Nonlinear Pólya Urn Models and Self-Organizing Processes

Tong Zhu

A Dissertation

in

Mathematics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2009

Robin Pemantle
Supervisor of Dissertation

Tony Pantev
Graduate Group Chairperson

Acknowledgments

First and foremost, I am deeply grateful to my advisor Robin Pemantle. His patient guidance and kind support made this work possible. I am thankful for the chance to study with him and appreciate all the help he has given to me for all these years. Thank you Robin!

I also want to thank Herman Gluck, Herbert Wilf and Tony Pantev for all the help they have given to me during my graduate study. I am so lucky to get your support.

A lot of thanks to Janet Burns, Monica Pallanti and Paula Scarborough, and everyone in the Math Department. Thank you for all the help and making the department like a family.

Last, but not least, I would like to express my deepest gratitude to my parents. Without their unconditional love and support, I would not be able to do what I have accomplished.

ABSTRACT

Nonlinear Pólya Urn Models and Self-Organizing Processes

Tong Zhu

Robin Pemantle, Advisor

We consider a class of nonlinear Pólya urn models, which are applied to model self-organizing processes. The first two chapters are devoted to a survey of urn models, theories and methods. In the third chapter, we study the nonlinear urns, which show monopoly properties. We found the asymptotic distributions of the minority color and the attraction time. We showed the asymptotic behaviors of the final monopoly stage. Our main method is the exponential embedding. We have also found a simple probabilistic way to prove combinatorial identities involving symmetric functions. In the last part, we introduce the applications of Pólya urn models and the future studies.

Contents

Chapter 1

Introduction

Urn models are a group of simple and useful probabilistic models ([68], [69]). One significant star in this family is the Pólya urn model and its generalizations.

In 1923, Eggenberger and Pólya ([39]) first introduced the original *Eggenberger-Pólya urn model* (usually simplified as *Pólya urn*) to study the contagious diseases. Through the past 80 years, people have generalized this model in many different ways. These models are commonly named *generalized Pólya urn models (GPU)*. Not only have the GPU models been widely developed in theories, but they have also been applied to numerous fields, for example, statistics, computer science, biomedicine and economics, etc.

1.1 Basic Models and Notations

A single-urn GPU model involves an urn containing balls of different colors (or types). At each discrete time (some generalizations consider continuous time intervals), some balls are added or removed according to probabilities depending on the composition of balls of different colors at the time.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The randomness we need on this space is defined by a sequence of IID random variables uniform on $[0,1]$, say $\{U_{nk} : n \geq 1, k \geq 1\}$. Let $\mathcal{F}_n$ $(n \geq 1)$ be the σ-field generated by $\{U_{ik} : 1 \leq 1 \leq n, k \geq 1\}$ and $\mathcal{F}_0$ be the trivial one. In most cases, $\{U_{n1}\}$ is enough, hence simplified as $\{U_n\}$. For an event $A \in \mathcal{F}$, the indicator function of A is

$$\mathbf{1}_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Let $\mathbb{Z}_+^d$ denote the d-dimensional nonnegative integer space. For any vector $\mathbf{r} = (r_1, ..., r_d) \in \mathbb{Z}_+^d$, define $|\mathbf{r}| = r_1 + ... + r_d$. Let $\mathbf{R}(n) = (R_1(n), ..., R_d(n))$ be a homogeneous Markov chain with state space $\mathbb{Z}_+^d$. In GPU models, $R_i(n)$ represents the number of balls of color i at time n. Define the normalized vector of $\mathbf{R}(n)$:

$$\mathbf{X}(n) := \begin{cases} \frac{\mathbf{R}(n)}{|\mathbf{R}(n)|} & \text{if } |\mathbf{R}(n)| \neq 0 \\ 0 & \text{if } |\mathbf{R}(n)| = 0, \end{cases}$$

which shows the composition of colors in the urn. Denote $\Delta^{d-1} := \{(x_1, ..., x_d) \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1, \text{ and } x_i \geq 0 \text{ for all } i\}$ to be the standard $(d-1)$-simplex. Hence,

$\mathbf{X}(n) \in \Delta^{d-1}$. Let $\Upsilon : \mathbb{Z}_+^d \times \mathbb{Z}_+^d \to [0,1]$ be the transition kernel of $\mathbf{R}(n)$, i.e. $\Upsilon(\mathbf{r}, \mathbf{r}') = \mathbb{P}(\mathbf{R}(n+1) = \mathbf{r}' | \mathbf{R}(n) = \mathbf{r})$. In most GPU models, the transition kernel Υ depends only on $\mathbf{X}(n)$. In this dissertation, we refer $\{\mathbf{R}(n)\}$ to be a *generalized (Pólya) urn process*.

Associated with $\{\mathbf{R}(n)\}$, there are some other random processes. Let $\{\xi_n\}$ be the drawn color sequence, with $\xi_n \in \{1, ..., d\}$ and $\mathbb{P}(\xi_{n+1} | \mathcal{F}_n) := \mu(\mathbf{X}(n))$. Also define $\delta(n) = (\delta_1(n), ..., \delta_d(n)) := \nu(\mathbf{X}(n))$ to be the adding sequence, i.e. $\delta_i(n) := R_i(n+1) - R_i(n)$, showing the replacement schemes, which could be either time-homogeneous or inhomogeneous.

In the following, we introduce some examples of GPU models:

Original Pólya Urn: Initially, the urn has $R_1(0)$ balls of color 1 and $R_2(0)$ balls of color 2. Each time, one ball is drawn out uniformly at random and returned to the urn with another ball of the same color. Formally, at time n or after the n-th operation,

$$\mathbb{P}(\xi_{n+1} = i | \mathcal{F}_n) = \frac{R_i(n)}{R_1(n) + R_2(n)}, \text{ for } i = 1, 2.$$

And $\delta_i(n+1) = \mathbf{1}_{\{\xi_{n+1} = i\}}$.

Friedman's Urn: In 1949, B. Friedman generalized the Pólya urn model in such a way that, each time one ball is drawn from the urn at random, it is returned to the urn with α balls of the drawn color, as well as β balls of the opposite color. Hence, the distribution of ξ_{n+1} is the same as in the original Pólya urn. And

$\delta_i(n+1) = \alpha \mathbf{1}_{\{\xi_{n+1}=i\}} + \beta \mathbf{1}_{\{\xi_{n+1}\neq i\}}$. This gives a matrix form to the replacement scheme, sometimes called a *replacement matrix*, which is important in determining the behavior of the urn process,

$$\mathbf{A} := \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}.$$

The replacement scheme can also be time-inhomogeneous. Generally, suppose the replacement matrix is

$$\mathbf{A}(n) = (a_{ij}(n))_{i,j=1,...,d}.$$

Then the adding sequence can be represented by

$$(\delta_1(n+1), ..., \delta_d(n+1)) = (\mathbf{1}_{\{\xi_{n+1}=1\}}, ..., \mathbf{1}_{\{\xi_{n+1}=d\}}) \cdot \mathbf{A}(n)$$

GPU with Urn Functions: In many cases, the measure $\mu(\mathbf{X}_n) = \mathbf{X}_n$. But for some generalizations, the draws depend on some *urn function*. This was first introduced in [57]. An *urn function* $\mathbf{F}(\mathbf{X}(n)) = (F_1(X_1(n)), ..., F_d(X_d(n)))$ is a function defined on the $(d-1)$-simplex

$$\mathbf{F} : \Delta^d \to \Delta^d.$$

In other words, $\mathbb{P}(\xi_{n+1} = i|\mathcal{F}_n) = F_i(X_i(n))$.

1.2 Classifications of GPU Models

Given a lot of generalizations, it is good to classify them. In any given iteration of the urn process, it is not hard to see that the drawing and replacing schemes play the key role. Hence, we may classify GPU models based on $\mu(\mathbf{X}_n)$ and $\nu(\mathbf{X}_n)$. For example, the replacement matrix $\mathbf{A}_n$ could be deterministic or random; a deterministic one could be classified by different forms (see for example [44]); even eigenvalues determine the probability behaviors (see for example [66]).

Remark 1.2.1. In Johnson and Kotz's book [68], they have pointed out that "The way in which the replacement is modified determines the particular subclass of distributions arising. In particular, the kind of replacement depends on the results of the sampling." (pp.176). See their book for more details.

These types of classifications are sometimes more or less physical. More essentially, classifying GPU models based on some characteristics is another idea, like (partial) exchangeability, which we will discuss more in later chapters, or asymptotic behaviors. But on the other hands, these are results of the physical operations. Knowing their relations might be another important job. For example, Pouyanne classified some GPU according to their asymptotic behaviors, as well as discussed the eigenvalue of the matrices.

Outline of this dissertation is as follows. In Section 2, we introduce several methods to study GPU models and their main results. Section 3 is devoted to the

discussion of one special type of GPU processes, which shows monopoly properties. Applications of GPU models are introduced in Section 4. And finally in Section 5, we will list some open problems.

Chapter 2

Methods and Theories

2.1 Exchangeability

Exchangeability is a basic property of the original Pólya urn. As a method, it is not so useful in the study of GPU models, but instead has some applications for statistics, which will be discussed more in later sections.

Definition 2.1.1. A sequence $\{X_n\}$ of random variables is **exchangeable** if for all $n \geq 2$,

$$(X_1, ..., X_n) \stackrel{\mathcal{D}}{=} (X_{\pi(1)}, ..., X_{\pi(n)})$$

for all $\pi \in S(n)$. And $S(n)$ is the symmetric group of $\{1, ..., n\}$.

Theorem 2.1.2 (de Finetti Theorem)**.** *A sequence $\{X_n\}$ of Bernoulli random variables is exchangeable if and only if there exists a distribution function $F(x)$ on $[0, 1]$*

such that for all n

$$\mathbb{P}(X_1 = x_1, ..., X_n = x_n) = \int_0^1 t^{\sum_{i=1}^n x_i}(1-t)^{n-\sum_{i=1}^n x_i} dF(t).$$

The drawn color sequence of the original Pólya urn model is exchangeable. For the simplicity of discussion, we will set $\xi_i = 1$ if color 1 is drawn and otherwise, $\xi_i = 0$.

Theorem 2.1.3. *The sequence* $\{\xi_n\}$ *is exchangeable. Hence, there is a random variable* Θ *with values in* $[0,1]$ *such that given* $\Theta = \theta$*, the variables* $\xi_1, \xi_2, ...$ *are independent Bernoulli(*θ*). And the de Finetti measure of* $\{\xi_n\}$*, i.e. the distribution of* Θ *is Beta(*$R_1(0)$*,*$R_2(0)$*). Conversely, the sequence* $\{\xi_n\}$ *can be constructed by a* Θ *randomly chosen from* $[0,1]$ *with distribution Beta(*$R_1(0)$*,*$R_2(0)$*) in the following step: each* ξ_n *is independently color-1 with probability* Θ *and is color-2 with probability* $1 - \Theta$*.*

Proof: Among any n draws, the probability that there are $m \leq n$ draws of color 1 is always:

$$\frac{R_1(0)(R_1(0)+1)\cdots(R_1(0)+m-1)\cdot R_2(0)(R_2(0)+1)\cdots(R_2(0)+n-m-1)}{(R_1(0)+R_2(0))(R_1(0)+R_2(0)+1)\cdots(R_1(0)+R_2(0)+n-1)}$$

This shows the exchangeability of $\{\xi_n\}$. The conclusion of the distribution follows from Theorem 2.1.4. ■

Theorem 2.1.4. *The process* $\{X_1(n)\}_{n\geq 0}$ *converges almost surely to some random limit* X_∞*. And*

(i) When $R_1(0) = 1$, $R_2(0) = 1$, X_∞ has a uniform distribution on $[0, 1]$;

(ii) In general, X_∞ has a beta distribution with parameters $R_1(0)$ and $R_2(0)$.

Proof:

(i) First, $\{X_1(n)\}_{n\geq 0}$ is a martingale:

$$\begin{aligned}\mathbb{E}(X_1(n+1)|\mathcal{F}_n) &= \mathbb{E}\Big(\frac{R_1(n) + \mathbf{1}_{\{\xi_{n+1}=1\}}}{R_1(n) + R_2(n) + 1}|\mathcal{F}_n\Big) \\ &= \frac{R_1(n)}{R_1(n) + R_2(n) + 1} + \frac{1}{R_1(n) + R_2(n) + 1}\frac{R_1(n)}{R_1(n) + R_2(n)} \\ &= X_1(n).\end{aligned}$$

Obviously, each $X_1(n)$ is bounded by 1. By Doob's martingale convergence theorem, $\{X_1(n)\}$ converges to a random variable X_∞ with probability one. The distribution of X_∞ will be a consequence of the general result in (ii).

(ii) In n draws, the probability that $\xi_i = 1$ for all $1 \leq i \leq n$ is

$$\mathbb{P}(\xi_i = 1, 1 \leq i \leq n) = \prod_{i=0}^{n-1} \frac{R_1(0) + i}{R_1(0) + R_2(0) + i}.$$

On the other hand, by Theorem 2.1.3 and part (i), the Θ in Theorem 2.1.3 is just X_∞. Since $\{\xi_n\}$ is exchangeable, from Theorem 2.1.2,

$$\mathbb{P}(\xi_i = 1, 1 \leq i \leq n) = \mathbb{E}(X_\infty^n).$$

■

In 1980, Diaconis and Freedman generalized de Finetti's theorem for Markov chains [30].

Definition 2.1.5. Suppose σ and τ are two finite strings on a state space I. We say σ and τ are **equivalent**, denoted as $\sigma \sim \tau$, if and only if σ and τ have the same initial state and the same number of transitions from i to j for every pair of states i and j in I.

With this, we can define the generalization of exchangeability, *partial exchangeability*.

Definition 2.1.6. A sequence $X = \{X_n\}_{n\geq 0}$ on state space I is called **partially exchangeable** if for any equivalent strings $\sigma = (\sigma_0, ..., \sigma_m)$ and $\tau = (\tau_0, ..., \tau_m)$ in I^{m+1},

$$\mathbb{P}(X_0 = \sigma_0, ..., X_m = \sigma_m) = \mathbb{P}(X_0 = \tau_0, ..., X_m = \tau_m).$$

Remark 2.1.7. Here we follow the definition of *partial exchangeability* in Diaconis and Freedman's paper [30], which is sometimes called **Markov exchangeability**.

Definition 2.1.8. Let $\mathcal{P}$ be the space of stochastic matrices on $I \times I$. X is **a mixture of Markov chains** if there is a probability μ on $\mathcal{P}$ such that for all $m \geq 1$ and all finite sequence $(i_0, ..., i_m)$ of elements in I,

$$\mathbb{P}(X_k = i_k : 0 \leq k \leq m) = \int_{\mathcal{P}} \prod_{k=0}^{m-1} p(i_k, i_{k+1}) \mu(dp).$$

Theorem 2.1.9 ([30])**.** *Let $X := \{X_n\}_{n\geq 0}$ be a sequence of random variables defined on a countable state space I. If X is recurrent, then X is a mixture of Markov chains*

if and only if X is partially exchangeable. And the mixing measure is uniquely determined.

In 2000, Muliere, Secchi and Walker constructed a *reinforced urn process* and showed that it is partially exchangeable. Let I be a countable state space. With every $i \in I$, an urn is associated with composition of balls $(R_{i1}(n), ..., R_{id}(n))$ of $d \geq 2$ colors at time $n \geq 0$. Given a function $q : I \times [d] \to I$, define a reinforced random walk Z_n on I as follows: fixed an initial state $i_0 \in I$, assume $Z_0 = i_0$ and for all $n \geq 1$, if $Z_{n-1} = i_{n-1} \in I$, from the urn associated with i_{n-1} a ball is drawn at random and returned to the urn with another ball of the same color, say color $m \in [d]$; then set $Z_n = q(i_{n-1}, m)$.

Theorem 2.1.10 ([82])**.** $\{Z_n\}$ *defined above is partially exchangeable.*

Proof: For all $n \geq 1$ and finite sequence $(i_0, ..., i_n)$ of I, the marginal distribution

$$\mathbb{P}(Z_0 = i_0, ..., Z_n = i_n) = \prod_{i \in I} \frac{\prod_{m \in [d]} \prod_{k=0}^{l_i(m)-1} (R_{im}(n) + k)}{\prod_{k=0}^{t(i)-1} \left(k + \sum_{m \in [d]} R_{im}(n)\right)},$$

where $l_i(m) := t(i, q(i, m))$ is the number of transitions in $(i_0, ..., i_n)$ from state i to state $q(i, m)$ and $t(i) = \sum_{i' \in I} t(i, i')$. By computing this probability and definition of partial exchangebility, the authors proved the theorem. ■

By Theorem 2.1.9, when $\{Z_n\}$ is recurrent, it is a mixture of Markov chains. Muliere, Secchi and Walker [82] also proved the unique mixing distribution on the stochastic matrices space, which is Dirichlet.

Reinforced urn processes are more convenient for constructing more general classes of priors commonly used in Polya trees and beta-Stacy processes.

Comparing with the Pólya urn model, not all Friedman's urns are exchangeable.

Theorem 2.1.11. *If $\{\xi_n\}$ is the color sequence generated by a Friedman's urn, then $\{\xi_n\}$ is infinite and exchangeable if and only if one of the following two conditions is satisfied:*

(a) $\alpha \geq 0$ *and* $\beta = 0$*;*

(b) $\alpha = \beta$ *and* $R_0 = W_0$.

Hill, Lane and Sufferth [58] studied the exchangeability property of GPU models with non-identity urn functions. They showed that for 2-color Polya urn models with urn function f, its drawn color sequence is exchangeable if and only if $f(x) = x$ (i.e. Pólya), or $f(x) \equiv p$ with $p \in (0, 1)$ (i.e. Bernoulli) or $f(x) = 1$ (i.e. deterministic, or always draw red balls).

2.2 Martingale Method and Stochastic Approximation

2.2.1 Martingale Method

B. Friedman [49] generalized the original Pólya urn in a way that every time α balls of the same color as the drawn ball, as well as β balls of the opposite color are added to the urn. In 1965, D. Freedman [47] used a martingale method and some moment calculations to show the convergence of $\{X_1(n)\}$, the sequence of proportions of color-1 balls in the urn.

Theorem 2.2.1 ([47]). *Let* $\rho = \frac{\alpha-\beta}{\alpha+\beta}$.

(i) If $\rho > \frac{1}{2}$ *and* $\alpha > 0$, $\beta \geq 0$, *then* $\lim_{n\to\infty} n^{-\rho}(R_1(n) - R_2(n)) = Z$ *with probability* 1, *where* Z *is some non-degenerate random variable;*

(ii) If $\rho = \frac{1}{2}$ *and* $\alpha, \beta > 0$, *then* $(R_1(n) - R_2(n))/\sqrt{n \log n} \xrightarrow{d} N(0, (\alpha-\beta)^2)$, *as* $n \to \infty$*;*

(iii) If $\rho < \frac{1}{2}$ *and* $\alpha \geq 0$, $\beta > 0$, *then* $(R_1(n) - R_2(n))/\sqrt{n} \xrightarrow{d} N(0, (\alpha-\beta)^2/(1-2\rho))$, *as* $n \to \infty$.

Proof:

(i) Let $s_0 = R_1(0) + R_2(0)$ and $a_n(i) = 1 + \frac{i(\alpha-\beta)}{s_0+(\alpha+\beta)n}$, $i \geq 0$. Define

$$Z_n = \frac{R_1(n) - R_2(n)}{\prod_{k=0}^{n-1} a_k(1)}.$$

We compute that

$$\begin{aligned}
\mathbb{E}(Z_{n+1}|\mathcal{F}_n) =& \big[\prod_{k=0}^{n} a_k(1)\big]^{-1}\mathbb{E}(R_1(n+1) - R_2(n+1)|\mathcal{F}_n) \\
=& \big[\prod_{k=0}^{n} a_k(1)\big]^{-1}[R_1(n) + \frac{R_1(n)\alpha + R_2(n)\beta}{R_1(0) + R_2(0) + (\alpha+\beta)n} \\
& - R_2(n) - \frac{R_1(n)\beta + R_2(n)\alpha}{R_1(0) + R_2(0) + (\alpha+\beta)n}] \\
=& \big[\prod_{k=0}^{n} a_k(1)\big]^{-1} a_n(1)(R_1(n) - R_2(n)) = Z_n
\end{aligned}$$

So Z_n is a martingale. Before showing that Z_n converges, we estimate $\prod_{k=0}^{n} a_k(i)$ for $i \geq 0$.

$$\begin{aligned}
\prod_{k=0}^{n} a_k(i) &= \prod_{k=0}^{n} \Big[1 + \frac{i(\alpha-\beta)}{s_0 + (\alpha-\beta)k}\Big] \\
&= \frac{\Gamma\big(s_0/(\alpha+\beta)\big)\Gamma\big((i(\alpha-\beta)+s_0)/(\alpha+\beta) + n + 1\big)}{\Gamma\big((i(\alpha-\beta)+s_0)/(\alpha+\beta)\big)\Gamma\big(s_0/(\alpha+\beta) + n + 1\big)}.
\end{aligned}$$

By Stirling's formula, $\Gamma(a+n)/\Gamma(b+n) \sim n^{a-b}$ for a, b real. Hence,

$$\prod_{k=0}^{n} a_k(i) \sim \frac{\Gamma\big(s_0/(\alpha+\beta)\big)}{\Gamma\big((i(\alpha-\beta)+s_0)/(\alpha+\beta)\big)} n^{i\rho}. \tag{2.2.1}$$

Then

$$\begin{aligned}
&\mathbb{E}\big[(R_1(n+1)-R_2(n+1))^2|\mathcal{F}_n\big]\\
=&\frac{R_1(n)}{s_0+(\alpha+\beta)n}(R_1(n)-R_2(n)+\alpha-\beta)^2+\\
&+\frac{R_2(n)}{s_0+(\alpha+\beta)n}(R_1(n)-R_2(n)-\alpha+\beta)^2\\
=&(R_1(n)-R_2(n))^2+(\alpha+\beta)^2+\frac{2(R_1(n)-R_2(n))^2(\alpha-\beta)}{s_0+(\alpha+\beta)n}\\
=&a_n(2)(R_1(n)-R_2(n))^2+(\alpha+\beta)^2 \qquad (2.2.2)
\end{aligned}$$

By induction,

$$\begin{aligned}
&\mathbb{E}\big[(R_1(n+1)-R_2(n+1))^2\big]\\
=&\prod_{k=0}^{n}a_k(2)(R_1(0)-R_2(0))^2+\Big[\sum_{k=0}^{n}\prod_{j=0}^{k}a_j(2)\Big](\alpha+\beta)^2.
\end{aligned}$$

In (2.2.1), let $i=2$, we get $\sup_n \mathbb{E}(Z_n^2)<\infty$. So Z_n converges to a finite limit with probability 1. By letting $i=1$ in (2.2.1), we complete the proof of part (i).

(ii) If $\rho=(\alpha-\beta)/(\alpha+\beta)=1/2$, we first compute the moments of $R_1(n)-R_2(n)$.

If $R_1(0)=R_2(0)$, $R_1(n)-R_2(n)$ has a symmetric distribution, hence $\mathbb{E}\big[(R_1(n)-R_2(n))^{2k+1}\big]=0$ for all k nonnegative integer. When $R_1(0)\neq R_2(0)$ and $k=0$, similar to the computation of (2.2.2),

$$\mathbb{E}(R_1(n+1)-R_2(n+1))=\prod_{j=0}^{n}a_j(1)(R_1(0)-R_2(0))\sim n^{\frac{1}{2}}.$$

When $k \neq 0$,

$$
\begin{aligned}
&\mathbb{E}\big[(R_1(n+1)-R_2(n+1))^{2k+1}|\mathcal{F}_n\big] \\
=&\frac{R_1(n)}{s_0+(\alpha+\beta)n}\big(R_1(n)-R_2(n)+\alpha-\beta\big)^{2k+1}+ \\
&+\frac{R_2(n)}{s_0+(\alpha+\beta)n}\big(R_1(n)-R_2(n)-\alpha+\beta\big)^{2k+1} \\
=&\frac{R_1(n)}{s_0+(\alpha+\beta)n}\sum_{j=0}^{2k+1}\binom{2k+1}{j}(R_1(n)-R_2(n))^j(\alpha-\beta)^{2k+1-j}+ \\
&+\frac{R_2(n)}{s_0+(\alpha+\beta)n}\sum_{j=0}^{2k+1}\binom{2k+1}{j}(R_1(n)-R_2(n))^j(-\alpha+\beta)^{2k+1-j} \\
=&a_n(2k+1)(R_1(n)-R_2(n))^{2k+1}+\sum_{j=1}^{k}\Big\{\Big[\binom{2k+1}{2j}(\alpha-\beta)^{2j}+ \\
&+\frac{\binom{2k+1}{2j+1}}{s_0+(\alpha+\beta)n}(\alpha-\beta)^{2j+1}\Big](R_1(n)-R_2(n))^{2k+1-2j}\Big\} \qquad (2.2.3)
\end{aligned}
$$

Let

$$
\begin{aligned}
b_n(2k+1)=\sum_{j=1}^{k}\Big\{\Big[\binom{2k+1}{2j}(\alpha-\beta)^{2j}+\frac{\binom{2k+1}{2j+1}}{s_0+(\alpha+\beta)n}(\alpha-\beta)^{2j+1}\Big] \\
(R_1(n)-R_2(n))^{2k+1-2j}\Big\}
\end{aligned}
$$

By induction, we can get $b_n(2k+1) \sim n^{k-\frac{1}{2}}(\log n)^{k-1}$. From the recursion (2.2.3) and (2.2.1),

$$
\mathbb{E}\big[(R_1(n+1)-R_2(n+1))^{2k+1}\big] \sim n^{k+\frac{1}{2}}(\log n)^k
$$

So

$$
\mathbb{E}\big[(R_1(n+1)-R_2(n+1))^{2k+1}\big] = o\big[(n\log n)^{k+\frac{1}{2}}\big].
$$

Similarly, we can get

$$\mathbb{E}\big[(R_1(n+1) - R_2(n+1))^{2k}\big] \sim \mu(2k)(n\log n)^k,$$

where $\mu(2k)$ is defined recursively by

$$\mu(2k+2) = \frac{(\alpha-\beta)^2}{k+1}\binom{2k+2}{2}\mu(2k),$$

with $\mu(0) = 1$.

Note that $\mu(2k)$ is the $2k$-th moment of a normal distribution with mean 0 and variance $(\alpha-\beta)^2$. Normal distribution is determined uniquely by its moments. Hence by moment convergence criterion, $(R_1(n) - R_2(n))/\sqrt{n\log n} \xrightarrow{d} N(0, (\alpha-\beta)^2)$.

(iii) The proof is similar to (ii), by estimating the moments of $(R_1(n) - R_2(n))$. ■

Corollary 2.2.2. *When $\beta \neq 0$, $\lim_{n\to\infty} R_1(n)/[R_1(n)+R_2(n)] = \frac{1}{2}$ with probability 1.*

2.2.2 Stochastic Approximation

A **stochastic approximation** process (algorithm) $\{\mathbf{X}(n)\}_{n\geq 0}$ is a discrete time stochastic process defined in Euclidean space $\mathbb{R}^d$ and in general can be written as

$$\mathbf{X}(n+1) = \mathbf{X}(n) + \epsilon_n\big(\mathbf{F}(\mathbf{X}(n)) + \zeta_{n+1}\big). \tag{2.2.4}$$

In (2.2.4), $\{\epsilon_n\}_{n\geq 1}$ are a given sequence called the *step sizes* and satisfy

$$\epsilon_n \geq 0 \text{ for all } n \ ; \sum \epsilon_n = \infty \text{ and } \lim_{n\to\infty} \epsilon_n = 0;$$

$\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ is a deterministic vector field on $\mathbb{R}^d$ and $\zeta_n \in \mathbb{R}^d$ are (deterministic or random) *noises*.

One of the oldest stochastic approximation algorithm was introduced by Robbins and Monro in 1951 [99]. They applied such a method to approximate the solution of an equation $M(x) = a$, which has a unique root, say $x = \theta$. In this equation, suppose $M(x)$ is unknown and there are only some experiment observations, and a is a given constant. They set up a recursive algorithm with (2.2.4) in a form

$$x_{n+1} = x_n + \epsilon_n(a - y_n),$$

where $\mathbb{E}(y_n) = M(x)$.

Theorem 2.2.3 ([99]). *Suppose*

(i) y_n *are bounded almost surely;*

(ii) ϵ_n *is of type* $1/n$*, i.e. there are constants* $0 < c_1 < c_2$ *such that* $c_1/n \leq \epsilon_n \leq c_2/n$ *for all* n*;*

(iii) M *is nondecreasing and* $M'(\theta) > 0$.

Then $x_n \to \theta$ *in probability.*

The Robbins-Monro approximation is a basis of stochastic approximation (algorithm). It has also been an important method to study GPU models, for example,

Hill, Lane and Sudderth [57], Arthur, Ermoliev and Kaniovski [3], Pemantle [90] and [91], Benaim and Hirsch [17], Higueras, Moler, Plo and Miguel, etc. Some of the results might be applied to general stochastic approximation processes as well.

For our GPU urn models, following the notation in Chapter 1, suppose an urn initially has d balls with d different colors. Each time the probability of drawing a color depends on an urn function $\mathbf{F}(\mathbf{X}(n))$, and then one new ball of the drawn color is added to the urn. The expected change of $\mathbf{X}(n)$ is:

$$\mathbb{E}(\mathbf{X}(n+1) - \mathbf{X}(n)|\mathcal{F}_n) = \frac{1}{d+n+1}\big[\mathbf{F}(\mathbf{X}(n)) - \mathbf{X}(n)\big].$$

Hence, $\{\mathbf{X}(n)\}$ is a stochastic approximation process:

$$\mathbf{X}(n+1) - \mathbf{X}(n) = \frac{1}{d+n+1}\big[\mathbf{F}(\mathbf{X}(n)) - \mathbf{X}(n) + \zeta_{n+1}\big], \tag{2.2.5}$$

where ζ_{n+1} is the martingale difference noise $\mathbf{X}(n+1) - \mathbb{E}(\mathbf{X}(n+))|\mathcal{F}_n)$ and obviously $\mathbb{E}(\zeta_{n+1}|\mathcal{F}_n) = 0$.

First, the limit of a stochastic process $\{\mathbf{X}(n)\}$ satisfying (2.2.5) lies in the **fixed point** set of $\mathbf{F}$, i.e. the set $S := \{\mathbf{p} : \mathbf{F}(\mathbf{p}) = \mathbf{p}\}$. In this section, S is only refered to $\mathbf{F}$'s fixed point set. For a two color case, $d = 2$, we have

Theorem 2.2.4. *If F is continuous, then $\mathbb{P}\big(\lim_{n\to\infty} X_1(n) \in S\big) = 1$.*

Proof: First, for $\forall \epsilon > 0$, let

$$U_\epsilon := \{x : F(x) - x > \epsilon\}$$

$$D_\epsilon := \{x : F(x) - x < -\epsilon\}.$$

If there is some $n_0 \geq 0$ such that $X_1(n_0) \in U_\epsilon$, then

$$\mathbb{P}(\lim_{n\to\infty} X_1(n) \notin U_\epsilon) = 1. \tag{2.2.6}$$

To prove (2.2.6), consider the first exit time τ of $\{X_1(n)\}$ from U_ϵ and let τ_n be the minimum of τ and $n \geq n_0$. Then the event $\{\tau_n \geq k\}$ is $\mathcal{F}_{k-1}$ measurable. By

$$\begin{aligned}\mathbb{E}(X_1(\tau_n)) &\geq \mathbb{E}\big(\sum_{k=n_0+1}^{n} (X_1(k) - X_1(k-1))\mathbf{1}_{\{\tau_n \geq k\}}\big) \\ &\geq \mathbb{E}\big(\sum_{k=n_0+1}^{n} \mathbb{E}(X_1(k) - X_1(k-1)|\mathcal{F}_{k-1})\mathbf{1}_{\{\tau=\infty\}}\big). \end{aligned} \tag{2.2.7}$$

Note that $\mathbb{E}(X_1(\tau_n)) \leq 1$ and by Markovian property and (2.2.5), if $X_1(k-1) \in U_\epsilon$, then

$$\begin{aligned}\mathbb{E}(X_1(k) - X_1(k-1)|\mathcal{F}_{k-1}) &= \mathbb{E}(X_1(k) - X_1(k-1)|X_1(k-1)) \\ &\geq \epsilon/(d+k). \end{aligned} \tag{2.2.8}$$

From (2.2.7) and (2.2.8), we can get for $\forall n \geq n_0$ and $\forall n_0 \geq 0$,

$$\epsilon \sum_{k=n_0+1}^{n} \mathbb{P}(\tau = \infty)/(d+k) \leq 1.$$

But $\sum_k 1/(d+k) \to \infty$, so $\mathbb{P}(\tau = \infty = 0$. Similarly,

$$\mathbb{P}(\lim_{n\to\infty} X_1(n) \notin D_\epsilon) = 1. \tag{2.2.9}$$

For $\forall \epsilon > 0$, (2.2.6) and (2.2.9) are true. Hence

$$\mathbb{P}(|\lim(X_1(n) - F(X_1(n))| \leq \epsilon) = 1.$$

So $\mathbb{P}(\lim X_1(n) \in S) = 1$. ■

Can all fixed points be the limit of $X_1(n)$? The answer is no.

In 1980, Hill, Lane and Sudderth, who first raised the concept of *urn function*, considered the two color case, $d = 2$.

Definition 2.2.5. A point $p_0 \in [0, 1]$ is called an **upcrossing point** (**downcrossing point**) for F if $p_0 \in S$ and for all p in some neighborhood of p_0, $p < p_0 \Rightarrow F(p) < p$ ($F(p) > p$, respectively) and $p > p_0 \Rightarrow F(p) > p$ ($F(p) < p$, resp.). And $p_0 \in S$ is a **touchpoint** if $F(p) > p$ (or $F(p) < p$) for all $p \neq p_0$ in a neighborhood of p_0.

Theorem 2.2.6 ([57], I)**.** *If F is continuous in a neighborhood of $p_0 \in [0, 1]$ and p_0 is a downcrossing point of F, then $\mathbb{P}(X_1(n) \to p_0) > 0$.*

Arthur, Ermoliev and Kaniovski generalized Hill, Lane and Sudderth's results to higher dimensional situations. For $d \geq 2$, suppose $\langle , \rangle$ is the inner product of $\mathbb{R}^d$. A classification of fixed points is:

Definition 2.2.7. A point $\mathbf{p}_0 \in S$ is called a **stable point** if for some neighborhood U of $\mathbf{p}_0$, there is a symmetric positive definite matrix A such that

$$\langle A \cdot [\mathbf{p} - \mathbf{F}(\mathbf{p})], \mathbf{p} - \mathbf{p}_0 \rangle > 0, \qquad \forall \mathbf{p} \in U \cap [0, 1]^d, \qquad \mathbf{p} \neq \mathbf{p}_0.$$

Similarly, a point $\mathbf{p}_0 \in S$ is called an **unstable point** if for some neighborhood U of $\mathbf{p}_0$, there is a symmetric positive definite matrix A such that

$$\langle A \cdot [\mathbf{p} - \mathbf{F}(\mathbf{p})], \mathbf{p} - \mathbf{p}_0 \rangle < 0, \qquad \forall \mathbf{p} \in U \cap [0, 1]^d, \qquad \mathbf{p} \neq \mathbf{p}_0.$$

Theorem 2.2.8 ([3], I)**.** *If* $\mathbf{p}_0$ *is a stable point of* $\mathbf{F}$*, then* $\mathbb{P}(\mathbf{X}(n) \to \mathbf{p}_0) > 0$.

Proof of Theorem 2.2.6: It can be shown that if there exists a continuous urn function G such that p_0 is the only fixed point of G and $F(p) \geq G(p)$ for $p < p_0$ and $f(p) \leq g(p)$ for $p > p_0$, then $X_1(n)$ converges to p_0 almost surely. Choose a neighborhood $\mathcal{N}$ of p_0 where F is continuous and $\{p_0\} = N \cap S$. Then we can construct an urn process with the same initial condition as the urn process for F and a proportion sequence $\{Y_1(n)\}$ and urn function G. By $\{Y_1(n)\}$ converging to p_0 almost surely and comparing $\{X_1(n)\}$ and $\{Y_1(n)\}$, $\mathbb{P}(X_1(n) \to p_0) > 0$.■

Proof of Theorem 2.2.8: By definition of stable points, there exists a symmetric positive definite matrix A such that

$$\langle A \cdot (\mathbf{p} - \mathbf{F}(\mathbf{p})), \mathbf{p} - \mathbf{p}_0 \rangle > 0,$$

for all $\mathbf{p} \in U \cap [0,1]^d$ and $\mathbf{p} \neq \mathbf{p}_0$. So define a function

$$\mathcal{L}(\mathbf{p}) := \langle A \cdot (\mathbf{p} - \mathbf{p}_0), \mathbf{p} - \mathbf{p}_0 \rangle.$$

Since A is positive definite, $\mathcal{L}$ is positive definite too. And $\mathcal{L}$ is twice differentiable and $\nabla \mathcal{L} \cdot (\mathbf{F}(\mathbf{p}) - \mathbf{p}) \leq 0$. Hence $\mathcal{L}$ is a Lyapunov function. Then by the following lemma, we can finish the proof. ■

Lemma 2.2.9. *Suppose an urn function* $\mathbf{F}$ *is continuous and its fixed point set* S *contains a finite connected component. If there exists a Lyapunov function* $\mathcal{L}$ *such that*

(i) $\mathcal{L}$ *is twice differentiable;*

(ii) $\mathcal{L}(\mathbf{p}) \geq 0$ *for all* $\mathbf{p} \in [0,1]^d$ *and*

(iii) $\langle \mathbf{F}(\mathbf{p}) - \mathbf{p}, \mathcal{L}_{\mathbf{p}}(\mathbf{p}) \rangle < 0$ *for* $\mathbf{p} \in [0,1]^d \subset \mathcal{N}(S)$*, which is an open neighborhood of* S*.*

Then $\{\mathbf{X}(n)\}$ *converges to a point of* S *or to the border of a connected component.*

Proof: First, $\mathcal{L}(\mathbf{X}(n))$ is a supermartingale. On $S \subset \mathcal{N}(S)$, the expected increment of $\mathcal{L}$ is always less than some negative value. Hence the process will exit $S \subset \mathcal{N}(S)$ in finite time, hence enters $\mathcal{N}(S)$ infinitely many times. It is also not hard to show that $\sum_{i=0}^{n}(1/d+i)\zeta_i$, where ζ_i is the noise in (4), is a martingale and converges. It means that at large time, there is no enough perturbation for the process to leave $\mathcal{N}(S)$. S has finite connected components. So $\{\mathbf{X}(n)\}$ converges to a single one of them or its border. Besides, it is impossible for $\{\mathbf{X}(n)\}$ to converge to different points inside S, because the expected increment in S is zero and the perturbation converges. ■

On the other hand, some fixed points will never be the convergent limits.

Theorem 2.2.10 ([57], II)**.** *If* F *is continuous in a neighborhood of* $p_0 \in [0,1]$ *and* p_0 *is a upcrossing point of* F*, then*

$$\mathbb{P}(X_1(n) \to p_0) = 0.$$

For higher dimensions,

Theorem 2.2.11 ([3], II)**.** *If* $\mathbf{p}_0$ *is an unstable point of* $\mathbf{F}$ *and is not a vertex of the d-simplex, then* $\mathbb{P}(\mathbf{X}(n) \to \mathbf{p}_0) = 0$*.*

In 1990, Pemantle showed the nonconvergence to linearly unstable points for a general class of stochastic approximation processes.

Go back to look at (2.2.4). Usually, ϵ_n is assumed to decrease to 0 as $n \to \infty$. Also, the noise is expected to be zero. So $\{\mathbf{X}(n)\}$ approximates the solution of the ODE

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}). \tag{2.2.10}$$

Associated to the differential flow defined by (2.2.6), there are some analogue to Definition 2.2.7:

Definition 2.2.12. A **critical point p** of $\mathbf{F}$. i.e. $\mathbf{F}(\mathbf{p}) = 0$, is **linearly stable** if all the eigenvalues of $D\mathbf{F}$ have negative real parts, where $D\mathbf{F}$ is the linear approximation of $\mathbf{F}$. If some eigenvalue has a positive real part, then $\mathbf{p}$ is called a **linearly unstable** point.

Theorem 2.2.13 ([90])**.** *Suppose* $\{\mathbf{X}(n)\}$ *satisfies* (2.2.4) *and lies in an open subset* $\Delta \subset \mathbb{R}^d$*. Assume* $\mathbf{F}$ *is smooth enough. Let* $\mathbf{p}_0 \in \Delta$ *be a critical point, i.e.* $\mathbf{F}(\mathbf{p}_0) = 0$ *and* U *be a neighborhood of* $\mathbf{p}_0$*. If* n *is large enough, whenever* $\mathbf{X}(n) \in U$ *the following conditions are satisfied:*

(i) $\mathbf{p}_0$ *is linearly unstable;*

(ii) $c_1/n^\gamma \le a_n \le c_2/n^\gamma$*;*

(iii) $\mathbb{E}\big(\langle \zeta_n, \theta\rangle^+ | \mathcal{F}_n\big) \ge c_3/n^\gamma$ *for every unit tangent vector* $\theta \in T\Delta$*;*

(iv) $|\zeta_n| \le c_4/n^\gamma$*,*

where $\gamma \in (1/2, 1]$ *and* c_1, c_2, c_3, c_4 *are positive constants. Then,*

$$\mathbb{P}\big(\mathbf{X}(n) \to \mathbf{p}_0\big) = 0.$$

Proof of Theorem 2.2.10: By the construction comparison in (1), without loss of generality, we may assume $F(p) \le p$ for $p < p_0$ and $F(p) \ge p$ for $p > p_0$. Suppose the statement is not true, i.e.

$$\mathbb{P}(X_1(n) \to p_0) = \delta > 0. \tag{2.2.11}$$

To show the contradiction, we first construct a process $\{(Y(n), q(n))\}$ associated with F in a way that: $Y(0) = 1/2$ and $q(0) = F(1/2)$; $Y(1) = 1/3$, with probability $1 - q(0)$, $Y(1) = 2/3$ with probability $q(0)$ and $q(1) = F(Y(1))$; in general $Y(n+1) = \big((2+n)Y(n) + \xi_{n+1}\big)/(n+3)$ with $\mathbb{P}(\xi_{n+1} = 1) = q(n) = 1 - \mathbb{P}(\xi_{n+1} = 0)$, and $q(n+1) = F(Y(n+1))$; and so on. Clearly, $\{Y(n)\}$ has the same distribution as the urn process $\{X_1(n)\}$ with urn function F.

Given two reals $\alpha > 0$ and $\beta > 0$, let $\varphi_{\alpha+1,\beta+1}(x)$ be the probability density function of a Beta distribution with parameters $(\alpha+1, \beta+1)$. For an F, define a utility function for $Y := \{Y_n\}$ associated with F

$$u(Y) := \int \lim_n \varphi_{\alpha+1,\beta+1}(Y_n) dY.$$

Lemma 2.2.14. *If $\alpha/(\alpha+\beta) = p_0$, then for $F(x) = x$, the associated process $\{Y_n\}$, which shares the same distribution as the original Pólya process, has the maximum utility function among all F's associated processes.*

The proof of this lemma contains some straightforward computation for Beta distributions. Since we only need its conclusion, the proof is omitted here.

Also, define two sequences $\{\alpha_n\}$ and $\{\beta_n\}$ such that $\alpha_n + \beta_n \to \infty$ and $\alpha_n/(\alpha_n + \beta_n) \to p_0$. Let $\varphi_{\alpha_n+1,\beta_n+1}(x)$ be the probability density function of a Beta distribution with parameters $(\alpha_n + 1, \beta_n + 1)$. Then

Lemma 2.2.15. *The sequence of Beta distributions with parameters (α_n+1, β_n+1) converges in distribution to a point mass at p_0. And*

$$\sup \varphi_{\alpha_n+1,\beta_n+1} \to \infty.$$

Proof of the lemma: By the mean of a beta $(\alpha_n + 1, \beta_n + 1)$ is $(\alpha_n + 1)/(\alpha_n + \beta_n + 2) \to p_0$ and the variance is $(\alpha_n+1)(\beta_n+1)/\left[(\alpha_n+\beta_n+2)^2(\alpha_n+\beta_n+1)\right] \to 0$, therefore the first convergence assertion is proved. Then since the variance goes to zero, the density functions can not remain bounded. QED

Now we are ready to prove our theorem. The urn function F is not the identity function. By (10), we have that the utility function of the associated process satisfies

$$u(Y) \geq \delta\varphi_{\alpha+1,\beta+1}(p_0).$$

Let $\alpha+\beta \to \infty$ and $\alpha/(\alpha+\beta) = p_0$. From Lemma 2, $u(Y) \to \infty$. On the other hand, the utility function associated with the identity function, i.e. the original Pólya urn process, approaches $\varphi_{1,1}(p_0) < \infty$, by direct computation. The conclusion of Lemma 1 gives the contradiction. ■

Before proving Theorem 2.2.11, a lemma for general stochastic approximation processes is needed.

Lemma 2.2.16. *If a process $\{\mathbf{X}(n)\}$ defined by (3) satisfies:*

(i) Let $S := \{\mathbf{p} : \mathbf{F}(\mathbf{p}) = 0\}$ and $S' \subset S$, for $\mathbf{p}_0 \in S'$ and $\mathbf{p}$ in a neighborhood of $\mathbf{p}_0$, there exists a symmetric positive definite matrix A such that $\langle A \cdot (\mathbf{p} - \mathbf{p}_0), \mathbf{F}(\mathbf{p})\rangle < 0$;

(ii) $\{\zeta_n\}$ has bounded fourth moments and for the matrix $B(\mathbf{p}, n) := \big(\mathbb{E}(\zeta_n^i(\mathbf{p}) \times \zeta_n^j(\mathbf{p}))\big)_{1\leq i,j\leq d}$, where ζ_n^i is the i-th component of ζ_n, there are constants $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 \leq tr(B) \leq c_2;$$

(iii) There exist some $k > 0$ and $r \in (0, 1)$ such that

$$|\mathbf{F}(\mathbf{p})|^2 + |tr\big(B(\mathbf{p}, n) - B(\mathbf{p}_0, n)\big)| \leq k|\mathbf{p} - \mathbf{p}_0|^r.$$

Then $\mathbf{P}(\mathbf{X}(n) \to \mathbf{p}_0) = 0$.

Proof of Theorem 2.2.11: All we need to do is to check that the urn process defined in (2.2.5) satisfies the three conditions in Lemma 4. Condition (i) is directly from the definition of unstable points. It is also easy to see that ζ_n has bounded fourth moments. Now the diagonal elements in the matrix $B(\mathbf{p}, n)$ are in the form:

$$\mathbb{E}(\zeta_n^i(\mathbf{p})^2) = F_i(p_i)(1 - F_i(p_i)).$$

At an unstable point $\mathbf{p}_0$, $\mathbf{F}(\mathbf{p}_0) = \mathbf{p}_0$. So $\mathrm{tr}(B(\mathbf{p}_0, n)) = \sum_{i=1}^{d} p_{0i}(1 - p_{0i})$. Since $\mathbf{p}_0$ is not a vertex of the simplex, $\mathrm{tr}(B)$ is bounded above and below. ■

Sketch of proof of Theorem 2.2.13: First, the stable manifold theorem helps to decompose the tangent subspace $D\mathbf{F}$.

Theorem 2.2.17 (Stable Manifold Theorem)**.** *Let* $\mathbf{F}$, $\vec{\mathbf{p}}_0$ *and* $D\mathbf{F}$ *be defined as in the theorem. Set* $\Delta = [0, 1]^d$. *Decompose the image of* $D\mathbf{F}(\Delta)$ *into two complementary subspaces* $D\mathbf{F}(\Delta) = U \oplus W$, *where* W *is the subspace spanned by eigenvectors of* $D\mathbf{F}$ *whose corresponding eigenvalues have positive real parts and* U *is spanned by eigenvectors whose corresponding eigenvalues have nonpositive real parts. Then there is a neighborhood* $\mathcal{N}$ *of* $\mathbf{0}$ *in* U *and a function* $g \in \mathcal{C}^2 : \mathcal{N} \to W$ *such that* $g(\mathbf{0}) = \mathbf{0}$ *and the derivative of* g *at* $\mathbf{0}$, $D_{\mathbf{0}}(g) = 0$. *Furthermore, let* $\mathcal{S} = \{\mathbf{p}_0 + \mathbf{u} + g(\mathbf{u}) | \mathbf{u} \in U\}$. $\mathbf{F}$ *is always tangent to* $\mathcal{S}$.

With the help of the stable manifold theorem, a function η is constructed, measuring the distance from the particle to $\mathbf{p}_0$ in an unstable direction, where we may

think of the movement of the process as a movement of a particle. The fact that not all eigenvalues of T have positive real parts implies the existence of U.

After the geometric decomposition and construction, there are two probabilistic estimations of the difference sequence of $\{\mathbf{X}(n)\}$ in order to show that the noise ζ_n always keeps the particle at some distance from $\mathbf{p}_0$ and never gets into a small neighborhood of $\mathbf{p}_0$. ■

In 1991, Pementle also found the conditions of convergence to touchpoints.

Theorem 2.2.18 ([91]). *Suppose F is continuous and p_0 is a touchpoint of F. If $F(p) > p$ and $\frac{F(p)-p}{p_0-p} \leq k$ for some $k < \frac{1}{2}$ and all $p \in (p_0 - \epsilon, p_0)$ for some $\epsilon > 0$ (or similarly, if $F(p) < p$, and $\frac{F(p)-p}{p_0-p} \leq k$ for some $k < \frac{1}{2}$ and all $p \in (p_0, p_0 + \epsilon)$), then $\mathbb{P}(X_1(n) \to p_0) > 0$.*

Proof of Theorem 2.2.18: Without loss of generality, we may assume that $F(p) > p$ for all $p \in (0,1)\ \{p_0\}$. From (2.2.7) and (2.2.10) in the proof of Theorem 10, it suffices to prove that with positive probability there is an $N > 0$ such that whenever $n > N$, $X_1(n) < p_0$. By Theorem 10, $X_1(n)$ converges to a fixed point of F. Hence the limit must be p_0.

Suppose for $k < 1/2$, $\frac{F(p)-p}{p_0-p} \leq k$ for all $p \in (p_0 - \epsilon, p_0)$. Choose k_1 and $\gamma > 1$ such that $k < k_1 < 1/2$ and $\gamma k_1 < 1/2$. Then, define a function $g(r) = re^{(1-r)/2k_1\gamma}$. Note that $g(1) = 1$ and $g'(1) = 1 - 1/2k_1\gamma < 0$. Hence there is a $r_0 \in (0,1)$ such that $g(r_0) > 1$. Also define $T(n) := e^{n(1-r_0)/\gamma k_1}$. So $g(r_0)^n = r_0^n\sqrt{T(n)} > 1$. Choose

an M big enough such that $\gamma r_0^M < \epsilon$. Let

$$\tau_M := \inf\{i > T(M) | X_1(i-1) < p_0 - r_0^M < X_1(i)\}$$

if there is such an i; if not, then let $\tau_M = -\infty$. Since $F : (0,1) \to (0,1)$, $\mathbb{P}(\tau_M > T(M)) > 0$. For all $n > M$, define

$$\tau_{n+1} := \inf\{i \geq \tau_n | X_1(i) > p_0 - r_0^{n+1}\}.$$

If for some $i_0 > T_M$, $X_1(i_0) \geq p_0$, then for all $n \geq M$, $\tau_n \leq i_0$.

Define

$$\eta_n = \begin{cases} (1 - F(X_1(n)))/(n+2) & \text{with probability } F(X_1(n)), \\ -F(X_1(n))/(n+2) & \text{with probability } 1 - F(X_1(n)) \end{cases}$$

Then for all $m > n$, define $Z_{n,m} := \sum_{i=n}^{m-1} \eta_i$.

Now for event $\mathfrak{A} := \{\inf_{i > \tau_n} X_1(n) \geq p_0 - \gamma r_0^n\}$, estimate

$$\begin{aligned} \mathbb{P}\big(\mathfrak{A}^c | \tau_n > T(N)\big) &\leq \mathbb{P}\big(\inf_{i > \tau_n} Z_{\tau_n, i} < -(\gamma_0 - 1) r_0^n | \tau_n > T(N)\big) \\ &\leq \mathbb{E}\big(Z_{\tau_n,\infty}^2 | \tau_n > T(N)\big)/[(\gamma_0 - 1) r_0^n]^2 \\ &\leq e^{-n(1-r_0)/k_1\gamma_0}/[(\gamma_0 - 1) r_0^n]^2 \\ &= [(\gamma_0 - 1) g(r_0)^n]^{-2}. \end{aligned}$$

Let $A_n = \mathbb{E}(X_1(n+1) - X_1(n) | \mathcal{F}_n)$. If $\mathfrak{A}$ holds, then

$$\begin{aligned} \sum_{T(n) < i < T(n+1)} A_i &= \sum_{T(n) < i < T(n+1)} [F(X_1(n)) - X_1(n)]/(n+2) \\ &< \frac{k}{k_1}(r_0^n - r_0^{n+1}) + \frac{k\gamma_0 r_0^n}{T(n)}. \end{aligned}$$

Note that if $\mathfrak{A}$ holds and $\tau_{n+1} \le T(n+1)$, then it must be

$$\begin{aligned} Z_{\tau_n,\tau_{n+1}} &= X_1(\tau_{n+1}) - X_1(\tau_n) - \sum_{i=\tau_n}^{\tau_{n+1}-1} A_i \\ &\ge r_0^n(1-r_0)(1-k/k_1) - \zeta_n - k\gamma_0 r_0^n/T(n), \end{aligned}$$

where ζ_n is bounded by $X_1(\tau_n) - X_1(\tau_n - 1) < 1/\tau_n < 1/T(n)$. By the discussion at the beginning of the proof, we know that $1/T(n) = o(r_0^{2n})$, so

$$l_0 := r_0^n(1-r_0)(1-k/k_1) - \zeta_n - k\gamma_0 r_0^n/T(n) \sim r_0^n(1-r_0)(1-k/k_1).$$

Then by the fact that $\mathbb{E}(Z^2_{\tau_n,\infty}|\tau_n > T(N)) < 1/T(n)$,

$$\begin{aligned} &\mathbb{P}(\tau_{n+1} \le T(n+1)|\tau_n > T(N)) \\ \le &\mathbb{P}(\mathfrak{A}^c|\tau_n > T(N)) \\ &+ \mathbb{P}\big(\mathfrak{A} \text{ and } \sup_{\tau_{n+1}} Z_{\tau_n,\tau_{n+1}} \ge l_0|\tau_n > T(N)\big) \\ \le &[(1-\gamma_0)g(r_0)^n]^{-2} + [(1-r_0)(1-k/k_1)g(r_0)^n]^{-2}. \end{aligned}$$

Taking the summation, we can get

$$\begin{aligned} &\mathbb{P}(\tau_n > T(n) \text{ for all } \quad n > M) \\ =&\mathbb{P}(\tau_M > T(M)) \prod_{n \ge M} \big(1 - \mathbb{P}(\tau_{n+1} \le T(n+1)|\tau_n > T(n))\big) > 0. \end{aligned}$$

This shows that $X_1(n)$ will eventually be $< p_0$. ■

In 1993, Benaïm and coauthors established a more comprehensive dynamical system approach to stochastic approximation. Since our focus is on urn models, here we will only state some related results.

In (3), suppose $\{\mathbf{X}_n\}$ have values in a compact set $\Delta \subset \mathbb{R}^d$ and $\mathbf{F}$ is a bounded $\mathcal{C}^m$ mapping with $m \geq 1$. Then the vector field of $\mathbf{F}$ can generate a $\mathcal{C}^m$ flow Φ in the following way:

$$\Phi : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$$

$$(t, \mathbf{X}) \mapsto \Phi_t(x)$$

defined by

$$\Phi_0 = Id$$

$$\frac{d\Phi_t(\mathbf{X})}{dt} = \mathbf{F}(\Phi_t(\mathbf{X})).$$

Then the asymptotic behavior of $\{\mathbf{X}_n\}$ can be approximated by the asymptotic behavior of Φ, which needs some topological definitions:

Definition 2.2.19. A subset $\Lambda \subset \mathbb{R}^d$ is an **attractor** of Φ if

(i) $\Lambda \neq \emptyset$ is compact and invariant, i.e. $\Phi_t(\Lambda) = \Lambda$;

(ii) There is a neighborhood $\mathcal{N}(\Lambda)$ of Λ such that as $t \to \infty$, the Euclidean distance $d(\Phi_t(\mathbf{x}), \Lambda) \to 0$ uniformly in $\mathbf{x} \in \mathcal{N}(\Lambda)$.

Applied to the urn processes in (4), Benaïm and Hirsch got:

Theorem 2.2.20 ([17], I)**.** *As in* (4)*, suppose* $\mathbf{F} : [0,1]^d \to [0,1]^d$ *is a* $\mathcal{C}^1$*. If*

(i) $\mathbf{F}([0,1]^d) \subset Int([0,1]^d)$*;*

(ii) The vector field $(\mathbf{F} - Id)$ *has an attractor* $\Lambda \subset Int([0,1]^d)$*.*

Then $\mathbb{P}(\{\lim \mathbf{X}_n\} \subset \Lambda) > 0$.

A better result can be derived for a special type of GPU model, called *Morse-Smale* urn, which means the vector field of the urn funciton $\mathbf{F}$ is **Morse-Smale**:

(1) All periodic orbits of $\mathbf{F}$ are hypernolic and all intersections of their stable and unstable manifolds are transverse;

(2) Every alpha or omega limit set is a periodic orbit;

(3) $\mathbf{F}$ is transverse to the boundary of the manifold where $\mathbf{F}$ is defined;

(4) $\mathbf{F}$ has a global attractor.

Theorem 2.2.21 ([17], II)**.** *Suppose the urn function* $\mathbf{F}$ *is* $\mathcal{C}^2$*. If*

(i) $\mathbf{F}([0,1]^d) \subset Int([0,1]^d)$*;*

(ii) $(\mathbf{F} - Id)$ *is a Morse-Smale vector field on* $\mathbb{R}^d$*,*

then

(1) For all the distinct hyperbolic periodic orbits Γ_i *of* $\mathbf{F}$*,* $\sum_i \mathbb{P}(\{\lim \mathbf{X}_n\} = \Gamma_i) = 1$*. In other words, the limits of* $\{\mathbf{X}_n\}$ *all fall in the hyperbolic periodic orbits of the urn function, and*

(2) $\mathbb{P}(\{\lim \mathbf{X}_n\} = \Gamma_i) > 0 \Leftrightarrow \Gamma_i$ *is linearly stable.*

Please see the original paper for proofs of the above theorems.

2.3 Embedding into Multitype Continuous Markov Branching Process (MCMBP)

A definition from Athreya and Ney's classical book on branching processes:

Definition 2.3.1. A stochastic process $\{\mathcal{X}(t) = (\mathcal{X}_1(t), ..., \mathcal{X}_d(t))\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **d-dimensional continuous time Markov branching process** if:

(i) Its state space is in $\mathbb{Z}^{d+} := \{\mathbf{x} = (x_1, ..., x_d) \in \mathbb{Z}^d | x^i \geq 0, i = 1, ...d\}$;

(ii) It is a strong Markov process with respect to $\mathcal{F}_t := \sigma(\mathcal{X}(s); s \leq t)$;

(iii) The transition probabilities $\mathbb{P}(\mathbf{i}, \mathbf{j}; t)$ satisfy

$$\sum_{\mathbf{j} \in \mathbb{Z}^{p+}} \mathbb{P}(\mathbf{i}, \mathbf{j}; t)\mathbf{s}^{\mathbf{j}} = \prod_{k=1}^{d} \big[\sum_{\mathbf{j} \in \mathbb{Z}^{d+}} \mathbb{P}(\mathbf{e}_k, \mathbf{j}; t)\mathbf{s}^{\mathbf{j}} \big]^{i_k} \tag{2.3.1}$$

where $\mathbf{i} = (i_1, ..., i_d) \in \mathbb{Z}^{d+}$, $\mathbf{s}$ is in the unit cube in $\mathbb{R}^{d+}$, $\mathbf{s}^{\mathbf{j}} = \prod_{k=1}^{d} s_k^{j_k}$ and $\{\mathbf{e}_k\}_{k=1}^{d}$ is the normal basis of $\mathbb{R}^d$, i.e. $\mathbf{e}_k = (0, ..., 1, ...0)$ with 1 in the k-th component.

In this definition, $\mathcal{X}_i(t)$ is referred to the number of particles of type i at time $t \geq 0$ and each particle has a random life time with exponential distribution. The lifetimes are independent. Hence the process is Markovian. Equation (11) characterizes its branching process property. Let

$$\mathbf{f}(\mathbf{s}) := (f_1(\mathbf{s}), ..., f_d(\mathbf{s})),$$

where

$$f_i(\mathbf{s}) = \sum_{\mathbf{j}\in\mathbb{Z}^{d+}} p_i(\mathbf{j})\mathbf{s}^{\mathbf{j}}$$

and $p_i(\mathbf{j})$ is the probability that a particle of type i produces j_k new particles of type k, $1 \le k \le d$. From a one dimensional branching process, it is not hard to see that for $\mathbf{i} = (0, ..., i_k, ..., 0)$ with the only non-zero i_k at the k-th component, $\sum_{\mathbf{j}} \mathbb{P}(\mathbf{i}, \mathbf{j}; t)\mathbf{s}^{\mathbf{j}} = \left(\sum_{\mathbf{j}} \mathbb{P}(\mathbf{e}_k, \mathbf{j}; t)\mathbf{s}^{\mathbf{j}}\right)^{i_k}$. Then the left hand side of (11) follows.

Associated with an urn process $\{\mathbf{X}(n)\}_{n\ge0}$ with a uniformly random drawing scheme, an MCMBP $\{\mathcal{X}(t) = (\mathcal{X}_1(t), ..., \mathcal{X}_d(t))\}_{t\ge0}$ can be defined with initial state $\mathcal{X}(0) = \mathbf{X}(0)$ and:

1) the lifetime of all types of particles, say $r_i \overset{\mathcal{D}}{\sim} \exp(1)$;

2) the probability generating function $\mathbf{f}(\mathbf{s})$ equals the p.g.f. of $\nu(\mathbf{R}(n))$ defined in Section 1.1;

3) all the r_i's and $\nu(\mathbf{R}(n))$ are independent.

Let $\{\tau_n\}_{n\ge0}$ with $\tau_0 = 0$ be the *split times* for this process. Then it is not hard to see that the two processes $\{\mathbf{X}(n)\}_{n\ge0}$ and $\{\mathcal{X}(\tau_n)\}_{n\ge0}$ are equivalent. So the works done for $\mathcal{X}(t)$ can be applied to $\mathbf{X}(n)$.

This method was introduced by Athreya and Karlin in 1968. It used the similarity between the drawing-replacing procedure in urn models and the splitting scheme in branching processes. As the authors noted, the Pólya urn model had been a basic

method to study branching process. On the other hand, branching processes can help "investigate the fluctuation behavior of urn scheme" by "exploiting properties of branching processes". Their main result is:

Theorem 2.3.2 ([7],I)**.** *For a d-dimensional MCMBP $\mathcal{X}(t)$, let A be its generating matrix (we may think of it as the replacement matrix for urn models). Suppose λ_1 is the largest eigenvalue of A, with eigenvector $\mathbf{v}_1$. For an eigenvector $\mathbf{v}$ of A with eigenvalue $\lambda \neq \lambda_1$. Define*

$$W\mathbf{u} := \lim_{t\to\infty} (X)(t)e^{-\lambda_1 t},$$

where W is some scalar random variable and $\mathbf{u}$ is a unit vector. And

$$Y_{\mathbf{v}}(t) := e^{\lambda t}\big(\mathbf{v}\cdot\mathcal{X}(t)\big), \qquad t \geq 0.$$

If $\lambda_1 > 2Re\lambda$ and the second moment of $\mathbf{f}$ exists, then

$$\lim_{n\to\infty} \mathbb{P}(0 < x_1 \leq W \leq x_2 < \infty, Y_{\mathbf{v}}(\tau_n)/\sqrt{\mathbf{u}\cdot\mathcal{X}(\tau_n)} \leq y) =$$

$$\mathbb{P}(0 < x_1 \leq W \leq x_2 < \infty|\mathcal{X}(0))\Phi(y/\sigma),$$

where

$$\Phi(x) := \frac{1}{2\pi}\int_{infty}^{x} e^{-t^2/2}dt$$

$$\sigma_i(t)^2 := e^{-\lambda_1 t}\mathbb{E}(Y_{\mathbf{v}}(t)^2|\mathcal{X}(0) = \mathbf{e}_i)$$

$$\sigma(t)^2 := \sum_{i=1}^{d} u_i\sigma_i(t)^2$$

$$\sigma := \lim_{t\to\infty} \sigma(t)^2.$$

As a corollary of this theorem, we can get a result for GPU models:

Corollary 2.3.3 ([7], II)**.** *As defined in section 1.1, let $A = (a_{ij})$ be the replacement matrix of a d-color Pólya process $\{\mathbf{X}(n)\}$. Suppose λ_1 is the largest eigenvalue of A and $\lambda \neq \lambda_1$ is another eigenvalue with eigenvector $\mathbf{v}$. If $2Re\lambda < \lambda_1$ and $\mathbb{E}(a_{ij}^2) < \infty$ for all i, j, then*

$$\frac{\mathbf{v} \cdot \mathbf{X}(n)}{\sqrt{n}} \xrightarrow{d} N(0, c), \text{ as } n \to \infty,$$

where c is a constant.

Until 2003, Janson used Athreya and Karlin's embedding method again. He proved some functional limit theorems for multitype branching processes, in which the GPU models can be embedded. As applications of the general theorems, we can get asymptotic normality theorems for $\{\mathbf{X}(n)\}$ as well as joint asymptotic normality for $(\mathbf{X}(n), \mathbf{N}(n))$, where $\mathbf{N}(n) = (N_1(n), ..., N_d(n))$ and $N_i(n) = \sum_{k=1}^{n} \mathbf{1}_{\{\xi_k = i\}}$. In [66], the author also gave a survey of related papers on GPU models, especially those getting similar results but using different methods, such as [108], [10], and [13], which we will introduce later.

In Janson's 2004's paper, he studied GPU models with two colors and triangular replacement matrices. So according to our notation, let $(R_1(n), R_2(n))$ be the composition of the urn at time n. The replacement matrix is:

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

Here, we allow $\alpha, \delta \geq -1$ and $\beta, \gamma \geq 0$. **Note:** in the 2-color case, when $\beta, \gamma > 0$, we call the replacement matrix *irreducible*. In Janson's paper, he assumed $\beta = 0$ and got:

Theorem 2.3.4 ([66], I)**.** *Suppose* $\delta, \gamma, R_2(0) > 0$.

(i) If $\alpha < \delta/2$, *then*

$$n^{-1/2}(R_2(n) - \delta\frac{\delta-\alpha}{\delta-\alpha+\gamma}n) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = \frac{\gamma\delta^3(\delta-\alpha)}{(\delta-2\alpha)(\delta-\alpha+\gamma)^2}$.

(ii) If $\alpha = \delta/2$, *then*

$$(n\ln n)^{-1/2}(R_2(n) - \delta\frac{\delta-\alpha}{\delta-\alpha+\gamma}n) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = \frac{\alpha\gamma\delta^2}{(\alpha+\gamma)^2}$.

(iii) If $\delta/2 < \alpha < \delta$, *then*

$$n^{-\alpha/\delta}(R_2(n) - \delta\frac{\delta-\alpha}{\delta-\alpha+\gamma}n) \xrightarrow{d} W := -\frac{\delta(\delta-\alpha)^{1+\alpha/\delta}}{\alpha(\delta-\alpha+\gamma)^{1+\alpha/\delta}}Z,$$

where Z *is a random variable with the characteristic function given by*

$$\mathbb{E}e^{itZ} = \frac{1}{\Gamma(R_2(0)/\delta)}\int_0^\infty (1 - i\alpha t x^{0-\alpha/\delta})^{-R_1(0)/\alpha} e^{g(t,x)} x^{R_2(0)/\delta - 1} dx, \qquad (\dagger)$$

with

$$g(t,x) := \int_x^\infty \varphi_2(-i\alpha t u^{-\alpha/\delta})du - i\frac{\gamma}{1-\alpha/\delta}tx^{1-\alpha/\delta} - x, \qquad (\ddagger)$$

where $\varphi_2(z) := (1+z)^{-\gamma/\alpha} - 1 + (\gamma/\alpha)z$.

(iv) If $\alpha = \delta$, *then*

$$\frac{\ln^2 n}{n}(R_2(n) - \frac{\alpha^2}{\gamma}\frac{n}{\ln n} - \frac{\alpha^2}{2}\frac{n\ln\ln n}{\ln^2 n}) \xrightarrow{d} W := \frac{\alpha^2}{\gamma^2}(\gamma\ln\frac{\gamma}{\alpha} + \gamma - \alpha - Z),$$

where Z *is a random variable with the characteristic function given by* (†) *with*

$$g(t,x) := \int_x^\infty \varphi_2(-i\alpha t u^{-1})du - i\gamma t\ln x - x.$$

(v) If $\alpha > \delta$, *then*

$$n^{\delta/\alpha}R_2(n) \xrightarrow{d} W := \delta\alpha^{\delta/\alpha}Z^{-\delta/\alpha},$$

where Z *is a random variable with the characteristic function given by* (†) *and* (‡) *or*

$$g(t,x) := \int_x^\infty \varphi_1(-i\alpha t u^{-\alpha/\delta})du - x$$

with $\varphi_1(z) := (1+z)^{-\gamma/\alpha} - 1$.

A special case on diagonal replacement matrices is

Theorem 2.3.5 ([66], II)**.** *Suppose* $\beta = \gamma = 0$, *and* $\alpha, \delta, R_1(0), R_2(0) > 0$. *Let* $U \sim \Gamma(R_1(0)/\alpha, 1)$ *and* $V \sim \Gamma(R_2(0)/\delta, 1)$ *be two independent random variables with Gamma distribution.*

(i) If $\alpha > \delta$, then

$$n^{-\alpha/\delta}(n\delta - R_2(0)) \xrightarrow{d} W := \delta U V^{-\alpha/\delta}$$

(ii) If $\alpha = \delta$, then

$$n^{-1}R_2(n) \xrightarrow{d} W := \delta \frac{V}{U+V}.$$

So W/δ has a Beta($R_2(0)/\delta$,$R_1(0)/\alpha$) distribution.

(iii) If $\alpha > \delta$, then

$$n^{-\delta/\alpha}R_2(n) \xrightarrow{d} W := \delta U^{-\delta/\alpha} V$$

Theorem 2.3.6 ([66], III)**.** *With the same assumption in Theorem 22, the moments of W can be given by,*

(i) If $\alpha < \delta$, then

$$\mathbb{E}W^s = \delta^s \frac{\Gamma(R_1(0)/\alpha + s)\Gamma(R_2(0)/\delta - s\alpha/\delta)}{\Gamma(R_1(0)/\alpha)\Gamma(R_2(0)/\delta)}, \qquad -R_1(0) < s\alpha < R_2(0).$$

(ii) If $\alpha = \delta$, then

$$\mathbb{E}W^s = \delta^s \frac{\Gamma((R_1(0) + R_2(0))/\alpha)\Gamma(R_2(0)/\delta + s)}{\Gamma(R_2(0)/\delta)\Gamma((R_1(0) + R_2(0) + s\delta)/\alpha)}, \qquad s \geq 0.$$

(iii) If $\alpha > \delta$, then

$$\mathbb{E}W^s = \delta^s \frac{\Gamma(R_1(0)/\alpha - s\delta/\alpha)\Gamma(R_2(0)/\delta + s)}{\Gamma(R_1(0)/\alpha)\Gamma(R_2(0)/\delta)}, \qquad -R_2(0) < s\delta < R_1(0).$$

If we consider the behavior of the urn after a fixed number of draws of color-2, let $X^{(m)}$ be the number of color-1 balls in the urn when m color-2 balls have been

drawn. Janson gave the probability generating function of $X^{(m)}$, as well as a limit theorem for it:

Theorem 2.3.7 ([66], IV)**.** *Suppose $\beta = 0$ and $\delta, R_2(0) > 0$. Let $m \to \infty$:*

(i) If $\alpha < \delta/2$, then

$$m^{-1/2}(X^{(m)} - \frac{\gamma}{1-\alpha/\delta}m) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = \frac{\alpha^2\gamma\delta(\delta+\gamma-\alpha)}{(\delta-2\alpha)(\delta-\alpha)^2}$.

(ii) If $\alpha = \delta/2$, then

$$(m \ln m)^{-1/2}(X^{(m)} - \frac{\gamma}{1-\alpha/\delta}m) \xrightarrow{d} N(0, \gamma(\gamma+\alpha)).$$

(iii) If $\delta/s < \alpha < \delta$, then

$$m^{-\alpha/\delta}(X^{(m)} - \frac{\gamma}{1-\alpha/\delta}m) \xrightarrow{d} Z,$$

where Z is the same as in Theorem(Janson, 2004, I) (iii).

(iv) If $\alpha = \delta$, then

$$m^{-1}(X^{(m)} - \gamma m \ln m) \xrightarrow{d} Z,$$

where Z is the same as in Theorem(Janson, 2004, I) (iv).

(v) If $\alpha > \delta$, then

$$m^{-\alpha/\delta}X^{(m)} \xrightarrow{d} Z$$

where Z is the same as in Theorem(Janson, 2004, I)(v).

In the 2004 paper, Janson also discussed the case in which the number of balls of one color stays fixed.

Theorem 2.3.8 ([66], V)**.** *Suppose $\beta = \delta = 0$.*

(i) If $\alpha > 0$, $\gamma > 0$ and $R_2(0) > 0$, then $\{R_1(n)\}$ is a persistent irreducible Markov chain with period $\gamma/|\alpha| + 1$. Assume $\alpha = -1$. Then

$$R_1(n) \equiv R_1(0) - n \mod (\gamma + 1).$$

For $\forall k \geq 0$,

$$\mathbb{P}(R_1(n) = k) - \mathbf{1}_{\{k \equiv R_1(0) - n \mod (\gamma+1)\}} \sum_{j=k-\gamma}^{k} \mathbb{P}(W = k) \to 0,$$

where W is a compound Poisson distributed random variable with probability generating function

$$\mathbb{E}(z^W) = \exp\big(R_2(0) \sum_{j=1}^{\gamma} \frac{z^j - 1}{j}\big).$$

(ii) If $\alpha = 0$, then

$$n^{-1/4}(R_1(n) - \sqrt{2\gamma R_2(0)} n^{1/2}) \xrightarrow{d} N(0, \sqrt{2\gamma^3 R_2(0)}/3).$$

(iii) If $\alpha > 0$ and either $R_1(0) > 0$ or $\gamma > 0$, then

$$(\ln n)^{-1/2}\big(R_1(n) - \alpha n - \frac{\gamma - \alpha}{\alpha} R_2(0) \ln n\big) \xrightarrow{d} N\big(0, \frac{(\gamma - \alpha)^2}{\alpha} R_2(0)\big).$$

A simple version of Athreya and Karlin's embedding method will be discussed more in Section 5.

2.4 Some Other Methods and Results

For completeness and convenience to apply, we state some theorems about GPU models which people obtained when they needed them for applications. In fact, many of them can be found in a more general version in Janson's paper [66].

In 1985, Bagchi and Pal gave a sufficient condition for the asymptotic normality behavior of balanced 2-color GPU models. And the replacement matrix should satisfy: 1) $\alpha+\beta=\gamma+\delta\geq 1$; 2) $\alpha\neq\gamma$; 3) $\beta,\gamma\geq 0$ and 4) If $\alpha<0$ then α divides γ and $R_1(0)$. Similarly, if $\delta<0$ then δ divides β and $R_2(0)$. An urn scheme satisfying these conditions is called a *tenable* urn. Intuitively, an urn is *tenable* if we can not draw more balls of a certain type than are present in the urn. By a method of moment, they got that:

Theorem 2.4.1 ([9]). *If* $\frac{\alpha-\gamma}{\alpha+\beta}\leq 2$ *and* $\alpha-\gamma\neq 0$, *then*

$$\mathbb{P}(\frac{R_1(n)-\mathbb{E}(R_1(n))}{\sigma(R_1(n))}<x)\to\Phi(x),$$

where $\sigma(R_1(n))$ *is the standard deviation of* R_n *and* $\Phi(x)$ *is the standard normal distribution.*

Motivated by this paper and with the same assumptions, Gouet (1993) [53] used the martingale technique to find some functional central limit theorems for $R_1(n)$:

Theorem 2.4.2 ([53], I). *In a tenable urn process with* $\beta\gamma>0$,

(i) If the ratio of eigenvalues of A, $\rho := (\alpha - \gamma)/(\alpha + \beta) > 1/2$, then

$$n^{-1/2}(R_1([nt]) + (nt)^{\rho}(b_{[nt]} - Z)) \Rightarrow t^{1-\rho}W \circ \varphi(t),$$

where $W(t)$ is the Wiener process, $\varphi(t) = \beta\gamma/(2\rho - 1)(\rho/(1-\rho))^2 t^{2\rho-1}$, Z is a nondegenerate random variable independent of W and b_k is defined by $a_n := \Gamma((R_1(n) + R_2(n))/(\alpha + \beta))/\Gamma(\rho + T_n/(\alpha + \beta))$ and $b_n := -\gamma \sum_{k=1}^{n} a_k$.

(ii) If $\rho = 1/2$, then

$$(n^t \ln n)^{-1/2}(R_1([n^t]) - \frac{\gamma}{\gamma + \beta}(R_1([n^t]) + R_2([n^t]))) \Rightarrow W \circ \varphi(t)$$

where $\varphi(t) = \beta\gamma t$.

(iii) If $\rho < 1/2$, then

$$n^{-1/2}(R_1([nt]) - \frac{\gamma}{\gamma + \beta}(R_1([nt]) + R_2([nt]))) \Rightarrow t^{\rho}W \circ \varphi(t),$$

where $\varphi(t) = \beta\gamma/(1 - 2\rho)(\rho/(1-\rho))^2 t^{1-2\rho}$.

Gouet also considered the triangular replacement matrix case:

Theorem 2.4.3 ([53], II)**.** *In a tenable urn process, suppose $\beta\gamma = 0$ and $max(\beta, \gamma) > 0$. Then*

$$n^{-\rho/2}(R_1([nt^{1/\rho}]) - n^{\rho}tZ) \Rightarrow W \circ \varphi(t),$$

where $\varphi(t) = \alpha Z t$ and Z is a nondegenerate positive random variable independent of W.

In 1996, Smythe generalized Bagchi and Pal's results and gave some central limit theorems for GPU models under conditions on eigenvalues and eigenvectors of *generating matrices*. By *generating matrix*, we mean if the entries in the replacement matrix are random numbers, i.e. the urn model is under a random environment, then take $\tilde{A} = [\tilde{a}_{ij}]$ with $\tilde{a}_{ij} = \mathbb{E}^{(i)}(a_{ij})$ where $\mathbb{E}^{(i)}$ is the expectation conditional on the drawing of s color i ball. For application purpose, the author also defined a type of GPU, *extended polya urns*(EPU): drawn balls not necessarily replaced, i.e. the diagonal elements in the replacement matrix could be negative. Several assumptions are needed for EPU's considered in this paper:

(1) $\tilde{a}_{ij} \geq 0$ and $\sum_{i=1}^{d} \tilde{a}_{ij}$ =a positive constant;

(2) $\mathbb{E}^{(i)}[a_{ij}^2] < \infty$ for all i, j;

(3) The maximal eigenvalue of $\tilde{A}$ is simple and has a strictly positive left eigenvector.

The asymptotic normality for both $d = 2$-color and $d > 2$-color cases are gotten:

Theorem 2.4.4 ([108])**.** *When $d = 2$, let λ_0 be the maximal eigenvalue of $\tilde{A}$. Suppose any other eigenvalue λ satisfies $2\lambda < \lambda_0$. Also, let ξ be a right eigenvector of the nonprincipal eigenvalue of $\tilde{A}$. Then*

(i) $(\xi \cdot X_1(n))/\sqrt{n}$ *and* $(\xi \cdot N_n)/\sqrt{n}$ *are each asymptotic normal.*

(ii) If $\xi \cdot (X_1(n+1) - X_1(n))$ *is* $\mathbb{P}^{(i)}$*-nondegenerate for* $i = 1, 2$*, where* $\mathbb{P}^{(i)}$ *is the*

probability conditional on a draw of color i, then $n^{-1/2}(\xi \cdot X_1(n), \xi \cdot N_n)$ has an asymptotic bivariate normal distribution.

(iii) Let $w = (w_1, w_2)$ be the left eigenvector of λ_0 and λ be the other eigenvalue of $\tilde{A}$. If for any nonzero vector $v = (v_1, v_2)$, $v \cdot (X_{n+1} - X_n)$ is not constant almost surely for $\mathbb{P}^{(i)}$, $i = 1, 2$, then $n^{-1/2}(X_{n1} - n\lambda_0 w_1, X_{n2} - n\lambda w_2)$ is asymptotically bivariate normal.

(iv) If for any nonzero vector $v = (v_1, v_2)$, $v \cdot (X_{n+1} - x_n)$ is $\mathbb{P}^{(i)}$-nondegenerate for $i = 1$ or 2, then $n^{-1/2}(X_{n1} - n\lambda_0 w_1, X_{n2} - n\lambda_0 w_2, N_{n1} - nw_1)$ is asymptotically trivariate normal.

The case for $d > 2$ needs the consideration of complex eigenvalues and eigenvectors. A few more assumptions are added. But the results are parallel, so we won't state them here.

Bai and Hu [10] generalized Smythe's results for nonhomogeneous generating matrices. Let $\tilde{A}_n = [\tilde{a}_{ij}^n] := [\mathbb{E}(a_{ij}^n)]$ to be the generating matrix at time n. If $\tilde{A}_n = \tilde{A}$ for all n, then we call the GPU model *homogeneous*. Athreya and Karlin [7] and Smythe [108] have all studied homogeneous GPU models. But in practical applications, the generating matrices are usually nonhomogeneous, like in the adaptive clinical trials. They kept the first one in Smythe's assumptions and supposed that there exists a positive regular matrix H such that $\sum_{n=1}^{\infty} \|\tilde{A}_n - H\|_\infty / n < \infty$, i.e. the generating matrices, though different from each other, are still close to some

H. Then they showed that

Theorem 2.4.5 ([10], I)**.** *$n^{-\lambda}(\mathbb{E}X_n)$ converges to a constant vector $\vec{v}$ of nonnegative entries, where λ is the unique maximal eigenvalue of H and $\vec{v}$ is the left eigenvalue of H corresponding to λ.*

Theorem 2.4.6 ([10], II)**.** *$n^{-\lambda}X_n$ converges in probability to $\vec{v}$ in the above theorem.*

Under a few more assumptions, they also proved the asymptotic normalities:

Theorem 2.4.7 ([10], III)**.** *Let τ be the maximum of zero and the real parts of all the non-maximal eigenvalues of H. If $\tau < 1/2$, then $n^{-1/2}(X_n - \mathbb{E}X_n)$ is asymptotically normal with mean vector 0. If $\tau = 1/2$, then $n^{-1/2}(\lg n)^{\nu-1/2}(X_n - \mathbb{E}X_n)$ is asymptotically normal with mean vector 0, where ν is the maximal order of the Jordan forms of H.*

These results apply for the homogeneous case.

Later, in 2002, Bai, Hu and Zhang gave the weak and strong approximation for $\{R_n\}$ by a Gaussian process for both homogeneous and nonhomogeneous in the 2-color GPU models based on Gouet's theorems.

In Athreya and Karlin's paper [7], they suggested an open question about the asymptotic normality of N_n. Bai and Hu [14] gave an answer to this question as well as an explicit formula for the asymptotic variance of N_n and showed the consistency of both X_n and N_n for both homogeneous and nonhomogeneous generating matrices.

There are more results by Bai and Hu, etc.. Please see the papers listed in the reference.

Kotz, Mahmoud and Robert [70] considered the case of 2-colored GPU's with nonnegative and fixed replacement matrix elements, not necessarily balanced. They calculated the exact form of the distribution of $\{N_{ni}\}$ by calculating some recurrence relations of the urn process.

Theorem 2.4.8 ([70]). *For* $1 \leq r \leq n$,

$$\mathbb{P}(N_{n1} = r) = \sum_{1 \leq i_1 < i_2 < \ldots < i_r \leq n} \prod_{k=1}^{r} (\phi(i_k, k) \prod_{i_{k-1} < j < i_k} (1 - \phi(j, k))),$$

where $\phi(x, y) = \frac{N_{01} + \beta(x-y) + \delta(y-1)}{N_{02} + N_{01} + (\alpha+\beta)(x-y) + (\gamma+\delta)(y-1)}$.

Motivated by understanding the $2-3$ tree, Flajolet, Gabarró and Pekari studied the balanced 2-colored GPU models, especially with negative diagonal entries in the replacement matrix. So the urn models are assumed to be tenable. First, they got the probability generating function of R_n:

Theorem 2.4.9 ([45], I). *Suppose the replacement matrix is of the form* $\begin{pmatrix} -\alpha & \alpha + s \\ \beta + s & -\beta \end{pmatrix}$, *with* $\alpha, \beta > 0$. *Then the probability generating function at time* n *of the urn's composition is*

$$p_n(u) := \frac{\Gamma(n+1)\Gamma((R_0 + B_0)/s)}{s^n \Gamma(n + \frac{(R_0+B_0)}{s})}([z^n]H(z, u)),$$

where, $[z^n]H(z,u)$ *means the coefficient of* z^n *in* H *and*

$$H(z,u) = \delta(u)^{R_0+B_0}\psi(z\delta(u)^s + I(u))$$

with $\delta(u) := (1-u^{\alpha+\beta+s})^{1/(\alpha+\beta+s)}$, $I(u) := \int_0^u \frac{t^{\alpha-1}}{\delta(t)^{\alpha+\beta}}dt$ *and* $\psi(I(u)) = \frac{u^{R_0}}{\delta(u)^{R_0+B_0}}$.

They also gave the Gaussian law and the convergence speed of X_{n1}, as well as calculation of the moments:

Theorem 2.4.10 ([45], II)**.** *For a GPU with negative diagonal entries in the replacement matrix,* R_n *is asymptotically Gaussian with speed of convergence to the limit* $O(n^{-1/2})$.

Theorem 2.4.11 ([45], 2004, III)**.** *For a GPU with negative diagonal entries in the replacement matrix and any* $r > 0$*, the* r*-th factorial moment of the distribution of* R_n *is of hypergeometric type: it's a finite linear combination of terms of the form*

$$\frac{\binom{n+(R_0+B_0)/s+l-k(\alpha+\beta+s)/s-1}{n}}{\binom{n+(R_0+B_0)/s-1}{n}}$$

with $0 \le l, k \le r$.

Remark 2.4.12. Note that Flajolet, Gabarró and Pekari devised a relatively new and purely analytic method. It starts with trying to find a combinatorial structure of the model and to find a PDE for the exponential generating function(egf) for the probability generating function(pgf) of the GPU. By solving the PDE, we can find an expression of the pgf at time n as the coefficient of the egf. Then it needs to build a complex-analytic structure with the complex Fermat curve and connect

the nature of the urn model with a conformal map, which leads to some probability consequences, like large deviations, convergence speed and moments. The analytic method seems to be able to produce plenty of results. However it needs the requirements of a balanced urn and can be applied mainly to 2-colored GPU's and some special cases of 3-colors, according to the authors.

Remark 2.4.13. Similar to the analytic method, Inoue and Aki [61] used another method based on the conditional probability generating functions to calculate the exact joint distribution of the number of balls with particular colors which are drawn in the first n draws. Their method can be applied to GPU models with $d \geq 2$ colors and nonrandom replacement matrices with nonnegative entries.

Chapter 3

A Close Look at the Monopoly Processes

This chapter studies a non-linear GPU model, which exhibits a monopoly property. What is different from the original Pólya urn process is the drawing probability, which depends not only on the current numbers of balls of different colors in the urn, but also on some given nonlinear functions, called *feedback functions*. The intuition of a feedback function is the same as an urn function. So we use $F(x) \geq 0$ as well for feedback function. With the setup in Section 1.1, the probability of drawing a ball of color j at time n is

$$\mathbb{P}(\text{color } j \text{ is drawn}|\mathcal{F}_n) = \frac{F(R_j(n))}{\sum_{k=1}^{d} F(R_k(n))}, \qquad j = 1, ..., d. \tag{3.0.1}$$

It is well known that when $F(x)$ satisfies certain conditions, monopoly happens, i.e. after finite times only one color can be drawn out. Hence if the system does

not stop, the monopoly color will have infinitely many balls and there are only a finite number of balls in the other colors. The simplest example of such a feedback function is $F(x) = x^\alpha$ with $\alpha > 1$. In this paper, we require that $F(x)$ should satisfy the following monopoly conditions:

(C1) $F(x) > 0$ for all $x > 0$;

(C2) $F(x) \geq F(y)$ if $x > y$;

(C3) $\sum_{n=1}^{\infty} 1/F(n) < \infty$.

Without loss of generality, in this paper we always assume the initial composition in the urn is $(1, ..., 1)$.

The asymptotic behaviors of the type of monopoly processes defined before are studied in the following two aspects:

1. The final number of balls of the minority colors in the urn.

2. After what time are all draws the same color?

3.1 Exponential Embedding Method

One of our main tools is the *exponential embedding* method, invented by Herman Robin and introduced by Burgess Davis in 1990 [28].

Define independent random variables $\{r_k(n), n \geq 1\}_{k=1}^{d}$. Each $r_k(n)$ is exponentially distributed with mean $1/F(n)$. For all k, define

$$\Sigma_k(N) := \sum_{n=1}^{N} r_k(n), \qquad N \geq 1.$$

Now write all of these $\{\Sigma_k(N)\}$ in an increasing order, say $t_1 < t_2 < ....$ Define another random variable sequence $\{\gamma_i\}_{i\geq 1}$ according to this order: if $t_i = \Sigma_{k_i}(N)$ for some $1 \leq k_i \leq d$ and $N \geq 1$, then let $\gamma_i = \xi_{k_i}$, $i = 1, 2, ...$, where ξ_{k_i} means the k_i-th color.

Theorem 3.1.1 ([28], Rubin's Theorem)**.** *The sequence $\{\gamma_i\}$ defined above is equivalent in distribution to the sampling color sequence to our nonlinear Polya urn process.*

This proof of this theorem can be found in [28]. Here we will state it in a more general pattern.

Proof: First, we will use the fact that if $X_1, ..., X_m$ $(m \geq 2)$ are independent exponential random variables with parameter $\lambda_1, ..., \lambda_m$, respectively, then $\min(X_1, ..., X_m)$ is also exponential with parameter $\lambda_1 + ... + \lambda_m$. Secondly, for two independent exponential random variables $X_1 \sim \exp(\lambda_1)$ and $X_2 \sim \exp(\lambda_2)$, $\mathbb{P}(X_1 < X_2) = \lambda_1/(\lambda_1 + \lambda_2)$.

By induction, in the $\{\gamma_i\}$ sequence when $i = 1$, since the initial values are all

one:

$$\begin{aligned}\mathbb{P}(\gamma_1 = \xi_{k_1}|\mathcal{F}_0) =& \mathbb{P}\big(r_{k_1}(1) < \min_{k\neq k_1}\{r_k(1)\}\big)\\ =& \frac{F(1)}{d\cdot F(1)}.\end{aligned}$$

Now suppose before γ_n, each Σ_k has been added up to N_k. Then

$$\begin{aligned}\mathbb{P}(\gamma_n = \xi_{k_n}) =& \mathbb{P}\big(r_{k_n}(N_{k_n}+1) < \min_{k\neq k_n}\{r_k(N_k+1)\}\big)\\ =& \frac{F(N_{k_n})}{\sum_{k=1}^{d} F(N_k)}.\end{aligned}$$

■

3.2 Summation of independent exponential random variables

In this section, we prove some combinatorial identities involving symmetric functions. These identities will be applied to simplify the computations in later sections.

Our method depends on the density function of the summation of independent exponential random variables. This density function can be found in Feller's book [41], but it does not provide proof.

Lemma 3.2.1. *For any sequence of independent exponentially distributed random variables* $\{X_1, ..., X_n\}$ $(n \geq 2)$*, with distinct parameters* $\lambda_1, ..., \lambda_n$ $(\lambda_i > 0)$*, respec-*

tively, the density function of their summation $\sum_{j=1}^{n} X_j$ *is*

$$f_n(x) = \Big[\prod_{j=1}^{n} \lambda_j\Big] \sum_{k=1}^{n} \frac{e^{-\lambda_k x}}{\prod\limits_{\substack{l=1 \\ l\neq k}}^{n} (\lambda_l - \lambda_k)}, \qquad x \geq 0. \tag{3.2.1}$$

The first method:

Proof: When $n = 2$, by convolution

$$f_2(x) = \lambda_1 \lambda_2 \frac{e^{-\lambda_1 x} - e^{-\lambda_2 x}}{\lambda_2 - \lambda_1}.$$

Suppose it is true for $n-1$. Then from convolution, we can get

$$f_n(x) = \Big[\prod_{j=1}^{n} \lambda_j\Big] \sum_{k=1}^{n-1} \frac{e^{-\lambda_k x} - e^{-\lambda_n x}}{\prod\limits_{\substack{l=1 \\ l\neq k}}^{n} (\lambda_l - \lambda_k)}$$

In order to finish the proof, we need to prove

$$-\sum_{k=1}^{n-1} \frac{1}{\prod\limits_{\substack{l=1 \\ l\neq k}}^{n} (\lambda_l - \lambda_k)} = \frac{1}{\prod\limits_{l=1}^{n-1} (\lambda_l - \lambda_n)}$$

Or equivalently,

$$\sum_{k=1}^{n} \frac{1}{\prod\limits_{\substack{l=1 \\ l\neq k}}^{n} (\lambda_l - \lambda_k)} = 0$$

By some calculation, this is equivalent to

$$\sum_{k=1}^{n} (-1)^k \prod_{\substack{1\leq j<l\leq n \\ l\neq k \\ j\neq k}} (\lambda_l - \lambda_j) = 0 \tag{3.2.2}$$

Note that the product is a Vandermonde determinant:

$$\begin{vmatrix} 1 & \lambda_1 & \lambda_1^2 & ... & \lambda_1^{n-2} \\ ... & ... & ... & ... & ... \\ 1 & \lambda_{k-1} & \lambda_{k-1}^2 & ... & \lambda_{k-1}^{n-2} \\ 1 & \lambda_{k+1} & \lambda_{k+1}^2 & ... & \lambda_{k+1}^{n-2} \\ ... & ... & ... & ... & ... \\ 1 & \lambda_n & \lambda_n^2 & ... & \lambda_n^{n-2} \end{vmatrix}$$

From this determinant, it is not hard to see that (3.2.1) is the determinant of

$$\begin{pmatrix} 1 & 1 & \lambda_1 & ... & \lambda_1^{n-2} \\ ... & ... & ... & ... & ... \\ 1 & 1 & \lambda_k & ... & \lambda_k^{n-2} \\ ... & ... & ... & ... & ... \\ 1 & 1 & \lambda_n & ... & \lambda_n^{n-2} \end{pmatrix},$$

which is obviously zero. ■

The second method to prove Lemma 3.2.1 is to use characteristic function, which makes the proof much simpler. We use it in the following lemma for the infinite sum of independent exponential random variables, with some constraint.

Lemma 3.2.2. *Let $\{X_1, X_2, ...\}$ be an infinite sequence of independent exponentially distributed random variables with distinct parameters $\lambda_j > 0$, $j \geq 1$. Suppose*

$\sum_{j=1}^{\infty} \frac{1}{\lambda_j} < \infty$. *Then the density function of* $\sum_{j=1}^{\infty} X_j$ *is*

$$f_\infty(x) = \Big[\prod_{j=1}^{\infty} \lambda_j\Big] \sum_{k=1}^{\infty} \frac{e^{-\lambda_k x}}{\prod_{\substack{l\geq 1 \\ l\neq k}} (\lambda_l - \lambda_k)}, \qquad x \geq 0. \tag{3.2.3}$$

Proof: The characteristic function of $\sum_{j=1}^{n} X_j$, $n \geq 1$, is

$$\begin{aligned} \phi_n(t) &= \prod_{j=1}^{n} \mathbb{E} e^{itX_j} \\ &= \prod_{j=1}^{n} \frac{1}{1 - \frac{it}{\lambda_j}}. \end{aligned}$$

Since $\sum_{j=1}^{\infty} \frac{1}{\lambda_j} < \infty$, $\{\phi_n(t)\}_{n\geq 1}$ converges to

$$\phi_\infty(t) = \prod_{j=1}^{\infty} \frac{1}{1 - \frac{it}{\lambda_j}}.$$

Note that the norm

$$\Big| \prod_{j=1}^{n} \frac{1}{1 - \frac{it}{\lambda_j}} \Big| = \prod_{j=1}^{n} \frac{1}{\sqrt{1 + \frac{t^2}{\lambda_j^2}}}$$

decreases as n increases. Hence

$$\int |\phi_\infty(t)| dt < \infty.$$

By the Fourier inversion, the density function of $\sum_{j\geq 1} X_j$ is

$$f_\infty(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_\infty(t) dt.$$

Now $\phi_\infty(t)$ has simple poles at $-i\lambda_j$, $j \geq 1$. And the residue of $e^{-itx}\phi_\infty(t)$ at $-i\lambda_j$ is

$$\text{res}(-i\lambda_j) = i\lambda_j e^{-\lambda_j x} \prod_{k\neq j} \frac{\lambda_k}{\lambda_k - \lambda_j}.$$

Finally the residue theorem gives us the explicit form of the density function. ■

3.2.1 Probability proof for some combinatorial identities

Summation of independent exponential random variables provides a much simpler way to prove some algebraic equations. Here we will state both the probabilistic method and the algebraic method, for comparison. These results are stated for the finite case, but are in fact true for the infinite one as well.

Corollary 3.2.3. *For any distinct $\lambda_1, ..., \lambda_n$, $n \geq 2$ and $\lambda_i > 0$, there is*

$$\sum_{k=1}^{n} \prod_{l\neq k} \frac{\lambda_l}{\lambda_l - \lambda_k} = 1. \tag{3.2.4}$$

Proof: **The probability method** is to construct a sequence of independent exponential random variables $\{X_k\}_{k=1}^n$ with distinct parameters λ_k. The integral of any density function is 1. From Lemma 3.2.1,

$$\begin{aligned}\int_0^\infty f_n(x)dx &= \Big[\prod_{j=1}^{n} \lambda_j\Big] \sum_{k=1}^{n} \frac{\int_0^\infty e^{-\lambda_k x}dx}{\prod_{\substack{l=1\\ l\neq k}}^{n}(\lambda_l - \lambda_k)} \\ &= \sum_{k=1}^{n} \prod_{l\neq k} \frac{\lambda_l}{\lambda_l - \lambda_k} = 1\end{aligned}$$

■

Proof: **The algebraic method** will use the Vandermonde determinant. Observe that (3.2.4) is true if and only if

$$\sum_{k=1}^{n}(-1)^{k-1}[\prod_{l\neq k}\lambda_l]\prod_{\substack{t\neq k\\ l<t}}(\lambda_t-\lambda_l)=\prod_{1\leq l<t\leq n}(\lambda_t-\lambda_l) \tag{3.2.5}$$

The left hand side of (3.2.5) is exactly the determinant of the following Vandermonder matrix, whose determinant has a formula in the form of the right hand side.

$$\begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{pmatrix}$$

■

Corollary 3.2.4. *For any* $\lambda_1, ..., \lambda_n$, $n \geq 2$, $\lambda_i > 0$,

$$\sum_{k=1}^{n}\frac{1}{\lambda_k}\prod_{\substack{l=1\\ l\neq k}}^{n}\frac{\lambda_l}{\lambda_l-\lambda_k}=\sum_{k=1}^{n}\frac{1}{\lambda_k} \tag{3.2.6}$$

Proof(**The probabilistic method**): same construction as in the previous proof. The expected value of $\{X_k\}_{k=1}^n$ is

$$\mathbb{E}(\sum_{k=1}^{n}X_k)=\sum_{k=1}^{n}\frac{1}{\lambda_k}.$$

On the other hand, from the density function (3.2.1) of $\sum_{k=1}^n X_k$, the expectation

is

$$\mathbb{E}(\sum_{j=1}^{n} X_j) = \int_0^\infty \Big[\prod_{j=1}^{n} \lambda_j\Big] \sum_{k=1}^{n} \frac{xe^{-\lambda_k x}}{\prod\limits_{\substack{l=1\\ l\neq k}}^{n} (\lambda_l - \lambda_k)} dx$$

$$= \sum_{k=1}^{n} \prod_{\substack{l=1\\ l\neq k}}^{n} \frac{\lambda_l}{\lambda_l - \lambda_k}$$

■

Proof (**The algebraic method**): (3.2.6) is true if and only if

$$\frac{\sum\limits_{k=1}^{n} (-1)^{k-1} \Big[\prod\limits_{l\neq k} \lambda_l^2\Big] \prod\limits_{\substack{t\neq k\\ l<t}} (\lambda_t - \lambda_l)}{\Big[\prod\limits_{k=1}^{n} \lambda_k\Big] \prod\limits_{1\leq l<t\leq n} (\lambda_t - \lambda_l)} = \frac{\sum\limits_{k=1}^{n} \prod\limits_{l\neq k} \lambda_l}{\prod\limits_{k=1}^{n} \lambda_k} \tag{3.2.7}$$

And (3.2.7) is equivalent to

$$\sum_{k=1}^{n} (-1)^{k-1} \Big[\prod_{l\neq k} \lambda_l^2\Big] \prod_{\substack{t\neq k\\ l<t}} (\lambda_t - \lambda_l) = \Big[\prod_{1\leq l<t\leq n} (\lambda_t - \lambda_l)\Big] \sum_{k=1}^{n} \prod_{l\neq k} \lambda_l \tag{3.2.8}$$

The left hand side of (3.2.8) is the determinant

$$D_1 := \begin{vmatrix} 1 & \lambda_1^2 & \lambda_1^3 & \dots & \lambda_1^n \\ 1 & \lambda_2^2 & \lambda_2^3 & \dots & \lambda_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_n^2 & \lambda_n^3 & \dots & \lambda_n^n \end{vmatrix}.$$

This is in fact a generalized Vandermonde matrix. Let us look at another (but

not the only) way to compute its determinant. Construct the following matrix:

$$\begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \dots & \lambda_n^n \\ 1 & x & x^2 & \dots & x^n \end{pmatrix}, \tag{3.2.9}$$

whose determinant is a polynomial of x, say $p(x)$. And it is easy to see that

$$p(x) = C_{n+1,1} + C_{n+1,2}x + ... + C_{n+1,n+1}x^n, \tag{3.2.10}$$

where C_{ij} is the cofactor. First, obviously $\lambda_1, ..., \lambda_n$ are the roots of $p(x) = 0$. Then from matrix (3.2.9), $C_{n+1,2} = (-1)^{n+3}D_1$ and

$$C_{n+1,n+1} = \prod_{1 \le l < t \le n} (\lambda_t - \lambda_l).$$

Next, by Viète's formula,

$$\begin{aligned} \sum_{k=1}^{n} \prod_{l \neq k} \lambda_l &= (-1)^{n-1} \frac{C_{n+1,2}}{C_{n+1,n+1}} \\ &= \frac{D_1}{\prod_{1 \le l < t \le n} (\lambda_t - \lambda_l)}. \end{aligned}$$

Hence, D_1 equals the right hand side of (3.2.8). ■

More generally, a corollary of the algebraic proof of Corollary 3.2.4:

Corollary 3.2.5. *For any $\lambda_1, ..., \lambda_n$, $n \geq 2$, $\lambda_i > 0$, let*

$$D_k := \begin{vmatrix} 1 & \lambda_1 & \dots & \lambda_1^{k-1} & \lambda_1^{k+1} & \dots & \lambda_1^n \\ 1 & \lambda_2 & \dots & \lambda_2^{k-1} & \lambda_2^{k+1} & \dots & \lambda_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{k-1} & \lambda_n^{k+1} & \dots & \lambda_n^n \end{vmatrix}, \tag{3.2.11}$$

where $1 \leq k \leq n-1$ (when $k = n$, it is the original Vandermonde determinant). Then

$$D_k = \prod_{1 \leq l < t \leq n} (\lambda_t - \lambda_l) \sum_{\substack{i_j = 1 \\ i_j \neq i_{j'} \text{ if } j \neq j'}}^{n} \prod_{j=1}^{n-k} \lambda_{i_j}. \tag{3.2.12}$$

Proof: In matrix (3.2.9) and the determinant polynomial (3.2.10), the coefficient of x^k is $C_{n+1,k+1}$, which is also $(-1)^{n+k+2} D_k$. By Viète's formula,

$$(-1)^{n-k} \frac{C_{n+1,k+1}}{C_{n+1,n+1}} = \sum_{\substack{i_j = 1 \\ i_j \neq i_{j'} \text{ if } j \neq j'}}^{n} \prod_{j=1}^{n-k} \lambda_{i_j}.$$

■

Corollary 3.2.6. *For any $\lambda_1, ..., \lambda_n$, $n \geq 2$, $\lambda_i > 0$, there is*

$$\sum_{k=1}^{n} \frac{1}{\lambda_k^2} \prod_{\substack{l=1 \\ l \neq k}}^{n} \frac{\lambda_l}{\lambda_l - \lambda_k} = \sum_{k=1}^{n} \frac{1}{\lambda_k^2} + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \frac{1}{\lambda_i \lambda_j}$$

Proof (**the probability method**): same construction of independent exponential random variables as before, then the second moment of their summation

is

$$\mathbb{E}\Big[\big(\sum_{k=1}^{n} X_k\big)^2\Big] = \mathbb{E}\Big(\sum_{k=1}^{n} X_k^2 + 2\sum_{k=1}^{n} X_i X_j\Big)$$
$$= \sum_{k=1}^{n} \frac{2}{\lambda_k^2} + 2\sum_{k=1}^{n} \frac{1}{\lambda_i \lambda_j}$$

On the other hand, from the density function (3.2.1),

$$\mathbb{E}\Big[\big(\sum_{k=1}^{n} X_k\big)^2\Big] = \int_0^\infty \Big[\prod_{j=1}^{n} \lambda_j\Big] \sum_{k=1}^{n} \frac{x^2 e^{-\lambda_k x}}{\prod\limits_{\substack{l=1 \\ l\neq k}}^{n} (\lambda_l - \lambda_k)}$$
$$= \sum_{k=1}^{n} \frac{2}{\lambda_k^2} \prod_{l\neq k} \frac{\lambda_l}{\lambda_l - \lambda k}.$$

■

Proof (**the algebraic method**): First,

$$\sum_{k=1}^{n} \frac{1}{\lambda_k^2} \prod_{l\neq k} \frac{\lambda_l}{\lambda_l - \lambda_k} = \frac{\sum\limits_{k=1}^{n} [\prod\limits_{l\neq k} \lambda_l^3](-1)^{k-1} \prod\limits_{\substack{1\leq l<t\leq n \\ t,l\neq k}} (\lambda_t - \lambda_l)}{[\prod\limits_{k=1}^{n} \lambda_k^2] \prod\limits_{1\leq l<t\leq n} (\lambda_t - \lambda_l)}. \tag{3.2.13}$$

Note that the top is the determinant

$$D := \begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1^3 & \lambda_2^3 & \dots & \lambda_n^3 \\ \lambda_1^4 & \lambda_2^4 & \dots & \lambda_n^4 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n+1} & \lambda_2^{n+1} & \dots & \lambda_n^{n+1} \end{vmatrix}.$$

In order to evaluate this determinant, set up a function

$$\mathcal{D}(x) := \begin{vmatrix} e^{x/\lambda_1^2} & e^{x/\lambda_2^2} & \dots & e^{x/\lambda_n^2} \\ \lambda_1 e^{x/\lambda_1^2} & \lambda_2 e^{x/\lambda_2^2} & \dots & \lambda_n e^{x/\lambda_n^2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1} e^{x/\lambda_1^2} & \lambda_2^{n-1} e^{x/\lambda_2^2} & \dots & \lambda_n^{n-1} e^{x/\lambda_n^2} \end{vmatrix}$$

$$= e^{\sum_{k=1}^{n} \frac{1}{\lambda_k^2} x} \begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_n^{n-1} \end{vmatrix}$$

Then the derivative of $\mathcal{D}(x)$ is

$$\frac{d\mathcal{D}}{dx} = \begin{vmatrix} \frac{1}{\lambda_1^2}e^{x/\lambda_1^2} & \frac{1}{\lambda_2^2}e^{x/\lambda_2^2} & \dots & \frac{1}{\lambda_n^2}e^{x/\lambda_n^2} \\ \lambda_1 e^{x/\lambda_1^2} & \lambda_2 e^{x/\lambda_2^2} & \dots & \lambda_n e^{x/\lambda_n^2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1}e^{x/\lambda_1^2} & \lambda_2^{n-1}e^{x/\lambda_2^2} & \dots & \lambda_n^{n-1}e^{x/\lambda_n^2} \end{vmatrix}$$

$$+ \begin{vmatrix} e^{x/\lambda_1^2} & e^{x/\lambda_2^2} & \dots & e^{x/\lambda_n^2} \\ \frac{1}{\lambda_1}e^{x/\lambda_1^2} & \frac{1}{\lambda_2}e^{x/\lambda_2^2} & \dots & \frac{1}{\lambda_n}e^{x/\lambda_n^2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1}e^{x/\lambda_1^2} & \lambda_2^{n-1}e^{x/\lambda_2^2} & \dots & \lambda_n^{n-1}e^{x/\lambda_n^2} \end{vmatrix}$$

$$+ \begin{vmatrix} e^{x/\lambda_1^2} & e^{x/\lambda_2^2} & \dots & e^{x/\lambda_n^2} \\ \lambda_1 e^{x/\lambda_1^2} & \lambda_2 e^{x/\lambda_2^2} & \dots & \lambda_n e^{x/\lambda_n^2} \\ e^{x/\lambda_1^2} & e^{x/\lambda_2^2} & \dots & e^{x/\lambda_n^2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1}e^{x/\lambda_1^2} & \lambda_2^{n-1}e^{x/\lambda_2^2} & \dots & \lambda_n^{n-1}e^{x/\lambda_n^2} \end{vmatrix} + \dots$$

$$+ \begin{vmatrix} e^{x/\lambda_1^2} & e^{x/\lambda_2^2} & \dots & e^{x/\lambda_n^2} \\ \lambda_1 e^{x/\lambda_1^2} & \lambda_2 e^{x/\lambda_2^2} & \dots & \lambda_n e^{x/\lambda_n^2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-3}e^{x/\lambda_1^2} & \lambda_2^{n-3}e^{x/\lambda_2^2} & \dots & \lambda_n^{n-3}e^{x/\lambda_n^2} \end{vmatrix}$$

Hence

$$\frac{d\mathcal{D}}{dx} = \begin{vmatrix} \frac{1}{\lambda_1^2}e^{x/\lambda_1^2} & \frac{1}{\lambda_2^2}e^{x/\lambda_2^2} & \dots & \frac{1}{\lambda_n^2}e^{x/\lambda_n^2} \\ \lambda_1 e^{x/\lambda_1^2} & \lambda_2 e^{x/\lambda_2^2} & \dots & \lambda_n e^{x/\lambda_n^2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1}e^{x/\lambda_1^2} & \lambda_2^{n-1}e^{x/\lambda_2^2} & \dots & \lambda_n^{n-1}e^{x/\lambda_n^2} \end{vmatrix}$$

$$+ \begin{vmatrix} e^{x/\lambda_1^2} & e^{x/\lambda_2^2} & \dots & e^{x/\lambda_n^2} \\ \frac{1}{\lambda_1}e^{x/\lambda_1^2} & \frac{1}{\lambda_2}e^{x/\lambda_2^2} & \dots & \frac{1}{\lambda_n}e^{x/\lambda_n^2} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-1}e^{x/\lambda_1^2} & \lambda_2^{n-1}e^{x/\lambda_2^2} & \dots & \lambda_n^{n-1}e^{x/\lambda_n^2} \end{vmatrix}$$

After some adjustment, we get

$$\frac{d\mathcal{D}}{dx} = \Big[\prod_{k=1}^{n}\frac{1}{\lambda_k^2}\Big]e^{\sum_{k=1}^{n}x/\lambda_k^2}\begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1^3 & \lambda_2^3 & \dots & \lambda_n^3 \\ \lambda_1^4 & \lambda_2^4 & \dots & \lambda_n^4 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n+1} & \lambda_2^{n+1} & \dots & \lambda_n^{n+1} \end{vmatrix}$$

$$- \Big[\prod_{k=1}^{n}\frac{1}{\lambda_k}\Big]e^{\sum_{k=1}^{n}x/\lambda_k^2}\begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \lambda_1^3 & \lambda_2^3 & \dots & \lambda_n^3 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^n & \lambda_2^n & \dots & \lambda_n^n \end{vmatrix} \tag{3.2.14}$$

By Corollary 3.2.5,

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \lambda_1^3 & \lambda_2^3 & \dots & \lambda_n^3 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^n & \lambda_2^n & \dots & \lambda_n^n \end{vmatrix} = \prod_{1\le l<t\le n}(\lambda_t-\lambda_l)\Big[\prod_{k=1}^{n}\lambda_k\Big]\sum_{i\neq j}\frac{1}{\lambda_i\lambda_j}$$

Besides, the derivative also equals

$$\frac{d\mathcal{D}}{dx} = \Big(\sum_{k=1}^{n}\frac{1}{\lambda_k^2}\Big)e^{\sum_{k=1}^{n}x/\lambda_k^2}\prod_{1\le l<t\le n}(\lambda_t-\lambda_l). \tag{3.2.15}$$

Equaling (3.2.12) and (3.2.13), we get

$$D = \Big[\prod_{k=1}^{n}\lambda_k^2\Big]\Big(\sum_{k=1}^{n}\frac{1}{\lambda_k^2}+\sum_{i\neq j}\frac{1}{\lambda_i\lambda_j}\Big)\prod_{1\le l<t\le n}(\lambda_t-\lambda_l). \tag{3.2.16}$$

Plugging (3.2.14) back to (3.2.11) proves the identity. ■

The following corollary is a general case of the previous one.

Corollary 3.2.7. *For any $\lambda_1, ..., \lambda_n$, $n \geq 2$, $m \geq 2$ and $\lambda_i > 0$, there is*

$$\sum_{k=1}^{n} \frac{1}{\lambda_k^m} \prod_{l \neq k} \frac{\lambda_l}{\lambda_l - \lambda_k} = \sum_{\substack{i_1+...+i_n=m \\ i_1,...,i_n \geq 0}} \frac{1}{\lambda_1^{i_1} ... \lambda_n^{i_n}}$$

Proof (**the probability proof**): similar construction of independent exponential random variables, then compute the m-th moment of $\sum X_k$.

$$\begin{aligned} \mathbb{E}\Big[(\sum_{k=1}^{n} X_k)^m\Big] &= \mathbb{E}\Big[\sum \binom{m}{i_1 \ ... \ i_n} X_1^{i_1} ... X_n^{i_n}\Big] \\ &= \sum_{i_1+...+i_n=m} \frac{m!}{\lambda_1^{i_1} ... \lambda_n^{i_n}} \end{aligned}$$

On the other hand, by integrating the density function eqrefequ 3.2.1, we have

$$\mathbb{E}\Big[(\sum_{k=1}^{n} X_k)^m\Big] = \sum_{k=1}^{n} \frac{m!}{\lambda_k^m} \prod_{l \neq k} \frac{\lambda_l}{\lambda_l - \lambda_k}$$

■

Compared to the probability proof, the algebraic method is too complicated to prove Corollary 3.2.7. As a consequence, it deduces a generalized Vandermonde determinant.

Corollary 3.2.8.

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1^{m+1} & \lambda_2^{m+1} & \dots & \lambda_n^{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{n-m-1} & \lambda_2^{n-m-1} & \dots & \lambda_n^{n-m-1} \end{vmatrix} = \prod_{1\le l<t\le n} (\lambda_t - \lambda_l) \sum_{\substack{i_1+\dots+i_n=m \\ i_1,\dots,i_n\ge 0}} \frac{\prod_{k=1}^n \lambda_k^m}{\lambda_1^{i_1}\dots\lambda_n^{i_n}}$$

Corollary 3.2.9. *For any $\lambda_1, ..., \lambda_n$, $n \ge 2$, $\lambda_i > 0$ and any $s \ne -\lambda_k$ (all $k = 1, ..., n$),*

$$\sum_{k=1}^n \frac{\lambda_k}{\lambda_k + s}\Big[\prod_{l\ne k} \frac{\lambda_l}{\lambda_l - \lambda_k}\Big] = \prod_{k=1}^n \frac{\lambda_k}{\lambda_k + s}$$

Or

$$\sum_{k=1}^n \frac{1}{\lambda_k + s}\Big[\prod_{l\ne k} \frac{1}{\lambda_l - \lambda_k}\Big] = \prod_{k=1}^n \frac{1}{\lambda_k + s}$$

Proof (**the probability method**): construct independent exponential random variables $\{X_k\}$ with parameter λ_k, then the Laplace transform of $\sum X_k$ is

$$\begin{aligned} \mathcal{L}(s) &= \mathbb{E}\big(e^{-s\sum_{k=1}^n X_k}\big) \\ &= \prod_{k=1}^n \mathbb{E}\big(e^{-sX_k}\big) \\ &= \prod_{k=1}^n \frac{\lambda_k}{\lambda_k + s} \end{aligned}$$

If we apply the density function (3.2.1),

$$\mathcal{L}(s) = \int_0^\infty \Big[\prod_{j=1}^n \lambda_j\Big] \sum_{k=1}^n \frac{e^{-sx}e^{-\lambda_k x}}{\prod\limits_{\substack{l=1\\ l\neq k}}^n (\lambda_l - \lambda_k)} dx$$

$$= \sum_{k=1}^n \frac{\lambda_k}{\lambda_k + s}\Big[\prod_{l\neq k} \frac{\lambda_l}{\lambda_l - \lambda_k}\Big]$$

■

Proof (**the algebraic method**):

$$\sum_{k=1}^n \frac{1}{\lambda_k + s}\Big[\prod_{l\neq k} \frac{1}{\lambda_l - \lambda_k}\Big] = \frac{\sum\limits_{k=1}^n \prod\limits_{l\neq k} (\lambda_l + s)(-1)^{k-1} \prod\limits_{\substack{1\leq l<k\leq n\\ l,t\neq k}} (\lambda_t - \lambda_l)}{\prod\limits_{k=1}^n (\lambda_k + s) \prod\limits_{1\leq l<t\leq n} (\lambda_t - \lambda_l)}$$

The numerator is a Vandermonde determinant

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 + s & \lambda_2 + s & \dots & \lambda_n + s \\ \vdots & \vdots & \vdots & \vdots \\ (\lambda_1 + s)^{n-1} & (\lambda_2 + s)^{n-1} & \dots & (\lambda_n + s)^{n-1} \end{vmatrix} = \prod_{1\leq l<t\leq n} (\lambda_t - \lambda_l)$$

■

In fact, Corollary 3.2.9 and its general case have a much simpler proof, as a consequence of Corollary 3.2.7.

Corollary 3.2.10. *For distinct* $\lambda_1, ..., \lambda_n$, $n \geq 2$, $\lambda_i > 0$ *and* $m \geq 1$, $s \neq \lambda_i$,

$$\sum_{k=1}^{n} \frac{1}{(\lambda_k + s)^m} \prod_{l \neq k} \frac{1}{\lambda_l - \lambda_k} = \frac{1}{\prod_{j=1}^{n}(\lambda_j + s)} \sum_{\substack{i_1+...+i_n=m-1 \\ i_1,...,i_n \geq 0}} \frac{1}{(\lambda_1 + s)^{i_1} ... (\lambda_n + s)^{i_n}}$$

Proof: Let $\tilde{\lambda_j} = \lambda_j + s$, $j = 1, ..., n$. Apply Corollary 3.2.7. ■

Recall that when the parameters are all the same, the summation of m independent exponential random variables with the parameter β has a density function

$$g_m(y) = \beta^m \frac{y^{m-1}}{(m-1)!} e^{-\beta y}, \qquad y \geq 0. \tag{3.2.17}$$

This is a $\Gamma(m, \beta)$ distribution.

The following lemma is a general formula.

Lemma 3.2.11. *Suppose* $\{X_1, ..., X_n, Y_1, ..., Y_m\}$, $n \geq 1$, $m \geq 2$, *are independent exponentially distributed random variables. The* X_j*'s have distinct parameters* λ_j, $j = 1, ..., n$. *And all the* Y_k*'s have the same parameter* β, $\beta \neq \lambda_j$ *for all* j. *Then the density function of* $\sum_{j=1}^{n} X_j + \sum_{k=1}^{m} Y_k$ *is*

$$h_{n,m}(s) = \beta^m e^{-\beta s} \Big[\prod_{j=1}^{n} \lambda_j \Big] \sum_{k=1}^{n} \left\{ \frac{1}{\prod_{\substack{l=1 \\ l \neq k}}^{n} (\lambda_l - \lambda_k)} \Big[\frac{e^{(\beta - \lambda_k)s}}{(\beta - \lambda_k)^m} - \sum_{l=1}^{m} \frac{s^{m-l}}{(m-l)!(\beta - \lambda_k)^l} \Big] \right\},$$

for $s \geq 0$.

Proof: By the convolution of the density functions (3.2.1) and (3.2.17),

$$\begin{aligned} h_{n,m}(s) &= \int_0^s f_n(s-y) g_m(y) dy \\ &= [\prod_{j=1}^{n} \lambda_j] \sum_{k=1}^{n} \frac{e^{-\lambda_k s}}{\prod_{l \neq k}(\lambda_l - \lambda_k)} \frac{\beta^m}{(m-1)!} \int_0^s e^{(\lambda_k - \beta)y} y^{m-1} dy. \end{aligned}$$

The integral

$$\begin{aligned}&\int_0^s e^{(\lambda_k-\beta)y}y^{m-1}dy\\ =&\frac{1}{(\lambda_k-\beta)}e^{(\lambda_k-\beta)s}s^{m-1}-\frac{m-1}{(\lambda_k-\beta)}\int_0^s e^{(\lambda_k-\beta)y}y^{m-2}dy\end{aligned}$$

Then, induction leads to the final formula. ■

3.3 Monopoly

Our first goal is to show that monopoly happens almost surely. The proof is for general feedback functions $F(x)$ and number of colors $K \geq 2$. And we use the exponential embedding method.

In the proof, we will need a concept *attraction time*, whose distribution will be discussed in later chapter. Let $\{X_n\}_{n\geq 1}$ be the drawing color sequence. Define the *attraction time* T_a by

$$T_a := \min\{k : X_{n+1} = X_n, \forall n \geq k\}.$$

Theorem 3.3.1. *If the feedback function $F(x)$ of a nonlinear Polya urn process satisfies conditions (C1)-(C3), then monopoly happens almost surely, i.e.*

$$\mathbb{P}(T_a < \infty) = 1.$$

Proof: By Theorem 3.1.1, after the exponential embedding, the time until the number of balls of color k, $k = 1, ..., K$, gets to infinity is

$$\Sigma_k(\infty) := \sum_{n=1}^{\infty} r_k(n).$$

Obviously,

$$\mathbb{E}(\Sigma_k(\infty)) = \sum_{n=1}^{\infty} 1/F(n) < \infty.$$

So $\mathbb{P}(\Sigma_k(\infty) < \infty) = 1$. Furthermore, all the variables $\Sigma_k(\infty)$ have continuous distribution, which implies that they are distinct with probability one. Let

$$\Sigma_{k_0}(\infty) = \min_{1 \le k \le K} \{\Sigma_k(\infty)\}.$$

Then for each $k \neq k_0$, there is a finite number $n(k)$ such that

$$\Sigma_k(n(k)) < \Sigma_{k_0}(\infty) < \Sigma_k(n(k)+1).$$

Let

$$T_a := \max_{k \neq k_0} \{n(k)\}.$$

Obviously, after time T_a only balls of color k_0 arrive. ■

There is another method to prove the existence of monopoly. It is based on the stochastic approximation method, which is an important tool for the study of reinforced random processes.

Some History: W. B. Arthur is the first one to introduce this type of nonlinear Polya urn processes to model the "lock-in" phenomena, i.e. monopoly, in economics. He got the idea from a paper by B. M. Hill, D. Lane and W. Sudderth in 1980, who first generalized Polya urn models via *urn functions* which are essentially the same to feedback functions. An urn function $f(x)$ is a continuous map from $[0,1]$ to $[0,1]$ (this is for the two-color case and can be generalized to higher dimensional).

Hill, at all proved that the proportion of one-color balls in the urn will converge almost surely to a fixed point of $f(x)$. A fixed point x_0 is a point satisfying the condition $f(x_0) = x_0$. In 1990, R Pemantle improved on their results by proving the nonconvergence to unstable fixed points and giving conditions on when convergence happens at a touchpoint. Interested readers please read the previous chapter, section 4.2, Theorem 8 to 11 or their original papers.

Briefly, the relation between urn functions and feedback functions is straight forward. Take the two-color case for example. Suppose $X_1(n)$ is the proportion of one color balls in the urn at time n, then $f(X_1(n)) = u(R_1(n))/[u(R_1(n)) + u(R_2(n))]$. From this, we can easily see that 0 and 1 are two fixed points for $f(x)$ and in fact are stable. Hence, finally $X_1(n)$ will converge to one of them, which means monopoly happens.

3.4 The Number of Balls of The Minority Color

First, let us look at the distribution of the number of balls of the minority color in the urn. These results currently are for two-color case and general $F(x)$. The special cases $F(x) = x^\alpha$, $\alpha > 1$ will be stated as corollaries. We will first discuss how many balls of the minority color wins in the end, or after its rival gets final monopoly. Then we will also consider the probability behavior of the minority color during the process of competition, or when it is not clear yet who will finally win,.

Note that during the process, the minority color is not fixed.

Definition 3.4.1. Let χ_k be the number of balls of the minority color at time k, i.e.

$$\chi_k = \min\{r_1(k), r_2(k)\}, \qquad k \geq 1.$$

Let χ_∞ be the number of balls of the minority color in the end. Then

Definition 3.4.2.

$$\chi_\infty = \lim_{k\to\infty} \min\{r_1(k), r_2(k)\}.$$

The following theorem estimates the tail distribution of χ_∞.

Lemma 3.4.3. *Let*

$$a_1(n) = \Big[\prod_{k=1}^{n} F(k)^2\Big] \sum_{j=1}^{n} \Big[\frac{1}{F(j)} \prod_{\substack{k=1\\k\neq j}}^{n} \frac{1}{F(k)^2 - F(j)^2}\Big].$$

Then

$$a_1(n) \sim \frac{1}{\sqrt{\pi n \xi_n}},$$

where $\xi_n \in (0, \frac{1}{F(1)^2})$ *depends on* F *and* n. *An upper and lower bound is*

$$\sqrt{\frac{1}{\pi \sum \frac{1}{F(j)}}} \leq a_1(n) \leq \sqrt{\frac{2}{\pi \sum \frac{1}{F(j)}}}$$

Proof: Here we will apply the Lagrange interpretation. Let

$$\lambda_j = \frac{1}{F(j)^2}, \qquad j \geq 1$$

and

$$f_n(x) = x^{n-\frac{1}{2}}, \qquad x \in [0, \lambda_1].$$

Give n distinct points $\{\lambda_1, ..., \lambda_n\}$, define the interpolation polynomial

$$p_n(x) = \sum_{k=1}^{n} \lambda_k^{n-\frac{1}{2}} \prod_{l \neq k} \frac{x - \lambda_l}{\lambda_k - \lambda_l}.$$

Then the interpolation error is

$$f_n(x) - p_n(x) = \frac{f_n^{(n)}(\xi_n)}{n!} \prod_{j=1}^{n} (x - \lambda_j), \qquad x \in [0, \lambda_1],$$

where $\xi_n \in (0, \lambda_1)$ depends on x. If we choose $x = 0$, then

$$(-1)^n \sum_{k=1}^{n} \frac{1}{F(k)^{2n-1}} \prod_{l \neq k} \frac{F(k)^2}{F(l)^2 - F(k)^2} = \frac{f_n^{(n)}(\xi_n)}{n!} (-1)^n \prod_{j=1}^{n} \frac{1}{F(j)^2}.$$

Hence

$$a_1(n) = \frac{f_n^{(n)}(\xi_n)}{n!}.$$

Besides,

$$\begin{aligned} f_n^{(n)}(\xi_n) &= (n - \frac{1}{2})...(1 - \frac{1}{2}) \xi_n^{-\frac{1}{2}} \\ &= \frac{(2n-1)!!}{2^n} \xi_n^{-\frac{1}{2}} \\ &= \frac{(2n)!}{2^{2n} n!} \xi_n^{-\frac{1}{2}}. \end{aligned}$$

By Stirling's formula, when $n \to \infty$,

$$a_1(n) \sim \frac{1}{\sqrt{\pi n \xi_n}}. \tag{3.4.1}$$

To estimate the bounds of $a_1(n)$, we need its integral representation, which will be shown in Theorem 3.4.4.

$$a_1(n) = \frac{1}{\pi} \int_{-\infty}^{\infty} \phi_n(t) dt,$$

where

$$\phi_n(t) = \prod_{j=1}^{n} \frac{1}{1 + \frac{t^2}{F(j)^2}}.$$

By

$$e^{-x} \leq \frac{1}{1+x} \leq e^{-x/2}$$

and

$$\sum \frac{1}{F(j)} < \infty,$$

we can get

$$\sqrt{\frac{1}{\pi \sum \frac{1}{F(j)}}} \leq a_1(n) \leq \sqrt{\frac{2}{\pi \sum \frac{1}{F(j)}}} \tag{3.4.2}$$

The form (3.4.1) is a simple representation of $a_1(n)$. On the other hand, (3.4.2) gives an estimation to the error point ξ_n. ■

Theorem 3.4.4. *As $n \to \infty$,*

$$\mathbb{P}(\chi_\infty \geq n) \sim a_1(n) \sum_{j=n+1}^{\infty} \frac{1}{F(j)},$$

where $a_1(n)$ is as defined in Lemma 3.4.3.

Proof: By the exponential embedding scheme we set up at the beginning, it is not hard to see that

$$\begin{aligned}\{\chi_\infty \geq n\} =&\Big\{\sum_{j=1}^{n} r_2(j) < \sum_{j=1}^{\infty} r_1(j) < \sum_{j=1}^{\infty} r_2(j)\Big\} \bigcup \Big\{\sum_{j=1}^{n} r_1(j) < \sum_{j=1}^{\infty} r_2(j) < \sum_{j=1}^{\infty} r_1(j)\Big\} \\ =&\Big\{-\sum_{j=n+1}^{\infty} r_1(j) < \sum_{j=1}^{n}(r_1(j) - r_2(j)) < \sum_{j=n+1}^{\infty}(r_2(j) - r_1(j))\Big\} \\ &\bigcup\Big\{-\sum_{j=n+1}^{\infty} r_2(j) < \sum_{j=1}^{n}(r_2(j) - r_1(j)) < \sum_{j=n+1}^{\infty}(r_1(j) - r_2(j))\Big\}\end{aligned}$$

Now, let

$$\begin{aligned}\Delta_n &:= \sum_{j=1}^{n}\big(r_1(j) - r_2(j)\big) \\ \Gamma_1(n+1) &:= \sum_{j=n+1}^{\infty} r_1(j) \\ \Gamma_2(n+1) &:= \sum_{j=n+1}^{\infty} r_2(j)\end{aligned}$$

Obviously, $\Gamma_1(n+1)$ and $\Gamma_2(n+1)$ are independent and identically distributed. Hence,

$$\begin{aligned}\{\chi_\infty \geq n\} =&\{-\Gamma_1(n+1) < \Delta_n < \Gamma_2(n+1) - \Gamma_1(n+1)\} \\ &\cup \{-\Gamma_2(n+1) < -\Delta_n < \Gamma_1(n+1) - \Gamma_2(n+1)\} \\ =&\{|\Delta_n| < \Gamma_1(n+1)\}\end{aligned}$$

So

$$\mathbb{P}(\chi_\infty \geq n) = \mathbb{P}(|\Delta_n| < \Gamma_1(n+1))$$

In order to estimate this probability, let us first look at the distribution of Δ_n.

The characteristic function of Δ_n is

$$\begin{aligned}\phi_n(t) &= \mathbb{E}(e^{it\Delta_n}) \\ &= \prod_{j=1}^{n} \frac{1}{1-\frac{it}{F(j)}} \times \prod_{j=1}^{n} \frac{1}{1+\frac{it}{F(j)}} \\ &= \prod_{j=1}^{n} \frac{1}{1+\frac{t^2}{F(j)^2}}\end{aligned}$$

Now we may apply the inversion formula to get the distribution of Δ_n, which is obviously continuous. For any $a \leq b$,

$$\lim_{T\to\infty} (2\pi)^{-1} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi_n(t)dt = \mathbb{P}(a \leq \Delta_n \leq b)$$

Suppose the density functions of Δ_n and $\Gamma_1(n+1)$ are $g(x)$ and $h(y)$, respectively. By independence,

$$\begin{aligned}\mathbb{P}(|\Delta_n| < \Gamma_1(n+1)) &= \int_0^\infty \int_{-y}^{y} g(x)h(y)dxdy \\ &= \int_0^\infty \mathbb{P}(-y \leq \Delta_n \leq y)h(y)dy \\ &= \int_0^\infty \Big[\frac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \phi_n(t) \frac{e^{ity} - e^{-ity}}{it} dt\Big] h(y)dy \\ &= \frac{1}{\pi} \int_0^\infty \Big[\int_{-\infty}^{\infty} \phi_n(t) \frac{\sin(ty)}{t} dt\Big] h(y)dy\end{aligned}$$

Applying the Taylor expansion of $\sin(ty)$,

$$\sin(ty) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} (ty)^{2k+1}.$$

Hence,

$$\mathbb{P}(|\Delta_n| < \Gamma_1(n+1)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \phi_n(t) \Big[\int_0^\infty yh(y)dy - \frac{1}{3!}t^2 \int_0^\infty y^3 h(y)dy + ...\Big] dt$$

First, note that

$$\int_{-\infty}^{\infty} t^k \phi_n(t) dt < \infty,$$

for all $k \geq 0$ and

$$\mathbb{E}\big[(\sum_{j \geq n+1} r_1(j))^m\big] = o\big(\mathbb{E}\big[\sum_{j \geq n+1} r_1(j)\big]\big),$$

for all $m \geq 2$.

Secondly, let us look at the integral,

$$\begin{aligned}\int_{-\infty}^{\infty} \phi_n(t) dt &= \int_{-\infty}^{\infty} \prod_{k=1}^{n} \frac{1}{1 + t^2/F(k)^2} dt \\ &= \int_{-\infty}^{\infty} \prod_{k=1}^{n} \frac{F(k)^2}{(F(k) + it)(F(k) - it)} dt.\end{aligned}$$

Here, $\{\pm iF(k)\}_{k=1}^{n}$ are simple poles of $\phi_n(t)$. And the residue of $iF(k)$ is

$$\mathrm{res}_{\phi_n}(iF(k)) = \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{F(j)^2}{F(j)^2 - F(k)^2} \cdot \frac{F(k)}{2i}.$$

So, by residue theorem

$$\int_{-\infty}^{\infty} \phi_n(t) dt = \pi\Big[\prod_{k=1}^{n} F(k)^2\Big] \sum_{k=1}^{n} \Big[\frac{1}{F(k)} \prod_{\substack{j=1 \\ j \neq k}}^{n} \frac{1}{F(j)^2 - F(k)^2}\Big].$$

Hence,

$$\mathbb{P}(|\Delta_n| < \Gamma_1(n+1)) \sim a_1(n) \sum_{j \geq n+1} \frac{1}{F(j)}.$$

■

Corollary 3.4.5. *For $F(x) = x^\alpha$, $\alpha > 1$,*

$$\mathbb{P}(\chi_\infty \geq n) \sim \frac{\tilde{a}_1(n)}{\alpha - 1}\Big(\frac{1}{n^{\alpha-1}}\Big),$$

where

$$\tilde{a}_1(n) \sim \frac{1}{\sqrt{\pi n \tilde{\xi}_n}}, \qquad \tilde{\xi}_n \text{ depends on } \alpha \text{ and } n.$$

and

$$\sqrt{\frac{1}{\pi \sum \frac{1}{j^\alpha}}} \leq \tilde{a}_1(n) \leq \sqrt{\frac{2}{\pi \sum \frac{1}{j^\alpha}}}.$$

Proof: This is a direct consequence of the theorem and the fact that as $n \to \infty$,

$$\sum_{x=n+1}^{\infty} \frac{1}{x^\alpha} \sim \frac{1}{(\alpha - 1)n^{\alpha-1}}.$$

■

Another way: There is another way to estimate the tail distribution of the number of the minor color balls, by using the useful? density function in Lemma 1. This method is more straightforward and much simpler. And both methods give the same asymptotic estimation.

First, obviously for any n, $f_n(x)$ is bounded by $\min\{\lambda_1, ..., \lambda_n\}$.

Lemma 3.4.6. *For any two independent random variables X and Y with density function $f(x)$ and $g(y)$ respectively. If $h(z)$ is the density function of $X + Y$, then*

$$\sup h(z) \leq \min\{\sup f(x), \sup g(y)\}$$

Proof:

$$h(z) = \int f(z-y)g(y)dy \leq \sup f(x) \int g(y)dy = \sup f(x).$$

Similarly, $\sup h(z) \leq \sup g(y)$. ■

Theorem 3.4.7.

$$\mathbb{P}(\chi_\infty \geq n) \sim a_1(n) \sum_{j=n+1}^{\infty} \frac{1}{F(j)},$$

where $a_1(n)$ is as defined in Lemma 3.4.3.

Proof: Let's look at $\mathbb{P}(\sum_{j=1}^{n} r_2(j) < \sum_{j=1}^{\infty} r_1(j) < \sum_{j=1}^{\infty} r_2(j))$. Suppose the density function of $\sum_{j=1}^{\infty} r_1(j)$ is $f_\infty(x)$ and the joint density function of $y = \sum_{j=1}^{n} r_2(j)$ and $z = \sum_{j=1}^{\infty} r_2(j)$ is $g(y,z)$. By the boundedness of $f_n(x)$ and independence, we have

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{j=1}^{n} r_2(j) < \sum_{j=1}^{\infty} r_1(j) < \sum_{j=1}^{\infty} r_2(j)\Big) \\
&= \int_0^\infty \int_0^z \int_y^z f_\infty(x)g(y,z)dxdydz \\
&= \lim_{n\to\infty} \int_0^\infty \int_0^z \Big[\int_y^z f_n(x)\Big] g(y,z)dxdydz \\
&= \lim_{n\to\infty} \int_0^\infty \int_0^z \left\{ \Big[\prod_{j=1}^{n} F(j)\Big] \sum_{k=1}^{n} \frac{(e^{-F(k)y} - e^{-F(k)z})/F(k)}{\prod_{\substack{l=1 \\ l\neq k}}^{n} (F(l) - F(k))} \right\} g(y,z)dydz. \qquad (3.4.3)
\end{aligned}$$

Inside the summation, the double integral gives

$$\begin{aligned}\mathbb{E}\big(e^{-F(k)\sum_{j=1}^{n} r_2(j)} - e^{-F(k)\sum_{j=1}^{\infty} r_2(j)}\big) &= \prod_{j=1}^{n} \frac{F(j)}{F(j)+F(k)} - \prod_{j=1}^{\infty} \frac{F(j)}{F(j)+F(k)} \\ &= \prod_{j=1}^{n} \frac{F(j)}{F(j)+F(k)} \Big(1 - \prod_{j=n+1}^{\infty} \frac{F(j)}{F(j)+F(k)}\Big).\end{aligned}$$

Next, we will show that

$$1 - \prod_{j=n+1}^{\infty} \frac{F(j)}{F(j)+F(k)} \sim F(k) \sum_{j=n+1}^{\infty} \frac{1}{F(j)}.$$

Since $F(x)$ is increasing,

$$\begin{aligned}\prod_{j=n+1}^{\infty} \frac{F(j)}{F(j)+F(k)} &= \prod_{j=n+1}^{\infty} \frac{1}{1+\frac{F(k)}{F(j)}} \\ &= \prod_{j=n+1}^{\infty} \Big[\sum_{s\geq 0} (-\frac{F(k)}{F(j)})^s\Big] \\ &= 1 - F(k) \sum_{j=n+1}^{\infty} \frac{1}{F(j)} + o\big(\sum_{j\geq n+1} \frac{1}{F(j)}\big).\end{aligned}$$

Finally,

$$\begin{aligned}&\mathbb{P}\Big(\sum_{j=1}^{n} r_2(j) < \sum_{j=1}^{\infty} r_1(j) < \sum_{j=1}^{\infty} r_2(j)\Big) \\ &\sim \lim_{n\to\infty} \big[\prod_{j=1}^{n} F(j)^2\big] \sum_{k=1}^{n} \frac{1}{2F(k)} \frac{1}{\prod_{l\neq k}(F(l)^2 - F(k)^2)} \Big[\sum_{j=n+1}^{\infty} \frac{1}{F(j)}\Big]\end{aligned}$$

■

Theorem 3.4.8.

$$\mathbb{P}(\chi_\infty = n) \sim \frac{a_1(n)}{F(n)},$$

where $a_1(n)$ is as defined in Lemma 3.4.3.

Proof: By exponential embedding, we can represent

$$\{\chi_\infty = n\} = \Big\{ \sum_{j=1}^{n-1} r_2(j) < \sum_{j=1}^{\infty} r_1(j) < \sum_{j=1}^{n} r_2(j) \Big\} \bigcup \Big\{ \sum_{j=1}^{n-1} r_1(j) < \sum_{j=1}^{\infty} r_2(j) < \sum_{j=1}^{n} r_1(j) \Big\}.$$

Suppose the density function of $\sum_{j=1}^{\infty} r_1(j)$ is $f(x)$ and the joint density function of $\sum_{j=1}^{n-1} r_2(j)$ and $\sum_{j=1}^{n} r_2(j)$ is $g(y,z)$. Then by Lemma 3.2.1,

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{j=1}^{n-1} r_2(j) < \sum_{j=1}^{\infty} r_1(j) < \sum_{j=1}^{n} r_2(j) \Big) \\
=& \int_0^\infty \int_0^z \Big[\int_y^z f(x)dx \Big] g(y,z)dydz \\
=& \Big[\prod_{j=1}^{\infty} F(j) \Big] \sum_{k=1}^{\infty} \frac{1}{F(k)} \frac{1}{\prod_{l \neq k} [F(l) - F(k)]} \int_0^\infty \int_0^z \big[e^{-F(k)y} - e^{-F(k)z} \big] g(y,z)dydz.
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{E}\Big[\exp\big(-F(k) \sum_{j=1}^{n-1} r_2(j) \big) - \exp\big(-F(k) \sum_{j=1}^{n} r_2(j) \big) \Big] \\
=& \prod_{j=1}^{n-1} \frac{F(j)}{F(j)+F(k)} - \prod_{j=1}^{n} \frac{F(j)}{F(j)+F(k)} \\
=& \frac{F(k)}{F(n)} \prod_{j=1}^{n} \frac{F(j)}{F(j)+F(k)}.
\end{aligned}$$

Hence,

$$\mathbb{P}(\chi_\infty = n) \sim \Big\{ 2\Big[\prod_{j=1}^{\infty} F(j) \Big] \sum_{k=1}^{\infty} \frac{1}{\prod_{l \neq k} [F(l) - F(k)]} \prod_{j=1}^{n} \frac{F(j)}{F(k)+F(j)} \Big\} \frac{1}{F(n)}.$$

■

Corollary 3.4.9. *When $F(x) = x^\alpha$, $\alpha > 1$,*

$$\mathbb{P}(\chi_\infty = n) \sim \frac{\tilde{a}_1(n)}{\alpha - 1} \frac{1}{n^\alpha},$$

where $\tilde{a}_1(n)$ is as defined in Corollary 3.4.5.

Our next job is to evaluate the number of minority balls at finite time. We first estimate some coefficients.

Lemma 3.4.10. *For $1 < n \leq \frac{k+2}{2}$, let*

$$a_2(k,n) := \Big[\prod_{j=1}^{n-1} F(j)\Big] \sum_{s=1}^{n-1} \frac{1}{\prod_{l\neq s}[F(l)-F(s)]} \prod_{j=1}^{k-n+2} \frac{F(j)}{F(j)+F(s)}$$

$$a_3(k,n) := \Big[\prod_{j=1}^{k-n+1} F(j)\Big] \sum_{t=1}^{k-n+1} \frac{1}{\prod_{l\neq t}[F(l)-F(t)]} \prod_{j=1}^{n} \frac{F(j)}{F(j)+F(t)}.$$

Then

$$a_2(k,n) = \binom{k+1}{n-1} \frac{\xi^{k-n+2}\mu^{n-1}}{(\mu+\xi)^{k+2}}$$

$$a_3(k,n) = \binom{k+1}{n} \frac{\tilde{\xi}^{n}\tilde{\mu}^{k-n+1}}{(\tilde{\mu}+\tilde{\xi})^{k+2}},$$

where ξ, μ, $\tilde{\xi}$ and $\tilde{\mu}$ are some constants, in $(0, \frac{1}{F(1)})$, depending on $F(x)$, k and n. Besides, $a_2(k,n)$ and $a_3(k,n)$ share the same bounds.

$$\frac{1}{2}\sqrt{\frac{1}{\pi\sum\frac{1}{F(j)}}} \leq a_2(k,n) \leq \sqrt{\frac{1}{2\pi\sum\frac{1}{F(j)}}}.$$

Proof: We will apply Lagrange interpolation estimation again.

(i) For $a_2(k,n)$, let $\lambda_s = \frac{1}{F(s)}$, $s = 1, ..., k-n+2$,

$$f(x) := \frac{x^k}{\prod_{j=1}^{k-n+2}(x+\lambda_j)} \qquad x \in [0, \lambda_1]$$

and $p(x)$ be the interpolation polynomial of $f(x)$ with interpolating points $\{\lambda_1, ..., \lambda_{n-1}\}$

$$p(x) := \sum_{s=1}^{n-1} \lambda_s^k \prod_{j=1}^{k-n+2} \frac{1}{\lambda_s + \lambda_j} \prod_{\substack{l \neq s \\ l=1}}^{n-1} \frac{x - \lambda_l}{\lambda_s - \lambda_l}$$

Then the error is

$$f(x) - p(x) = \frac{f^{(n-1)}(\xi)}{(n-1)!} \prod_{j=1}^{n-1} (x - \lambda_j), \qquad x \in [0, \lambda_1]$$

with some $\xi \in (0, \lambda_1)$ depending on x.

If choosing $x = 0$, we get

$$-p(0) = \frac{f^{(n-1)}(\xi)}{(n-1)!} (-1)^{n-1} \prod_{j=1}^{n-1} \lambda_j$$

where in our problem ξ depends on F, k and n.

Hence

$$a_2(k, n) = \frac{f^{(n-1)}(\xi)}{(n-1)!}$$

In order to evaluate $f^{(n-1)}(\xi)$, we first use the partial fraction to simplify $f(x)$. We get

$$\frac{x^k}{\prod_{j=1}^{k-n+2} (x + \lambda_j)} = \sum_{j=1}^{k-n+2} \frac{(-\lambda_j)^k}{x + \lambda_j} \prod_{l \neq j} \frac{1}{\lambda_l - \lambda_j}.$$

Then

$$f^{(n-1)}(x) = (-1)^{n-1}(n-1)! \sum_{j=1}^{k-n+2} \frac{(-\lambda_j)^k}{(x + \lambda_j)^n} \prod_{l \neq j} \frac{1}{\lambda_l - \lambda_j}.$$

So

$$a_2(k,n) = (-1)^{k+n-1} \sum_{j=1}^{k-n+2} \frac{\lambda_j^k}{(\xi+\lambda_j)^n} \prod_{l\neq j} \frac{1}{\lambda_l - \lambda_j}$$
$$= \sum_{j=1}^{k-n+2} \frac{\lambda_j^k}{(\xi+\lambda_j)^n} \prod_{l\neq j} \frac{1}{\lambda_j - \lambda_l}$$

Apply the interpolation technique again by letting

$$g(x) := \frac{x^{k+1}}{(\xi+x)^n}, \qquad x \in [0, \lambda_1]$$

Interpolated at $\{\lambda_1, ..., \lambda_{k-n+2}\}$, the interpolation polynomial is

$$q(x) = \sum_{j=1}^{k-n+2} \frac{\lambda_j^{k+1}}{(\xi+\lambda_j)^n} \prod_{l\neq j} \frac{x-\lambda_j}{\lambda_j - \lambda_l}.$$

The error is

$$g(x) - q(x) = \frac{g^{(k-n+2)}(\mu)}{(k-n+2)!} \prod_{j=1}^{k-n+2} (x-\lambda_j),$$

where $\mu \in (0, \lambda_1)$ depends on x. Let $x = 0$. We can get

$$a_2(k,n) = \frac{g^{(k-n+2)}(\mu)}{(k-n+2)!}.$$

Partial fraction gives

$$g(x) = \sum_{s=1}^{n} \binom{k+1}{n-s} (-\xi)^{k+1-n+s} \frac{1}{(x+\xi)^s}.$$

The derivative is

$$g^{(k-n+2)}(\mu) = \sum_{s=1}^{n} \binom{k+1}{n-s} (-\xi)^{k+1-n+s} (-1)^{k-n+2} \frac{(s+k-n+1)!}{(s-1)!} \frac{1}{(\mu+\xi)^{k-n+2+s}}$$
$$= \frac{(k+1)!}{(n-1)!} \frac{\xi^{k-n+2}}{(\mu+\xi)^{k+2}} \sum_{s=1}^{n} \frac{(n-1)!}{(n-s)!(s-1)!} (-\xi)^{s-1} (\mu+\xi)^{n-s}$$
$$= \frac{(k+1)!}{(n-1)!} \frac{\xi^{k-n+2}\mu^{n-1}}{(\mu+\xi)^{k+2}}.$$

Hence

$$a_2(k,n) = \binom{k+1}{n-1} \frac{\xi^{k-n+2}\mu^{n-1}}{(\mu+\xi)^{k+2}}.$$

(ii) With similar methods, we get the estimation of $a_3(k,n)$. ■

Theorem 3.4.11. *After time k, the number of minority balls in the urn has probability behavior:*

1) If $1 < n \leq \frac{k+2}{2}$,

$$\mathbb{P}(\chi_k = n) = \frac{2a_2(k,n)}{F(k-n+2)} + \frac{2a_3(k,n)}{F(n)},$$

where $a_2(k,n)$ and $a_3(k,n)$ are defined in Lemma 3.4.10.

2) If $n = 1$,

$$\mathbb{P}(\chi_k = 1) = \prod_{j=2}^{k} \frac{F(j)}{F(j)+F(1)}.$$

Proof:

1) When $n > 1$, there are two possibilities to make the number of balls of the minority color at time k to be n. First, the n-th minority ball is added at the k-th pick. Second, all the $n-1$ minority balls are added before the k-th pick. Hence, exponential embedding gives

$$\begin{aligned}\{\chi_k = n\} =& \Big\{ \sum_{j=1}^{k-n+1} r_2(j) < \sum_{j=1}^{n-1} r_1(j) < \sum_{j=1}^{k-n+2} r_2(j) \Big\} \\ & \bigcup \Big\{ \sum_{j=1}^{n-1} r_1(j) < \sum_{j=1}^{k-n+1} r_2(j) < \sum_{j=1}^{n} r_1(j) \Big\}\end{aligned}$$

Applying the density function in Lemma 3.2.1,

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{j=1}^{k-n+1} r_2(j) < \sum_{j=1}^{n-1} r_1(j) < \sum_{j=1}^{k-n+2} r_2(j)\Big) \\
=&\Big[\prod_{j=1}^{n-1} F(j)\Big] \sum_{s=1}^{n-1} \frac{1}{F(s)} \frac{1}{\prod_{l\neq s}[F(l)-F(s)]} \Big[\prod_{j=1}^{k-n+1} \frac{F(j)}{F(j)+F(s)} - \prod_{j=1}^{k-n+2} \frac{F(j)}{F(j)+F(s)}\Big] \\
=&\Big\{\Big[\prod_{j=1}^{n-1} F(j)\Big] \sum_{s=1}^{n-1} \frac{1}{\prod_{l\neq s}[F(l)-F(s)]} \prod_{j=1}^{k-n+2} \frac{F(j)}{F(j)+F(s)}\Big\} \frac{1}{F(k-n+2)}.
\end{aligned}$$

And

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{j=1}^{n-1} r_1(j) < \sum_{j=1}^{k-n+1} r_2(j) < \sum_{j=1}^{n} r_1(j)\Big) \\
=&\Big\{\Big[\prod_{j=1}^{k-n+1} F(j)\Big] \sum_{t=1}^{k-n+1} \frac{1}{\prod_{l\neq t}[F(l)-F(t)]} \prod_{j=1}^{n} \frac{F(j)}{F(j)+F(t)}\Big\} \frac{1}{F(n)}.
\end{aligned}$$

2) When $n = 1$, it means that the minority color is never added until at least time $k+1$. Represented in the exponential embedding, it is

$$\{\chi_k = 1\} = \Big\{r_1(1) > \sum_{j=1}^{k} r_2(j)\Big\} \bigcup \Big\{r_2(1) > \sum_{j=1}^{k} r_1(j)\Big\}.$$

Let $f(x)$ and $g(y)$ be the probability density function of $r_1(1)$ and $\sum_{j=1}^{k} r_2(j)$,

respectively. Then,

$$\begin{aligned}
&\mathbb{P}\Big(r_1(1) > \sum_{j=1}^{k} r_2(j)\Big) = \int_0^\infty f(x) \int_0^x g(y)dydx \\
=&\Big[\prod_{j=1}^{k} F(j)\Big] \sum_{s=1}^{k} \frac{F(1)}{\prod_{l\neq s}(F(l)-F(s))} \Big[\int_0^\infty e^{-F(1)x} \int_0^x e^{-F(s)y} dydx\Big] \\
=&\Big[\prod_{j=1}^{k} F(j)\Big] \sum_{s=1}^{k} \frac{1}{\prod_{l\neq s}(F(l)-F(s))} \frac{1}{F(1)+F(s)} \\
=&\prod_{j=1}^{k} \frac{F(j)}{F(j)+F(1)}
\end{aligned}$$

The last equality is from Corollary 3.2.9. In fact, we can get this probability directly from the probabilistic intuition of the adding-ball scheme. But we want to show the comparison of $n = 1$ and $n > 1$ by exponential embedding method. ■

Corollary 3.4.12. *When $F(x) = x^\alpha$, $\alpha > 1$, for $1 < n \leq \frac{k+2}{2}$,*

$$\mathbb{P}(\chi_k = n) = \frac{2\tilde{a}_2(k,n)}{(k-n+2)^\alpha} + \frac{2\tilde{a}_3(k,n)}{n^\alpha},$$

where

$$\begin{aligned}
\tilde{a}_2(k,n) &= \binom{k+1}{n-1} \frac{\xi^{k-n+2}\mu^{n-1}}{(\mu+\xi)^{k+2}} \\
\tilde{a}_3(k,n) &= \binom{k+1}{n} \frac{\tilde{\xi}^{n}\tilde{\mu}^{k-n+1}}{(\tilde{\mu}+\tilde{\xi})^{k+2}},
\end{aligned}$$

where ξ, μ, $\tilde{\xi}$ and $\tilde{\mu}$ are some constants, in $(0,1)$, depending on α, k and n. And the bounds for them are

$$\frac{1}{2}\sqrt{\frac{1}{\pi\sum\frac{1}{j^\alpha}}} \leq \tilde{a}_2(k,n) \leq \sqrt{\frac{1}{2\pi\sum\frac{1}{j^\alpha}}}.$$

One of our goals to estimate $\mathbb{P}(\chi_k = n)$ is for the following attraction time. Before that, we need to show that, when the urn function $F(x)$ is not growing too fast, like exponential functions, then the above coefficients $a_2(k,n)$ and $a_3(k,n)$ are asymptoticly equal to each other.

Lemma 3.4.13. *As defined in Lemma 3.4.10,*

$$a_2(k,n) \sim a_3(k,n).$$

Proof: The main idea is also to apply the Lagrange interpolation. Consider the function $f(x) = x^k$. Let $\lambda_j = \frac{1}{F(j)}$. Interpolating $f(x)$ at

$$\{\lambda_1, ..., \lambda_{k-n+1}, -\lambda_1, ..., -\lambda_n\},$$

the interpolation polynomial is

$$p(x) = \sum_{t=1}^{k-n+1} \lambda_t^k \prod_{\substack{l\neq t \\ l=1}}^{k-n+1} \frac{x-\lambda_l}{\lambda_t - \lambda_l} \prod_{j=1}^{n} \frac{x+\lambda_j}{\lambda_t + \lambda_j} + \sum_{t=1}^{n} (-\lambda_t)^k \prod_{\substack{l\neq t \\ l=1}}^{n} \frac{x+\lambda_l}{-\lambda_t + \lambda_l} \prod_{j=1}^{k-n+1} \frac{x-\lambda_j}{-\lambda_t - \lambda_j}$$

There are $k+1$ interpolation points. In the formula of the error term, $f^{(k+1}(\xi) =$

0. So there is

$$
\begin{aligned}
a_3(k,n) &= \Big[\prod_{j=1}^{n} F(j)\Big]\Big[\prod_{j=1}^{k-n+1} F(j)\Big]\sum_{t=1}^{n}\prod_{\substack{l\neq t\\ l=1}}^{n}\frac{1}{F(l)-F(t)}\prod_{j=1}^{k-n+1}\frac{1}{F(j)+F(t)}\\
&= \Big[\prod_{j=1}^{n-1} F(j)\Big]\Big[\prod_{j=1}^{k-n+2} F(j)\Big]\frac{F(n)}{F(k-n+2)}\\
&\sum_{t=1}^{n-1}\Big[\prod_{\substack{l\neq t\\ l=1}}^{n-1}\frac{1}{F(l)-F(t)}\prod_{j=1}^{k-n+2}\frac{1}{F(j)+F(t)}\frac{F(k-n+2)+F(t)}{F(n)-F(t)}\Big]\\
&+ F(n)\prod_{j=1}^{n-1}\frac{F(j)}{F(j)-F(n)}\prod_{j=1}^{k-n+1}\frac{F(j)}{F(j)+F(n)}\\
&> a_2(k,n) + F(n)\prod_{j=1}^{n-1}\frac{F(j)}{F(j)-F(n)}\prod_{j=1}^{k-n+1}\frac{F(j)}{F(j)+F(n)}
\end{aligned}
$$

Now

$$
\begin{aligned}
&F(n)\prod_{j=1}^{n-1}\frac{F(j)}{F(j)-F(n)}\prod_{j=1}^{k-n+1}\frac{F(j)}{F(j)+F(n)}\\
\sim&(-1)^{n-1}F(n)e^{2\sum_{j=1}^{n-1}(\frac{F(j)}{F(n)})^2-(n-1)-F(n)\sum_{j=n}^{k-n+1}\frac{1}{F(j)}}
\end{aligned}
$$

Similarly, we can get

$$
a_2(k,n) > a_3(k,n) + (-1)^{k-n+1}F(k-n+2)e^{\frac{2}{F(k-n+2)}\sum_{j=1}^{k-n+1}F(j)-(k-1)}
$$

■

3.5 Attraction time

In this section, let us look back at the *attraction time*, T_a. From T_a, only one-color balls can be drawn out.

Before estimating the distribution of T_a, we will simplify some definitions first, since in this section we only care about the two-color case. Let

$$\xi_n := \begin{cases} 1, & \text{if the minor color is drawn at time } n \\ 0, & \text{if the major color is drawn at time } n \end{cases}$$

Consequently,

Definition 3.5.1.

$$T_a := \min\{k : \xi_n = 0, \forall n \geq k\}.$$

Then the distribution of T_a can be given in

Theorem 3.5.2. *The distribution of the attraction time is*

$$\mathbb{P}(T_a = k) = 2a_2(k,n) \sum_{n=1}^{k} \frac{1}{F(k-n+1)} \prod_{i \geq k-n+1} \frac{F(i)}{F(n)+F(i)},$$

where $a_2(k,n)$ is defined in Lemma 3.4.10.

Proof: In order to estimate $\mathbb{P}(T_a = k)$, we need the condition of the number of the minority balls at time $k-1$, i.e χ_{k-1}. And given $\chi_{k-1} = n$, $(1 \leq n \leq k)$, $\{T_a = k\}$ means that the $(k-1)$-th pick picks the minority color and from then on, no more minority color balls are picked; hence $\chi_\infty = \chi_{k-1}$. So

$$\{T_a = k | \chi_{k-1} = n\} = \{\xi_{k-1} = 1, \chi_\infty = n | \chi_{k-1} = n\}.$$

Summed over n,

$$\begin{aligned}\mathbb{P}(T_a = k) &= \sum_{n=1}^{k} \mathbb{P}(T_a = k|\chi_{k-1} = n)\mathbb{P}(\chi_{k-1} = n) \\ &= \sum_{n=1}^{k} \big[\mathbb{P}(\xi_{k-1} = 1|\chi_{k-1} = n)\mathbb{P}(\chi_\infty = n|\chi_{k-1} = n)\big]\mathbb{P}(\chi_{k-1} = n).\end{aligned}$$

Without loss of generality, we may suppose color 1 is the minority color for the computation. By the independence of $\{r_1(i)\}$ and $\{r_2(j)\}$ and the memoryless property of exponential processes, we have

$$\{\xi_{k-1} = 1|\chi_{k-1} = n\} = \{r_1(n-1) < r_2(k-n+1)\}.$$

Hence,

$$\mathbb{P}(\xi_{k-1} = 1|\chi_{k-1} = n) = \frac{F(n-1)}{F(n-1) + F(k-n+1)}.$$

Similarly,

$$\{\chi_\infty = n|\chi_{k-1} = n\} = \{\sum_{i=k-n+1}^{\infty} r_2(i) < r_1(n)\}.$$

By Lemma 3.2.2, the density function of $\sum_{i=k-n+1}^{\infty} r_2(i)$ is

$$f(x) = [\prod_{i=k-n+1}^{\infty} F(i)] \sum_{i=k-n+1}^{\infty} \frac{e^{-F(i)x}}{\prod\limits_{\substack{l\geq k-n+1 \\ l\neq i}} (F(l) - F(i))}.$$

Hence,

$$\begin{aligned}\mathbb{P}(\chi_\infty = n|\chi_{k-1} = n) &= [\prod_{i=k-n+1}^{\infty} F(i)] \sum_{i=k-n+1}^{\infty} \frac{\int_0^\infty \int_0^y e^{-F(i)x} F(n) e^{-F(n)y} dxdy}{\prod\limits_{\substack{l\geq k-n+1 \\ l\neq i}} (F(l) - F(i))} \\ &= [\prod_{i=k-n+1}^{\infty} F(i)] \sum_{i=k-n+1}^{\infty} \Big[\prod_{l\neq i} \frac{1}{F(l) - F(i)}\Big] \frac{1}{F(n) + F(i)}.\end{aligned}$$

By Corollary 3.2.9,

$$\mathbb{P}(\chi_\infty = n|\chi_{k-1} = n) = \prod_{i=k-n+1}^{\infty} \frac{F(i)}{F(n)+F(i)}.$$

■

Corollary 3.5.3. *When $F(x) = x^\alpha$, $\alpha > 1$,*

$$\mathbb{P}(T_a = k) = 2\tilde{a}_2(k,n) \sum_{n=1}^{k} \frac{1}{(k-n+1)^\alpha} \prod_{i \geq k-n+1} \frac{i^\alpha}{n^\alpha + i^\alpha},$$

where $\tilde{a}_2(k,n)$ is defined in Corollary 3.4.12.

Theorem 3.5.4 ([26])**.** *When $F(x) = x^\alpha$, $\alpha > 1$,*

$$\mathbb{P}(T_a = k) = \Theta\left(\frac{1}{k^{\alpha-\alpha'}}\right),$$

where $\alpha' = (\alpha-1)/\alpha$.

Corollary 3.5.5. *When $F(x) = x^\alpha$, $\alpha > 1$,*

$$\mathbb{P}(T_a = k) \sim \frac{2\tilde{a}_2(k,n)}{k^{\alpha-\alpha'}},$$

where $\alpha' = (\alpha-1)/\alpha$ and $\tilde{a}_2(k,n)$ is defined in Corollary 3.4.12.

3.6 Large Time Minority Has Little Chance to Win Back

There are vibrations at the beginning of the evolution. It is hard to tell which color will finally win. We will show in the following that after some time point, the

leading side has an overwhelming chance to dominate. Here, our method is simpler and the result is more accurate than the one in [85].

Suppose at some point, when the total number of balls in the urn is $n_0 >> 1$, color 1 has $p_0 n_0$ balls where $0 < p_0 < \frac{1}{2}$. In the following, we estimate the probability that there are pn $(0 < p_0 < p < 1)$ color 1 balls in the urn when there are $n > n_0$ balls in total. For simplicity, we assume that $p_0 n_0$ and pn are integers. With the help of exponential embedding, this event can be represented by

$$\Big\{ \sum_{i=p_0 n_0}^{pn-1} r_1(i) < \sum_{i=q_0 n_0}^{qn-1} r_2(i) \Big\},$$

where $q_0 = 1 - p_0$ and $q = 1 - p$.

Theorem 3.6.1 ([85])**.**

$$\mathbb{P}\Big(\sum_{i=p_0 n_0}^{pn-1} r_1(i) < \sum_{i=q_0 n_0}^{qn-1} r_2(i) \Big) < 8e^{2-(p_0 n_0)^{1/4}}.$$

Theorem 3.6.2. *As defined above,*

$$\mathbb{P}\Big(\sum_{i=p_0 n_0}^{pn-1} r_1(i) < \sum_{i=q_0 n_0}^{qn-1} r_2(i) \Big) < \sum_{k=p_0 n_0}^{pn-1} \frac{1}{F(k)}$$

Proof: Let $f(x)$ and $g(y)$ be the density functions of $\sum_{i=p_0 n_0}^{pn-1} r_1(i)$ and $\sum_{i=q_0 n_0}^{qn-1} r_2(i)$, respectively. Then

$$f(x) = \Big[\prod_{i=p_0 n_0}^{pn-1} F(i) \Big] \sum_{k=p_0 n_0}^{pn-1} \frac{e^{-F(k)x}}{\prod_{l \neq k} (F(l) - F(k))}$$

and

$$g(y) = \Big[\prod_{i=q_0n_0}^{qn-1} F(i)\Big] \sum_{k=q_0n_0}^{qn-1} \frac{e^{-F(k)y}}{\prod_{l\neq k}(F(l)-F(k))}.$$

So

$$\mathbb{P}\Big(\sum_{i=p_0n_0}^{pn-1} r_1(i) < \sum_{i=q_0n_0}^{qn-1} r_2(i)\Big) = \int_0^\infty \Big[\int_0^y f(x)dx\Big] g(y)dy$$

Similar to the proof of the previous theorems, we can get

$$\begin{aligned}&\mathbb{P}\Big(\sum_{i=p_0n_0}^{pn-1} r_1(i) < \sum_{i=q_0n_0}^{qn-1} r_2(i)\Big)\\ =&\Big[\prod_{i=p_0n_0}^{pn-1} F(i)\Big] \sum_{k=p_0n_0}^{pn-1} \frac{1}{F(k)} \frac{1}{\prod_{l\neq k}(F(l)-F(k))} \Big[1 - \prod_{j=q_0n_0}^{qn-1} \frac{F(j)}{F(j)+F(k)}\Big]\end{aligned}$$

By Corollary 3.2.3,

$$\Big[\prod_{i=p_0n_0}^{pn-1} F(i)\Big] \sum_{k=p_0n_0}^{pn-1} \frac{1}{F(k)} \frac{1}{\prod_{l\neq k}(F(l)-F(k))} = 1.$$

Hence, the probability is

$$\mathbb{P}\Big(\sum_{i=p_0n_0}^{pn-1} r_1(i) < \sum_{i=q_0n_0}^{qn-1} r_2(i)\Big) = 1 - \sum_{k=p_0n_0}^{pn-1} \prod_{\substack{l\neq k\\ l=p_0n_0}}^{pn-1} \frac{F(l)}{F(l)-F(k)} \prod_{j=q_0n_0}^{qn-1} \frac{F(j)}{F(j)+F(k)}$$

Take the logarithm of the product, for

$$\begin{aligned}\log \prod_{j=q_0n_0}^{qn-1} \frac{F(j)}{F(j)+F(k)} &= \sum_{j=q_0n_0}^{qn-1} \log\Big(\frac{1}{1+\frac{F(k)}{F(j)}}\Big)\\ &\sim -F(k) \sum_{j=q_0n_0}^{qn-1} \frac{1}{F(j)}\end{aligned}$$

Let $s = \sum_{j=q_0n_0}^{qn-1} \frac{1}{F(j)}$. Then

$$\mathbb{P}\Big(\sum_{i=p_0n_0}^{pn-1} r_1(i) < \sum_{i=q_0n_0}^{qn-1} r_2(i)\Big) = 1 - \sum_{k=p_0n_0}^{pn-1} e^{-F(k)s} \prod_{\substack{l\neq k \\ l=p_0n_0}}^{pn-1} \frac{F(l)}{F(l)-F(k)}$$

$$=1 - \sum_{k=p_0n_0}^{pn-1} \Big(1 - F(k)s + \frac{F(k)^2s^2}{2} - \cdots\Big) \prod_{\substack{l\neq k \\ l=p_0n_0}}^{pn-1} \frac{F(l)}{F(l)-F(k)}$$

$$= \sum_{k=p_0n_0}^{pn-1} \Big(F(k)s - \frac{F(k)^2s^2}{2} + \cdots\Big) \prod_{\substack{l\neq k \\ l=p_0n_0}}^{pn-1} \frac{F(l)}{F(l)-F(k)}$$

Since $\sum \frac{1}{F(j)}$ is convergent, we can choose s small enough so that

$$\mathbb{P}\Big(\sum_{i=p_0n_0}^{pn-1} r_1(i) < \sum_{i=q_0n_0}^{qn-1} r_2(i)\Big) < \sum_{k=p_0n_0}^{pn-1} \frac{1}{F(k)} \prod_{\substack{l\neq k \\ l=p_0n_0}}^{pn-1} \frac{F(l)}{F(l)-F(k)}$$

$$= \sum_{k=p_0n_0}^{pn-1} \frac{1}{F(k)},$$

by Corollary 3.2.4. ■

3.7 The Decay/Growth is Slow

Although at large time point the minority color has little chance to win back, it still takes a long time for the majority to finally announce the victory. We will show this by estimating that the probability of the minority color has more than p percentage of the balls in the urn, which is equivalent to say that the majority color has less than $q = 1 - p$ percentage, for some $0 < p < \frac{1}{2}$. Let χ_n be the number of the minority balls when the total number of balls in the urn is n.

Theorem 3.7.1.

$$\mathbb{P}(\chi_n \geq pn) = a_4(p,n) \sum_{j=pn}^{qn-1} \frac{1}{F(j)},$$

where

$$a_4(p,n) \sim \frac{1}{\sqrt{\pi pn\xi_n}},$$

where $\xi_n \in (0, \frac{1}{F(1)})$ *is some constant depending on* $F(x)$, n *and* p. *And* $a_4(p,n)$ *has the same bounds as* $a_1(n)$.

Proof: When the minority gets pn, it means the majority has got pn too. The event can be represented by

$$\begin{aligned}\{\chi_n \geq pn\} =& \{\sum_{j=1}^{pn-1} r_2(j) < \sum_{i=1}^{pn-1} r_1(i) < \sum_{j=1}^{qn-1} r_2(j)\} \\ & \bigcup \{\sum_{i=1}^{pn-1} r_1(i) < \sum_{j=1}^{pn-1} r_2(j) < \sum_{i=1}^{qn-1} r_1(i)\}\end{aligned}$$

Let $f(x)$ be the density function of $\sum_{i=1}^{pn-1} r_1(i)$ and $g(y,z)$ be the joint density function of $\sum_{j=1}^{pn-1} r_2(j)$ and $\sum_{j=1}^{qn-1} r_2(j)$. Then

$$\mathbb{P}\Big(\sum_{j=1}^{pn-1} r_2(j) < \sum_{i=1}^{pn-1} r_1(i) < \sum_{j=1}^{qn-1} r_2(j)\Big) = \int_0^\infty \int_0^z \big[\int_y^z f(x)dx\big] g(y,z)dydz$$

Similar to above,

$$\begin{aligned}&\mathbb{P}\Big(\sum_{j=1}^{pn-1} r_2(j) < \sum_{i=1}^{pn-1} r_1(i) < \sum_{j=1}^{qn-1} r_2(j)\Big) \\ =&\Big[\prod_{i=1}^{pn-1} F(i)\Big] \sum_{k=1}^{pn-1} \frac{1}{F(k)} \frac{1}{\prod_{l\neq k}(F(l)-F(k))} \prod_{j=1}^{pn-1} \frac{F(j)}{F(j)+F(k)} \Big[1 - \prod_{j=pn}^{qn-1} \frac{1}{1+\frac{F(k)}{F(j)}}\Big]\end{aligned}$$

And

$$1-\prod_{j=pn}^{qn-1}\frac{1}{1+\frac{F(k)}{F(j)}}=1-\prod_{j=pn}^{qn-1}\Big[\sum_{s\geq 0}\big(-\frac{F(k)}{F(j)}\big)^s\Big]$$
$$=F(k)\sum_{j=pn}^{qn-1}\frac{1}{F(j)}+o\big(\sum_{j=pn}^{qn-1}\frac{F(k)}{F(j)}\big)$$

By the same method in Lemma 4, we finish the proof. ■

Corollary 3.7.2. *When $F(x)=x^\alpha$ $(\alpha>1)$,*

$$\mathbb{P}(\chi_n\geq pn)\sim\tilde{a}_4(p,n)\frac{p^{1-\alpha}-q^{1-\alpha}}{\alpha-1}\frac{1}{n^{\alpha-1}},$$

where

$$\tilde{a}_4(p,n)\sim\frac{1}{\sqrt{\pi pn\tilde{\xi}_n}},$$

where $\tilde{\xi}_n\in(0,1)$ is some constant depending on α, n and p. And $\tilde{a}_4(p,n)$ has the same bounds as $\tilde{a}_1(n)$.

Chapter 4

Applications of GPU Models

With the development of theories and the need of applications, people have paid a lot of attention to GPU models since 1960's. It has been successfully applied in many fields. Usually, GPU models are applied to stochastic processes in which early events determine the ultimate outcomes. Here I will give a brief survey of some main recent work.

4.1 Reinforced Random Walk (RRW)

Coppersmith and Diaconis invented this concept in 1987. There are two types of RRW: one is edge-reinforced random walk (ERRW) and the other is vertex-reinforced random walk (VRRW). Given an undirected graph with all the edges having an initial weight, for simplicity, say 1, a rabbit jumps on the graph starting

from a specific point, say v_0. And suppose there're k edges leading out from v_0. The probability that the rabbit chooses one of these edges to go is $1/k$. After running through an edge, say e_0, its weight is increased by a fixed (or random) number α. Then choose the next edge. And the probability to run through e_0 again becomes $(1+\alpha)/(k'+\alpha)$, if there're k' edges leading out from the current vertex. This precess is continuous. This is a concrete description of *edge-reinforced random walk*. Similarly, a *vertex-reinforced random walk* is a random walk on an undirected graph with all the vertices having an initial weight. At each time, passing a vertex will increase its weight by a fixed (or random) number. Choosing the next vertex to move is with probability proportional to the weights of the vertices adjoining to the current one. For the mathematical definitions of edge(vertex)-reinforced random walk, see Pemantle [88] or Davis [28].

How to apply GPU models to study RRW's? Pemantle [88] has pointed out that the ERRW has similar behaviors with the Polya urn and the VRRW has similar behaviors with the Friedman's urn. Let's look at an example.

Example 4.1.1 (An ERRW on $\mathbb{Z}$). Let $\{X_n\}_{n\geq 0}$ denote the process. So it takes values in $\mathbb{Z}$. At time n, the weight of the edge (or interval) $(i, i+1)$ is one plus the number of times of crossing it, if the initial weight is one. This RRW was introduced by Diaconis. And he showed that it's equivalent to having an independent Polya's urn (the original one) at each integer vertex and the urn at difference vertices are independent, so drawing a red ball means moving to the increasing

direction and drawing a black in the other way. The exchangeability property of Polya's urn shows that this motion is equivalent to a random walk in a random environment(RWRE), where the environment is distributed as the limiting beta random variables at Polya's urn processes associated with each integer vertex. So the study of this ERRW is deduced to the study of the corresponding RWRE.

Similarly, for VRRW, we place a Polya urn at each vertex. The classical Polya's urn model doesn't work well for all the RRW's. So in some specific cases, we need to modify it or generalize the original Polya urn model.

How about more than one rabbit jumping on the graph? Kovchegov [71] modified(generalized) the classical Polya's urn model to study a multi-particle ERRW, say two rabbits jumping on $\mathbb{Z}$. Their positions can be one left(right) to the other or at the same integer vertex. He modified the Polya's urn model in this way: at each integer vertex, we set an urn with red balls, black balls AND a **magic** ball such that when the left rabbit jumps to the vertex, the magic ball is assumed as a red ball, and when the right rabbit jumps to the vertex, the magic ball is assumed to be black. If the magic ball is drawn, we add two balls of the color that the magic ball is assumed to be at that time. Still the rabbits move according the color drawn. For more mechanism of this process, see Kovchegiv's paper.

Remark 4.1.2. In fact, we can study directed-edge-reinforced random walk, i.e. on directed graph, in the same way with the assistance of Polya urns. See Keane and

Rolles [101] for details.

Remark 4.1.3. There is another similar characterization that the Polya's urns and the ERRW's share. It's well known that the Polya's urn arise as naturally as Dirichlet distribution (see Mauldin, Sudderth and Williams [77]). Diaconis conjectured that so does ERRW. Rolles [101] proved that it's true for a modified ERRW. See also Diaconis and Rolles [31].

Remark 4.1.4. From the description above, it seems there should usually be some connections between all the reinforced random processes, which will provide methods to study other similar random process. Pemantle wrote a survey on this in 2001.

4.2 Statistics

Blackwell and MacQueen [22] generalized Pólya urn model in a very general way, which allows "continuum" colors and a given measure similar to the urn function. Let $\mathcal{M}$ be a separable and complete metric space, which is in fact the space of the colors in the urn. Let υ be a finite positive measure on $\mathcal{M}$. A sequence $\{\xi_n : n \geq 1\}$ of random variables with values in $\mathcal{M}$ is a *Pólya sequence with parameter* υ if for every $S \subset \mathcal{M}$,

1. $\mathbb{P}(\xi_1 \in S) = \upsilon(S)/\upsilon(\mathcal{M})$ and

2. $\mathbb{P}(\xi_{n+1} \in S|\xi_1, ..., \xi_n) = (\upsilon(S) + \sum_{i=1}^{n} \mathbf{1}_{\{\xi_i \in S\}})/(\upsilon(\mathcal{M}) + n)$.

Theorem 4.2.1 ([22]). *Let $\{\xi_n\}$ be a Pólya sequence on $\mathcal{M}$ with parameter υ. Then:*

(i) $v_n := (\upsilon + \sum_{i=1}^{n} \mathbf{1}_{\{\xi_i \in S\}})/(\upsilon(\mathcal{M}) + n)$ *converges almost sure to a discrete probability distribution υ^*;*

(ii) υ^* *has a* Ferguson distribution *with parameter υ, i.e. for every partition $(M_1, ..., M_k)$ of $\mathcal{M}$, the vector $< \upsilon^*(M_1), ..., \upsilon^*(M_k) >$ has a Dirichlet distribution with parameter $(\upsilon(M_1), ..., \upsilon(M_k))$.*

(iii) $\{\xi_n\}$ *is exchangeable and its de Finetti measure is μ^*.*

Proof: When $\mathcal{M}$ is finite,

Their goal was to generate an exchangeable sequence of random variables whose de Finetti measure is a Dirichlet process.

Remark 4.2.2. Since Blackwell and MacQeen's construction, many authors have agreed that Polya urn model might be the simplest and most concrete way to construct infinite exchangeable sequences. At the same time, it has been one of the foundations of GPU models' applications on statistics.

The Pólya urn model was introduced to define a finite population Bayesian bootstrap(FPBB) by Lo in 1988. The FPBB is a Bayesian analogue of finite population bootstrap(FPB). From the name, we can see that FPB is the bootstrap method for a finite population. For a detail description of this method, see Gross. Suppose there's a finite population $\{y_1, ..., y_N\}$ and a sample from it is given $\{x_1, ..., x_n\}$ with

$n < N$. So the FPRR method can be described as a Polya sampling algorithm: 1) given an urn with n balls and each a type(color), $\{x_1, ..., x_n\}$, draw $m := N - n$ balls successfully from the urn and every time the drawn ball is replaced with another ball of the same type into the urn. This is a *Polya sample of size m* procedure. Say $z_1, ..., z_m$ are sampled; 2) define a random distribution function H_{mn} by the sample empirical distribution function F_n of $\{x_1, ..., x_n\}$ and the empirical distribution G_m of $\{z_1, ..., z_m\}$. Evaluate the functional θ of H_{mn} and F_n; 3) Repeat the previous steps a large number of times, say L, getting $G_{m1}, ...G_{mL}$ and $\theta_1, ..., \theta_L$; 4) Use the empirical distribution of $\theta_1, ..., \theta_L$ to approximate the posterior distribution which a Bayesian might be interested in (see [72]).

In 1981, Lo defined a *censored date Bayesian bootstrap*(CDBB).Eleven years later, Muliere and Walker introduced a new bootstrap method, the finite censored Bayesian bootstrap(FCBB), for a finite population with censored data in terms of a genralized Polya urn scheme. Their sample procedure involves a multi-urn algorithm, each a Polya. Each urn contains something-we-want balls and something-else balls. If we sample from the first urn with the usual Polya urn drawing procedure, then the next urn, until a something-we-want ball is sampled. For details, see the paper. This is actually from the two authors studied in 1997 which will be explained in the following.

Cohen extended Lo's idea of the FPBB method to the uneuqal probability Bayesian bootstrap, i.e. the sample of each unit depends on some different probabilities. I think, this could be connected to the idea of urn function in our GPU models.

Related to the Bayesian bootstrap is the *Polya posterior*, which is a noninformative Bayesian procedure used when little or no prior information is available. In 1998, Nelson and Meeden studied how to use the Polya posterior to estimate a function of the population when we know some prior information about the population. Their method is as follows: they divide the population into different strata according to the prior information we've known, then at each stratum the Polya posterior is used. One year later, Meeden modified the Polya sample procedure again to study two-stage cluster sampling. His method is to place "big" balls in an urn and in each ball there's a set of units. See details in their paper. Also see [84] for further progress.

Since in 1973 when Blackwell and MacQueen found the connection between the original Polya urn model and the Dirichlet process, many authors have agreed that the Polya urn model might be the simplest and most concrete way to construct prior distributions and to construct exchangeable sequences. So people generalized Polya urn models according to their needs. For example, Muliere and Walker constructed a genralized Polya urn model motivated by the beta-Stacy process, which illustrates

the existence of a more general scheme for generating an exchangeable sequence than Blackwell and MacQueen's construction. And they showed that this urn process is exchangeable. This is what we mentioned before. Another example to use the connection between Polya urn model and Dirichlet process is the "Gibbs sampler" algorithm in Escobar [40], which influences the later simulation-based computation method in nonparametric and semiparametric Bayesian movement (like [40] and [38]).

Remark 4.2.3. Note that the exchangeability of Polya urn model plays a key role in most of the applications. But as we know, not all the GPU's have exchangeability. But sometimes, we depend more on other advantages of GPU's, like in Ishwaran's paper [62] when the author used GPU to model a finite mixture distribution by depending more on the mechanism of GPU generating right number of distinct values of the population.

Remark 4.2.4. In 2000, O'Cinneide and Pokrovskii used the Poisson approximation method for matches in a multi-Polya-urn model to prove a limit law of nonuniform random transformations. See their paper for details.

4.3 In Computer Science

The first application of GPU models in computer science is in the computer aided geometric design(CAGD). Goldman [50] used the original Polya urn model to gen-

erate blending functions. In computer aided geometric design, given a collection of points $P_0, ..., P_n$, to construct a curve $P(t) = \sum_i B_i^n(t)P_i$, $t \in [0, 1]$ will use the predefined *blending functions*, $\{B_i^n(t)\}_{i=0}^n$. The requirement for curves, consequently for the blending functions, of CAGD is high. There're 13 desirable properties: 1. well-defined, 2. convex hull, 3. smooth, 4. interpolates end points, etc. For the whole list and descriptions of the 13 properties, see the paper for details. So we can see that to find suitable blending functions satisfying all of the 13 properties, even half of them, is not an easy job. Motivated by the observation that a function satisfying the properties should first be a discrete distribution function and the fact that Polya urn model is usually a good choice to construct discrete distributions, Goldman used the Polya distribution functions, $D_i^n(t)$, i.e. the probability of drawing exactly i balls in the first n trials, as the blending function with $\delta/(R_0 + B_0)$ fixed and $t := R_0/(R_0 + B_0)$ changing and showed that they satisfy 11 desirable properties. Also, the author showed that if we want to relax some end point condition for some curve, then the corresponding distribution functions for the original Friedman's urn model can be used as blending functions. Specifically, Goldman constructed the GPU models for two special blending functions, the *B-spline* and *Beta-spline* functions and use the urn model to study properties of Beta-splines. See the paper ([50] and [52]) for details. There is a good survey in this field ([52]). In that paper, some basic properties of 2-colored GPU models, the comparison of urn models and approximation schemes and their application in splines are discussed.

See also Barry and Goldman [16] for a brief description of using Polya tensor product and Polya triangular patch to approximate surfaces in CAGD, as well as a thorough study of Polya curves and Polya polynomials based on the earlier works of Goldman. Until 2003, Chen expanded Goldman's result [52] to construct a class of approximation operators for which the Weierstrass Approximation Theorems by using the original Friedman's urn. It seems that using GPU models to construct blending functions to approximate surfaces has not been studied systematically and higher dimensional urn models haven't been used in CAGD.

The second application is in image segmentation and labelling. In the model, each pixel is represented by an urn with balls of different colors, one color for each class label. The initial composition of each urn can be determined by the initial measures of similarity of the pixels assigned by an arbitrary segmentation technique. At each stage, all the urns in the neighborhood of the pixel we're looking at are combined together and one ball is drawn randomly from the combined urn, then a fixed number of balls of the drawn color are added to the urn for the pixel; OR drawing on ball from each urn in the neighborhood and add a fixed number of balls of the majority color in theses balls to the urn for the pixel. This shows the influence of the neighborhood of a pixel in image segment. This urn model was modified by Banerjee, Burlina and Alajaji in 1999. They used urn models because first they can generate Markov chains and second they provide a natural probability representation for image labels. The authors also showed that compared to some

other methods for image segmentation, the urn models yields better segmentations in relatively small numbers of iterations.

Remark 4.3.1. Another application of GPU models in digital images is in the blending technology, which studies how to hide a secret image with several public images. See Zhao and Gong [116] for details. Their idea is similar to generate the blending functions in Goldman's method, except that they looked at the probability of the first black ball at the n-th draw.

The third application is in random trees. In our previous chapters, there're several authors whose results of GPU models are motivated by the goal to study trees, like Bachi and Pal [9]. In 1991, Mahmoud and Smythe used a GPU model to study the distribution of leaves in rooted subtrees of recursive trees. The model has replacement matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ and the colors represent different positions of the nodes. More generally, Janson [66] proved the asymptotic degree distribution in random recursive trees by considering GPU models with infinitely many types of colors, which can be truncated into finite case. Mahmoud [75] used a bottom-up approach and a GPU model to study rotations in fringe-balanced binary trees. The replacement matrix there is $\begin{pmatrix} -2 & 1 & 2 \\ 4 & -1 & -2 \\ 4 & -1 & -2 \end{pmatrix}$ where we need an extension that

the entries in the replacement matrix could be less than -1. For completeness, comparing to the urn method, later Panholzer and Prodinger used an analytic approach to study the rotations in fringe-balanced trees.

Remark 4.3.2. For more applications in random trees, see Mórí [80]

Remark 4.3.3. Another application is in constructing communication channel. Alajaji and Fuja (1994) used the original Polya urn model to construct discrete binary additive communication channel. They focused on the drawn color sequence.

4.4 In Clinical Trials

Clinical trial is one of the classical application fields of GPU models. It was first introduced by Wei in 1977. The original Friedman's urn was used. To compare the effectiveness of K ($K \geq 2$) treatments in a process, suppose eligible patients arrive sequentially and must be treated immediately. The statistical problem here is how to assign the treatment to each patient. Also, there're several factors known or thought to affect the patients' ability to respond to treatment and each factor has several levels. And each patient could not have all the factors. Use a generalized Friedman's urn. There're K colors of balls in the urn and initially each color has the same number of balls. If a color i ball is drawn at a stage, then α balls of the same color and β balls of each of the different colors are added to the urn with the drawn ball. Here $\alpha, \beta > 0$. So each time a patient is waiting for a treatment, if the color

i is drawn then assign treatment i to the patient. To consider influencing factors, use an urn at each level of a factor and measure the 'unbalance' of each urn. The more unbalanced the urn is, the higher probability the drawn result of that urn will be used. (See Wei, 1978). Due to the ethical problem, in a medical trial there're two goals: one is to get information about the effectiveness of the K treatments; the other one is to treat each current patient in the best way the experimenter can. So for the ethical reason, Wei and Durham raised the GPU model in 1978: initially there're w_i balls of each color in the urn, $i = 1, ...K$; when the first patient arrives, a ball is drawn randomly from the urn and the treatment of the corresponding color, say i, is assigned to the patient; suppose we can get the response of the previous treatment: if it's successful, then add $\alpha > 0$ balls of color i to the urn; if it's failed, then add $\beta > 0$ balls of each color j ($j \neq i$) to the urn. This is also called a *randomized play-the-winner* rule. (The play-the-winner rule was first introduced by Zelen (1969)). This type of designs are called *adaptive designs*. Schouten (1995) extended Wei's GPU model to those without replacement to study small strata case when blinding is impossible. And in 1998, Durham, Flournoy and Li proposed a *randomized Polya urn for selecting optima* in which a ball is drawn from the urn at each time and if the response is a success then another ball of the same color is added; if the response is a fail, then no new balls are added. A stopping rule is associated with this process and the urn contains some additional information that pertains only to the stopping time.

People also allow balls to be taken out of the urn. Ivanova, Rosenberger, Durham and Flournoy extended Durham, Flournoy and Li's design (1998) in a way that after each draw the drawn ball is returned, if response is success, then one ball of the same color is added; if fail, then one ball of the same color is removed from the urn. They then embedded the urn model into a continuous birth and death process, motivated by Athreya and Karlin's embedding method. So they call the model a *birth and death urn*.

Notice that the above models don't have a lot of flexibility. In other words, in real experiments, there're usually time trends, which correspond to the replacement matrices of the GPU models *nonhomogeneous*. Some of Bai and Hu and some other authors' papers considered this situations as a type of GPU models.

In probability design, there is always a goal people care about: finding the optimal design. In clinical trials, it is more important. In 2001, Bai, Chen and Hu gave an optimal adaptive design which is based on the idea to make the error probability minimal and construct an asymptotic optima. Chen and Zhu (2004) gave several optimal designs represented by GPU models. See also Durham, Flournoy and Li (1998).

Remark 4.4.1. See also Smythe and Wei (1983), Wei, Smythe, Lin and Park (1990), Rosenberger (1993), Bai, hu and Rosenberger (2002). for further discussions of applications of GPU models in clinical trial. There're two survey papers about

adaptive designs, Rosenberger (1996) and Dirienzo (2000).

Remark 4.4.2. Similar to the adaptive design, Rosenberger and Grill (1997) applied a randomized GPU model to psychophysical experiments, since the the convergence properties of the urn model allow selection of design points in a distribution that approximates the optimal design and parameters of the urn model can be adjusted to control the spread of this distribution.

4.5 In Biology

Applications of GPU models in biology have a long history. See Johnson and Kotz (1977) for some earlier work. Let's mention some relatively new results.

The first subfield is in biological population statistics. In 1982, Janardan used the classical Polya urn model to study correlation between the numbers of male and female children in a family. The famous Ewens Sampling Fornula (see Ewens, 1972) can also be generated by a GPU model, which is more Friedman-like. See Hoppe (1984), Donnelly (1986) and a relatively complete survey in this field, Hoppe (1987).

The second subfield is in evolutionary theory. Current papers in this field are more mathematical, see Schreiber (2001), Benaim, Schreiber and Tarres (2001) and McKenzie and Steel (2000). Schreiber generalized the Polya urn model in a way

that two balls are drawn from the urn at each step and the replacement scheme becomes more complicated, which depends on the composition of the chosen colors. He used this to model the "replicator processes" in evolutionary theory.

Remark 4.5.1. Another 'indirect' application subfield is in phylogeny, since random trees can be used in the null models and Polya urn models and random trees have some connections. See Maddison and Slatkin (1991) and Barrett, Donoghue and Sober (1991).

4.6 In Social Science

4.6.1 In Psychology:

The application of GPU models in the learning processes in mathematical psychology might be since 1919. There's a relatively complete introduction of this field in Johnson and Kotz's book (1977). See Pemantle and Skyrms (2003) for some recent development. I will only introduce some recent papers in psychology, since they have a close relation to the applications of GPU models in statistics. Navarro, Griffiths, Steyvers and Lee (2005) tried to model individual differences by Dirichlet process, which uses Blackwell and Macqueen's result about Polya urn model. Xue (2005) used Polya urn models to study the conformity of individual behavior.

4.6.2 In Economics:

The application of Polya urn model in economics can be traced back to 1958. Janson used the original Polya urn to model brand choice's contagion characteristic. Arthur and Lane called this kind of phenomenons *information contagion*. See their paper (1993) for a relatively complete discussion of these processes and how they are related to GPU models with urn functions. Arthur, Ermoliev and Kaniovski (1984 and 1987) introduced the nonlinear Polya urn models to study path-dependent processes for applications in chemical kinetics, industrial location theory and the emergence of technological structure in economy. See also Arthur (1989). Motivated by the evolutionary processes, Samuel tried to use the original Polya urn model to microdata disclosure risk assessment in 1998. Based on this, Fienberg and Makov modified the Polya urn a little to continue study disclosure risk. For the relation between Polya distribution and Bayesian Learning and consequently application in pricing system, see Jonssan(2003).

Remark 4.6.1. For more, see Martin and Yo (2002), Pemantle and Skyrms (2000).

Chapter 5

Further Development

We have seen that many people proved theoretical results of GPU models at the same time applying them to some field. This is decided by the strong applicable property of GPU models. And this trend will continue. But from some point of view, I think, we need to "focus" on purely and systematically theoretical studies, because the dependence on applications makes the structure of GPU models scattered and being a little more closed to other mathematical fields may introduce new methods to our study.

For applications, there're still a lot more potential. First, some theoretical results of GPU models have not been used, especially those for higher dimensions. Second, with the development of randomization in applied mathematics in recent years, GPU models are waiting to be employed and teamed with other stochastic models. Last, in some specific fields: 1) GPU models will continue working for biology

and biomedical trials. They might depend on more development in applications in statistics and the deeper study of nonhomogeneous GPU models seems having a bright future; 2) GPU models will be applied more in recent-developed fields in computer science. For example, the application in CAGD was exploited by Goldman in 1980's. But after almost 20 years, people picked it up again. So it should have a lot of potential, especially that higher dimensional GPU models might be able to be used in approximate images in higher dimensional space.

Bibliography

[1] Alajaji, F. and Fuja, T.: A communication channel modeled on contagion, IEEE Trans. on Info. Theo., vol.40, no.6 (1994)

[2] Arthur, B., Ermol'ev, Y. and Kaniovskii, Y.: A generalized urn problem and its applications, Cybernetics, vol. 19 (1983)

[3] Arthur, W. B., Ermoliev. Y. M. and Kaniovski, Y. M.: Strong laws for a class of path-dependent stochastic processes with applications, Proc. of International conference on stoch. optim. (1984)

[4] Arthur, W. B., Ermoliev. Y. M. and Kaniovski, Y. M.: Path-dependent processes and the emergence of macro-structure, European J. of Operat. Res., 30, pp.294-303 (1987)

[5] Arthut, W. B.: Competing technologies, increasing returns and lock-in by historical events, The Econ. J, 99, pp.116-131 (1989)

[6] Arthur, W. B. and Lane, D. A.: Information contagion, Stru. Chan. & Econ. Dyn., vol.4, no.1 (1993)

[7] Athreya, K. B. and Karlin, S.: Embedding of urn schemes into continuous time Markov branching processes and related limit theorems, Ann. of Math. Stat, vol.39, no.6, pp.1801-1817 (1968)

[8] Athreya, K. B. and Ney, P. E.: Branching Processes, Springer-Verlag (1972)

[9] Bagchi, A. and Pal. K.: Asymptotic normality in the generalized Polya-Eggenberger urn model, with an application to computer data structures, SIAM. J. Alg. Disc. Meth, vol.6, no.3 (1985)

[10] Bai, Z. D. and Hu, F.: Asymptotic theorems for urn models with nonhomogeneous generating matrices, Stoch. Proc. and Appl., vol.80, pp.87-101 (1998)

[11] Bai, Z. D., Chen, G. J. and Hu. F.: Limit theorems for urn models in time-dependent sequential trials, Chinese Ann. of Math., vol.22 A, pp.89-96 (2001)

[12] Bai, Z. D., Hu, F. and Rosenberger, W. F.: Asymptotic properties of adaptive designs for clinical trials with delayed response, Ann. of Stat., vol.30, no.1, pp. 122-139 (2002)

[13] Bai, Z. D., Hu, F. and Zhang, L. X.: Gaussian approximation theorems for urn models and their applications, Ann. of Appl. Prob., vol.12, no.4, pp.1149-1173 (2002)

[14] Bai, Z. D. and Hu, F.: Asymptotics in randomized urn models, Ann. of Appl. Prob., vol.15, no.1B, pp.914-940 (2005)

[15] Banerjee, A., Burlina, P. and Alajaji, F.: Image segmentation and labeling using the Polya urn model, IEEE Trans on Image Proc., vol.8, no.9 (1999)

[16] Barry, P. L. and Goldman, R. N.: Interpolation and approximation of curves and surfaces using Polya polynomials, Graphical Models and Image Proc., vol.53, no.2, pp.137-148 (1991)

[17] Benaim, M and Hirsch, M. W.: Dynamics of Morse-Smale irn processes, Ergod. th. & Dyna,. Sys., vol.15, pp.1005-1030 (1995)

[18] Benaim, M.: Vertex-reinforced random walks and a conjecture of Pemantle, Ann. of Prob., vol.25, no.1, pp.361-392 (1997)

[19] Benaim, M., Schreiber, S. and Tarrés, P.: Generalized urn models of evolutionary processes, Ann. of Appl. Prob., vol.14, no.3, pp.1455-1478 (2004)

[20] Billingsley, P.: Probability and Measure, 3rd ed. Wiley (1995)

[21] Blackwell, D. and Kendall, D.: The martin boundary for Polya urn scheme and an application to stochastic population growth, J Appl. Prob., vol.1, pp.284-296 (1964)

[22] Blackwell, D. and MacQueen, J. B.: Ferguson distribution via Polya urn scheme, Ann. of Stat., vol.1, no.2, pp.353-355 (1973)

[23] Blom, G. and Holst, L.: Random walks of ordered elements with applications, Amer. Stat., vol.40, no.4 (1986)

[24] Chappell, R., Czado, C. and Newton, M. A.: Bayesian inference for semiparametric binary regression, J. of the Amer. Stat. Ass., vol.91, no.433 (1996)

[25] Chen, F.: An extended urn model with application to approximation, Trans. of the Amer. Math. Soc., vol.356, no.9, pp.3505-3515 (2003)

[26] Cotar, C. and Limic, V.: Attraction time for strongly reinforced walks, arXiv, math.PR/0612048 (2006)

[27] Cohen, M. P.: The Bayesian bootstrap and multiple imputation for unequal probability smaple designs, Proceedings of the Survey Research Methods Section, ASA (1997)

[28] Davis, B.: Reinforced random walk, Prob. Th. Rel. Fields, vol.84, pp.203-229 (1990)

[29] Devroye, L.: Limit laws for local counters in random binary search trees, Random Struc. and Algo., vol.2, no.3 (1991)

[30] Diaconis, P. and Freedman, D.: De Finetti's Theorem for Markov Chains, Ann. of Prob., vol. 8, no. 1 (1980)

[31] Diaconis, P. and Rolles, S.: Bayesian analysis for reversible Markov chains, *online* (2004)

[32] Dirienzo, A. G.: Using urn models for the design of clinical trials, Indian J. of Stat., vol.62, Series B, pp.43-69 (2000)

[33] Donnelly, P.: Partition structures, Polya urns, the Ewens sampling formula and the ages of alleles, Theor. Popul. Bio., no.30, pp.271-288 (1985)

[34] Drmota, M. and Vatutin, V.: In K. B. Athreya and P. Jagers, editors, Classical and modern branching processes (Minneapolis, MN, 1994), volume 84 of IMA Vol. Math. Appl., pages 89-110, Springer, New York (1997)

[35] Dubins, L. E. and Savage, L. J.: How to gamble if you must: Inequalities for stochastic processes, Dover (1976)

[36] Dunson, D. B.: Bayesian Density Regression, *online* (2005)

[37] Durham, S., Flournoy, N. and Li, W.: A sequential design for maximizing the probability of a favorable response, Can. J. Statist., vol.26, pp.479-495 (1998)

[38] Durrett, R.: Probability: Theory and Examples, 3rd edition, Thomson (2005)

[39] Eggenberger, F. and Pólya, G: $\ddot{U}$ber die statistik vorketter vorgänge, Zeit. Angew. Math. Mech., 3:279-289 (1923)

[40] Escobar, M. D. and West, M.: Computing nonparametric hierarchial models, *online* (1992)

[41] Ewens, W. J. and Tavare, S.: The Ewens sampling formula, Multivariate discrete distributions by Johnson, N. L., Kotz, S. and Balakrishnan. Wiley (1995)

[42] Feller, W.: An introduction to probability theory and its applications, vol. I and II, 2nd edition, Wiley (1966)

[43] Fienberg, S. E. and Makov, U. E.: Uniqueness, urn models and disclosure risk, Research in official stat. (2001)

[44] Flajolet, P., Dumas, P., Puyhaubert, V.: Some exactly solvable models of urn process theory, Disc. Math. & Comp. Sci., vol. AG, 59-118 (2006)

[45] Flajolet, P., Gabarro, J. and Pekari, H.: Analytic urns, Ann. of Prob., to appear (2004)

[46] Frank, R. E.: Brand choice as a probability process, J. of Business, vol.35, no.1, pp.43-56 (1962)

[47] Freedman, D. A.: Bernard Friedman's urn, Ann. of Math. Stat., vol.36, no.3, pp.956-970 (1965)

[48] Freedman, D. A.: Equivalence and singularity for Friedman urns, Ann. of Math. Stat., vol.37, no.1, pp.268-270 (1966)

[49] Friedman, B.: A simple urn model, Comm. Pure Appl. Math. vol.29, pp.59-70 (1949)

[50] Goldman, R. N.: Polya's urn model and computer aided geometry design, SIAM J. Alg. Disc. Meth., vol.6, no.1 (1985)

[51] Goldman, R. N.: Urn models and Beta-splines, IEEE Comp. Geom. and Comp. Aided Design (1986)

[52] Goldman, R. N.: Urn models, approximations and splines, J. of Appr. Theory, vol.54, pp.1-66 (1988)

[53] Gouet, R.: Martingale functional central limit theorems for a generalized Polya urn, Ann. of Prob., vol.21, no.3, pp.1624-1639 (1993)

[54] Griffiths, T. L., Navarro, D. J., M. Steyvers and Lee, M. D.: Modeling individual differences using Dirichlet process, preprint of J. of Math. Psychology (2005)

[55] Higueras, I, Moler, J., Plo, F. and San Miguel, M.: Urn models and differential algebraic equations, J. Appl. Prob., vol.40, pp.401-412 (2003)

[56] Higueras, I, Moler, J., Plo, F. and San Miguel, M.: An urn model with non-deterministic total replacement, Monograffas del Semin. Matem Garcia de Galdeano, vol.27, pp.337-343 (2003)

[57] Hill, B. M., Lane, D. and Sudderth, W.: A strong law for some generalized urn processes, Ann. of Prob., vol.8, no.2, pp.214-226 (1980)

[58] Hill, B. M., Lane, D. and Sudderth, W.: Exchangeable urn processes, Ann. of Prob., vol.15, no.4, pp.1586-1592 (1987)

[59] Holst, L.: Unified approach to limit theorems for urn models, J. Appl. Prob., vol.16, pp.154-162 (1979)

[60] Hoppe, F. M.: The sampling theory of neutral alleles and an urn model in population genetics. J. of Math. Bio., no.25, pp.123-159 (1987)

[61] Inoue, K. and Aki, S.: Polya urn models under general replacement schemes, J. Japan Stat. Soc., vol.31, no.2, pp.193-205 (2001)

[62] Ishwaran, H.: Exponential posterior consistency via generalized Polya urn schemes in finite semiparametric mixtures, Ann. of Stat., vol.26, no.6, pp.2157-2178 (1998)

[63] Ishwaran, H and Zarepour, M.: Random probability measures wia Polya sequences: revisiting the Blackwell-MacQueen urn scheme, *online* (2003)

[64] Ivanova, A., Rosenberger, W. F., Durham, S. D. and Flournoy, N.: A birth and death urn for randomized clinical trials: asymptotic methods, Indian J. of Stat., vol.62, Series B, pp.104-118 (2000)

[65] Janardan, K. G.: Correlation between the numbers of two types of children in a family, using the Markov-Polya model, Math. Bios., no.62, pp.123-136 (1982)

[66] Janson, S.: Functional limit theorems for multitype branching processes and generalized Polya urns, Stoch. Proc. Appl., vol.110, no. 2, pp.177-245 (2004)

[67] Janson, S.: Limit theorems for triangular urn schemes, preprint in Prob. Th. Rel. Fields (2004)

[68] Johnson, N. L. and Kotz, S.: Urn models and their application, John Wiley & Sons (1977)

[69] Kotz, S. and Balakrishnan, N.: Advances in Urn models during the Past Two Decades, Advances in Combinatorial Methods and Applications to Probability and Statistics, Birkhäuser, 1997

[70] Kotz, S., Mahmoud, H. and Robert, P.: On generalized Polya urn models, Stat. & Prob. Letters, vol.49, pp.163-173 (2000)

[71] Kovchegov, Y.: Multi-particle processes with reinforcements, *online* (2005)

[72] Lo, A. Y.: A Bayesian bootstrap for a finite population, Ann. of Stat., vol.16, no.4, pp.1684-1695 (1988)

[73] Mahmoud, H. M. and Smythe, R. T.: On the distribution of leaves in rooted subtrees of recursive trees, Ann. of Appl. Prob., vol.1, no.3, pp.406-418 (1991)

[74] Mahmoud, H. M.: Polya urn models in random trees, *online* (1997)

[75] Mahmoud, H. M.: On rotations in fringe-balanced binary trees, Info. Proc. Letters, vol.65, pp.41-46 (1998)

[76] Martin, C. F. and Ho, Y. C.: Value of information in the Polya urn process, Info. Scie., vol.147, pp.65-90 (2002)

[77] Mauldin, R. D., Sudderth, W. D. and Williams, S. C.: Polya trees and random distributions, Ann. of Stat., vol.20, no.3, pp.1203-1221 (1992)

[78] Mckenzie, A. and Steel, M.: Distributions of cherries for two models of trees, Math. Biosciences, 164, pp. 81-92 (2000)

[79] Meeden, G.: A noninformative Bayesian approach for two-stage cluster sampling, Indian J. of Stat., vol.61, Series B, pp.133-144 (1999)

[80] Mori, T. F.: The maximum degree of the Barabasi-Albert random tree, Math. Sub. Class. (1991)

[81] Muliere, P. and Walker, S.: Extending the family of Bayesian bootstraps and exchangeable urn schemes, J. R. Stat. Soc. B, vol.66, pp.175-182 (1998)

[82] Muliere, P., Secchi, P. and Walker, S. G.: Urn schemes and reinforced random walks, Stoch. proc. and their appl., no.88, pp.59-78 (2000)

[83] Muliere, P. and Secchi, P.: Urn schemes for constructing priors, from 'Bayesian theory in a completely predictive perspective' Wiley (2000)

[84] Nelson, D. and Meeden, G.: Using prior information about population quantiles in finite population sampling, Indian J. of Stat., vol.60, Series A, pp.426-445 (1998)

[85] Oliveira, R: The onset of dominance in balls-in-bins processes with feedback, arXiv, math.PR/0510415 (2006)

[86] Panholzer, A. and Prodiger, H.: An analytic approach for the analysis of rotations in fringe-balanced binary search trees, Ann. of Comb., vol.2, pp.173–184 (1998)

[87] Pemantle, R.: Random processes with reinforcement, Ph. D. thesis, M.I.T. (1988)

[88] Pemantle, R.: Phase transition in reinforced random walk and RWRE on trees, Ann. of Prob., vol.16, no.3, pp.1229-1241 (1988)

[89] Pemantle, R.: A time-dependent version of Polya's urn, J. of Th. Prob., vol.3, no.4 (1988)

[90] Pemantle, R.: Nonconvergence to unstable points in urn models and stochastic approximations, Ann. of Prob., vol.18, no.2, pp698-712 (1990)

[91] Pemantle, R.: When are touchpoints limits for generalized Polya urns? Proc. of Amer. Math. Soc., vol.113, no.1, pp.235-243 (1991)

[92] Pemantle, R. and Skyrms, B.: A dynamic model of social network formation, PNAS, vol.97, no.16 (2000)

[93] Pemantle, R.: Random processes with reinforcement, preprint (2001)

[94] Pemantle, R. and Skyrms, B.: Network formation by reinforcedment learning: the long and medium run, Math. Soc. Sci. to appear (2003)

[95] Pemantle, R. and Volkov, S.: Vertex-reinforced random walk on $\mathbb{Z}$ has finite range, Ann. of Prob., vol.27, no.3, pp.1368–1388 (1999)

[96] Rosenberger, W. F.: Asymptotic inference with response-adaptive treatment allocation designs, Ann. of Stat., vol.21, no.4, pp.2098-2107 (1993)

[97] Rosenberger, W. F.: New directions in adaptive designs, Stat. Sci., vol.11, no.2, pp.137-149 (1996)

[98] Rosenberger, W. F. and Grill, S. E.: A sequential design for psychophysical experiments: an application to estimating timing of sensory events, Stat. in Med., vol.16, pp.2245-2260 (1997)

[99] Robbins, H. and Monro, S.: A stochastic approximation method, Ann. of Math. Stat., vol.22, no.3, pp.400-407 (1951)

[100] Rogers, L. C. G. and Williams, D.: Diffusions, Markov Processes, and Martingales, vol.1, 2nd edition, Wiley (1994)

[101] Rolles, S. W. W.: How edge-reinforced random walk arises naturally, Prob. Th. Rel.Fields, vol.126, no.2, pp.243-260 (2003)

[102] Rubin, D. B.: The Bayesian bootstrap, Ann. of Stat., vol.9, no.1, pp.130-134 (1981)

[103] Samuel, A. M.: A Bayesian, species-sampling-inspired approach to the uniques problem in microdata disclosure risk assessment, J. of official Stat., vol.14, no.4, pp373-383 (1998)

[104] Schouten, H. J.: Adaptive biased urn randomization in small strata when blinding is impossible. Biometrics, wol.51, no.4, pp.1529-1535 (1995)

[105] Schreiber, S. J.: Urn models, rreplicator processes and random genetic drift, SIAM J. Appl. Math., vol.61, no.6, pp.2148-2167 (2001)

[106] Simon, R.: Adaptive treatment assignment methods and clinical trials. Biometrics. vol.33. no.4. pp.743-749 (1977)

[107] Smythe, R. T. and Wei, L. J.: Significance tests with restricted randomization design, Biometrika, vol.70, no.2, pp.496-500 (1983)

[108] Smythe, R. T.: Central limit theorems for urn models, Stoc. Proc. and Their Appl., vol.65, pp.115-137 (1996)

[109] Walker, S. and Muliere, P.: Beta-Stacy processes and a generalization of Polya-urn scheme, Ann. of Stat., vol.25, no.4, pp.1762-1780 (1997)

[110] Wei, J. L.: A class of designs for sequential clinical trials, J. of Amer. Stat. Asso., vol.72, no.358, pp.382-386 (1977)

[111] Wei, L. J.: An application of an urn model to the design of sequential controlled clinical trials, J. of Amer. Stat. Asso., vol.73, no.363, pp.559-563 (1978)

[112] Wei, L. J.: The generalized Polya urn design for sequential medical trials, Ann. of Stat., vol.7, no.2, pp.291-296 (1979)

[113] Wei, L. J., Smythe, R. T., Lin, D. Y. and Park, T. S.: Statistical inference with data-dependent treatment allocation rules, J. of Amer. Stat. Asso., vol.85, no.409 (1990)

[114] Xue, J.: A Polya urn model for conformity. *online* (2005)

[115] Zaman, A.: Urn models for Markov exchangeability. Ann. of Prob., vol.12, no.1, pp.223-229 (1984)

[116] Zhao, X. F. and Gong, Z. T.: Digital image blending technology based on Polya model, Research on Comp. Appl., vol.7 (2003)

[117] Zhu, C. and Chen. G.: Some optimal adaptive designs in clinical trials, Chinese J. of Appl. Prob., vol.21, no.1 (2005)

[118] Zou, G. H. and Liu, X. J.: A generalization of Polya Urn Model, J. of Math. for Tech., vol. 10, no. 4 (1994)