n, U_{n-1}, \cdots, U_0\right) = \frac{U_n(ds)}{n+1}.$$

- We will refer to the process $(U_n)_{n \geq 0}$ as the *urn model* with colors index by S, initial configuration U_0 and replacement kernel R.

Random Configurations

- **Random configuration of the urn:** With slight abuse of terminology, we will call the random probability measure $\frac{U_n}{n+1}$, as the *random configuration of the urn*.

Random Configurations

- **Random configuration of the urn:** With slight abuse of terminology, we will call the random probability measure $\frac{U_n}{n+1}$, as the *random configuration of the urn*.

- In fact,

$$\mathbf{P}\left(Z_n \in A \,\middle|\, U_n, U_{n-1}, \cdots, U_0\right) = \frac{U_n(A)}{n+1}, \;\; A \in \mathcal{S}.$$

Random Configurations

- **Random configuration of the urn:** With slight abuse of terminology, we will call the random probability measure $\frac{U_n}{n+1}$, as the *random configuration of the urn*.

- In fact,
$$\mathbf{P}\left(Z_n \in A \,\middle|\, U_n, U_{n-1}, \cdots, U_0\right) = \frac{U_n(A)}{n+1}, \;\; A \in \mathcal{S}.$$

- In other words, the n-th random configuration of the urn is the conditional distribution of the $(n+1)$-th selected color, given $U_0, U_1, \ldots, U_n$.

Expected Configurations

- **Expected configuration of the urn:** We define the probability measure $\frac{\mathbf{E}[U_n]}{n+1}$, as the *expected configuration of the urn*.

Expected Configurations

- **Expected configuration of the urn:** We define the probability measure $\frac{\mathbf{E}[U_n]}{n+1}$, as the *expected configuration of the urn*.

- In fact,

$$\mathbf{P}\left(Z_n \in A\right) = \frac{\mathbf{E}\left[U_n\right](A)}{n+1}, \;\; A \in \mathcal{S}.$$

Expected Configurations

- **Expected configuration of the urn:** We define the probability measure $\frac{\mathbf{E}[U_n]}{n+1}$, as the *expected configuration of the urn*.
- In fact,
$$\mathbf{P}\left(Z_n \in A\right) = \frac{\mathbf{E}\left[U_n\right](A)}{n+1}, \;\; A \in \mathcal{S}.$$
- In other words, the n-th expected configuration of the urn is the marginal distribution of the $(n+1)$-th selected color.

Branching Markov Chains on Random Recursive Tree

- For $n \geq -1$, let $\mathcal{T}_n$ be the *random recursive tree* on $(n+2)$ vertices labeled by $\{-1; 0, 1, 2, \ldots, n\}$, where the vertex labeled by -1 is considered as the *root*.

Branching Markov Chains on Random Recursive Tree

- For $n \geq -1$, let $\mathcal{T}_n$ be the *random recursive tree* on $(n+2)$ vertices labeled by $\{-1; 0, 1, 2, \ldots, n\}$, where the vertex labeled by -1 is considered as the *root*.

- We define
$$\mathcal{T} := \bigcup_{n \geq -1} \mathcal{T}_n,$$
and call it the *(infinite) random recursive tree*.

Branching Markov Chains on Random Recursive Tree

Definition: Branching Markov Chains on Random Recursive Tree

A stochastic process $(W_n)_{n \geq -1}$ taking values in $\hat{S} := \{\Delta\} \cup S$ is called a *branching Markov chain* on $\mathcal{T}$ starting at the root -1 and at a position $W_{-1} = \Delta \notin S$ if for any $n \geq 0$ and $A \in \mathcal{S}$,

$$\mathbf{P}\left(W_n \in A \,\middle|\, W_{n-1}, W_{n-2}, \ldots, W_{-1}; \mathcal{T}_n\right) = \begin{cases} U_0(A) & \text{if } \ W_{\overleftarrow{n}} = \Delta; \\ R\left(W_{\overleftarrow{n}}, A\right) & \text{otherwise,} \end{cases}$$

where $\overleftarrow{n}$ is the parent of the vertex labeled by n in $\mathcal{T}_n$.

Grand Representation Theorem

Grand Representation Theorem [B. and Thacker (2016)]

Consider an urn model with colors indexed by a Polish space $S \subseteq \mathbb{R}^d$ endowed with the Borel σ-algebra $\mathcal{S}$. Let R be the replacement kernel and U_0 be the initial configuration. For $n \geq 0$, let Z_n be the random color of the $(n+1)$-th draw. Let $(W_n)_{n \geq -1}$ be the *branching Markov chain* on $\mathcal{T}$ as defined above. Then

$$(Z_n)_{n \geq 0} \stackrel{d}{=} (W_n)_{n \geq 0} .$$

Marginal Representation Theorem

Mrginal Representation Theorem [B. and Thacker (2016)]

Consider an urn model with colors indexed by Polish space $S \subseteq \mathbb{R}^d$ endowed with the Borel σ-algebra $\mathcal{S}$. Let R be the replacement kernel and U_0 be the initial configuration. For $n \geq 0$, let Z_n be the random color of the $(n+1)$-th draw. Let $(X_n)_{n\geq 0}$ be the associated Markov chain on S with transition kernel R and initial distribution U_0. Then there exists an increasing non-negative sequence of stopping times $(\tau_n)_{n\geq 0}$ with $\tau_0 = 0$, which are independent of the Markov chain $(X_n)_{n\geq 0}$, such that,

$$Z_n \stackrel{d}{=} X_{\tau_n},$$

for any $n \geq 0$. Moreover, as $n \to \infty$,

$$\frac{\tau_n}{\log n} \longrightarrow 1 \text{ a.s.}$$

and

$$\frac{\tau_n - \log n}{\sqrt{\log n}} \stackrel{d}{\longrightarrow} N(0,1).$$

Proof of the Grand Representation Theorem

- Recall that the fundamental recursion is

$$U_n = U_{n-1} + R\left(Z_n, \cdot\right)$$

Proof of the Grand Representation Theorem

- Recall that the fundamental recursion is

$$U_n = U_{n-1} + R\left(Z_n, \cdot\right)$$

- Thus we can write

$$U_n = U_{n-1} + R\left(Z_{n-1}, \cdot\right)$$

Proof of the Grand Representation Theorem

- Recall that the fundamental recursion is

$$U_n = U_{n-1} + R\left(Z_n, \cdot\right)$$

- Thus we can write

$$\begin{aligned} & U_n = U_{n-1} + R\left(Z_{n-1}, \cdot\right) \\ \Leftrightarrow \quad & U_n = U_0 + R\left(Z_0, \cdot\right) + R\left(Z_1, \cdot\right) + \cdots + R\left(Z_{n-1}, \cdot\right) \end{aligned}$$

Proof of the Grand Representation Theorem

- Recall that the fundamental recursion is

$$U_n = U_{n-1} + R\left(Z_n, \cdot\right)$$

- Thus we can write

$$\begin{aligned} & U_n = U_{n-1} + R\left(Z_{n-1}, \cdot\right) \\ \Leftrightarrow \quad & U_n = U_0 + R\left(Z_0, \cdot\right) + R\left(Z_1, \cdot\right) + \cdots + R\left(Z_{n-1}, \cdot\right) \\ \Leftrightarrow \quad & \frac{U_n}{n+1} = \frac{U_0}{n+1} + \sum_{k=0}^{n-1} \frac{R\left(Z_k, \cdot\right)}{n+1}. \end{aligned}$$

Proof of the Grand Representation Theorem

- Recall that the fundamental recursion is

$$U_n = U_{n-1} + R\left(Z_n, \cdot\right)$$

- Thus we can write

$$\begin{aligned} & U_n = U_{n-1} + R\left(Z_{n-1}, \cdot\right) \\ \Leftrightarrow \quad & U_n = U_0 + R\left(Z_0, \cdot\right) + R\left(Z_1, \cdot\right) + \cdots + R\left(Z_{n-1}, \cdot\right) \\ \Leftrightarrow \quad & \frac{U_n}{n+1} = \frac{U_0}{n+1} + \sum_{k=0}^{n-1} \frac{R\left(Z_k, \cdot\right)}{n+1}. \end{aligned}$$

- Also recall that the conditional distribution of Z_n given $U_0; Z_0, Z_1, \cdots, Z_{n-1}$ is nothing but $\frac{U_n}{n+1}$.

Proof of the Grand Representation Theorem

A Sampling Scheme: So given $U_0; Z_0, Z_1, \cdots, Z_{n-1}$ the color Z_n can be selected as follow:

Proof of the Grand Representation Theorem

A Sampling Scheme: So given $U_0; Z_0, Z_1, \cdots, Z_{n-1}$ the color Z_n can be selected as follow:

(I) First select a (discrete) uniform random variable D_n on index set $\{0, 1, 2, \cdots, n-1\}$;

Proof of the Grand Representation Theorem

A Sampling Scheme: So given $U_0; Z_0, Z_1, \cdots, Z_{n-1}$ the color Z_n can be selected as follow:

(I) First select a (discrete) uniform random variable D_n on index set $\{0, 1, 2, \cdots, n-1\}$;

(II) Next, if $D_n \geq 1$ then select Z_n from the probability distribution $R\left(Z_{D_n}, \cdot\right)$, that is, a move by R-chain from the position Z_{D_n};

Proof of the Grand Representation Theorem

A Sampling Scheme: So given $U_0; Z_0, Z_1, \cdots, Z_{n-1}$ the color Z_n can be selected as follow:

(I) First select a (discrete) uniform random variable D_n on index set $\{0, 1, 2, \cdots, n-1\}$;

(II) Next, if $D_n \geq 1$ then select Z_n from the probability distribution $R\left(Z_{D_n}, \cdot\right)$, that is, a move by R-chain from the position Z_{D_n};

(III) Otherwise, select Z_n from the initial configuration U_0.

Proof of the Grand Representation Theorem

Another Sampling Scheme: Let $P_{-1}; P_0, P_1, P_2, \cdots$ be i.i.d. Poisson point processes of unit intensity.

Proof of the Grand Representation Theorem

Another Sampling Scheme: Let $P_{-1}; P_0, P_1, P_2, \cdots$ be i.i.d. Poisson point processes of unit intensity. We start at say, -1 which we call the root. We endowed -1 with a state, say $\Delta \notin S$, and the Poisson clock P_{-1}.

Proof of the Grand Representation Theorem

Another Sampling Scheme: Let $P_{-1}; P_0, P_1, P_2, \cdots$ be i.i.d. Poisson point processes of unit intensity. We start at say, -1 which we call the root. We endowed -1 with a state, say $\Delta \notin S$, and the Poisson clock P_{-1}.

(I) When the clock rings a new vertex labeled by 0 appears and gets attached to the root -1. It is then endowed with a state Z_0 which is a sample from U_0 and also receives the Poisson clock P_0.

Proof of the Grand Representation Theorem

Another Sampling Scheme: Let $P_{-1}; P_0, P_1, P_2, \cdots$ be i.i.d. Poisson point processes of unit intensity. We start at say, -1 which we call the root. We endowed -1 with a state, say $\Delta \notin S$, and the Poisson clock P_{-1}.

(I) When the clock rings a new vertex labeled by 0 appears and gets attached to the root -1. It is then endowed with a state Z_0 which is a sample from U_0 and also receives the Poisson clock P_0.

(II) Now a new vertex labeled 1 appears when one of the Poisson clocks rings and it gets attached to the vertex for which the clock ringed. It is then endowed with a state Z_1 which is a sample from U_0 if it is attached at -1, otherwise it is a move by R-chain from Z_0. It also receives its Poisson clock P_1.

Proof of the Grand Representation Theorem

Another Sampling Scheme: Let $P_{-1}; P_0, P_1, P_2, \cdots$ be i.i.d. Poisson point processes of unit intensity. We start at say, -1 which we call the root. We endowed -1 with a state, say $\Delta \notin S$, and the Poisson clock P_{-1}.

(I) When the clock rings a new vertex labeled by 0 appears and gets attached to the root -1. It is then endowed with a state Z_0 which is a sample from U_0 and also receives the Poisson clock P_0.

(II) Now a new vertex labeled 1 appears when one of the Poisson clocks rings and it gets attached to the vertex for which the clock ringed. It is then endowed with a state Z_1 which is a sample from U_0 if it is attached at -1, otherwise it is a move by R-chain from Z_0. It also receives its Poisson clock P_1.

(III) Having constructed the vertices $-1; 0, 1, \cdots, n-1$ endowed with states $Z_{-1} \equiv \Delta; Z_0, Z_1, \cdots, Z_{n-1}$ respectively, we bring a new vertex n when one of the clocks $P_0; P_1, P_2, \cdots, P_{n-1}$ rings. It gets attached to the vertex for which the clock ringed. It is then endowed with a state Z_n which is either a sample from U_0 (if it got attached to 0) or a move by R-chain from the state of the vertex it got attached to.

Few Remarks on the Representation Theorems

- The *Grand Representation Theorem* links the sequence of chosen colors to a branching Markov chain on the random recursive tree.

Few Remarks on the Representation Theorems

- The *Grand Representation Theorem* links the sequence of chosen colors to a branching Markov chain on the random recursive tree.
- The *Marginal Representation Theorem* is an immediate consequence of it.

Few Remarks on the Representation Theorems

- The *Grand Representation Theorem* links the sequence of chosen colors to a branching Markov chain on the random recursive tree.
- The *Marginal Representation Theorem* is an immediate consequence of it.
- Now by evoking known properties of the random recursive tree, we can try to prove results for either of the two processes.

An Assumption on the Replacement Kernal

$(X_n)_{n\geq 0}$ denotes a Markov chain with state space S, transition kernel R and starting distribution U_0.
We now make the following assumption:

(A) There exists a (non-random) probability Λ on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$ and a vector $\mathbf{v} \in \mathbb{R}^d$, and two functions $a : \mathbb{R}_+ \to \mathbb{R}$ and $b : \mathbb{R}_+ \to \mathbb{R}_+$, such that, for any initial distribution U_0,

$$\frac{X_n - a(n)\mathbf{v}}{b(n)} \xrightarrow{d} \Lambda. \tag{1}$$

Asymptotic of the Random Configuration of the Urn

Define $\mathcal{F}_n := \sigma\left(Z_0, Z_1, \cdots, Z_n\right)$, $n \geq 0$. Let P_n be a version of the regular conditional distribution of Z_n given $\mathcal{F}_n$. Note by construction $P_n = \frac{U_n}{n+1}$ almost surely.

Asymptotic of the Random Configuration of the Urn

Theorem 1 [B. and Thacker (2017)]

Suppose assumption **(A)** holds. Let P_n^{cs} is the conditional distribution of $\frac{Z_n - a(\log n)\mathbf{v}}{b(\log n)}$ given $\mathcal{F}_n$, that is, a scaled and centered version of P_n with centering by $a(\log n)\,\mathbf{v}$ and scaling by $b(\log n)$, then

(a) If $a = 0$ and $b = 1$, then

$$P_n^{\text{cs}} = P_n \xrightarrow{p} \Lambda \text{ in } \mathcal{P}(S). \tag{2}$$

Asymptotic of the Random Configuration of the Urn

Theorem 1 [B. and Thacker (2017)]

Suppose assumption **(A)** holds. Let P_n^{cs} is the conditional distribution of $\frac{Z_n - a(\log n)\mathbf{v}}{b(\log n)}$ given $\mathcal{F}_n$, that is, a scaled and centered version of P_n with centering by $a\,(\log n)\,\mathbf{v}$ and scaling by $b\,(\log n)$, then

(a) If $a = 0$ and $b = 1$, then

$$P_n^{cs} = P_n \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(S\right). \tag{2}$$

(b) Suppose $a = 0$ and b is regularly varying function, then

$$P_n^{cs} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3}$$

Asymptotic of the Random Configuration of the Urn

Theorem 1 [B. and Thacker (2017)]

(c) Suppose a is differentiable and $\lim_{x\to\infty} a'(x) = \tilde{a} < \infty$. Also assume b is regularly varying and $\lim_{x\to\infty} \frac{\sqrt{x}}{b(x)} = \tilde{b} < \infty$, then

$$P_n^{cs} \xrightarrow{p} \Xi \text{ in } \mathcal{P}\left(\mathbb{R}^d\right), \tag{4}$$

where Ξ is Λ if $\tilde{a} = 0$ or $\tilde{b} = 0$, otherwise, it is given by the convolution of Λ and Normal $\left(0, \tilde{a}^2 \tilde{b}^2\right) \mathbf{v}$.

Asymptotic of the Expected Configuration of the Urn

Theorem 2 [B. and Thecker (2017)]

Suppose assumption **(A)** holds, then

(a) If $a = 0$ and $b = 1$, then

$$Z_n \Rightarrow \Lambda. \tag{5}$$

(b) Suppose $a = 0$ and b is regularly varying function, then

$$\frac{Z_n - a(\log n)\,\mathbf{v}}{b(\log n)} \Rightarrow \Lambda, \tag{6}$$

(c) Suppose a is differentiable and $\lim_{x\to\infty} a'(x) = \tilde{a} < \infty$. Also assume b is regularly varying and $\lim_{x\to\infty} \frac{\sqrt{x}}{b(x)} = \tilde{b} < \infty$, then

$$\frac{Z_n - a(\log n)\,\mathbf{v}}{b(\log n)} \Rightarrow \Xi, \tag{7}$$

where Ξ is Λ if $\tilde{a} = 0$ or $\tilde{b} = 0$, otherwise, it is given by the convolution of Λ and Normal $\left(0, \tilde{a}^2\tilde{b}^2\right)\mathbf{v}$.

Classical Set Up: Finite/Countable Color Set

Theorem 3

Suppose S is countable, $\mathcal{S} = \wp(S)$, R is ergodic with stationary distribution π on S. Then as $n \to \infty$,

$$\frac{U_n}{n+1} \xrightarrow{p} \pi \text{ in } \mathcal{P}(S). \tag{8}$$

In particular,

$$\frac{\mathbf{E}[U_n]}{n+1} \xrightarrow{w} \pi, \tag{9}$$

as $n \to \infty$.

Classical Set Up: Block Diagonal Replacement Matrix

Theorem 4 [B. and Thacker (2017)]

Consider an urn model with colors indexed by a set S and replacement kernel given by

$$R = \begin{pmatrix} R_{11} & 0 & 0 & \cdots & 0 \\ 0 & R_{22} & 0 & \cdots & 0 \\ 0 & 0 & R_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R_{kk} \end{pmatrix},$$

Then for every initial configuration U_0, as $n \to \infty$,

$$\frac{U_n}{n+1} \xrightarrow{p} \Pi \text{ in } \mathcal{P}(S), \tag{10}$$

where Π is a random probability measure on $(S, \mathcal{S})$ given by

$$\Pi(A) = \sum_{i \in \mathcal{I}} \pi_i (A \cap C_i)\, \nu_i, \;\; A \in \mathcal{S}, \tag{11}$$

and ν has *Ferguson Distribution* on the countable set $\mathcal{I}$ with parameter $U_0 \circ \phi^{-1}$.

Non-Classical: Infinite Colors with Kernal as the Random Walk

Theorem 5 [B. and Thacker (2015)]

Consider an infinite color urn model with colors indexed by $S = \mathbb{Z}^d$, and kernel R is simple symmetric random walk. Suppose the starting configuration is U_0. We define,

$$P_n^{cs}(A) := \frac{U_n}{n+1}\left(\sqrt{\log n}A\right), \quad A \in \mathcal{B}_{\mathbb{R}^d},$$

then, as $n \to \infty$,

$$P_n^{cs} \xrightarrow{p} \Phi_d \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{12}$$

In particular,

$$\frac{Z_n}{\sqrt{\log n}} \Rightarrow \text{Normal}_d(0, \mathbf{I}_d), \tag{13}$$

as $n \to \infty$.

Component Sizes of Random Recursive Tree

Theorem 6 [B. and Thacker (2017)]

Let $\mathcal{T}_n$ be the random recursive tree on $n+2$ vertices labeled as $\{-1; 0, 1, 2, \ldots, n\}$ with -1 as the root. Let N_n be the degree of -1 in $\mathcal{T}_n$ and $S_1, S_2, \cdots, S_{N_n}$ be the sizes of the subtrees rooted at the children of the root -1. Let Ξ_n be the (finite) point process on $(0,1)$ obtained from the random points $\left(\frac{S_1}{n+1}, \frac{S_2}{n+1}, \cdots, \frac{S_{N_n}}{n+1}\right)$. Then almost surely,

$$\Xi_n \xrightarrow{d} \text{Dirichlet}\,(dx)\,.$$

Thank You