JIRSS (2003)
Vol. 2, No. 1, pp 53 - 114

Pólya Urn Models and Connections to Random Trees: A Review

Hosam M. Mahmoud

Department of Statistics, The George Washington University, Washington, D.C. 20052, U.S.A. (hosam@gwu.edu)

Abstract. This paper reviews Pólya urn models and their connection to random trees. Basic results are presented, together with proofs that underly the historical evolution of the accompanying thought process. Extensions and generalizations are given according to chronology:

- Pólya-Eggenberger's urn
- Bernard Friedman's urn
- Generalized Pólya urns
- Extended urn schemes
- Invertible urn schemes

Connections to random trees are surveyed. Numerous applications to trees common in computer science are discussed, including:

Received: February 2003

Key words and phrases: Pólya urns, Poissonization, random trees, stochastic process.

- Binary search trees
- Fringe-balanced trees
- m-ary search trees
- 2–3 trees
- Paged binary trees
- Bucket quad trees
- Bucket k–d trees

The applications also include various types of recursive trees:

- Standard recursive trees
- Pyramids
- Plane-oriented recursive trees
- Phylogenetic trees
- Bucket recursive trees
- Sprouts

Limit distributions, and phase changes therein are presented within the unifying theme of Pólya urn models.

1 Introduction

Pólya urn models are urns of colored balls with replacement schemes. Balls are sampled at random from the urn and, depending on the color of the ball withdrawn, balls of various colors are replaced in the urn.

Initially, these urns were intended to model contagion (Eggenberger and Pólya, 1923). Epidemics and other such spreading phenomena have a branching nature within a population. Thus, steadily Pólya urn schemes acquired importance in all branching phenomena, such as chain letters, and many phenomena with an underlying random tree structure.

The intent of this article is to review Pólya urn schemes. We present basic results as they came by chronologically, and we sketch their original proof to provide hints on the broad array of methods

employed, and the evolutionary thought process leading to the current state of the art. Numerous applications in random trees are discussed. Being a review article, it is our intention to provide a wide survey of the associated literature, too.

The sections of the paper are organized as follows. In Section 2 the notation for a working language is specified. We define precisely the class of urns the paper is dealing with in Section 3. A word on tenability is mentioned in Section 4. In Section 5 we portray the genesis and classical foundations of Pólya urn theory. In Section 6 we sketch the more modern developments of this still-burgeoning theory. In Section 7 we give several applications to trees arising in a variety of computer science applications involving random trees. Other practical settings such as pyramid schemes are taken up in Section 8. Other directions and possible extensions for future research are outlined in Section 9.

2 Notation

The number $H_n = \sum_{i=1}^{n} 1/i$ is the nth harmonic number. The notation $\stackrel{\mathcal{D}}{=}$ is for exact equality in distribution, whereas $\stackrel{\mathcal{D}}{\longrightarrow}$ is for convergence in distribution. Likewise, $\stackrel{P}{\longrightarrow}$ and $\stackrel{a.s.}{\longrightarrow}$ are respectively for convergence in probability and almost surely.

The following abbreviations will be used for standard random variables:

$\beta(a, b)$	Beta with parameters (a, b)
$B(n, p)$	Binomial on n trials with rate of success p
$Ber(p)$	Bernoulli with rate of success p
$Geo(p)$	Geometric with rate of success p per trial
$\mathcal{N}(\mu, \sigma^2)$	Normal variate with mean μ and variance σ^2
$\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate normal vector in d dimensions with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

Vectors and matrices are represented in boldface. For economy of space a column vector $\mathbf{V}$ will be written horizontally in the transposed form $\mathbf{V}^T$. The vector of k ones is written as $\mathbf{J}_k$. The notation $\mathbf{o}_k(n)$ stands for a k-component vector all the components of which are $o(n)$ in the usual scalar sense. If $\lambda_1, \ldots, \lambda_k$ are the roots of the characteristic equation of a $k \times k$ matrix, they will be arranged according to their decreasing real parts, that is, $\Re\,\lambda_1 \geq \Re\,\lambda_2 \geq \ldots \geq \Re\,\lambda_k$.

3 Growing Pólya urns

A *Pólya urn* is an urn containing balls of up to k different colors. The urn evolves in discrete time steps. At each step, a ball is sampled uniformly at random (all balls being equally likely). The color of the ball withdrawn is observed, and the ball is returned to the urn. If at any step the color of the ball withdrawn is i, $i = 1, \ldots, k$, then A_{ij} balls of color j are placed in the urn, $j = 1, \ldots, k$, where A_{ij} follows a discrete distribution on a set of integers. Generally speaking, the entries A_{ij} can be deterministic or random, positive or negative.

It is customary to represent the urn scheme by a square ball addition matrix, or *schema*:

$$\mathbf{A} = [A_{ij}], \qquad i, j = 1, \ldots, k,$$

the rows of which are indexed by the color of the ball picked, and the columns are indexed by the color of the balls added. We call the expected value $\mathbf{E}[\mathbf{A}]$ the *generator*. The primary interest lies in the long-term composition of the urn and in the stochastic path leading to it. So, the number of balls of each color and the number of *splits* of a particular color (the number of times a ball of that color is drawn) are examples of important parameters.

Remark: *In the case of 2×2 schemata, we shall always think of the two colors as white and blue, with the top row corresponding to additions upon drawing a white ball, and the first column corresponding to the number of white balls added. In this case, we shall use W_n to denote the number of white balls after n draws, and B_n to denote the number of blue balls after n draws; with W_0 and B_0 being the initial conditions. The total number of balls after n draws is $\tau_n = W_n + B_n$.*

Toward an asymptotic theory, we need our urn to withstand the test of time. We shall deal only with *tenable* schemes—urns that remain feasible no matter which stochastic path is being followed. In a tenable urn, it is always possible to indefinitely draw balls according to the rules. For example, the instance

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

of Bernard Friedman's urn is tenable, whichever nonempty initial state it starts in. (In fact an urn is tenable, if all A_{ij} are nonnegative,

under any nonempty starting conditions.) By contrast, an urn scheme of white and blue balls with the schema

$$\begin{pmatrix} -1 & -X \\ 3 & 4 \end{pmatrix}, \tag{1}$$

with X being a $Ber(\frac{4}{7})$ random variable, may or may not be tenable, depending on the initial conditions. This urn is not tenable if it starts out with more white balls than blue: for instance, if the urn starts out with three white balls and two blue balls, it is not possible to perpetuate drawing according to the rules along some stochastic paths; if the event that a white ball is drawn and $X = 1$ persists three times, which is one possible stochastic path, on the third draw the urn cannot progress. On the other hand, if $W_0 < B_0$ the urn is obviously tenable. Even on the most resistant path to growth, when $X = 1$ always persists whenever there is a chance to pick white balls, the number of white balls cycles in the set $\{0, 1, 2, 3\}$, and the number of blue balls after n draws is $\frac{1}{4}n + O(1)$, as $n \to \infty$.

Toward asymptotics, we shall also consider only *growing* urns, or urns the size of which grows to infinity on all possible stochastic paths. It does not mean that the urn grows to infinite size that the number of balls necessarily increases after each draw. The urn size is allowed to decrease occasionally, but to grow the scheme will reverse a transient decreasing streak. The term "growing urn" is a measure of size at infinite time. The schema (1), growing with $W_0 < B_0$, is one such urn.

4 A word on tenability

Not much attention has been devoted to tenability issues. Most research effort was spent on proving results for urns that are already known to be tenable, but not much has been said about when an urn is tenable. Balaji and Mahmoud (2003+), is a modest attempt to characterize the tenability of 2×2 schemes of the general deterministic form

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Generally speaking, any 2×2 generator with two negatives in the same column is not tenable, because we are always depleting the color corresponding to that column. So, a matrix with four or

three negatives cannot be tenable, because two negatives must be in the same column. The tenuous case is that with two negatives. If a generator has two negatives, they cannot be in the same column. This leaves the cases

$$\underset{(i)}{\begin{pmatrix} - & - \\ \oplus & \oplus \end{pmatrix}} \qquad \underset{(ii)}{\begin{pmatrix} \oplus & \oplus \\ - & - \end{pmatrix}} \qquad \underset{(iii)}{\begin{pmatrix} - & \oplus \\ \oplus & - \end{pmatrix}},$$

where $-$ indicates a negative entry, and $\oplus$ indicates an entry that is positive or 0. The cases (i) and (ii) are symmetric through the renaming of the colors. The observation in Mahmoud and Balaji (2003+), is that the urn (i) is tenable if:

- W_0 and c are both multiples of $|a|$.
- $\det(\mathbf{A}) \leq 0$.
- $\det \begin{pmatrix} a & b \\ W_0 & B_0 \end{pmatrix} < 0$.

Case (iii) requires less stringent conditions. This is a 2×2 special case of the $k \times k$ tenable urn scheme studied in Gouet (1997). It suffices to have

- W_0 and c are both multiples of $|a|$.
- B_0 and b are both multiples of $|d|$.
- At least one of the entries b or c is positive.

The tenability case of only one negative is easy to characterize.

The characterization in all cases is argued by considering the "most critical path" which depletes the urn whenever possible. The core of the argument is to find conditions to return recursively to a critical state where one color is depleted.

5 Classical development

By classical development we refer to all the relevant materiel that can be found in textbooks. Johnson and Kotz (1977) is a classic in this field. Kotz and Balakrishnan (1997) is a companion survey of Pólya urn models that goes into many more offshoots and derived urns (not discussed in the present paper), rather than getting into connections

to random trees as in the current survey. Kotz and Balakrishnan (1997) also has a more pronounced combinatorial flavor; the current survey gets more into asymptotics. Athreya and Ney (1972) puts a generalized model in the perspective of the branching process. More recent contributions are dubbed "modern" and relegated to the next section of this article. Pólya urns also appear in some classic books such as Fréchet (1943) and Feller (1971).

5.1 The Pólya-Eggenberger Urn

The earliest studies of Pólyaurns focused on 2×2 schemata. One of the very first studies is Eggenberger and Pólya(1923), but it is reported that the model had been considered in Markov (1917) and Tchuprov (1922). The model was discussed further in Pólya(1931). Eggenberger and Pólya(1923) is concerned with the the fixed schema

$$\begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix}, \tag{2}$$

where one adds to the urn s (a positive integer) balls of the same color as the ball withdrawn. This urn is commonly known as the Pólya-Eggenberger urn (sometimes referred to in casual writing as Pólya's urn). Much of the rest of the ensuing theory is generalization in many different directions.

It was natural in the first approach to the problem to seek discrete distributions underlying the process in exact form (in the style of 19th Century research). Indeed, the discrete distribution found in the Pólya-Eggenberger defined a fundamentally new distribution.

Theorem 5.1.1. *(Eggenberger and Pólya, 1923). Let $\tilde{W}_n$ be the number of white splits in the Pólya-Eggenberger urn after n draws. Then,*

$$P\{\tilde{W}_n = k\} = \frac{W_0(W_0+s)\ldots\big(W_0+(k-1)s\big)B_0(B_0+s)\ldots\big(B_0+(n-k-1)s\big)}{\tau_0(\tau_0+s)\ldots\big(\tau_0+(n-1)s\big)}\binom{n}{k}.$$

Proof (sketch). The standard proof bears an idea similar to the derivation of the binomial law on n independent trials. The difference

is in that in the binomial case the trials are identical, but here the probabilities are adaptive in time.

In a string of n draws achieving k white splits, there has to be $n-k$ blue draws. Suppose $1 \le i_1 \le i_2 \le \ldots \le i_k \le n$ are the time indexes of the white draws. The probability of this particular string is

$$\begin{aligned}\frac{B_0}{\tau_0} \times \frac{B_0+s}{\tau_1} \times \frac{B_0+2s}{\tau_2} \times \cdots \times \frac{B_0+(i_1-2)s}{\tau_{i_1-2}} \times \frac{W_0}{\tau_{i_1-1}} \\ \times \frac{B_0+(i_1-1)s}{\tau_{i_1}} \times \cdots \times \frac{B_0+(i_2-3)s}{\tau_{i_2-2}} \\ \times \frac{W_0+s}{\tau_{i_2-1}} \times \frac{B_0+(i_2-1)s}{\tau_{i_2}} \times \cdots \times \frac{B_0+(n-k-1)s}{\tau_{n-1}}.\end{aligned}$$

Note that this expression does not depend on the string of indexes. The time indexes can be chosen in $\binom{n}{k}$ ways. □

Calculations involving combinatorial reductions of the forms found in Riordan (1968) yield the moments.

Corollary 5.1.1. *(Eggenberger and Pólya, 1923). Let W_n be the number of white balls in the Pólya-Eggenberger urn after n draws. Then,*

$$\begin{aligned}\mathbf{E}[W_n] &= \frac{W_0}{\tau_0} sn + W_0. \\ \mathbf{Var}[W_n] &= \frac{W_0 B_0 s^2 n(sn+\tau_0)}{\tau_0^2(\tau_0+s)}.\end{aligned}$$

Theorem 5.1.2. *(Eggenberger and Pólya, 1923). Let $\tilde{W}_n$ be the number of white splits in the Pólya-Eggenberger urn after n draws. Then,*

$$\frac{\tilde{W}_n}{n} \xrightarrow{a.s.} \beta\Big(\frac{W_0}{s}, \frac{B_0}{s}\Big).$$

Proof (sketch). Assume both W_0 and B_0 to be greater than 0. If either one is 0, we have a degenerate urn, progressing on only on one color, with no randomness. In this case Theorem 5.1.1 remains valid through the appropriate boundary interpretation. The present

theorem also remains valid through the appropriate interpretation of a beta distribution, when one of its parameters is 0.

The proof is based on the appropriate passage to the limit. Rewrite the exact distribution of Theorem 5.1.1 as

$$P\{\tilde{W}_n = k\} = \frac{\Gamma(k + W_0/s)\,\Gamma(n - k + B_0/s)}{\Gamma(W_0/s)\,\Gamma(B_0/s)\,\Gamma(n + \tau_0/s)\,/\,\Gamma(\tau_0/s)} \binom{n}{k}.$$

So, for $x \in [0, 1]$, the distribution function of the white splits is

$$P\{\tilde{W}_n \le nx\} = \sum_{k=0}^{\lfloor nx \rfloor} \frac{\Gamma(k + W_0/s)\,\Gamma(n - k + B_0/s)}{\Gamma(W_0/s)\,\Gamma(B_0/s)\,\Gamma(n + \tau_0/s)\,/\,\Gamma\big((W_0 + B_0)/s\big)} \binom{n}{k}.$$

Using Stirling's approximation to the gamma function and factorials in the binomial coefficient, proceed to the limit as $n \to \infty$ with

$$P\Big\{\frac{\tilde{W}_n}{n} \le x\Big\} \to \frac{\Gamma((W_0 + B_0)/s)}{\Gamma(W_0/s)\,\Gamma(B_0/s)} \int_0^x u^{W_0/s - 1}(1 - u)^{B_0/s - 1}\,du;$$

the right hand side is the distribution function of the $\beta(W_0/s, B_0/s)$ random variable. □

It is curious that the limiting properties of a Pólya-Eggenberger urn depend critically on the initial conditions.

5.2 Bernard Friedman's urn

Bernstein (1940), Savkevich (1940), and Bernard Friedman (1949) generalize the basic model (2) to one where one adds s balls of the same color, and a balls of the antithecal color:

$$\begin{pmatrix} s & a \\ a & s \end{pmatrix}. \tag{3}$$

For mathematical convenience, as well as æsthetics, Bernard Friedman (1949) (and most ensuing classical studies) stayed with the case of constant row sum.

The reason why this is convenient will be discussed when we get into the more modern approaches to the problem. Bernard Friedman (1949) develops a functional equation for the number of white balls. Recalling that $\tau_n = W_n + B_n = \tau_0 + sn$, we have a steady linear nonrandom rate of increase. Of course, the case $s = a$ is *degenerate*, where $W_n = W_0 + sn$. This degenerate case is of no interest, as there is no randomness in it.

Theorem 5.2.1. *(Friedman, 1949). Let W_n be the number of white balls in a nondegenerate Bernard Friedman's urn after n draws. The moment generating function $\phi_n(t) = \mathbf{E}[e^{W_n t}]$ satisfies the difference-differential equation*

$$\phi_{n+1}(t) = e^{at}\Big[\phi_n(t) + \frac{e^{(s-a)t}-1}{\tau_n}\phi_n'(t)\Big].$$

Proof (sketch). Let $\mathbf{1}_n^W$ and $\mathbf{1}_n^B \equiv 1 - \mathbf{1}_n^W$ be respectively the indicators of the events of drawing a white or a blue ball at the nth step. The number of white balls after $n+1$ draws is what it was after n steps, plus the addition (possibly negative) incurred by the ball sampled at step $n+1$:

$$W_{n+1} = W_n + s\mathbf{1}_n^W + a\mathbf{1}_n^B = W_n + (s-a)\mathbf{1}_n^W + a.$$

Then

$$\mathbf{E}[e^{W_{n+1}t} \,|\, W_n] = e^{(W_n+a)t}\,\mathbf{E}[e^{(s-a)\mathbf{1}_n^W t} \,|\, W_n]. \tag{4}$$

Further, we have the conditional expectation

$$\begin{aligned}\mathbf{E}[e^{(s-a)\mathbf{1}_n^W t} \,|\, W_n] &= \mathbf{E}[e^{(s-a)\mathbf{1}_n^W t} \,|\, W_n, \mathbf{1}_n^W = 0]\,\mathbf{Prob}\{\mathbf{1}_n^W = 0 \,|\, W_n\} \\ &+ \mathbf{E}[e^{(s-a)\mathbf{1}_n^W t} \,|\, W_n, \mathbf{1}_n^W = 1]\,\mathbf{Prob}\{\mathbf{1}_n^W = 1 \,|\, W_n\} \\ &= \Big[1 - \frac{W_n}{\tau_n}\Big] + \frac{W_n}{\tau_n}\,e^{(s-a)t}.\end{aligned}$$

Plug this into (4) and take expectations. □

The functional equation in Theorem 5.2.1 is not particularly easy to solve for any arbitrary combination of values of s and a. Nevertheless, explicit solutions are available for special cases, such as $a = 0$ (Pólya-Eggenberger's urn), and $s = 0$. Friedman (1949) suggested the transformation

$$\psi_n(t) = (1 - e^{-t(s-a)})^{\delta+\gamma n}\,\phi_n(t), \tag{5}$$

with

$$\delta := \frac{\tau_0}{s-a}, \qquad \gamma := \frac{s+a}{s-a}.$$

This gives a slightly simpler recurrences:

$$\psi_{n+1}(t) = \frac{e^{st}}{\tau_n}(1 - e^{-t(s-a)})^{\gamma+1}\,\psi_n'(t). \tag{6}$$

We shall discuss a solvable instance of this functional equation in recursive trees. The solution in this special case should give us general hints on how to approach the functional equation of Theorem 5.2.1.

Twentieth century research paid attention to simplifying results by focusing on the essential elements or "asymptotics." David Freedman (1965) develops an asymptotic theory for Bernard Friedman's urn.

Theorem 5.2.2. *(Freedman, 1965). Let W_n be the number of white balls in a nondegenerate Bernard Friedman's urn after n draws. Let $\rho = (s-a)/(s+a)$. If $\rho < \frac{1}{2}$, then*

$$\frac{W_n - \frac{1}{2}(s+a)n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{(s-a)^2}{4(1-2\rho)}\Big).$$

A proof will be given when we present results on the more general Bagchi-Pal urn. Freedman (1965) gives an expository section on Bernard Friedman's urn, where the results of Friedman (1949) are mostly presented in terms of the difference $W_n - B_n$: For $\rho < 1/2$ the limiting distribution is normal under the $\sqrt{n}$ scale, and it is interesting to note that in the case $\rho = \frac{1}{2}$, one needs a different norming factor to obtain a Gaussian limit distribution:

$$\frac{W_n - B_n}{\sqrt{n \log n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, (s-a)^2).$$

For $\rho > 1/2$ the behavior is radically different:

$$\frac{W_n - B_n}{n^\rho} \xrightarrow{\mathcal{D}} \beta\Big(\frac{W_0}{s}, \frac{B_0}{s}\Big).$$

It is curious to note that in the case $\rho < \frac{1}{2}$, the effect of any initial condition is washed out asymptotically. The urn balances itself in the long run—each color constitutes half the urn content on average. Contrast this with the case $\rho > \frac{1}{2}$, where the asymptotic proportion of colors depends critically on the initial conditions.

5.3 Poissonization

The next rather important, and most natural, generalization was developed in Athreya and Karlin (1968), where the scope extends to $k \times k$ generators for fixed as well as random schemata. These results are surveyed in Athreya and Ney (1972), which provides a comprehensive study of branching processes.

One main contribution in Athreya and Karlin (1968) is the idea of embedding the discrete urn process in a continuous-time Poisson Process. This connection later came to be known by the name *poissonization*. Athreya and Karlin use poissonization to derive a number of important results. They also rederive the results of Freedman (1965). Smythe (1996) extends the scope of Athreya and Karlin (1968) to schemata with negative diagonal elements. We shall present the results when we get to Smythe's *extended urn*. We focus first on explaining poissonization, or the equivalence of a continuous-time process (at certain opportune moments) to the discrete urn process.

Embedding a discrete stochastic processes in a continuous-time process had been utilized for some time prior to Athreya and Karlin (1968). Poissonization can be traced back to rudiments in Kac (1949). Translating the result back in terms of the discrete process can be done in principle, as stated in Athreya and Karlin (1968), but has always been fraught with difficulty in practice. In later decades, this translation was dubbed the term *depoissonization*. (Other forms of poissonization appear in informatics to solve functional equations and have a distinct analytic flavor (see Jacquet and Régnier, 1986, and Jacquet and Szpankowski, 1998), or is treated as a bona fide mathematical transform (see Gonnet and Munro, 1984, and Poblete, 1987).

To explain poissonization, we use an analogy from racing (Mahmoud, 2002). Suppose $\mathbf{A} = [A_{ij}]$ is the schema of a $k \times k$ urn. Endow every ball with an independent Poisson process with intensity 1. The process is compounded by a reward system that emulates the urn discrete process. A ball of color i produces an independent realization of A_{ij} children of color j, for $j = 1, \ldots, k$, at its points of renewal. Let us view the balls as contestants in a race. The runners are categorized into teams wearing shirts of the color of the ball they represent. When a runner of team i wins the race, an independent realization of A_{ij} runners (balls) wearing shirts of color j are added to the jth team, $j = 1, \ldots, k$. Each new runner carries a new independent Poisson process with the same reward system. At any point in time, given

the last renewal point, the lack of memory in the Poisson processes appears as if it resets all the Poisson processes afresh. We can view this Markovian system as if whenever a race is won, the referee's whistle is immediately blown to restart a race among all the existing runners—if a runner has covered a certain portion of the course in a race, the runner is not allowed to carry over any gain to the next race; the runner's remaining time to cover the rest of the course remains exponentially distributed (with parameter 1), as a result of resetting the race. Let $R_n^{(i)}$ be the number of runners in team i after n races, and let $\mathbf{R}_n = (R_n^{(1)}, \ldots, R_n^{(k)})^T$. By the independent identical distribution of running times, any of the runners is equally likely to win the race, that is

$$\mathbf{Prob}\{\text{team } i \text{ wins the } (n+1)\text{st race} \,|\, \mathbf{R}_n\} = \frac{R_n^{(i)}}{\sum_{\ell=1}^{k} R_n^{(\ell)}},$$

and an independent realization of A_{ij} runners of color j will be added, $j = 1, \ldots, k$, constituting a growth rule in the number of runners identical to that of the urn's growth under random sampling from a $k \times k$ Pólya urn with schema $\mathbf{A}$. In other words, $\mathbf{R}_n \overset{\mathcal{D}}{=} \mathbf{X}_n$, if the two processes start with identical initial conditions $\mathbf{R}_0 = \mathbf{X}_0$. However, $\mathbf{R}_n$ is only a discretized form of a continuous renewal process with rewards, the renewals of which are the starting whistle of the races, and the rewards of which at every renewal are determined by an independent realization of $\mathbf{A}$. It is helpful to think of $\mathbf{R}_n$, for $n = 1, 2, \ldots$, as a series of snapshots in time of the continuous process at the moments when a renewal takes place.

From this equivalence principle between the continuous-time and discrete -time processes Athreya and Karlin (1968) obtain a myriad of results, including those earlier results of Freedman (1965).

6 Modern developments

It was natural to think of breaking the perfect symmetry of Bernard Friedman's urn (3). Bagchi and Pal, 1985 considered the more general case

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{7}$$

for any four integers, so long as the choice is tenable, with the exception of a few intricate cases. Namely, they require a constant row

sum ($a+b=c+d=:K$), and to guarantee tenability, $b>0, c>0$, ad if $a<0$, then a divides W_0 and c, and if $d<0$, then d divides B_0 and b. We exclude degenerate cases: the case $b=c=0$, which is the Pólya-Eggenberger urn, and the case $a=c$, such a case has no randomness. The case where one minor diagonal element is zero ($bc=0$, $\max(b,c)>0$), is also excluded.

Theorem 6.1. *(Bagchi and Pal, 1985). Let W_n be the number of white balls after n draws from a nondegenerate Bagchi-Pal urn. If $a-c<\frac{1}{2}K$,*

$$W_n^* := \frac{W_n - \frac{c}{(b+c)}Kn}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{bcK(a-c)^2}{(b+c)^2(K-2(a-c))}\Big),$$

If $a-c=\frac{1}{2}K$,

$$W_n^* := \frac{W_n - \frac{c}{(b+c)}Kn}{\sqrt{n\ln n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, bc).$$

Proof (sketch). This theorem is proved by showing that the moments of W_n^* converge to those of the normal distribution specified; the normal distribution being uniquely characterized by its moments.

We illustrate the proof only for the mean and variance. Higher moments follow similar principles, but the calculations are more complex. We have the recurrence

$$P\{W_{n+1}=W_n+a\,|\,W_n\} = \frac{W_n}{\tau_n}; \tag{8}$$

$$P\{W_{n+1}=W_n+c\,|\,W_n\} = 1-\frac{W_n}{\tau_n}. \tag{9}$$

This gives conditionally

$$\mathbf{E}[W_{n+1}\,|\,W_n] = \Big(1+\frac{a-c}{\tau_n}\Big)W_n + c. \tag{10}$$

The unconditional expectation is therefore

$$\mathbf{E}[W_{n+1}] = \Big(1+\frac{a-c}{\tau_n}\Big)\mathbf{E}[W_n] + c,$$

for $n\geq 0$. The setting

$$Y_n = W_n - \frac{c}{b+c}\tau_n \tag{11}$$

puts the equation into an iteratable form:

$$\mathbf{E}[Y_{n+1}] = \Big(1 + \frac{a-c}{\tau_n}\Big)\mathbf{E}[Y_n].$$

The solution is obtained by unwinding the recurrence all the way back to $n = 0$:

$$\begin{aligned}\mathbf{E}[Y_n] &= \Big(W_0 - \frac{c}{b+c}\tau_0\Big)\prod_{j=0}^{n-1}\Big(1 + \frac{a-c}{\tau_j}\Big)\\ &= \Big(W_0 - \frac{c}{b+c}\tau_0\Big)\frac{\Gamma(\tau_0/K)\,\Gamma((\tau_n + (a-c))/K)}{\Gamma((\tau_0 + (a-c))/K)\,\Gamma(\tau_n/K)}.\end{aligned}$$

By the Stirling approximation of the Gamma function, one finds $\mathbf{E}[Y_n] = O(n^{(a-c)/K})$. Because $a - c \le \frac{1}{2}K$, we have

$$\mathbf{E}[W_n] \sim \frac{c}{b+c}\tau_n + O(\sqrt{n}\,);$$

the linear term is dominant.

Although we shall not develop the variance in detail, we shall take a brief look into its structure because the phase change in the theorem is worthy of notice. When $a - c = \frac{1}{2}K$, a different scale factor is required to produce a Gaussian limit law. This is because the variance has an essentially different form. To see this, write (8) and (9) in the form

$$\begin{aligned}P\{W_{n+1}^2 = (W_n + a)^2 \,|\, W_n\} &= \frac{W_n}{\tau_n};\\ P\{W_{n+1}^2 = (W_n + c)^2 \,|\, W_n\} &= 1 - \frac{W_n}{\tau_n}.\end{aligned}$$

With the transformation (11), we have a recurrence

$$\mathbf{E}[Y_{n+1}^2] = \Big(1 + \frac{2(a-c)}{\tau_n}\Big)\mathbf{E}[Y_n^2] + \frac{(b-c)(a-c)^2}{(b+c)\tau_n}\mathbf{E}[Y_n] + \frac{bc(a-c)^2}{(b+c)^2)}.$$

If $a - c < \frac{1}{2}K$, this recurrence asymptotically has a linear solution. By contrast, If $a - c = \frac{1}{2}K$, this recurrence simplifies to

$$\mathbf{E}[Y_{n+1}^2] = \frac{\tau_{n+1}}{\tau_n}\mathbf{E}[Y_n^2] + \frac{b^2 - c^2}{\tau_n}\mathbf{E}[Y_n] + bc.$$

the complementary solution of which is linear (in fact it is τ_n), but the recurrence in this case has a superlinear particular solution. □

Corollary 6.1. *In a nondegenerate Bagchi-Pal urn, if* $a - c \leq \frac{1}{2}K$, *then*

$$\frac{W_n}{n} \xrightarrow{a.s.} \frac{cK}{b+c}.$$

Proof (sketch). By Chebychev's inequality, for any fixed $\varepsilon > 0$,

$$P\Big\{\Big|\frac{W_n}{n} - \frac{cK}{b+c}\Big| > \varepsilon\Big\} \leq \frac{O(n \ln n)}{\varepsilon^2 n^2} \to 0.$$

Therefore,

$$\frac{W_n}{n} \xrightarrow{P} \frac{cK}{b+c}.$$

Almost sure convergence follows as the probability space is countable. (See the discussion in Chow and Teicher, 1978). □

The case where one minor diagonal element is zero ($bc = 0$, $\max(b, c) > 0$), has been excluded. It is handled via an elegant martingale technique in Gouet (1989) to show a strong law for W_n/n, and via the functional central limit theorem in Gouet (1993) to show convergence to a normal law modulated by a multiplicative independent random variable.

6.1 Extended urn schemes

The Bagchi-Pal urn provided a prelude that motivated the studies of Gouet (1997) of deterministic $k \times k$ analogues, and Smythe (1996) of random schemata with similar constraints.

Gouet (1997) considers $k \times k$ deterministic analogues of the Bagchi-Pal urn. In Gouet's urn, with schema $\mathbf{A} = [a_{ij}]$, the row sum is constant. Negative entries on the diagonal are allowed, so long as the urn remains tenable. Let $X_n^{(i)}$ be the number of nodes of color i after n draws. Tenability is guaranteed if certain divisibility conditions are satisfied: if $a_{jj} < 0$, then a_{jj} divides $X_0^{(j)}$, and a_{ij}, $i = 1, \ldots k$. A first-order theory is found in Gouet, 1997, according to which $n^{-1}X_n^{(i)}$ converges almost surely to a beta distribution, and jointly these beta distributions are marginals of a Dirichlet distribution.

Smythe (1996) goes one step further with the nondeterministic analog of Gouet's urn. Moreover, Smythe (1996) finds the second-order theory for these urns, that is the rate of convergence in Gouet's strong laws.

An urn is an *extended Pólya urn* if it is tenable (requiring some divisibility conditions) and its generator $\mathbf{E}[\mathbf{A}] := [a_{ij}]$ satisfies the conditions:

(i) $a_{ij} > 0$, for $i \neq j$.

(ii) The row sums are equal: for each $1 \leq i \leq k$, $\sum_{j=1}^{k} a_{ij} = \lambda_1 > 0$.

(iii) Each random entry has finite second moment.

(iv) For any nonprincipal eigenvalue λ_i, $i = 2, \ldots, k$, we have $\Re\lambda_i < \frac{1}{2}\lambda_1$.

(v) All eigenvalues are simple.

(vi) No two distinct complex eigenvalues have equal real part, except for conjugate pairs.

(vii) The eigenvectors are linearly independent.

(viii) There are no purely imaginary eigenvalues.

Under conditions (i)–(viii) the generator has a number of appealing properties that makes it possible to derive results.

Theorem 6.1.1. *(Smythe, 1996). Suppose* $\mathbf{A}$ *is the* $k \times k$ *schema of a tenable extended urn, with a generator* $\mathbf{E}[\mathbf{A}]$ *that has principal eigenvalue* λ_1*, and corresponding* k*-component left (row) eigenvector* $\mathbf{V}_1^T$*. Let* $X_n^{(i)}$ *be the number of balls of color* i *after* n *draws from the urn, and* $\mathbf{X}_n := (X_n^{(1)}, \ldots, X_n^{(k)})^T$*. Then*

$$\frac{1}{\sqrt{n}}(\mathbf{X}_n - \lambda_1 n \mathbf{V}_1) \xrightarrow{\mathcal{D}} \mathcal{N}_k(\mathbf{0}, \mathbf{\Sigma}),$$

for some limiting covariance matrix $\mathbf{\Sigma}$*.*

Proof (sketch). Under conditions (i)—(ii), $\mathbf{E}[\mathbf{A}] = [a_{ij}]$ is a Metzler-Leontieff matrix, and enjoys certain properties, such as, for example, having one principal eigenvalue that equals the sum across any row, and the components of the corresponding eigenvector are all nonnegative. Suppose $\mathbf{V}_1 = (v_1, \ldots, v_k)^T$. Let the total number of balls

after n draws be $\tau_n = \lambda_1 n + \tau_0$. Consider color 1; according to the ball addition rules, conditionally we have

$$\begin{aligned}\mathbf{E}[X_n^{(1)} \,|\, \mathbf{X}_{n-1}] &= X_{n-1}^{(1)} + \mathbf{E}\Big[A_{11}\frac{X_{n-1}^{(1)}}{\tau_{n-1}} + \cdots + A_{k1}\frac{X_{n-1}^{(k)}}{\tau_{n-1}} \,\Big|\, \mathbf{X}_{n-1}\Big] \\ &= X_{n-1}^{(1)} + \frac{1}{\tau_{n-1}}\Big(\mathbf{E}[A_{11}]\, X_{n-1}^{(1)} + \cdots + \mathbf{E}[A_{k1}]\, X_{n-1}^{(k)}\Big).\end{aligned}$$

Similar recurrence equations can be written for the other colors, and we can put them together in the matric form

$$\mathbf{E}[\mathbf{X}_n \,|\, \mathbf{X}_{n-1}] = \Big(\mathbf{I} + \frac{1}{\tau_{n-1}}\mathbf{E}[\mathbf{A}^T]\Big)\mathbf{X}_{n-1}. \tag{12}$$

When this recurrence is solved, one gets asymptotically

$$\mathbf{E}[\mathbf{X}_n] \sim \lambda_1 n \mathbf{V}_1.$$

As a hint on how the dominant eigenvector comes into the picture, the various conditions on the schema, particularly condition (vii), tell us that $\mathbf{E}[\mathbf{A}^T]$, can be represented as $\mathbf{M}\,\mathbf{diag}(\lambda_1, \ldots, \lambda_k)\,\mathbf{M}^{-1}$, where **diag** is a digonal matrix with the specified elements on its diagonal, and $\mathbf{M}$ is a modal matrix of $\mathbf{E}[\mathbf{A}^T]$. [1] When this representation is plugged into (12) and the equation is itereated, it gives the average result after a lengthy, but straighforwad, computation.

Let us asymptotically center the ball counts by setting

$$\tilde{X}_n^{(i)} = X_n^{(i)} - \lambda_1 v_i n.$$

Then

$$\begin{aligned}\mathbf{E}[\tilde{X}_n^{(i)} - \tilde{X}_{n-1}^{(i)} \,|\, \mathbf{X}_{n-1}] &= \mathbf{E}[X_n^{(i)} - X_{n-1}^{(i)} \,|\, \mathbf{X}_{n-1}] - \lambda_1 v_i \\ &= \frac{1}{\tau_{n-1}}\sum_{r=1}^{k}\mathbf{E}[A_{ri}]\, X_{n-1}^{(r)} - \lambda_1 v_i.\end{aligned}$$

Thus,

$$q_n^{(i)} := \tilde{X}_n^{(i)} - \tilde{X}_{n-1}^{(i)} - \frac{1}{\tau_{n-1}}\sum_{r=1}^{k}\mathbf{E}[A_{ri}]\, X_{n-1}^{(r)} + \lambda_1 v_i$$

[1] A *modal matrix* of a given $k \times k$ matrix with k linearly independent eigenvectors is the $k \times k$ matrix, the columns of which are the k eigenvectors of the given matrix.

is a martingale difference. For any constants $b_{jn}^{(r)}$, the combinaion

$$R_n := \sum_{j=1}^{n} \sum_{i=1}^{k} b_{ji}^{(i)} q_j^{(i)},$$

is a martingale. The main idea in the rest of the proof is to take an arbitrary linear combination $W_n = \sum_{i=1}^{k} \alpha_i \tilde{X}_n^{(i)}$ for any constants $\alpha_1, \ldots, \alpha_k$ (not all equal to zero) and approximate it by a true martingale via asymptotically negligible adjustments. This steers the proof toward the Cramér-Wold device (see Billingsley, 1968). That is, we approximate W_n by R_n by setting in R_n the coefficients of $X_m^{(i)}$ to 0, for $m < n$, and the coefficients of $X_n^{(i)}$ to α_i, for $i = 1, \ldots, k$. These coefficients are determined by a recursive system of equations that depends on the α_i's. One finds that $W_n = R_n + o(\sqrt{n}\,)$. The martingale $R_n/\sqrt{n}$ checks out Lindeberg's conditional condition and the conditional variance condition (see Hall and Heyde, 1980 for the required technique), and the martingale central limit theorem holds for $R_n/\sqrt{n}$, and consequently normality holds for $W_n/\sqrt{n}$. □

The proof sketch of Theorem 6.1.1 does not give insight into why $\Re\lambda_2 < \frac{1}{2}\lambda_1$ is required for a central limit theorem. We shall say only a few words on this. A look into the variance structure shows that an eigenvalue λ_i contributes a term of the order $n^{2\lambda_i/\lambda_1}$ in the second moment. The two leading components in the variance are the linear component and a component of the exact order $n^{2\lambda_2/\lambda_1}$. So long as $2\Re\lambda_2 < \lambda_1$, the linear component dominates asymptotically, and scaling by $\sqrt{n}$ results in a nontrivial random variable. However, if $2\Re\lambda_2 > \lambda_1$, the nonlinear component dominates, and the random variable scaled by $\sqrt{n}$ blows up. The form of Theorem 6.1.1 does not remain valid in this case. The discussion of the variance in the proof of Theorem 6.1 gives a glimpse into this matter in the 2×2 case.

Remark: Smythe (1996) notes that there is no known convenient way for getting the covariance $\mathbf{\Sigma}$.

6.2 Invertible urn schemes

The case of constant row sum corresponds to a constant rate of increase. For example, consider a Bagchi-Pal urn with schema of the general form (7). Under the constraint of constant row sum

$K := a + b = c + d$, the total number of balls after n draws is deterministic; $\tau_n = W_0 + B_0 + Kn$. Upon plugging this into (10), the conditional equation can be transformed into a martingale relation. Moments are given by linear recurrences, for instance, taking expectation of (10),

$$\begin{aligned}\mathbf{E}[W_n] &= \mathbf{E}\Big[\Big(\frac{\tau_{n-1}+a-c}{\tau_{n-1}}\Big)W_{n-1}\Big] + c \\ &= \Big(\frac{\tau_{n-1}+a-c}{\tau_{n-1}}\Big)\mathbf{E}[W_{n-1}] + c.\end{aligned}$$

Contrast this with the case of nonconstant row sum. In this case, τ_n is a random variable, and it is not as easy to get a recurrence; in the expectation $\mathbf{E}[W_{n-1}/\tau_{n-1}]$ one cannot take out τ_{n-1}. One needs to seek an alternative approach. Some real-world applications (a few are considered in this review) required the handling of urn schemes with nonconstant row sum.

Difficulties with nonconstant row sum in the generator have been known since Rosenblatt (1940), who looked into asymmetric Pólya-Eggenberger urns with the schemata

$$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix},$$

with $a \neq d$. Kotz, Mahmoud and Robert (2000) discuss the difficulties arising in the case of nonconstant row sum. Bona fide nonlinear asymptotics come into play. The specific example discussed in Kotz, Mahmoud and Robert (2000) is

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

where it is shown that

$$\mathbf{E}[B_n] \sim \frac{n}{\ln n}.$$

Mahmoud (2002, 2003+), handles the long-term averages in the case of nonconstant row sum by a depoissonization method based on stochastic differential equations and elementary linear algebra. The operations involved require matrix inversion. Therefore these references deal only with *invertible urn schemes*, or schemes the generators of which are invertible, that is schemes with generator $\mathbf{E}[\mathbf{A}]$, where $(\mathbf{E}[\mathbf{A}])^{-1}$ exists.

Let $\mathbf{A}$ be a $k \times k$ schema of random entries (allowing degenerate deterministic distributions). Recall the discussion and notation used

in poissonization (cf. 5.3). In particular, recall t_n, the renewal times. The main line of proof in Mahmoud (2002, 2003+), is to depoissonize by finding an asymptotic representation for the renewal times.

Let $\mathbf{B} = \mathbf{E}[\mathbf{A}^T]$, and suppose it has $k' \le k$ distinct eigenvalues, $\lambda_1 = \lambda'_1, \ldots, \lambda'_{k'}$, with corresponding multiplicities $\nu_1, \ldots, \nu_{k'}$. It is known in linear algebra that the matric function $e^{\mathbf{B}t_n}$ has an expansion in terms of the exponential functions $e^{\lambda'_j t_n}$ (see Smiley, 1965). There are fixed nonzero matrices $\boldsymbol{\mathcal{E}}_1^{(0)}, \boldsymbol{\mathcal{E}}_1^{(1)}, \ldots, \boldsymbol{\mathcal{E}}_1^{(\nu_1-1)}, \boldsymbol{\mathcal{E}}_2^{(0)}, \ldots, \boldsymbol{\mathcal{E}}_2^{(\nu_2-1)}, \ldots, \boldsymbol{\mathcal{E}}_{k'}^{(\nu_{k'}-1)}$ associated with $\mathbf{B}$ for which

$$e^{\mathbf{B}t_n} = \sum_{s=0}^{\nu_1-1} \frac{t_n^s}{s!} e^{\lambda'_1 t_n} \boldsymbol{\mathcal{E}}_1^{(s)} + \cdots + \sum_{s=0}^{\nu_{k'}-1} \frac{t_n^s}{s!} e^{\lambda'_{k'} t_n} \boldsymbol{\mathcal{E}}_{k'}^{(s)}.$$

As $t_n \xrightarrow{a.s.} \infty$,

$$e^{\mathbf{B}t_n} \sim \frac{t_n^{\nu_1-1}}{(\nu_1-1)!} e^{\lambda_1 t_n} \boldsymbol{\mathcal{E}}_1^{(\nu_1-1)}. \tag{13}$$

For any fixed coefficients $\alpha_r^{(s)}$, the *scalar* function $\mathbf{E}\Big[\sum_{r=1}^{k'} \sum_{s=0}^{\nu_r-1} \alpha_r^{(s)} \frac{t_n^s}{s!} e^{\lambda'_r t_n}\Big]$ can be computed from the generalized mean-value theorem—if $f_{t_n}(x)$ is the density of t_n, there exists a unique value $\bar{t}_n$, for which

$$\begin{aligned}
\mathbf{E}\Big[\sum_{r=1}^{k'} \sum_{s=0}^{\nu_r-1} \alpha_r^{(s)} \frac{t_n^s}{s!} e^{\lambda'_r t_n}\Big] &= \int_0^\infty \Big(\sum_{r=1}^{k'} \sum_{s=0}^{\nu_r-1} \alpha_r^{(s)} \frac{x^s}{s!} e^{\lambda'_r x}\Big) f_{t_n}(x)\, dx \\
&= \Big(\sum_{r=1}^{k'} \sum_{s=0}^{\nu_r-1} \alpha_r^{(s)} \frac{\bar{t}_n^s}{s!} e^{\lambda'_r \bar{t}_n}\Big) \int_0^\infty f_{t_n}(x)\, dx \\
&= \sum_{r=1}^{k'} \sum_{s=0}^{\nu_r-1} \alpha_r^{(s)} \frac{\bar{t}_n^s}{s!} e^{\lambda'_r \bar{t}_n}. \qquad (14)
\end{aligned}$$

The main result in Mahmoud (2003+), deals with a more general situation than the one presented below. The original paper deals with degenerate cases and cases with multiple principal eigenvalue. We present a simplified form of these results for the sake of transparency. The nondegenerate case we present is one where the coefficient $\alpha_1^{(\nu_1-1)}$ in (14) is nonzero. The relation (14) then simplifies to

$$\mathbf{E}\Big[\sum_{r=1}^{k'} \sum_{s=0}^{\nu_r-1} \alpha_r^{(s)} \frac{t_n^s}{s!} e^{\lambda'_r t_n}\Big] \sim \alpha_1^{(\nu_1-1)} \frac{\bar{t}_n^{\nu_1-1}}{(\nu_1-1)!} e^{\lambda_1 \bar{t}_n}.$$

Therefore, for k–component vectors $\mathbf{C}$ and $\mathbf{D}$,

$$\mathbf{C}^T\mathbf{E}[e^{\mathbf{B}t_n}]\mathbf{D} \sim \alpha_1^{(\nu_1-1)}\frac{\bar{t}_n^{\nu_1-1}}{(\nu_1-1)!}e^{\lambda_1\bar{t}_n}\mathbf{C}^T\boldsymbol{\mathcal{E}}_1^{(\nu_1-1)}\mathbf{D}, \qquad (15)$$

in the nondegenerate case $\mathbf{C}\boldsymbol{\mathcal{E}}_1^{(\nu_1-1)}\mathbf{D} \neq 0$. It is well known that a Poisson process with parameter λ gives $\lambda\Delta t$ average number of renewals in a period of length Δt. Consider contributions to color 1 in an incremental period Δt. The balls are independent Poisson processes running in parallel, with rewards at the renewal points, giving

$$\begin{aligned}\mathbf{E}[R^{(1)}(t+\Delta t)\,|\,\mathbf{R}(t)] &= \mathbf{E}\Big[R^{(1)}(t) + (A_{11}R^{(1)}(t) + \cdots \\ &\qquad + A_{k1}R^{(k)}(t)\Big)\Delta t\,|\,\mathbf{R}(t)\Big].\end{aligned}$$

Similar equations can be written for the other colors. We have the incremental equations

$$\begin{aligned}\mathbf{E}\Big[R^{(1)}(t+\Delta t)\,|\,\mathbf{R}(t)\Big] &= R^{(1)}(t) + \Big(\mathbf{E}[A_{11}]R^{(1)}(t) + \cdots + \mathbf{E}[A_{k1}]R^{(k)}(t)\Big)\Delta t,\\ \mathbf{E}\Big[R^{(2)}(t+\Delta t)\,|\,\mathbf{R}(t)\Big] &= R^{(2)}(t) + \Big(\mathbf{E}[A_{12}]R^{(1)}(t) + \cdots + \mathbf{E}[A_{k2}]R^{(k)}(t)\Big)\Delta t,\\ &\vdots\\ \mathbf{E}\Big[R^{(k)}(t+\Delta t)\,|\,\mathbf{R}(t)\Big] &= R^{(k)}(t) + \Big(\mathbf{E}[A_{1k}]R^{(1)}(t) + \cdots + \mathbf{E}[A_{kk}]R^{(k)}(t)\Big)\Delta t.\end{aligned}$$

This set of simultaneous equations can be represented compactly in matrix form as

$$\frac{1}{\Delta t}\mathbf{E}[\mathbf{R}(t+\Delta t) - \mathbf{R}(t)\,|\,\mathbf{R}(t)] = \mathbf{B}\,\mathbf{R}(t),$$

where $\mathbf{B} = \mathbf{E}[\mathbf{A}^T]$. Take expectations, and let $\Delta t \to 0$ to get the differential equation

$$\frac{d}{dt}\mathbf{E}[\mathbf{R}(t)] = \mathbf{B}\,\mathbf{E}[\mathbf{R}(t)],$$

the solution of which is easily seen to be

$$\mathbf{E}[\mathbf{R}(t)] = e^{\mathbf{B}t}\mathbf{R}(0).$$

This equation is for fixed t. To depoissonize, we intend to use it at the random renewal time t_n. By the generalized mean-value theorem it can be shown that

$$\mathbf{E}[\mathbf{R}(t_n)] = e^{\mathbf{B}\bar{t}_n}\mathbf{R}(0),$$

for some t_n. Giving the average

$$\mathbf{E}[\mathbf{R}(t_n)] = \mathbf{E}[\mathbf{X}_n] = e^{\mathbf{B}\bar{t}_n}\mathbf{X}_0. \tag{16}$$

Theorem 6.2.1. *(Mahmoud, 2003+). Let* $\mathbf{A}$ *be the* $k\times k$ *schema of a growing urn with invertible generator* $\mathbf{B}^T$*. Let* $X_n^{(i)}$ *be the number of balls of color* i *after* n *draws, and let the composition vector be* $\mathbf{X}_n = (X_n^{(1)}, \ldots, X_n^{(k)})^T$*. Assume the scheme is nondegenerate, in the sense that* $\mathbf{J}_k^T\mathbf{B}^{-1}\boldsymbol{\mathcal{E}}_1^{(\nu_1-1)}\ \mathbf{X}_0\ \neq 0$*. The composition vector satisfies*

$$\mathbf{E}[\mathbf{X}_n] = \frac{1}{\mathbf{J}_k^T\mathbf{B}^{-1}\boldsymbol{\mathcal{E}}_1^{(\nu_1-1)}\mathbf{X}_0}\mathbf{X}_0\boldsymbol{\mathcal{E}}_1^{(\nu_1-1)}\, n + \mathbf{o}_k(n).$$

Proof (sketch). Let $\tilde{R}^{(i)}(t)$ be the number of times team i has won a race by time t, and let $\tilde{\mathbf{R}}(t) = (\tilde{R}^{(1)}(t), \ldots, \tilde{R}^{(k)}(t))^T$. Suppose the $\tilde{R}^{(j)}(t)$ independent realizations of A_{j1} are $A_{j1}^{(r)}$, for $r = 1, \ldots, \tilde{R}^{(j)}(t)$. It is clear that

$$R^{(1)}(t) = R^{(1)}(0) + \sum_{r=1}^{\tilde{R}^{(1)}(t)} A_{11}^{(r)} + \cdots + \sum_{r=1}^{\tilde{R}^{(k)}(t)} A_{k1}^{(r)}.$$

Applying Wald's equation (see Ross, 1983), we get

$$\mathbf{E}[R^{(1)}(t)] = R^{(1)}(0) + \mathbf{E}[\tilde{R}^{(1)}(t)]\mathbf{E}[A_{11}] + \cdots + \mathbf{E}[\tilde{R}^{(k)}(t)]\mathbf{E}[A_{k1}].$$

Similar equations can be written for all the other colors, and we can put them in a matric equation form:

$$\mathbf{E}[\mathbf{R}(t)] = \mathbf{B}\,\mathbf{E}[\tilde{\mathbf{R}}(t)] + \mathbf{R}(0).$$

Thus, at the random time t_n,

$$\mathbf{B}\,\mathbf{E}[\tilde{\mathbf{R}}(t_n)\,|\,t_n] + \mathbf{X}_0 = \mathbf{E}[\mathbf{R}(t_n)\,|\,t_n] = e^{\mathbf{B}t_n}\mathbf{X}_0,$$

the expectation of which is

$$\mathbf{B}\,\mathbf{E}[\tilde{\mathbf{R}}(t_n)] + \mathbf{X}_0 = e^{\mathbf{B}\bar{t}_n}\mathbf{X}_0.$$

On any stochastic path whatever, we have n races by time t_n; the components of $\tilde{\mathbf{R}}(t_n)$ must add up to n. So,

$$\begin{aligned} n &= \mathbf{E}[\mathbf{J}_k^T\,\tilde{\mathbf{R}}(t_n)] \\ &= \mathbf{J}_k^T\mathbf{B}^{-1}\Big(e^{\mathbf{B}\bar{t}_n} - \mathbf{I}\Big)\mathbf{X}_0. \end{aligned}$$

Recalling that we are in the nondegenerate case $\mathbf{J}_k^T \mathbf{B}^{-1} \boldsymbol{\mathcal{E}}_1^{(\nu_1-1)} \mathbf{X}_0 \neq 0$, from (15) we have

$$n \sim \frac{\bar{t}_n^{(\nu_1-1)}}{(\nu_1-1)!} e^{\lambda_1 \bar{t}_n} \mathbf{J}_k^T \mathbf{B}^{-1} \boldsymbol{\mathcal{E}}_1^{(\nu_1-1)} \mathbf{X}_0. \tag{17}$$

By (13) and (16) we finally have

$$\mathbf{E}[\mathbf{X}_n] = e^{\mathbf{B}\bar{t}_n} \mathbf{X}_0 \sim \frac{\bar{t}_n^{(\nu_1-1)}}{(\nu_1-1)!} e^{\lambda_1 \bar{t}_n} \boldsymbol{\mathcal{E}}_1^{(\nu_1-1)} \mathbf{X}_0,$$

which gives the result when compared with (17). □

While the leading asymptotics in Theorem 6.2.1 do not depend on the initial conditions, the lower-order asymptotics do (see Inoue and Aki, 2001, where a particular urn with a fixed schema was analyzed).

7 Connection of Pólya urns to search trees

The binary search tree is a fundamental construct of computer science that is used for efficient storage of data, and the modeling of many algorithms. For definitions and combinatorial properties see Mahmoud (1992); for applications in sorting see Knuth (1998) or Mahmoud (2000).

Under the constraints of modern technology, external computer memory is much cheaper than internal memory. That is why internal memory is usually small. Speedwise, data in internal memory are accessed much faster than data residing outside. In applications involving large volumes of data special data structures are preferred to store the bulk of data outside the computer on (slow) secondary storage. Fragments therein are brought into internal memory upon request for fast local searching. These data structures involve generalizations of the binary search tree into data structures the nodes of which are blocks or buckets. Two such generalizations are presented. The analysis of both is amenable to urn models results. Surprisingly, they both exhibit an interesting phase transition in their distribution.

To handle data in high dimensions, the binary tree is generalized in yet another direction. Quad trees and k–d trees are suitable storage media for geometry algorithms, such as nearest neighbor queries. The analysis of these trees, too, is amenable to Pólya urn models.

The results presented below had all been obtained mostly by other methods, but it has recently become clear that they can all be obtained via various urn models. Mahmoud (2002) provides many of these alternative proofs.

7.1 Binary search trees

A *binary tree* is a structure of nodes each having no children, one left child, one right child, or two children (one left and one right). In practice, the tree carries labels (data), and is endowed with a *search property*. According to the search property, the label of any node is larger than the labels in its left subtree and no greater than any label in its right subtree, and this property permeates the structure recursively.

Several models of randomness are commonly used on binary trees. The uniform model in which all trees are equally likely is useful in formal language studies, compilers, computer algebra, etc. Kemp (1984) is a good source for this subject. The *random permutation model* comports more faithfully to sorting and data structures. In the random permutation probability model, we assume that the tree is built from permutations of $\{1, \ldots, n\}$, where a uniform probability model is imposed on the permutations instead of the trees. When all $n!$ permutations are equally likely or *random*, binary search trees are not equally likely. It is well known in enumerative combinatorics that the number of distinct binary trees is $(n+1)^{-1}\binom{2n}{n} < n!$ (see Knuth, 1998). By the pigeon-hole principle, several permutations must "collide" and correspond to the same tree, but there are permutations that each correspond to precisely one tree. The random permutation model does not give rise to a uniform distribution on binary trees. Generally the model is biased toward shorter trees, which is a desirable properly for fast searching applications (see Mahmoud, 1992). The random permutation model covers a wide variety of real-world applications, such as sampling data from *any* continuous distribution, where the data ranks almost surely form a random permutation.

The term *random tree* will refer to a tree built from a random permutation. A tree grows from a permutation $(\pi_1, \ldots, \pi_n)$ of $\{1, \ldots, n\}$ as follows. The first element π_1 is inserted in an empty tree, a root of a new nonempty tree is allocated for it. Each subsequent element π_j is taken to the left or right subtree according as whether it is smaller or larger than the root. In the subtree, the element is inserted recur-

sively. The search continues until an empty subtree is found where the element is inserted into a new node, just like π_1 was initially inserted in the root.

A binary search tree is often *extended* by supplying each node with a sufficient number of distinguished nodes called *external* to uniformly make the outdegree of each original tree node (now called *internal*) equal to 2. Figure 1 shows an extended binary search tree grown from the permutation $3, 7, 4, 2, 6, 1, 5$.

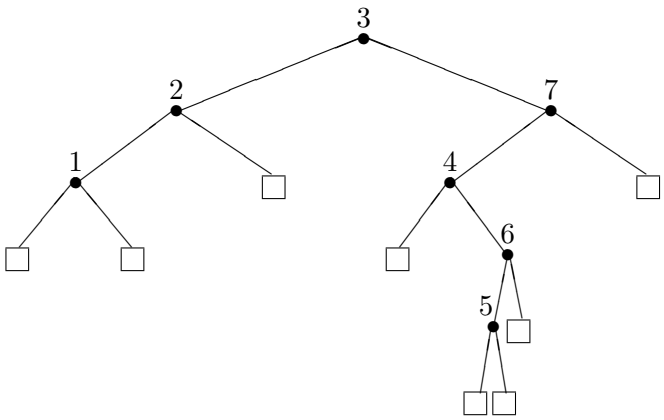


Figure 1: An extended binary tree on 7 internal nodes.

The number of internal nodes with k external children, $k = 0, 1, 2$, provides certain space efficiency measures on binary search trees. For example, leaves in the original tree (nodes with two external children in the extension) are wasteful nodes that are allocated pointers that are not used (in a real implementation they carry nill value). This efficiency measure can be found from the following.

Theorem 7.1.1. *(Devroye, 1991). Let L_n be the number of leaves in a binary search tree of size n. Then,*

$$\frac{L_n - \frac{1}{3}n}{n} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{2}{45}\Big).$$

Proof (sketch). It is well known that the random permutation model on binary search trees gives an evolutionary stochastic process in which the external nodes are equally likely (see Knuth, 1998). That

is, if a tree is grown in steps by turning a randomly chosen external node into an internal node, the resulting binary tree shapes follow the random permutation probability distribution.

Color each external node with an external sibling with white, and color all the other external nodes with blue. When insertion hits a white external node, the node is turned into an internal one, with two white external children; its sibling turns blue. If insertion hits a blue external node, it turns into an internal node with two white external children. The schema

$$\begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix}$$

underlies this process. If W_n is the number of white balls in the urn after n draws, it follows from Theorem 6.1 that

$$\frac{W_n - \frac{2}{3}n}{n} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{8}{45}\Big).$$

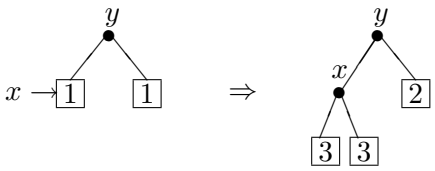
But then $W_n = 2L_n$. □

Theorem 6.1.1 gives a bit more; it specifies a multivariate joint distribution for the profile of nodes of various outdegrees in the tree, of which the distribution of the leaves is only one marginal.

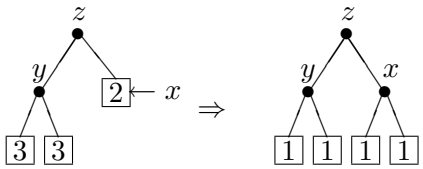
7.2 Fringe-balanced trees

Binary trees are quite efficient (see Mahmoud, 1992 for numerous applications). To improve the speed of retrieval even more, height reduction heuristics known as *balancing* are employed. One such heuristic is the local balancing of the fringe (Poblete and Munro, 1985). Fringe-balancing is a local "rotation" that compresses subtrees near the leaves into shorter shrubs. The operation is performed, when the insertion of a new node falls under an internal node that happens to be the only child of its parent in the binary tree. When this occurs a compression operation called *rotation* promotes the median of these three nodes to become the parent and repositions the other two nodes under it in a manner preserving the search property. Figure 2 illustrates some cases of fringe balancing. The cases not shown in Figure 2 are only mirror images of those depicted. The introduction of these rotations reduces the total path length, but the cost of rotation becomes a factor in the construction of the tree.

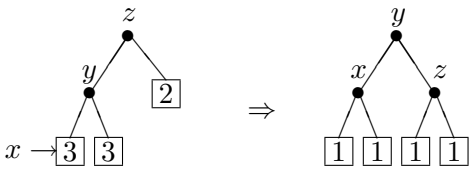
To model rotations, we color the leaves according to a scheme in which the number of times the leaves of a particular color are chosen for insertion corresponds to rotations. Color 1 generally represents balance. If insertion falls at a leaf of color 1, a tolerable degree of imbalance appears in the tree and is coded by colors 2 and 3: color 2 for the leaf, the sibling of which is internal; the two children of this internal node are colored with color 3 (Figure 2(a)). If insertion falls at a leaf of color 2, the insertion fills out a subtree forming a perfectly balanced shrub of height 2. The four balanced leaves of this perfectly balanced shrub are colored with color 1 (Figure 2(b)). If instead insertion falls at a leaf of color 3, more imbalance is present, and balance is restored by a rotation and again a perfectly balanced shrub of height 2 appears; the four leaves of this shrub are colored with color 1 (Figure 2(c)).



(a) Imbalance appears when insertion hits a leaf of color 1.



(b) Insertion at a leaf of color 2 fills out a subtree.



(c) A rotation is executed when insertion hits a leaf of color 3.

Figure 2: Insertions in a fringe-balanced binary search tree.

After n insertions in an empty tree, the number of times insertion falls at a leaf of color 3 is the number of rotations, to be denoted by R_n. The underlying urn has the schema

$$\begin{pmatrix} -2 & 1 & 2 \\ 4 & -1 & -2 \\ 4 & -1 & -2 \end{pmatrix}.$$

This tenable urn scheme is not exactly an extended urn à la Smythe (1996), because it has negative elements off the main diagonal. However, Smythe's proof can be adapted, to work here, too, indicating that Smythe's results can be generalized to a superclass of extended urn models.

Theorem 7.2.1. *(Mahmoud, 1998; Panholzer and Prodinger, 1998). Let R_n be the number of rotations after n insertions in an initially empty fringe-balanced binary search tree. Then,*

$$\frac{R_n - \frac{2}{7}n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{66}{637}\Big).$$

As mentioned before, there is no known convenient method to find variances in Smythe-like schemes. So far, each instance of such an urn scheme has been handled in an ad-hoc mannar in the original references. The variance in Theorem 7.2.1 was obtained by recurrence methods (in fact in exact form). (The same exact result is presented in Hermosilla and Olivos, 1985). Panholzer and Prodinger, 1998 report on an analytic approach to Theorem 7.2.1.

7.3 m-ary Search Trees

Search speed is reduced with increased branching, as data are distributed among more subtrees. The m-ary tree has branching factor m, and each node holds up to $m-1$ keys. The tree grows from n keys according to a recursive algorithm: The first insertion falls in an empty tree; a root node is allocated for it. The second key joins, but the two keys are arranged in increasing order from left to right. Subsequent keys (up to the $(m-1)$st insertion) are likewise placed in the root node, and after each insertion the keys are sorted in increasing order. The $(m-1)$st insertion fills out the root node; a subsequent key goes to the ith subtree if it ranks i among the first

m keys. Within the subtree the insertion procedure is applied recursively. Figure 3 shows a ternary tree grown from the permutation $6, 4, 1, 5, 3, 9, 2, 8, 7$.

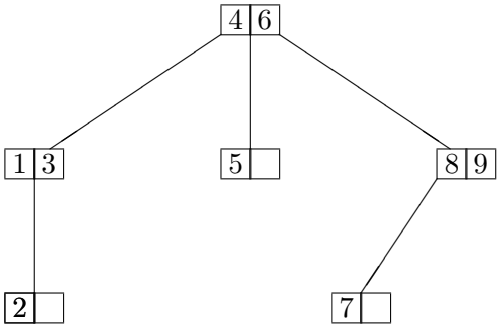


Figure 3: A ternary search tree ($m = 3$).

Unlike the binary case of deterministic size, for $m \geq 3$ the size of the m-ary tree is random. The size is a measure of the space efficiency of the algorithm. The tree growth can be modeled by an urn consisting of balls of $m-1$ different colors. Color i corresponds to a "gap" (an insertion position) between two keys of a leaf node with $i-1$ keys, for $i = 1, \ldots, m-1$ ($i = 1$ corresponds to insertion in an empty tree). A leaf node containing $i-1$ keys has i gaps ($i-1$ real numbers cut up the real line into i intervals). A key falling in a node with $i \leq m-2$ keys will adjoin itself in the node, increasing its number of keys to i, and consequently the number of gaps to $i+1$. The corresponding rule for the growth of the associated urn is to promote i balls of color i into $i+1$ balls of color $i+1$; this insertion affects colors i and $i+1$ and no other.

The urn rule corresponding to filling out a node is a little different. The node already contains $m-1$ gaps corresponding to $m-1$ balls of color $m-1$. The insertion falls in a gap and the node fills out defining m empty subtrees (at the next level in the tree). The m empty subtrees are represented by m balls of color 1 in the urn . The

schema is

$$\begin{pmatrix} -1 & 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -2 & 3 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \ddots & & \vdots & \vdots \\ 0 & 0 & & & \dots & & -(m-2) & m-1 \\ m & 0 & & & \dots & & 0 & -(m-1) \end{pmatrix}.$$

The characteristic equation $|\mathbf{A}^T - \lambda\mathbf{I}| = 0$ expands into

$$(\lambda+1)(\lambda+2)\dots(\lambda+m-1) = m!.$$

This scheme is an extended urn scheme à la Smythe (1996). The eigenvalues of which have the property that $\Re\,\lambda_2 < \frac{1}{2}\lambda_1 = \frac{1}{2}$, for m up to 26, $\Re\,\lambda_2 > \frac{1}{2}$, for $m \geq 27$.

Theorem 7.3.1. *(Chern and Hwang, 2001). Let S_n be the size of an m–ary search tree grown from a random permutation of $\{1,\dots,n\}$. If the branch factor $3 \leq m \leq 26$, then*

$$\frac{S_n - n/(2(H_m-1))}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma_m^2),$$

for some effectively computable constant σ_m^2.

The limiting variance σ_m^2 can be computed by recurrence (Mahmoud, 1992 gives an exact formula). Chern and Hwang (2001) prove asymptotic normality in the range $3 \leq m \leq 26$ by a method of recursive moments. If $m > 26$, the associated urn is no longer an extended urn in the sense of Smythe (1996). Chern and Hwang (2001) argue that no limit distribution for the size of the m-ary tree exists under the central limit norming because of too large oscillations in all higher moments.

7.4 2–3 trees

To guarantee fast retrieval, m–ary trees are balanced by operations that grow the tree by first accessing an insertion position, and if that causes imbalance, keys are sent up, recursively, and if necessary, they go all the way back, forcing a new root to appear. We illustrate this

balancing in 2–3 trees. The nodes here are of two types, type 1 holds one key, and type 2 holds two keys (the corresponding branching is respectively 2 and 3). The search phase for the insertion of a new key follows the same general principles as in the m-ary tree. When insertion falls in a node of type 1, it is promoted into type 2. When insertion of a key K overflows a leaf of type 2, already carrying x and y, the median of the three keys x, y, K is sent up, and the type-2 leaf is split into two type-1 leaves. If the median promotion puts it in a type-1 internal node, the node is expanded into type-2 to accommodate the new key, and the two new type-1 leaves, together with the sibling of the old type-2 leaf are arranged according to order as in the usual ternary search tree. If instead, the median promotion tends to put it in an already filled type-2 node, the node is split, recursively, etc. Figure 4 illustrates the insertion of a key into a 2–3 tree. All the leaves in a 2–3 trees are at the same level, guaranteeing that any search will be done in $O(\log n)$.

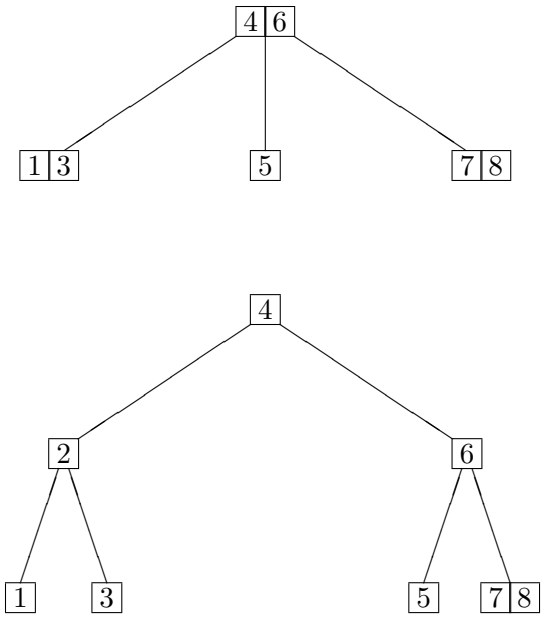


Figure 4: A 2–3 tree before and after inserting the key 2.

The fringe balancing principle, established in Yao (1978), states that most of the nodes are leaves—the number of leaves in the tree

well approximates the size. More precisely, Yao (1978) argues that bounds on the expected size can be found from the expected number of leaves.

Theorem 7.4.1. *(Yao, 1978, Bagchi and Pal, 1985). Let S_n be the number of nodes after n insertions in an initially empty 2–3 tree. Then,*

$$\frac{9}{14}\,n + O(1) \;\le\; \mathbf{E}[S_n] \;\le\; \frac{6}{7}\,n + O(1).$$

Proof (sketch). We first determine L_n, the number of leaves, on average. Each key in a type-1 leaf defines two insertion gaps; associate two white balls with these gaps. Each pair of keys in a type-2 leaf defines three insertion gaps; associate three blue balls with these gaps. From the evolution rules of the tree, the schema

$$\begin{pmatrix} -2 & 3 \\ 4 & -3 \end{pmatrix}$$

underlies the urn of white and blue balls. By Theorem 6.1, we have

$$\mathbf{E}[W_n] = \frac{4}{7}\,n + O(1), \qquad \mathbf{E}[B_n] = \frac{3}{7}\,n + O(1).$$

On average, the number of leaves is given by

$$\mathbf{E}[L_n] = \frac{1}{2}\mathbf{E}[W_n] + \frac{1}{3}\mathbf{E}[B_n] \sim \frac{3}{7}\,n. \tag{18}$$

It is well known that the number of leaves in an m-ary tree on n nods is related to X_n the number of internal nodes by

$$L_n = (m-1)X_n + 1.$$

The size of the 2–3 tree with L_n leaves is maximized if all internal nodes are of type-1, in which case $L_n = X_n + 1$, and is minimized if all internal nodes are of type-2, in which case $L_n = 2X_n + 1$. Thus, the size, $S_n = L_n + X_n$, is sandwiched:

$$L_n + \frac{L_n - 1}{2} \;\le\; S_n \;\le\; L_n + L_n - 1.$$

The result follows upon taking expectations and plugging in (18). □

Yao (1978) describes a series of refinements by which the bounds in Theorem 7.4.1 can be improved. These refinements consider larger and larger shrubs on the fringe of the tree. They can be modeled by urns with more colors.

7.5 Paged binary trees

Another generalization of binary search trees, suitable for internal memory algorithms working in conjunction with a slow external memory, is the paged binary tree, the leaves of which are buckets that can carry up to c keys. Each internal node carries one key and has a right and a left child. When a searching algorithm climbs along a path starting at the root, and reaches the leaf at the end of the path, the block of data in the leaf is fetched from external memory, thus a batch of data is brought into internal memory at one time, where they can be subjected to fast searching. Usually the capacity c is chosen to coincide with a "page" of memory, the standard unit of size of a data block in computer memory.

The tree grows from n keys according to a splitting policy. Up to c keys go into the root node. In practice the keys are kept in sorted order to enhance fast local searching. Toward a balanced split, $c = 2b$ is usually an even number. When the $(2b + 1)$st key appears, the tree is restructured: The root bucket is turned into an internal node retaining only the median of the $2b + 1$ keys. Two new pages are created and adjoined as the right and left children of the internal node. Each of the two leaves is half filled. The b keys below the median, are placed in the left page, the rest in the right. This branching facilitates later search and insertion. A subsequent key is guided to the left subtree if it is less than the root key, otherwise it goes into the right subtree. Within the appropriate subtree the insertion procedure is applied recursively. Figure 5 illustrates a paged tree with page capacity 2, arising from the permutation $6, 3, 1, 4, 2, 5$; the process experiences two splits: First when the key 1 is inserted, the median 3 of the set $\{6, 3, 1\}$ is kept in the root, then again when 5 is inserted, the node containing 4 and 6 splits, keeping the median 5 at the root. In Figure 5 the pages are shown as boxes and the internal nodes as circles.

The associated urn consists of balls of $b + 1$ different colors, and has a $(b + 1) \times (b + 1)$ schema. Color i corresponds to a "gap" (an insertion position) between two keys of a leaf node carrying $i + b - 1$ keys, for $i = 1, \ldots, b+1$. A key falling in an unfilled leaf with $i+b-1$ keys ($i \leq b$) will increase its keys to $i+b$ and consequently the number of gaps to $i + b + 1$. The corresponding growth rule in the associated urn is to promote i balls of color i into $i + 1$ balls of color $i + 1$; this insertion affects colors i and $i + 1$ and no other. The urn rule corresponding to splitting is distinguished. A splitting node contains $2b$ keys with $2b + 1$ gaps of color $b + 1$. The insertion falls in one of

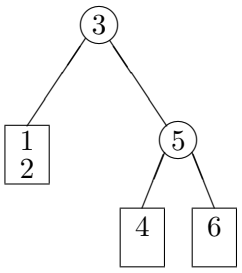


Figure 5: A paged binary tree with page capacity 2.

these gaps, and an overflow occurs forcing a splitting: Two half-filled leaves appear, each containing b keys, thus having $b+1$ insertion gaps each; the urn gains $2b+2$ balls of color 1. The schema is

$$\mathbf{A} = \begin{pmatrix} -(b+1) & b+2 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -(b+2) & b+3 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -(b+3) & b+4 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \ddots & & & \vdots & \vdots \\ 0 & 0 & & \dots & & & -2b & (2b+1) \\ 2b+2 & 0 & & \dots & & & 0 & -(2b+1) \end{pmatrix}.$$

The characteristic equation $|\mathbf{A}^T - \lambda \mathbf{I}| = 0$ expands into

$$(\lambda + b + 1)(\lambda + b + 2) \dots (\lambda + 2b + 1) = \frac{(2b+2)!}{(b+1)!}.$$

The simple roots of this characteristic equation satisfy $\Re\,\lambda_2 < \frac{1}{2}\Re\,\lambda_1 = \frac{1}{2}$, for $b \leq 58$. For all $b \geq 59$, $\Re\,\lambda_2 > \frac{1}{2}$. This scheme is an extended urn scheme à la Smythe, 1996 for b up to 58 (last capacity admitting normality is $c = 116$).

Theorem 7.5.1. *(Chern and Hwang, 2001). Let S_n be the size of a paged binary tree (with page capacity $c = 2b$), after n insertions. Then*

$$S_n^* = \frac{S_n - n/((2b+1)(H_{2b+1} - H_{b+1}))}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_b^2),$$

for some effectively computable constant σ_b^2.

The limiting variance σ_b^2 can be obtained with quite a bit of tedious recurrence computation. Chern and Hwang (2001) prove

asymptotic normality for $1 \leq b \leq 58$ by a method of moments (the last capacity for which normality holds is 116). Chern and Hwang (2001) argue that for $b > 58$, the normed random variable S_n^* no longer has a normal limit.

7.6 Bucket quad trees

The quad tree, introduced in Finkel and Bentley (1974), is a suitable data structure for d-dimensional data (see Samet, 1990 or Yao, 1990 for applications). A point in d dimensions defines 2^d quadrants; other data are grouped according to the quadrant they belong to. The 2^d quadrants are canonically numbered $1, 2, \ldots, 2^d$. To expedite work with blocks of data, the nodes are turned into buckets to hold c points. Devroye and Vivas (1998) suggest a balancing policy that associates with each bucket a guiding *index*, which is a dummy point composed of d coordinates, the ith coordinate of which is the median of the c keys' ith coordinates. As median extraction of c numbers is involved in each coordinate, it is therefore customary to take $c = 2b + 1$, an odd number.

The bucket quad tree has 2^d subtrees, numbered $1, \ldots, 2^d$, from left to right. It grows by putting up to c keys in the root node. When a bucket is filled to capacity, its index is computed and stored with the data. A subsequent point belonging to the qth quadrant goes to the qth subtree, where it is subjected to the same algorithm recursively to be inserted in the subtree. Figure 6 shows six points in the unit square, and the corresponding two-dimensional quad tree with bucket capacity 3; the index for the first three point is indicated by the symbol $\times$; the canonical count goes from the bottom left quadrant of the index and proceeds clockwise.

A node holding j keys in this tree has $j+1$ insertion positions, $j = 1, \ldots, c-1$. We shall consider an idealized probability distribution on quad trees induced by growing them by choosing an insertion position at random (all insertion positions being equally likely). This probability model is most suited for simulation.

The balls in the associated $(2b+1) \times (2b+1)$ urn scheme have $2b+1$ different colors. Color i accounts for $i-1$ keys in a leaf node, for $i = 1, \ldots, 2b+1$. A key falling in an unfilled leaf with $i-1$ keys will be placed in the leaf and the number of attraction positions increases from i to $i+1$. The corresponding growth rule in the associated urn is to promote i balls of color i into $i+1$ balls of color $i+1$; this

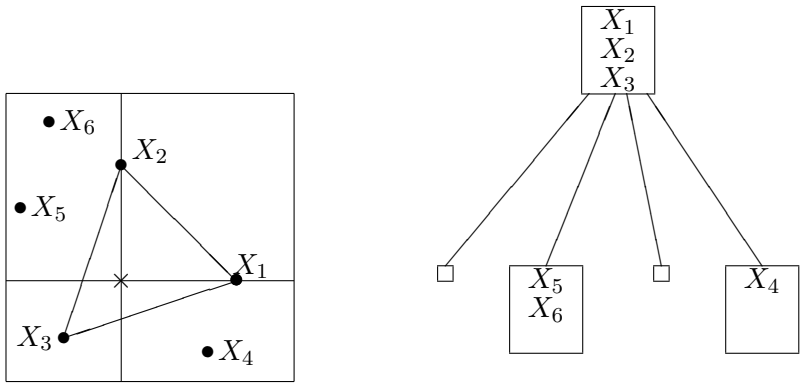


Figure 6: Points in the unit square and their corresponding quad tree with bucket capacity 3. The symbol $\times$ is the index of X_1, X_2 and X_3.

insertion affects colors i and $i+1$ and no other. When the $(2b+1)$st key joins a node containing $2b$ keys, 2^d insertion positions appear on the next level. The complete schema is

$$\mathbf{A} = \begin{pmatrix} -1 & 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -2 & 3 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \ddots & & \vdots & \vdots \\ 0 & 0 & & & \dots & & -2b & (2b+1) \\ 2^d & 0 & & & \dots & & 0 & -(2b+1) \end{pmatrix}.$$

Let $X_n^{(i)}$ be the number of balls of color i, for $i = 1, \dots, 2b+1$. In this schema the row sums may not be the same, but it is an invertible scheme, and the results of Theorem 6.2.1 apply. The characteristic equation $|\mathbf{A}^T - \lambda \mathbf{I}| = 0$ expands into

$$(\lambda+1)(\lambda+2)\dots(\lambda+2b+1) = (2b+1)!\, 2^d.$$

The principal eigenvalue λ_1 is clearly a function of both the capacity and dimension. For instance, with $c = 3$ (that is, $b = 1$) and $d = 4$, the bucket 4–dimensional quad tree has a principal eigenvalue $\lambda_1 = 2.651649306\dots$, and the profile

$$\mathbf{E}[X_n^{(1)}] \quad \sim \quad 1.599129368\dots\, n,$$

$$\begin{aligned}\mathbf{E}[X_n^{(2)}] &\sim 0.6875537100\ldots\, n,\\ \mathbf{E}[X_n^{(3)}] &\sim 0.3649662279\ldots\, n.\end{aligned}$$

From these average measures one can deduce the average number of leaves, and consequently the size of the tree to be

$$\mathbf{E}[S_n] \sim 0.6030697064\ldots\, n.$$

7.7 Bucket k-d trees

The k-d tree was introduced by Bentley in, 1975 as a means of data storage amenable to computational geometry algorithms, such as range searching (Samet, 1990 surveys applications of this tree). The bucket k-d tree is a tree for d dimensional data. It is a *binary* tree of bucketed nodes of capacity k each. Up to k keys (usually an odd number, say $2b+1$) go into the root. When $2b+1$ keys aggregate in the root node, a dummy index is computed and stored with the keys. The index is the median of the first coordinates of the points in the node. A subsequent key goes into the left subtree if the key's first coordinate is less than the root's index; otherwise the key goes into the right subtree. In whichever subtree the key goes, it is subjected recursively to the same insertion algorithm, but at level ℓ in the tree, the $(\ell+1)$st coordinate of the inserted key is used in a comparison against an index that is the median of the $(\ell+1)$ coordinates of the c keys of a filled node (if the node is not filled, the key adjoins itself to that node). The process cycles through the coordinates with the interpretation that $d+1$ wraps around into the first coordinate. Figure 7 shows eight points in the unit square and the corresponding 3–2 tree; the indexes are indicated by the symbol $\times$.

A node holding j keys in this tree has $j+1$ equally likely insertion positions, $j = 1, \ldots, c-1$. The k-d tree's $(2b+1)\times(2b+1)$ associated urn grows just like that of the quad tree. The balls in the associated urn have $2b+1$ different colors. The difference concerns the policy of handling an overflow. An insertion hitting a leaf containing $2b$ keys, fills out the leaf; and two insertion positions appear on the next level.

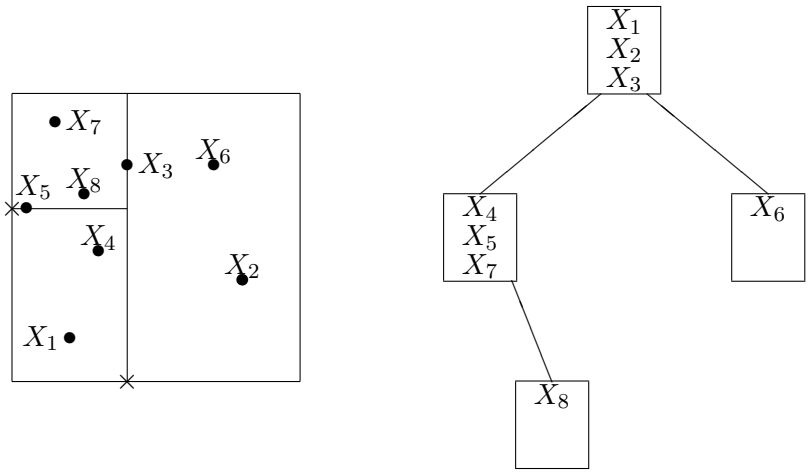


Figure 7: Points in the unit square and their corresponding 3–2 tree.

The complete schema is

$$\begin{pmatrix} -1 & 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -2 & 3 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \ddots & & \vdots & \vdots \\ 0 & 0 & & & \dots & & -2b & (2b+1) \\ 2 & 0 & & & \dots & & 0 & -(2b+1) \end{pmatrix}.$$

The k-d urn scheme coincides with that of a bucket 1-dimensional quad tree. For instance, with $k = 3$ (that is, $b = 1$), in the random bucket 3-2 tree the average number of nodes allocated after n key insertions is

$$\mathbf{E}[S_n] = 0.3938823094\ldots\, n,$$

which can be found from Theorem 6.2.1.

8 Connection to recursive trees

The random recursive tree is a combinatorial structure used to model a variety of applications such as contagion, chain letters, philology, etc. (see the survey in Smythe and Mahmoud, 1996, and the 40 plus references therein).

The model has been generalized in a number of directions to take into account the experience of the chain letter holder or to fit operation on computers. These generalizations are sketched below. The associated results were mostly derived by other methods, such as recurrence and moments methods. Bergeron, Flajolet and Salvy, 1992 give a completely analytic approach to the subject. It has recently become clear that a unifying approach via urn models ties all these results.

8.1 Standard recursive trees

Starting out with one root node, *the random recursive tree* grows in discrete time steps. At each step a node in the tree is chosen uniformly at random to be the parent of a new entrant. If the nth node is labeled with n, all root-to-leaf paths will correspond to increasing sequences. Note that there is no restriction on node degrees in recursive trees. In fact they can grow indefinitely (Javanian and Vahidi-Asl, 2003+, find a central limit theorem for an appropriately normed version of a node degree). Figure 8 shows all the recursive trees of order 4.

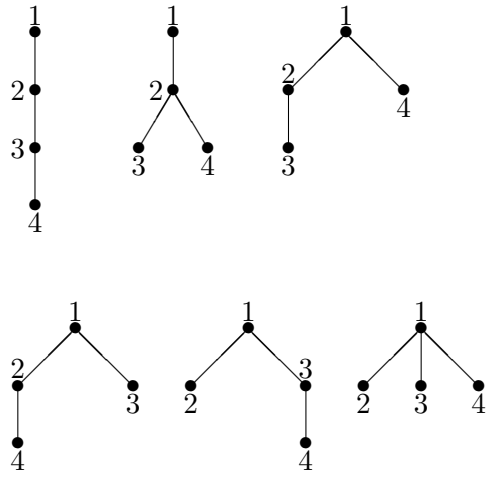


Figure 8: Recursive trees of order 4.

Recursive trees have been proposed as a model for chain letters (Gastwirth, 1977), where a company is founded to spread a particular item (lottery tickets, good luck charm, etc.). The initial recruiter

looks for a willing participant to buy a copy of the letter. The recruiter and the new letter holder compete with equal chance to recruit a new participant. The process proceeds in this way, where at each stage all existing participants compete with equal chance to attract the next participant in the scheme.

There is a profit associated with selling a copy of the letter. The promise is that everybody will make profit. Nevertheless, when the number of participants grows to n, it is unavoidable that many letter holders purchased the letter but did not sell it. These are frustrated participants in the scheme (leaves in the tree).

How many frustrated participants are there? This is the question that Najock and Heyde (1982) address, and find out that about half of the participants will be frustrated, on average. Najock and Heyde (1982) characterize the exact and asymptotic distribution of the number of leaves in a random recursive tree, using recurrence methods; we give a proof based on urns.

Theorem 8.1.1. *(Najock and Heyde, 1982). Let L_n be he number of leaves in a random recursive tree of size n. This random variable has the exact distribution*

$$\mathbf{Prob}\{L_n = k\} = \frac{1}{(n-1)!} \left\langle {n-1 \atop k-1} \right\rangle,$$

where $\left\langle {n \atop k} \right\rangle$ is the Eulerian number of the first kind, for $k = 1, \ldots, n-1$. The number of leaves also exhibits Gaussian tendency:

$$\frac{L_n - \frac{1}{2}n}{\sqrt{n}} = \mathcal{N}\Big(0, \frac{1}{12}\Big).$$

Proof (sketch). Color the leaves of a recursive tree with white, the rest of the nodes with blue. When a white leaf recruits, it is turned into an internal blue node and acquires a new node as a white child. When a blue node recruits, it attracts a white leaf as a child. This is a Friedman's urn with the schema

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

starting out with one white ball. If W_n is the number of white balls after n draws, then $L_n = W_{n-1}$.

With $\gamma = -1$, $\delta = -1$, and $\tau_n = n+1$, the functional equation (6) is specialized here to

$$\psi_{n+1}(t) = \frac{\psi_n'(t)}{n+1} = \frac{\psi_{n-1}''(t)}{(n+1)n} = \ldots = \frac{1}{(n+1)!} \times \frac{d^{n+1}}{dt^{n+1}}\psi_0(t).$$

Reverting to the moment generating function $\phi_n(t)$, cf. (5), with the boundary condition $\phi_0(t) = e^t$, we have the solution

$$\phi_n(t) = \frac{1}{n!}(1-e^t)^{n+1}\frac{d^n}{dt^n}\Big(\frac{e^t}{1-e^t}\Big),$$

From this form it is easy to show by induction that the probability generating function, $\phi_n(\ln t)$ coincides with the well known generating function of the sequence of Eulerian numbers of the first kind (see Graham, Knuth and Patashnik, 1994).

The central limit tendency is an immediate application of Theorem 5.2.2. □

Gastwirth and Bhattacharya (1984) address a variation regarding the profit. In some chain letter schemes the success of the recruits of a letter holder adds to his own profit. A letter holder gets a commission whenever anyone in his own subtree recruits. A measure of this success is the size (number of nodes) of the subtree rooted at k.

Theorem 8.1.2. *(Gastwirth and Bhattacharya, 1984). Let S_{nk} be the size of the subtree rooted at the kth entrant to a recursive tree. As both k and n increase to infinity in such a way that $k/n \to \gamma$, the distribution of this size approaches a geometric law:*

$$S_{nk} \xrightarrow{\mathcal{D}} \mathit{Geo}(\gamma).$$

Proof (sketch). At the kth stage, color the entrant k with white, all else ($k-1$ nodes) with blue. Whenever a node recruits, paint its child with the same color. The Pólya-Eggenberger urn

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

underlies this process; the urn starts out with the initial conditions $W_0 = 1$ and $B_0 = k-1$. So, $S_{nk} = W_{n-k}$, which has the exact Pólya-Eggenberger distribution. Here the number of white splits $\tilde{W}_n =$

$W_n - 1$. Write the distribution of the white splits (see Theorem 5.1.1) as

$$P\{\tilde{W}_{n-k} = j\} = \frac{(k-1)\,\Gamma(n-j-1)\,\Gamma(n-k+1)}{\Gamma(n)\,\Gamma(n-k-j+1)}.$$

Passage to the limit, via Stirling's approximation to the Gamma function, as $k, n \to \infty$ in such a way that $k/n \to \gamma$, gives

$$P\{\tilde{S}_{nk} - 1 = j\} \to kn^{-j-1}(n-k)^j \to (1-\gamma)^j\gamma. \ \square$$

Meir and Moon (1988) brought the notion of the number of nodes of a certain outdegree in a recursive tree into the spotlight. They calculated the means of the number of nodes of outdegree 0 (the leaves), 1, and 2, and came up with a partial computation of the variance covariance matrix. Mahmoud and Smythe (1992) approached the subject via urn modeling.

Theorem 8.1.3. *(Mahmoud and Smythe, 1992). Let $X_n^{(j)}$ be the number of nodes of outdegree j (for $j = 0, 1, 2$) in a random recursive tree. The vector $\mathbf{X}_n = (X_n^{(0)}, X_n^{(1)}, X_n^{(2)})^T$ converges in distribution to a multivariate normal*

$$\frac{1}{\sqrt{n}}(\mathbf{X}_n - \boldsymbol{\mu} n) = \mathcal{N}_3(\mathbf{0}, \boldsymbol{\Sigma}).$$

where $\boldsymbol{\mu} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8})^T$, and the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{12} & -\frac{7}{72} & -\frac{5}{432} \\ -\frac{7}{72} & \frac{71}{432} & -\frac{37}{864} \\ -\frac{5}{432} & -\frac{37}{864} & \frac{473}{5184} \end{pmatrix}.$$

Proof (sketch). Suppose the leaves (nodes of outdegree 0) are colored with white, nodes of outdegree 1 are colored with blue, nodes with outdegree 2 with red, all else with black. By the same combinatorial arguments used before to connect the tree to an urn, one reasons that the urn scheme associated with the development is

$$\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

This is an extended urn, so the rest of the proof follows from Theorem 6.1.1. □

In principle, one can continue to refine the coloring of the nodes of a recursive tree, to reflect the number of nodes of outdegree $0, 1, \ldots, m$. One needs $m+2$ colors, one for each outdegree up to m, and one additional color for all the nodes of outdegree higher than m. The result would be a multivariate normal in $m+1$ dimensions, the $(m+1) \times (m+1)$ covariance matrix would require formidable linear algebra computations.

There is interest also in the structure of the branches or the subtrees rooted at node k. For example, Javanian and Vahidi-Asl (2003+), determine a central limit theorem for the root outdegree of the subtree rooted at k, as $n \to \infty$. While the size of the subtree rooted at k measures the success of the kth entrant in a chain letter scheme, the number of leaves is a measure of his legal liability. The leaves in the subtree are unsuccessful participants in the scheme, who purchased the letter but were not able to sell it. These leaves represent the potential number of complaints against the kth participant.

Theorem 8.1.4. *(Mahmoud and Smythe, 1991). Let $L_n^{(k)}$ be the number of leaves in the subtree rooted at k in a random recursive tree. For k fixed,*

$$\frac{L_n^{(k)}}{n} \xrightarrow{P} \frac{1}{2}\beta(1, k-1),$$

as $n \to \infty$.

Proof (sketch). A color code renders the colored nodes balls in an urn. Color the leaves of the subtree with white, the internal nodes of the subtree with blue, the nodes outside the subtree with red, to come up with the urn schema:

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

This extended urn scheme follows Gouet's, 1997 first order theory. □

It is worth noting that in the original proof of Mahmoud and Smythe (1991), the (scaled) difference between the number of white and blue balls is shown to be a mixture of normals, the mixing density is that of $\beta(1, k-1)$.

8.2 Pyramids

There is no restriction on the outdegree of the nodes of a recursive tree. Some chain letter schemes put a restriction on the number of copies a letter holder can sell. If a letter holder wishes to continue after he makes a certain number of sales, he must reenter (purchase a new copy). A reentry is a new participant (a new node in the tree). We illustrate the urn and associated type of result in the binary case. The tree modeling in this scheme is a *pyramid*, a recursive tree where outdegrees are restricted. A binary pyramid grows out of a root. When a node attracts its second child, it is saturated, and not allowed to recruit. The unsaturated nodes are equally likely recruiters. Only the first five trees in Figure 8 are binary pyramids.

Theorem 8.2.1. *(Gastwirth and Bhattacharya, 1984). Let L_n be the number of leaves in a binary pyramid of size n. Then, as $n \to \infty$,*

$$\frac{1}{n}\mathbf{E}[L_n] \to \frac{1}{2}(3-\sqrt{5}).$$

Proof (sketch). Color the leaves with white, the rest of the unsaturated nodes with blue. Leave the saturated nodes uncolored. When a leaf recruits, it remains unsaturated (of outdegree 1), and acquires a leaf under it; when a blue node recruits, it becomes saturated (colorless), but also acquires a leaf under it. From the point of view of the urn, colorless balls are tossed out. The schema is

$$\begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}$$

the result follows upon completing the eigenvalue computation outlined in Theorem 6.2.1. □

The original proof of Theorem 8.2.1 is established via diffusion theory, where additionally a central limit theorem is obtained.

8.3 Plane-Oriented recursive trees

Orientation in the plane was not taken into account in the definition of recursive trees. The two labeled trees in Figure 9 are only two drawings of the *same* recursive tree. If different orientations are taken to

represent different trees, we arrive at a definition of a *plane-oriented recursive tree*. In such a tree, if a node has outdegree Δ; there are Δ children under it, with $\Delta + 1$ "gaps." The leftmost and rightmost insertion positions are gaps, too. If these gaps are represented by external nodes, we obtain an extended plane-oriented recursive tree. Figure 10 shows one of the plane-oriented recursive trees of Figure 9 after it has been extended; the external nodes are shown as squares in Figure 10. As a stochastic process, the plane-oriented recursive tree grows by choosing one of the gaps at random, all gaps in the tree being equally likely. This uniform distribution on gaps gives rise to a uniform distribution on plane-oriented recursive trees.

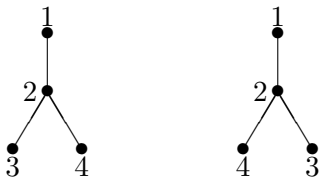


Figure 9: Two different plane-oriented trees.

Mahmoud, Smythe and Szymański (1993) characterized the exact and limit distribution of the number of leaves in a random plane-oriented recursive tree, using recurrence methods. Let L_n be the number of leaves in a random plane-oriented recursive tree. The exact distribution of L_n can be treated by counting methods. This random variable has the exact distribution

$$\mathbf{Prob}\{L_n = k\} = \frac{1}{(2n-3)!!} \left\langle\left\langle {n-1 \atop k-1} \right\rangle\right\rangle,$$

where $\left\langle\left\langle {n \atop k} \right\rangle\right\rangle$ is the Eulerian number of the second kind (see Graham, Knuth and Patashnik, 1994), and for odd j, the double factorial $j!!$ is $j(j-2) \times \ldots \times 5 \times 3 \times 1$.

Theorem 8.3.1. *(Mahmoud, Smythe and Szymański, 1993). In a plane-oriented recursive tree of size n, the number of leaves, L_n, exhibits the Gaussian tendency*

$$\frac{L_n - \frac{2}{3}n}{n} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{1}{9}\Big).$$

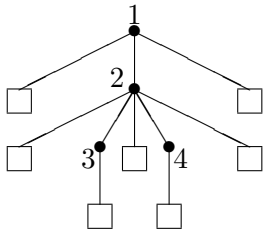


Figure 10: An extended plane-oriented recursive tree.

Proof (sketch). The equally likely objects in this model are the insertion positions or gaps (external nodes in the extended tree). Color each gap underneath a leaf with white, the rest of the gaps with blue. When a leaf recruits, it is turned into an internal node and acquires a new child as a leaf, with a white external node under it. Two blue gaps appear underneath the new internal node, as right and left siblings of the new leaf. When insertion hits a blue gap, it turns into a leaf, with one white gap under it; two blue gaps appear as siblings of the new leaf (net gain of only one blue gap). This is an urn process with the schema

$$\begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix},$$

The central limit tendency is an immediate application of Theorem 6.1. □

8.4 Phylogenetic trees

Certain forms of recursive trees have been used as models for phylogeny (see Aldous, 1995). An explicit form was suggested in McKenzie and Steel (2000) as a model for evolutionary relations allowing simple testing of the similarity between actual and simulated trees.

The phylogenetic tree is a binary tree with labeled leaves. The edges connected to the leaves are called *pendants*. The tree grows under some probability model. One appealing model is the Yule process, in which a species modeled by this stochastic system evolves by having a uniformly chosen pendant "split." This coincides precisely with the extended binary search tree (under the random permutation model), the external nodes of which are equally likely insertion position; Theorem 7.1.1 applies.

Other flavors of phylogenetic trees are discussed in McKenzie and Steel (2000), such as the unrooted version of the Yule model, and rooted and unrooted versions under a uniform probability model (all combinatorial trees of a particular size are equally likely). Similar results are derived, also with the aid of urns.

Figure 11 shows a phylogenetic tree on four species (leaves) and one on five species evolving from it by the splitting of an edge (pointed to by a down arrow). A pair of leaves at distance 2 (two leaves that are adjacent to a common internal node) are called a *cherry*. The number of cherries represents the number of species that are "cousins" or most similar on the hereditary scale.

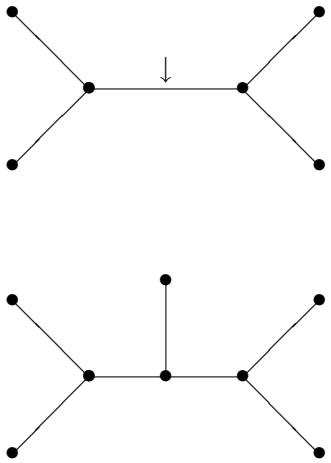


Figure 11: A phylogenetic tree on four species, and a tree on five species arising from it by the splitting of the edge pointed to.

Theorem 8.4.1. *(McKenzie and Steel, 2000). Let C_n be he number of cherries in a rooted uniform phylogenetic tree on n species. Then*

$$\frac{C_n - \frac{n}{4}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{1}{16}\Big).$$

Proof (sketch). What is equally likely in this tree model are the edges. Color the edges in a cherry with white, and all the other pendants with blue. Nonpendant edges are colored with red. When

a white pendant splits, the number of cherries remains the same, with two associated white pendants as before the splitting. But the new edge connected to the common internal vertex of the cherry is not pendant; the number of red edges increases by one. The old white sibling pendant of the splitting edge is no longer involved in a cherry, and must be turned blue. When a blue pendant splits, a new cherry appears, with two associated white edges, the nonpendant part of edge becomes red. When a red edge splits, it is partitioned into two red edges (net increase of one red edge), and a blue pendent is attached.

The colored edges evolve as the balls of an extended urn with the scheme

$$\begin{pmatrix} 0 & 1 & 1 \\ 2 & -1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

An eigenvalue analysis, as required in Theorem 6.1.1 gives a multivariate central limit result. The marginal distribution for the number of white pendants satisfies

$$\frac{W_n - \frac{n}{2}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{1}{4}\Big).$$

But then $W_n = 2C_n$. □

8.5 Bucket recursive trees

Bucket recursive trees were introduced in Mahmoud and Smythe, 1995 to generalize the ordinary recursive tree into one with nodes carrying multiple labels.

Every node in a bucket recursive tree has capacity b. The tree grows by adding labels (recruiting agents) from $\{1, \ldots, n\}$; at the jth step j is inserted. The affinities of all the recruiters are equal. The first agent goes into the root and recruits a second agent for the same office, then the two compete with equal chance to recruit a third, and so on. Up to b recruiters go into the root, filling it to full capacity. The b agents of the root office compete with equal chance to recruit the $(b+1)$st agent, who starts a new subordinate office (node attached as a child of the root). For each new entrant, the existing members of the tree compete with equal probabilities (adaptive in time) to recruit the entrant. If an agent belonging to an unfilled office succeeds in attracting the new entrant, the entrant joins

the recruiter's office. If the recruiter's office is full, the new entrant starts a new subordinate office attached as a child of the recruiter's office. Figure 12 shows a bucket recursive tree with bucket capacity 2, after 8 recruits join the system.

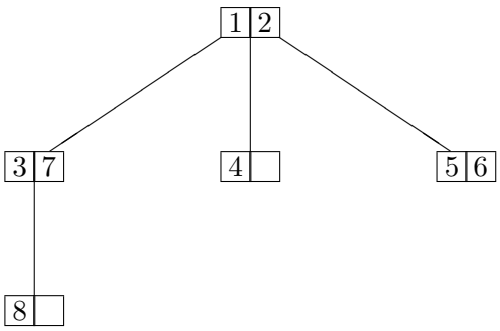


Figure 12: A bucket recursive tree with bucket capacity 2.

After n agents are in the system, the number of offices is S_n, the size of the tree. An urn of balls of b colors corresponds to the recruiting process. The affinity of an office with i agents in it is i and is represented in the urn by i balls of color i. When an office of capacity $i < b$ recruits, the urn loses i balls of color i and gains $i + 1$ balls of color $i + 1$. When a full office recruits, we only add a ball of color 1 to the urn. The schema is

$$\begin{pmatrix} -1 & 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -2 & 3 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \ddots & & \vdots & \vdots \\ 0 & 0 & & & \dots & & -(b-1) & b \\ 1 & 0 & & & \dots & & 0 & 0 \end{pmatrix}.$$

The characteristic equation $|\mathbf{A}^T - \lambda \mathbf{I}| = 0$ expands into

$$(\lambda + 1)(\lambda + 2) \dots (\lambda + b) = b!.$$

Again, this urn scheme is Smythe's, 1996 extended urn scheme, the eigenvalues of which have the property that $\Re\, \lambda_2 < \frac{1}{2}\lambda_1 = \frac{1}{2}$, for b up to 26 (see Mahmoud and Smythe, 1995 for a proof). For $b \leq 26$,

one obtains a multivariate normal law among the number of balls of different kinds. Through graph-theoretic considerations, one obtains the average size of the tree from the profile of colors (see Mahmoud and Smythe, 1995 for details).

Theorem 8.5.1. *(Mahmoud and Smythe, 1995). Let S_n be the size of a bucket recursive tree after n insertions. If the node capacity b satisfies $3 \le b \le 26$, then*

$$\frac{S_n - n/H_b}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_b^2),$$

where σ_b^2 is an effectively computable constant.

The limiting variance σ_b^2 requires some rather lengthy computation. If $b \ge 27$, the associated urn no longer satisfies the sufficient condition for asymptotic normality in an extended urn in the sense of Smythe (1996). It is argued in Mahmoud and Smythe (1995) that $(S_n - n/H_b)/\sqrt{n}$ does not converge to a normal limit, if $b \ge 27$. (Mahmoud and Smythe (1995) show that under a different scale, a nonnormal limit exists.)

8.6 Sprouts

A *random recursive sprout*, introduced in Mahmoud (2003+), is a random tree that grows just like the ordinary recursive tree, except that the recruiter attracts a random number of nodes instead of just 1. If the recruiter is internal, $Y \ge 1$ leaves are adjoined to it; and if a leaf is chosen as parent, $X \ge 1$ leaves are attached to it as children, and the parent leaf turns into an internal node. The random variables X and Y are of course discrete on positive integers, and may generally be different. The standard recursive tree is a case where deterministically $X \equiv Y \equiv 1$.

The random sprout extends the domain of application of recursive trees. For example, it may model a chain letter scheme, where the company starts out with one recruiter who seeks buyers for a copy of a letter he is holding. At some point the recruiter will sell a copy of his letter to X persons whom he may meet socially. The initial founder of the company remains in contention and all $X + 1$ letter holders go competing for the next batch of buyers with equal chance. The

chain letter scheme propagates in society, however, whenever someone who sold copies of the letter before finds buyers he sells according to the distribution Y (presumably stochastically larger than X). The model takes experience into account. Those who have the experience of selling before can generally sell more than new inexperienced letter holders.

Suppose the leaves of the sprout are colored white and the internal nodes are colored blue, and think of these nodes as balls in an urn. The progress of the tree corresponds to the growth of an urn with the schema

$$\mathbf{A} = \begin{pmatrix} X-1 & 1 \\ Y & 0 \end{pmatrix}.$$

Let S_n be the size of the sprout, Let $\mu_x := \mathbf{E}[X]$, and $\mu_Y := \mathbf{E}[Y]$. Here, $\mathbf{X}_0 = (1,0)^T$, and the transposed generator is

$$\mathbf{B} = \begin{pmatrix} \mu_x - 1 & \mu_Y \\ 1 & 0 \end{pmatrix},$$

with eigenvalues

$$\lambda_{1,2} = \frac{(\mu_X - 1) \pm \sqrt{(\mu_X - 1)^2 + 4\mu_Y}}{2}.$$

As both $\mu_x \geq 1$ and $\mu_Y \geq 1$, the two eigenvalues are real and distinct. By Theorem 6.2.1,

$$\mathbf{E}[\mathbf{X}_n] = \frac{\mu_Y}{\mu_Y - \lambda_2} \begin{pmatrix} \mu_X - 1 - \lambda_2 \\ 1 \end{pmatrix} n + \mathbf{o}_2(n).$$

Adding up the two components of $\mathbf{X}_n$, we arrive at the average size of a random sprout:

$$\mathbf{E}[S_n] = \mu_Y \frac{\mu_X - \lambda_2}{\mu_Y - \lambda_2} n + o(n).$$

9 Extensions and future directions

Starting with a simple 2×2 urn model, the theory was developed in many directions and was used in many applications. There are still numerous promising directions under investigation.

For example, Pemantle (1990) looked into schemata that are adaptive in time. That is, the ball addition matrix itself is changing with

time. This adaptive urn process is proposed as a model for the opinion polls in the American presidential election, where voters may be changing their mind throughout the campaign.

Aldous, Flannery and Palacios (1988) considered a generalized model where the drawing probabilities of a certain color are not just the proportion of balls of that color in the urn, but more generally a proportional factor of it—for a $k \times k$ scheme, with $X_n^{(i)}$ being the number of balls of color i after n draws, the probability of picking color i at the nth draw is $c_i X_{n-1}^{(i)} / \sum_{j=1}^{k} c_j X_{n-1}^{(j)}$, for a vector of proportionality constants $(c_1, \ldots, c_k)^T$. The results remain pretty much the same; for example, we have strong convergence of proportions to a constant vector with components related to the principal eigenvectors. Janson (2003+), presents a broader context.

Certain variations of Pólya urns have become popular models in clinical trials (see Wei, 1977, 1978a, 1978b, and contributions in Bai, Hu and Rosenberger, 2002, and Rosenberger, 2002). Bai and Hu (1999) recognize the need for adaptive schemata in clinical trials. Under impositions of slow change, they still get results resembling those of Smythe (1996). Namely, if $\mathbf{A}_n$ is the schema used at the nth step, and there exists a matrix $\mathbf{A}$ such that $\sum_{n=1}^{\infty} n^{-1} ||\mathbf{A}_n - \mathbf{A}||_\infty < \infty$, then central limit theorems similar to those in Theorem 6.1.1 apply (still, subject to an eigenvalue structure like Smythe's with second largest real part less than half of the principal eigenvalue).

For the spreading of several epidemics simultaneously, Kriz (1972) proposed a model of several parallel generalized Pólya urns leading naturally to the notion of a joint vector of independent Pólya distributions, which was called the polytomic Pólya distribution.

We portrayed growing tenable urns in the review. There are classes of tenable urns that do not grow. For example, the urn with the scheme

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \tag{19}$$

remains of the same size as the starting urn, after any number of draws. This is the Ehrnfest urn (Ehrenfest and Ehrenfest, 1907), used for modeling diffusion of gas particles; Karlin and McGregor (1965) provide a contemporary view. Initially the model comprises two chambers of particles. A particle is chosen at random from among all the particles in the two chambers, and moved to the other chamber. This process is equivalent to one Pólya urn, where all the balls are put together, those of one urn are colored white, those of the other

are colored blue, then the urn evolves according to the schema (19). The long-term trend reveals the degree of mixing between the two gases.

We briefly give the reader the flavor of the kind of question one might address, and the sort of answer one gets in a nongrowing urn by one example in the Ehrnfest model. The total number of balls in the combined urn is fixed, say τ_0, but the number of balls of one particular color has a discrete distribution over nonnegative integers in the range $\{0, 1, \ldots, \tau_0\}$.

Theorem 9.1. *(Ehrenfest and Ehrenfest, 1907). Suppose the Ehrenfest model starts out with W_0 particles in one chamber and B_0 in the other. Let W_n be the number of particles in the first chamber after n exchanges. Then*

$$W_n \xrightarrow{\mathcal{D}} B\Big(W_0 + B_0, \frac{1}{2}\Big).$$

Proof (sketch). After n exchanges, there are W_n white and B_n blue balls respectively in the combined Pólya urn. For W_n to be k after n draws, we must either have $k+1$ white balls in the previous step, and draw white, thus decreasing the white balls by one, or have $k-1$ white balls in the previous step, and increase the white balls by drawing blue:

$$\begin{aligned} P\{W_n = k\} &= \frac{k+1}{W_0+B_0}\, P\{W_{n-1} = k+1\} \\ &\quad + \frac{W_0+B_0-k+1}{W_0+B_0}\, P\{W_{n-1} = k-1\}. \end{aligned}$$

After justifying the existence of a limit $P\{W_n = k\} \to P(k)$, we find that the limit must satisfy

$$P(k) = \frac{k+1}{W_0+B_0}\, P(k+1) + \frac{W_0+B_0-k+1}{W_0+B_0}\, P(k-1).$$

Reorganize as

$$(W_0+B_0)(P(k) - P(k-1)) = (k+1)\, P(k+1) - (k-1)\, P(k-1).$$

From this write down a system of equations for $k, k-1, \ldots, 0$ (with the natural interpretation that $P(-1) = 0$), and add them up to obtain

$$P(k+1) = \frac{W_0+B_0-k}{k+1}\, P(k).$$

By induction this gives

$$P\{W_n = k\} \to P(k) = \frac{1}{2^{W_0+B_0}}\binom{W_0+B_0}{k},$$

for each $k = 0, \ldots, W_0 + B_0$. □

The common theme in all the forgoing material is that *one* ball is sampled at random. Applications warrant modeling urns subject to multiple draws. For instance, we may consider an urn of white and blue balls where at each stage a pair of distinct balls is withdrawn. The pair is returned to the urn, and depending on the multiset outcome, we may decide to add a certain number of white balls and a certain number of blue balls. In this case we have a *rectangular* schema. The instance

$$\begin{array}{c} \\ WW \\ WB \\ BB \end{array} \begin{array}{c} \begin{array}{cc} W & B \end{array} \\ \begin{pmatrix} -1 & 2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

appears in Tsukiji and Mahmoud, 2001 in an application to circuits. The binary recursive circuit can be viewed as the second member in a hierarchy of random graphs, with the recursive forest being the first in the hierarchy. The Fth member of the hierarchy grows from an initial set of nodes (inputs in the context of circuits); at each stage, F distinct parents (fan-in of the circuit) are chosen to parent a new child. A forest of recursive trees ($F = 1$) is the simplest member of the hierarchy. Figure 13 shows all binary ($F = 2$) circuits after two insertions into an initial graph of two isolated nodes of outdegree 0 (fan-out 0 in the context of circuits).

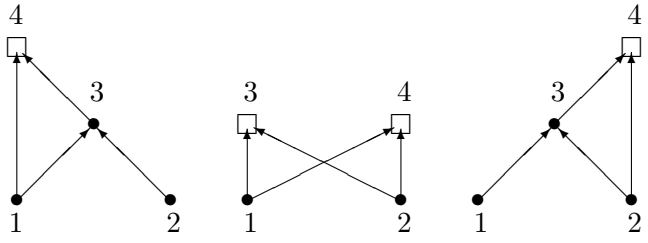


Figure 13: All binary circuits of size 4 grown from two inputs. Square nodes are outputs.

The rectangular urn schema above is obtained by coloring outputs (nodes of fan-out 0) with white, all else with blue, and arguing as usual the various changes in the colors upon withdrawing two parents. Via martingale difference formulation, Tsukiji and Mahmoud, 2001 find the central limit tendency

$$\frac{L_n - \frac{1}{3}n}{\sqrt{n}} = \mathcal{N}\Big(0, \frac{2}{45}\Big)$$

for L_n, the number of outputs in the random circuit after n insertions.

The eigenvalue theory used for results such as Smythe (1996), Mahmoud (2003+), and Bai and Hu (1999) would not be a natural course of investigation for rectangular urns. However, martingale theory remains viable.

Very recently, an analytic attack by Flajolet, Gabarró and Pekari (2003+), has been successful for 2×2 schemes. In this approach various types of functional equations are set up and solved, in some cases exactly. The approach yields explicit large deviation rates, and detects elliptic functions. The case

$$\begin{pmatrix} s+1 & 0 \\ 1 & s \end{pmatrix},$$

with $s \geq 1$, declared problematic in Bagchi and Pal, 1985, is shown to have a stable, but nonnormal law. (In this case $s = \lambda_2 \geq \frac{1}{2}\lambda_1 = s{+}1$.) This is a continuing effort, and the results will soon be announced.

References

Aldous, D. (1995), Probability distributions on cladograms. In Random Discrete Structures, Eds. Aldous, D. and Pemantle, R., IMA Volumes in Mathematics and its Applications, **76**, 1–18.

Aldous, D., Flannery, B. and Palacios, J. (1988), Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains. Probability in Engineering and Information Sciences, **2**, 293–307.

Athreya, K. and Karlin, S. (1968), Embedding of urn schemes into continuous time Markov branching process and related limit theorems. Annals of Mathematical Statistics, **39**, 1801–1817.

Athreya, K. and Ney, P. (1972), Branching Processes. New York: Springer-Verlag.

Bagchi, A. and Pal, A. (1985), Asymptotic normality in the generalized Pólya-Eggenberger urn model with applications to computer data structures. SIAM Journal on Algebraic and Discrete Methods, **6**, 394–405.

Bai, Z. and Hu, F. (1999), Asymptotic theorems for urn models with nonhomogeneous generating matrices. Stochastic Processes and their Applications, **80**, 87–101.

Bai, Z., Hu, F. and Rosenberger, W. (2002), Asymptotic properties of adaptive designs for clinical trials with delayed response. Annals of Statistics, **30**, 122–139.

Balaji, S. and Mahmoud, H. (2003+), Tenability of Pólya urns. Journal of College Mathematics (Problem), submitted.

Bentley, J. (1975), Multidimensional binary search trees used for associative searching. Communications of the ACM, **18**, 509–517.

Bergeron, F., Flajolet, P. and Salvy, B. (1992), Varieties of increasing trees. Lecture Notes in Computer Science, Ed. J. Raoult, **581**, 24–48.

Bernstein, S. (1940), Sur un problème du schéme des urnes à composition variable. C. R. (Dokl.) Acad. Sci. URSS, **28**, 5–7.

Billingsley, P. (1968), Convergence of Probability Measures. New York: Wiley.

Chern, H. and Hwang, H. (2001), Phase change in random m-ary search trees and generalized quicksort. Random Structures and Algorithms, **19**, 316–358.

Chow, Y. and Teicher, H. (1978), Probability Theory. New York: Springer-Verlag.

Devroye, L. (1991), Limit laws for local counters in random binary search trees. Random Structures and Algorithms, **2**, 303–316.

Devroye, L. and Vivas, A. (1998), Random fringe-balanced quadtrees, (manuscript).

Eggenberger, F. and Pólya, G. (1923), Über die statistik verketetter vorgäge. Zeitschrift für Angewandte Mathematik und Mechanik, **1**, 279–289.

Ehrenfest, P. and Ehrenfest, T. (1907), Über zwei bekannte einwände gegen das Boltzmannsche H-theorem. Physik, Z., **8**, 311–314.

Feller, W. (1971), An Introduction to Probability Theory and Its Applications. Volume II, New York: Wiley.

Finkel, R. and Bentley, J. (1974), Quad trees: a data structure for retrieval on composite keys. Acta Informatica, **4**, 1–9.

Flajolet, P., Gabarró, J. and Pekari, H. (2003+), Analytic urns, (manuscript).

Fréchet, M. (1943), Les Probabilités Associées a un Système d'Evenements Compatibles et Dépendants. Paris: Herman.

Freedman, D. (1965), Bernard Friedman's urn. Annals of Mathematical Statistics, **36**, 956–970.

Friedman, B. (1949), A simple urn model. Communications of Pure and Applied Mathematics, **2**, 59–70.

Gastwirth, J. (1977), A probability model of pyramid schemes. American Statistician, **31**, 79–82.

Gastwirth, J. and Bhattacharya, P. (1984), Two probability models of pyramids or chain letter schemes demonstrating that their promotional claims are unreliable. Operations Research, **32**, 527–536.

Gonnet, G. and Munro, J. (1984), The analysis of linear probing sort by the use of a new mathematical transform. Journal of Algorithms, **5**, 451–470.

Gouet, R. (1989), A martingale approach to strong convergence in a generalized Pólya-Eggenberger urn model. Statistics and Probability Letters, **8**, 225–228.

Gouet, R. (1993), Martingale functional central limit theorems for a generalized Pólya urn. Annals of Probability, **21**, 1624–1639.

Gouet, R. (1997), Strong convergence of proportions in a multicolor Pólya urn. Journal of Applied Probability, **34**, 426–435.

Graham, R., Knuth, **E**. and Patashnik, O. (1994), Concrete Mathematics. Massachusetts: Addison-Wesley, Reading.

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

Hermosilla, L. and Olivos, J. (1985), A bijective approach to single rotation trees. Presented at The Fifth Conferencia Internacional en Ciencia de la Computacion, Santiago, Chile.

Inoue, K. and Aki, S. (2001), Pólya urn model under general replacement schemes. Journal of Japan Statistical Society, **31**, 193–205.

Jacquet, P. and Régnier, M. (1986), Trie Partitioning process: limiting distributions. Lecture Notes in Computer Science, **214**, 196–210.

Jacquet, P. and Szpankowski W. (1998), Analytical depoissonization and its applications. Theoretical Computer Science, **201**, 1–62.

Janson, S. (2003+), Functional limit theorems for multitype branching processes and generalized Pólya urns, (manuscript).

Javanian, M. and Vahidi-Asl, M. (2003+), Note on the outdegree of a node in random recursive trees. Journal of Applied Mathematics and Computing (to appear).

Johnson, N. and Kotz, S. (1977), Urn models and Their Applications. New York: Wiley.

Kac, M. (1949), On deviations between theoretical and empirical distributions. Proc. N. A. S., **35**, 252–257.

Karlin, S. and McGregor, J. (1965), Ehrenfest urn models. Journal of Applied Probability, **2**, 352–376.

Kemp, R. (1984), Fundamentals of the Average Case Analysis of Particular Algorithms. Wiley-Teubner Series in Computer Science, New York: Wiley.

Knuth, D. (1998), The Art of Computer Programming. Vol. 3: Sorting and Searching, 2nd ed. Massachusetts: Addison-Wesley, Reading.

Kotz, S. and Balakrishnan, N. (1997), Advances in urn models during the past two decades. In Advances in Combinatorial Methods and Applications to Probability and Statistics, **49**, 203–257. Birkhäuser, Boston.

Kotz, S., Mahmoud, H. and Robert, P. (2000), On generalized Pólya urn models. Statistics and Probability Letters, **49**, 163–173.

Kriz, J. (1972), Die PMP-vertielung. Stat. Hefte, **13**, 211–224.

Mahmoud, H. (1992), Evolution of Random Search Trees. New York: Wiley.

Mahmoud, H. (1998), On rotations in fringe-balanced binary trees. Information Processing Letters, **65**, 41–46.

Mahmoud, H. (2000), Sorting: A distribution Theory. New York: Wiley.

Mahmoud, H. (2002), The size of random bucket trees via urn models. Acta Informatica, **38**, 813–838.

Mahmoud, H. (2003+), Invertible growing urn schemes with an application to random sprouts. Acta Informatica, tentatively accepted.

Mahmoud, H. and Smythe, R. (1991), On the distribution of leaves in rooted subtrees of recursive trees. The Annals of Applied Probability, **1**, 406–418.

Mahmoud, H. and Smythe, R. (1992), Asymptotic joint normality of outdegrees of nodes in random recursive trees. Random Structures and Algorithms, **3**, 255–266.

Mahmoud, H. and Smythe, R. (1995), Probabilistic analysis of bucket recursive trees. Theoretical Computer Science, **144**, 221–249.

Mahmoud, H., Smythe, R. and Szymański, J. (1993), On the structure of plane-oriented recursive trees and their branches. Random Structures and Algorithms, **4**, 151–176.

Markov, A. (1917), Generalization of a problem on a sequential exchange of balls (in Russian). Collected Works, Meeting of Physico-Mathematical Society of the Academy of Sciences.

McKenzie, A. and Steel, M. (2000), Distributions of cherries for two models of trees. Mathematical Biosciences, **164**, 81–92.

Meir, A. and Moon, J. (1988), Recursive trees with no nodes of out-degree one. Congressus Numerantium, **66**, 49–62.

Najock, D. and Heyde, C. (1982), On the number of terminal vertices in certain random trees with an application to stemma construction in philology. Journal of Applied Probability, **19**, 675–680.

Panholzer, A. and Prodinger, H. (1998), An analytic approach for the analysis of rotations in fringe-balanced binary search trees. Annals of Combinatorics, **2** , 173–184.

Pemantle, R. (1990), A time-dependent version of Pólya's urn. Journal of Theoretical Probability, **3**, 627–637.

Poblete, P. (1987), Approximating functions by their Poisson transform. Information Processing Letters, **23**, 127–130.

Poblete, P. and Munro, J. (1985), The analysis of a fringe heuristic for binary search trees. Journal of Algorithms, **6**, 336–350.

Pólya, G. (1931), Sur quelques points de la théorie des probabilités. Annals of the Institute of Henri Poincaré, **1**, 117–161.

Riordan, J. (1968), Combinatorial Identities. New York: Wiley.

Rosenberger, W. (2002), Randomized urn models and sequential design. Sequential Analysis, **21**, 1–41.

Rosenblatt, A. (1940), Sur le concept de contagion de M. G. Pólya dans le calcul des probabilités. Proc. Acad. Nac. Cien. Exactas, Fis. Nat., Peru (Lima), **3**, 186–204.

Ross, S. (1983), Stochastic Processes. New York: Wiley.

Samet, H. (1990), Applications of Spatial Data Structures. Massachusetts: Addison-Wesley, Reading.

Savkevich, V. (1940), Sur le schéma des urnes à composition variable. C. R. (Dokl.) Acad. Sci. URSS, **28**, 8–12.

Smiley, M. (1965), Algebra of Matrices. Allyn and Bacon, Inc., Boston, Massachusetts.

Smythe, R. (1996), Central limit theorems for urn models. Stochastic Processes and Their Applications, **65**, 115–137.

Smythe, R. and Mahmoud, H. (1996), A survey of recursive trees. Theory of Probability and Mathematical Statistics, **51**, 1–29 (appeared in Ukrainian in (1994)),

Tchuprov, A. (1922), Ist die normale stabilität empirisch nachweisbar? Nord Stat. Tidskr., **B/I**, 373–378.

Tsukiji, T. and Mahmoud, H. (2001), A limit law for outputs in random circuits. Algorithmica, **31**, 403–412.

Wei, L. (1977), A class of designs for sequential clinical trials. Journal of the American Statistical Association, **72**, 382–386.

Wei, L. (1978a), The adaptive biased coin design for sequential experiments. The Annals of Statistics, **6**, 92–100.

Wei, L. (1978b), An application of an urn model to the design of sequential controlled clinical trials. Journal of the American Statistical Association, **73**, 559–563.

Yao, A. (1978), On random 2–3 trees. Acta Informatica, **9**, 159–170.

Yao, F. (1990), Computational Geometry. In Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity, Eds. van Leeuwen, J., 343–389. Amsterdam: MIT Press.