Central Limit Theorem for the Size of the Range of a Renewal Process

Paweł Hitczenko [1,2,3]
Robin Pemantle [4,5]

Abstract: We study the range of a Markov chain moving forward on the positive integers. For every position, there is a probability distribution on the size of the next forward jump. Taking a scaling limit as the means and variances of these distributions approach given continuous functions of position, there is a Gaussian limit law for the number of sites hit in a given rescaled interval.

We then apply this to random coupling. At each time, n, a random function f_n is applied to the set $\{1, \ldots, N\}$. The range R_n of the composition $f_n \circ \cdots \circ f_1$ shrinks as n increases. A Gaussian limit law for the total number of values of $|R_n|$ follows from the limit law together with an extension to non-compact rescaled ranges.

Keywords: iterated function, random function, Markov chain, coupling

Subject classification: 60K05, 60J10, 60F05

Running headline: CLT for the range of a renewal process

[1]corresponding author
[2]Research supported in part by NSA grant # MSPF-02G-043
[3]Department of Mathematics, Drexel University, 32nd and Chestnut Streets, Philadelphia, PA 19104, phitczen@mcs.drexel.edu, fax: (215)895-1582
[4]Research supported in part by National Science Foundation grant # DMS 0103635
[5]University of Pennsylvania Department of Mathematics, 209 S. 33rd Street, Philadelphia, PA 19104, pemantle@math.upenn.edu

1 Introduction

Let $\{f_n : n = 1, 2, 3, \ldots\}$ be random functions chosen independently and uniformly from the set of all N^N functions from the set $[N] := \{1, \ldots, N\}$ to itself. The composition $g_n := f_n \circ \cdots \circ f_1$ has been studied in various contexts. It models the coalescence of ancestry as one goes backwards in time in a simple population genetics model usually referred to as the Wright-Fisher model. (We refer to [7] for a detailed discussion of population genetics, here we mention only that our work is most directly related to the Wright-Fisher model for haploid populations that goes back to [4]; for further information see e.g. [10, 15].) It also can be viewed as a toy model for studying the complete coupling time for a Markov chain, when the coupling is chosen at random ([16]). Several quantities are of interest here. For instance, the size of the range of g_n is a non-increasing process eventually absorbed at 1. One may ask for the time necessary to reach this state ([5, 12, 15]). Also of interest is the number of distinct sizes that the range of g_n takes on.

The size of the range of g_n is a Markov chain in n. It is not hard explicitly to compute the transition probabilities, nor surprising that they are approximated by continuous functions as $n, N \to \infty$. One may then compute means and variances for the jump from one size of range to the next, sum over jumps, pass to the limit where the sum becomes an integral, and quickly arrive at a plausible Gaussian limit. The main point of this paper is to do just this: prove a general CLT for the number of sites hit by a rescaled renewal process and use this to derive a Gaussian limit law for the number of distinct sizes taken on by the range of g_n.

This is in some sense straightforward. The general CLT for the size of the range of a renewal process is straightforward to prove. We have explicit knowledge of the transition probabilities. On the other hand, we find two aspects of this endeavor compelling. First, we were surprised that a statement of a Gaussian limit law for the size of the range of a renewal process under this type of rescaling could not be found in the literature. We would like to correct this omission, especially since the formula for the rescaled variance is not immediately obvious. Secondly the application to iterated functions requires some special care because of the non-compactness of the rescaled range.

The next section states and proves the CLT for renewal processes whose jump means and variances converge to a function of a continuous location parameter. The subsequent section contains the application to the collapsing random function chain and the last section contains a few further remarks.

2 Central limit theorem

We are concerned with the number of sites hit by a renewal process. More vividly, imagine a board game where you roll the dice but can only move in one direction. If the distribution of the die-roll can be an arbitrary function of the present position then one has a general time-inhomogeneous renewal process; of course it is difficult to say much at this level. Let us suppose we have a family of games whose board size N goes to infinity, and that the mean and variance $\mu_N(\lceil xf(N) \rceil)$ and $V_N(\lceil xf(N) \rceil)$, of the die-roll when in position $\lceil xf(N) \rceil$ converge to some limits $\mu(x)$ and $V(x)$. We would then expect scaling limits for derived quantities, such as the number of sites hit in a rescaled range $[f(N)a, f(N)b]$. Indeed, one readily sees that a proportion of $1/\mu(x)$ of the sites near $xf(N)$ are hit, and believes that the difference between the number hit and its mean should have a normal limit.

For the remainder of this section, let functions f on $\mathbb{Z}^+$ and μ, V on $\mathbb{R}^+$ be given. Suppose that $\{P_N(a,\cdot)\}$ are a family of transition probabilities, so that in the N^{th} process, the probability of a jump from a to $a+k$ is $P_N(a,k)$. Define

$$\mu_N(a) := \sum_x x P_N(a,x), \qquad V_N(a) := \sum_x (x-\mu_N(a))^2 P_N(a,x)$$

to be the mean and variance of P_N. We assume that there are continuous functions μ and V, and an increasing f going to infinity, such that $\mu_N(\lceil \cdot\, f(N)\rceil) \to \mu(\,\cdot\,)$ and $V_N(\lceil \cdot\, f(N)\rceil) \to V(\,\cdot\,)$. In order to avoid writing ceilings, from now on we will assume that the functions μ_N and V_N are defined on the whole real line (and are constant between consecutive integers).

Let $a < b$ be positive numbers and define $R_N = R_N(a,b) := |R \cap [f(N)a, f(N)b]|$ to be the size of the range of the process P_N intersected with the interval $[f(N)a, f(N)b]$. Then

Theorem 1 *(i) Under the above assumptions we have:*

$$\mathbb{E}R_N = (1+o(1)) \cdot f(N) \int_a^b \frac{dx}{\mu(x)},$$

(ii) For $\delta > 0$ set $\nu_{N,\delta}(a) := \sum_x |x|^{2+\delta} P_N(a,x)$ and assume that there exist $\delta > 0$ and a continuous function ν_δ such that

$$\nu_{N,\delta}(\lceil x f(N)\rceil) \le \nu_\delta(x), \quad N \ge 1,\ a \le x \le b. \tag{1}$$

Then

$$\frac{R_N - f(N)\int_a^b (1/\mu(x))dx}{\sqrt{V_N}} \Longrightarrow N(0,1)$$

where

$$V_N = V_N(a,b) := f(N) \int_a^b \frac{V(x)}{\mu^3(x)}\, dx\,;$$

and $N(0,1)$ denotes a standard Gaussian random variable.

Proof: To prove (i), let $\{X_j\}$ be the values of the Markov chain and let

$$\tau_a = \inf\{j :\ X_j \ge f(N)a\} \quad \text{ and } \quad \tau_b = \inf\{j \ge \tau_a :\ X_j \ge f(N)b\}.$$

We will show that (under the assumption that $\mu_N(xf(N))$ is sufficiently close to $\mu(x)$) we have

$$\mathbb{E}R_N = \cdot f(N) \int_a^b \frac{dx}{\mu(x)} + O\left(\frac{\ln f(N)}{f(N)}\right).$$

Consider the martingale difference sequence array

$$d_{N,k} = 1 - \frac{X_k - X_{k-1}}{\mu_N(X_{k-1})}, \quad \mathcal{F}_k = \sigma(X_0, \dots, X_k). \tag{2}$$

Our proof rests on an observation that each of the "1" in the definition of $d_{N,k}$ corresponds to a visited state. Thus to analyze the total number of states visited within the interval $[f(N)a, f(N)b]$ we focus on the sequence $d_{N,k}$ started at τ_a and stopped at τ_b, $d_{N,k}I(\tau_a < k \le \tau_b)$. Since this new sequence preserves martingale difference property we have

$$\mathbb{E}R_N = \mathbb{E}\sum_{k=1}^{N} I(\tau_a < k \le \tau_b)\frac{X_k - X_{k-1}}{\mu_N(X_{k-1})} = \mathbb{E}\sum_{k=\tau_a+1}^{\tau_b} \frac{X_k - X_{k-1}}{\mu_N(X_{k-1})}.$$

We now note that the expression within the expectation on the right is roughly a Riemann sum approximation of the integral

$$\int_{af(N)}^{bf(N)} \frac{dx}{\mu_N(x)} = f(N)\int_a^b \frac{dx}{\mu_N(xf(N))} = f(N)\left(\int_a^b \frac{dx}{\mu(x)} + o(1)\right)$$

where a (random) partition $\{f(N)a, X_{\tau_a}, \dots, X_{\tau_b-1}, f(N)b\}$ of the interval $[f(N)a, f(N)b]$ is used. Thus, (i) will be proven once we show that the expected error in that approximation is sufficiently small. To this end write

$$\sum_{k=\tau_a+1}^{\tau_b} \frac{X_k - X_{k-1}}{\mu_N(X_{k-1})} = \int_{af(N)}^{bf(N)} \frac{dx}{\mu_N(x)} - \int_{af(N)}^{X_{\tau_a}} \frac{dx}{\mu_N(x)} + \frac{X_{\tau_b} - bf(N)}{\mu_N(X_{\tau_b-1})} \tag{3}$$

$$+ \sum_{k=\tau_a+1}^{\tau_b} \left\{\frac{X_k \wedge bf(N) - X_{k-1}}{\mu_N(X_{k-1})} - \int_{X_{k-1}}^{X_k\wedge bf(N)} \frac{dx}{\mu_N(x)}\right\}. \tag{4}$$

We will show that the expectation of the sum of the all the terms but the first is $O(\ln(f(N))/f(N))$. Set

$$A_0^j = \{X_0 = j\} \quad \text{and} \quad A_r^j = \{X_0 < j, \dots, X_{r-1} < j, X_r = j\}, \quad r \ge 1.$$

Then $\sum_{r\ge 0}\mathbb{P}(A_r^j) = \mathbb{P}(j \in R) \le 1$ and thus by Markov property, Cauchy-Schwarz, and Chebyshev's inequality we see that the expectation of the last term in (3) is bounded above by

$$\begin{aligned}
\mathbb{E}\frac{X_{\tau_b} - X_{\tau_b-1}}{\mu_N(X_{\tau_b-1})} &= \sum_{r\ge 0}\sum_{j<bf(N)} \mathbb{E}I_{A_r^j} I(X_{r+1} > bf(N))\frac{X_{r+1} - j}{\mu_N(j)} \\
&= \sum_{r\ge 0}\sum_{j<bf(N)} \mathbb{E}\frac{I_{A_r^j}}{\mu_N(j)}\mathbb{E}((X_{r+1} - j)I(X_{r+1} > bf(N)|A_r^j) \\
&= \sum_{r\ge 0}\sum_{j<bf(N)} \mathbb{E}\frac{I(j\in R)}{\mu_N(j)}\mathbb{E}((X_1 - j)I(X_1 > bf(N))|X_0 = j) \\
&\le \sum_{j<bf(N)} \mathbb{E}\frac{I(j\in R)}{\mu_N(j)}\frac{\mathbb{E}((X_1 - j)^2|X_0 = j)}{bf(N) - j} \le \sum_{j<bf(N)} \frac{V_N(j) + \mu_N^2(j)}{\mu_N(j)}\cdot\frac{1}{bf(N) - j} \\
&= O\left(\sup_{0\le x\le b}\left\{\frac{V(x) + \mu^2(x)}{\mu(x)}\right\}\right)\cdot \sum_{1\le j<bf(N)} \frac{1}{j} = O(\ln f(N))
\end{aligned}$$

Next, the expectation of the second integral in (3) is

$$f(N) \cdot \mathbb{E} \int_a^\infty \frac{I(X_{\tau_a}/f(N) \geq x)}{\mu_N(xf(N))} dx = f(N) \int_a^\infty \frac{\mathbb{P}(X_{\tau_a} \geq xf(N))}{\mu_N(xf(N))} dx, \tag{5}$$

and, for any $\ell > af(N)$,

$$\begin{aligned}
\mathbb{P}(X_{\tau_a} \geq \ell) &= \sum_{j<af(N)} \sum_{r \geq 0} \mathbb{P}(A_r^j, X_{r+1} \geq \ell) = \sum_{j<af(N)} \sum_{r \geq 0} \mathbb{P}(A_r^j) \cdot \mathbb{P}(X_{r+1} \geq \ell | X_r = j) \\
&\leq \sum_{j<af(N)} \mathbb{P}(X_1 - j \geq \ell - j | X_0 = j) \leq \sum_{j<af(N)} \frac{\mathbb{E}((X_1 - j)^2 | X_0 = j)}{(\ell - j)^2} \\
&\leq \sup_{1 \leq k < af(N)} \{V_N(k) + \mu_N^2(k)\} \sum_{j=1}^\infty \frac{1}{(\ell - af(N) + j)^2} \\
&= O\left(\sup_{0 \leq x \leq a} \{V(x) + \mu^2(x)\} \right) \cdot \frac{1}{\ell + 1 - af(N)}.
\end{aligned}$$

Let $\gamma_N = o(f(N))$ satisfy $\gamma_N \to \infty$. Splitting the integral in (5) according whether $x > a + \gamma_N/f(N)$ or not, and using the above bound, we see that the quantity in (5), up to a multiplicative factor of $O\left(\sup_{x<a}(V(x) + \mu^2(x))\right)$, is bounded above by

$$\begin{aligned}
& f(N) \int_a^{a+\gamma_N/f(N)} \frac{\mathbb{P}(X_{\tau_a} \geq xf(N))}{\mu_N(xf(N))} dx + f(N) \mathbb{P}(X_{\tau_a} \geq af(N) + \gamma_N) \cdot \int_a^\infty \frac{dx}{\mu_N(xf(N))} \\
&= O\left(\sup_{a \leq x \leq a + \gamma_N/f(N)} \frac{1}{\mu(x)} \right) \cdot \int_{af(N)}^{af(N)+\gamma_N} \frac{dy}{y + 1 - af(N)} \\
&\qquad + O(f(N)) \cdot \mathbb{P}(X_{\tau_a} \geq af(N) + \gamma_N) = O(\ln \gamma_N) + O\left(\frac{f(N)}{\gamma_N} \right) = O(\ln f(N)),
\end{aligned}$$

provided we choose $\gamma_N = O(f(N)/\ln f(N))$.

Next, the expectation of the absolute value of (4) is bounded by

$$\mathbb{E} \left\{ \sum I(\tau_a < k \leq \tau_b) \mathbb{E}_{\mathcal{F}_{k-1}} \left| \frac{X_k \wedge bf(N) - X_{k-1}}{\mu_N(X_{k-1})} - \int_{X_{k-1}}^{X_k \wedge bf(N)} \frac{dx}{\mu_N(x)} \right| \right\}. \tag{6}$$

Each term within the conditional expectation is bounded above by

$$\begin{aligned}
& (X_k \wedge bf(N) - X_{k-1}) \sup_{X_{k-1} \leq j \leq X_k \wedge bf(N)} \left| \frac{1}{\mu_N(j)} - \frac{1}{\mu_N(X_{k-1})} \right| \\
&\leq \frac{(X_k \wedge bf(N) - X_{k-1})^2}{f(N)} \sup_{X_{k-1}/f(N) \leq x \leq (X_k/f(N)) \wedge b} \frac{|\mu_N'(xf(N))|}{\mu^2(xf(N))} \\
&= O\left(\sup_{a \leq x \leq b} \left| \frac{\mu'(x)}{\mu^2(x)} \right| \right) \frac{(X_k \wedge bf(N) - X_{k-1})^2}{f(N)} \leq \frac{C}{f(N)} \mathbb{E}_{\mathcal{F}_{k-1}} (X_k - X_{k-1})^2,
\end{aligned}$$

where C is a constant depending only on μ, a, and b. Since

$$\mathbb{E}_{\mathcal{F}_{k-1}}\left(X_k - X_{k-1}\right)^2 = V_N(X_{k-1}) + \mu_N^2(X_{k-1}) \leq \sup_{af(N)\leq k\leq bf(N)}\left\{\frac{V_N(k)+\mu_N^2(k)}{\mu_N(k)}\right\}\mu_N(X_{k-1})$$
$$= O\left(\sup_{a\leq x\leq b}\frac{V(x)+\mu^2(x)}{\mu(x)}\right)\mathbb{E}_{\mathcal{F}_{k-1}}(X_k - X_{k-1}),$$

we see that (6) is bounded by

$$\frac{K}{f(N)}\mathbb{E}\sum I(\tau_a < k \leq \tau_b)\mathbb{E}_{\mathcal{F}_{k-1}}(X_k - X_{k-1}) = \frac{K}{f(N)}\mathbb{E}\sum I(\tau_a < k \leq \tau_b)(X_k - X_{k-1})$$
$$\leq \frac{K}{f(N)}\left((b-a)f(N) + \mathbb{E}(X_{\tau_b} - bf(N))\right) = O(1),$$

where K depends on μ, V, a, and b. This proves part (i).

Proof of (ii): A proof of (ii) follows the same idea; we will show that our martingale satisfies the assumptions of the Lindeberg's central limit theorem for martingales (see e.g. [3, Theorem 35.12]). Specifically, denoting $s_N^2 = \mathbb{E}\sum_{k=1}^N I(\tau_a < k \leq \tau_b)d_{N,k}^2$ we will show that

$$\frac{1}{s_N^2}\sum_{k=1}^N I(\tau_a < k \leq \tau_b)\mathbb{E}_{\mathcal{F}_{k-1}}d_{N,k}^2 \xrightarrow{P} 1, \tag{7}$$

and

$$\forall\ \epsilon > 0 \quad \frac{1}{s_N^2}\sum_{k=1}^N I(\tau_a < k \leq \tau_b)\mathbb{E}_{\mathcal{F}_{k-1}}d_{N,k}^2 I(|d_{N,k}| > \epsilon s_N) \xrightarrow{P} 0. \tag{8}$$

In order to check these conditions we first show that

$$s_N^2 = f(N)\int_a^b \frac{V(x)}{\mu^3(x)}dx + o(f(N)).$$

Define a sequence (W_k) by

$$W_k = (X_k - X_{k-1})\frac{V_N(X_{k-1})}{\mu_N^3(X_{k-1})}.$$

Then

$$\mathbb{E}_{\mathcal{F}_{k-1}}d_{N,k}^2 = \mathbb{E}_{\mathcal{F}_{k-1}}W_k = \frac{V_N(X_{k-1})}{\mu_N^2(X_{k-1})},$$

and thus, by the elementary properties of the conditional expectation,

$$s_N^2 = \mathbb{E}\sum I(\tau_a < k \leq \tau_b)d_{N,k}^2 = \mathbb{E}\sum_{k=\tau_a+1}^{\tau_b}\frac{V_N(X_{k-1})}{\mu_N^3(X_{k-1})}(X_k - X_{k-1}).$$

Once again, we recognize the expression within the last expectation as a Riemann sum and the same argument as before shows that

$$s_N^2 = \int_{af(N)}^{bf(N)}\frac{V_N(x)}{\mu_N^3(x)}dx + o(f(N)) = f(N)\int_a^b\frac{V(x)}{\mu^3(x)}dx + o(f(N)).$$

We will now prove (7) and (8). Expressing (7) in terms of W_k's means

$$\frac{\sum \mathbb{E}_{\mathcal{F}_{k-1}} W_k}{\sum \mathbb{E} W_k} \xrightarrow{P} 1. \tag{9}$$

Now write $\mathbb{E}_{\mathcal{F}_{k-1}} W_k = \mathbb{E}_{\mathcal{F}_{k-1}} W_k - W_k + W_k$ and notice that

$$\mathbb{E}_{\mathcal{F}_{k-1}} W_k - W_k = \frac{V_N(X_{k-1})}{\mu_N^2(X_{k-1})} \left(1 - \frac{X_k - X_{k-1}}{\mu_N(X_{k-1})}\right) = \frac{V_N(X_{k-1})}{\mu_N^2(X_{k-1})} d_{N,k}$$

is a martingale transform of $d_{N,k}$'s by a bounded, predictable sequence $V_N(X_{k-1})/\mu_N^2(X_{k-1})$. By Chebyshev's inequality and orthogonality of martingale differences we get

$$\mathbb{P}\left(\left|\frac{\sum(\mathbb{E}_{\mathcal{F}_{k-1}} W_k - W_k)}{\mathbb{E}\sum W_k}\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \frac{\mathbb{E}\left(\sum(\mathbb{E}_{\mathcal{F}_{k-1}} W_k - W_k)\right)^2}{(\mathbb{E}\sum W_k)^2} \leq \frac{C}{\epsilon^2} \frac{\mathbb{E}\sum d_{N,k}^2}{(\mathbb{E}\sum d_{N,k}^2)^2} = O\left(\frac{1}{f(N)}\right).$$

Thus, to complete a proof of (9) it suffices to show that $\forall\ \epsilon > 0$

$$\mathbb{P}\left(\left|\frac{\sum W_k - \mathbb{E}\sum W_k}{\mathbb{E}\sum W_k}\right| > \epsilon\right) \longrightarrow 0. \tag{10}$$

To this end, we write

$$\begin{aligned} \mathbb{E}\left|\sum W_k - \mathbb{E}\sum W_k\right| \quad &\leq \quad \mathbb{E}\left|\sum \frac{V_N(X_{k-1})}{\mu_N^3(X_{k-1})}(X_k - X_{k-1}) - \int_{af(N)}^{bf(N)} \frac{V_N(x)}{\mu_N^3(x)} dx\right| \\ &\quad + \left|\int_{af(N)}^{bf(N)} \frac{V_N(x)}{\mu_N^3(x)} dx - \mathbb{E}\sum \frac{V_N(X_{k-1})}{\mu_N^3(X_{k-1})}(X_k - X_{k-1})\right|, \end{aligned}$$

which is $o(f(N))$ by the same arguments as used in part (i). Since $\mathbb{E}\sum W_k$ is of order $f(N)$ we obtain (10) and thus also (9) and (7). We now turn to (8). Let δ satisfy (1). For $\tau_a < k \leq \tau_b$,

$$\begin{aligned} \mathbb{E}_{\mathcal{F}_{k-1}} |d_{N,k}|^{2+\delta} \quad &= \quad \mathbb{E}_{\mathcal{F}_{k-1}} \left|1 - \frac{X_k - X_{k-1}}{\mu_N(X_{k-1})}\right|^{2+\delta} \leq 2^{2+\delta}\left(1 + \mathbb{E}_{\mathcal{F}_{k-1}} \frac{(X_k - X_{k-1})^{2+\delta}}{\mu_N^{2+\delta}(X_{k-1})}\right) \\ &\leq \quad 2^{2+\delta}\left(1 + \sup_{a \leq x \leq b} \frac{\nu_\delta(x)}{\mu^{2+\delta}(x)}\right) = O(1). \end{aligned}$$

Hence, by the usual argument, each of the terms in the sum (8) is bounded by K/s_N^δ, and since there are no more than $(b-a)f(N)$ terms in that sum we obtain

$$\frac{1}{s_N^2} \sum I(\tau_a < k \leq \tau_b) \mathbb{E}_{\mathcal{F}_{k-1}} d_{N,k}^2 I(|d_{N,k}| \geq \epsilon s_N) \leq \frac{K(b-a)f(N)}{f^{1+\delta}(N)} \longrightarrow 0,$$

which implies (ii). □

3 Random functions on a finite set

The statement of Theorem 1 can be extended in various directions. Part of the argument relies on a compactness of $[a, b]$ so, perhaps the most immediate question is to examine what happens if a and b are allowed to depend on N, in particular let $b \to \infty$. In principle this presents no difficulty, but one would have to take into account a relationship between the growth of $f(N)$ and $\mu_N(xf(N))$. There are many ways of doing this, and rather than attempt to give a general condition, we work out how to do this for the specific example at hand, namely compositions of random functions.

Let $f_1, f_2, f_3, \ldots$ be a sequence of functions chosen independently and uniformly randomly from the N^N functions on $\{1, \ldots, N\}$. Let $g_1 = f_1$, and for $k > 1$ let $g_k = f_k \circ g_{k-1}$ be the composition of the first k random functions. Define T to be the smallest m for which g_m is a constant function. (i.e. $g_m(i) = g_m(j)$ for all $i \neq j$.) The natural set–up is to consider a Markov chain $\{Y_k : \ k = 0, 1 \ldots\}$ where Y_k is the cardinality of the range of g_k (thus the initial state is N and the state 1 is absorbing; the transition probabilities for other states can be easily computed [10, Appendix II], [5, 18] (the last paper also contains references to earlier work). The expected value and the asymptotic distribution of T have been studied in various contexts ([12, 15, 14, 5, 8, 9]). Here we will be interested in the number of states visited before the absorption. We prove

Theorem 2 *We have:*

(i)

$$\mathbb{E}R_N \sim \sqrt{2\pi N}.$$

(ii)

$$\frac{R_N - \sqrt{2\pi N}}{\sigma N^{1/4}} \Longrightarrow N(0,1), \quad \textit{where} \quad \sigma^2 = \frac{2-\sqrt{2}}{3}\sqrt{\pi}.$$

Proof: We will consider a jump process (X_k) associated with (Y_k) (see [2]); that is we define a sequence of stopping times by

$$J_0 = 0, \quad J_{n+1} = \inf\{k > J_n : \ Y_k \neq Y_{J_n}\}$$

and then a sequence of holding times $T_k = J_k - J_{k-1}$, $k = 1, 2, \ldots$. We then set $X_k = Y_{J_k}$ for $k = 0, 1, \ldots$. In order to determine scaling $f(N)$ and the functions $\mu(x)$ and $V(x)$ for the chain X it will be convenient to use the following description of transition rules for Y: the state 1 is absorbing and for any other state k, the law of Y_{m+1} given that $Y_m = k$ is the law of the number of occupied urns when k balls are dropped uniformly and independently into N urns. We let $E_{k,p}$ and var_k denote $\mathbb{E}((k - Y_1)^p | Y_0 = k)$ and $\mathrm{var}(k - Y_1 | Y_0 = k))$, respectively. Writing, for simplicity $E_k = E_{k,1}$, we have

Lemma 3 *(i)* $\frac{k^2}{2N}\left(1 - \frac{k}{2N}(1 + \frac{3N}{k^2})\right) \leq E_k \leq \frac{k^2}{2N}$,

(ii) $\frac{k^2}{N} - O(k^3/N^2) - O(k^5/N^3) \leq \mathrm{var}_k \leq \frac{k^2}{N} + O(k^3/N^2)$,

(iii) for $p \geq 1$, $\exists\, C_p$ such that $E_{k,p} \leq C_p \left(\frac{k^2}{2N}\right)^p$,

Proof of Lemma 3: Let I_j, $j = 2, \ldots, k$ be the event that the jth ball falls in an occupied urn. For the simplicity of notation we will not distinguish between events and their indicators. Since at most $j-1$ urns are occupied when the jth ball is dropped, we clearly have $\mathbb{P}(I_j) \leq (j-1)/N$. On the other hand, if $I_{j,\ell}$ is the event that the jth ball falls into an urn containing the ℓth ball, then

$$\begin{aligned}
\mathbb{P}(I_j) &= \mathbb{P}(\bigcup_{\ell=1}^{j-1} I_{j,\ell}) \geq \sum_{\ell=1}^{j-1} \mathbb{P}(I_{j,\ell}) - \sum_{1\leq m<\ell<j} \mathbb{P}(I_{j,\ell} \cap I_{j,m}) \\
&= \sum_{\ell=1}^{j-1} \frac{1}{N} - \sum_{1\leq m<\ell<j} \mathbb{P}(I_{j,\ell} \cap I_{j,m} | I_{\ell,m}) \cdot \mathbb{P}(I_{\ell,m}) \geq \frac{j-1}{N} - \frac{1}{N} \sum_{1\leq m<\ell<j} \mathbb{P}(I_{\ell,m}) \\
&\geq \frac{j-1}{N} - \frac{1}{N^2}\binom{j-1}{2}.
\end{aligned}$$

Since $k - Y_1$ is the number of times a balls falls into an already occupied urn we have

$$E_k = \mathbb{E}\sum_{j=2}^{k} I_j = \sum_{j=2}^{k} \mathbb{P}(I_j).$$

Hence (i) follows by a simple summation. For (ii) we write

$$\mathbb{E}(\sum_{j=2}^{k} I_j)^2 = \sum_{j=2}^{k} \mathbb{P}(I_j) + 2\sum_{2\leq i<j\leq k} \mathbb{P}(I_i \cap I_j) = \sum_{j=2}^{k} \mathbb{P}(I_j) + 2\sum_{2\leq i<j\leq k} \mathbb{P}(I_j|I_i)\cdot\mathbb{P}(I_i).$$

But conditioning on I_i amounts to removing the ith ball from the consideration. Thus,

$$\frac{j-2}{N} - \frac{1}{N^2}\binom{j-2}{2} \leq \mathbb{P}(I_j|I_i) \leq \frac{j-2}{N}.$$

This gives

$$\sum_{2\leq i<j\leq k} \frac{j-2}{N}\left(1-\frac{j-3}{2N}\right)\cdot\frac{i-1}{N}\left(1-\frac{i-2}{2N}\right) \leq \sum_{2\leq i<j\leq k} \mathbb{P}(I_i \cap I_j) \leq \sum_{2\leq i<j\leq k} \frac{j-2}{N}\cdot\frac{i-1}{N},$$

and, upon summation, implies (ii). Finally, to prove (iii) note that for $2 \leq i_1 < i_2 < \ldots < i_r \leq k$ we have

$$\begin{aligned}
&\mathbb{P}(I_{i_1} \cap \ldots \cap I_{i_r}) = \mathbb{P}(I_{i_r}|I_{i_{r-1}} \cap \ldots \cap I_{i_1}) \cdot \ldots \cdot \mathbb{P}(I_{i_2}|I_{i_1}) \cdot \mathbb{P}(I_{i_1}) \\
&\leq \frac{i_r - 1 - (r-1)}{N} \cdot \ldots \cdot \frac{i_2 - 2}{N} \cdot \frac{i_1 - 1}{N} \leq \prod_{\ell=1}^{r} \frac{i_\ell - 1}{N},
\end{aligned}$$

so that

$$\mathbb{P}(\sum_{j=2}^{k} I_j \geq r) = \mathbb{P}(\exists\, 1 \leq i_1 < i_2, \ldots i_r \leq k : I_{i_1} \cap \ldots \cap I_{i_r}) \leq \sum_{2\leq i_1<\ldots<i_r\leq k} \prod_{\ell=1}^{r} \frac{i_\ell - 1}{N}$$

$$\le \frac{1}{r!} \sum_{\substack{2\le i_1,i_2,\dots,i_r\le k \\ \text{all}}} \prod_{\ell=1}^{r} \frac{i_\ell - 1}{N} \le \frac{1}{r!}\left(\sum_{j=2}^{k}\frac{j-1}{N}\right)^r \le \left(\frac{e}{r}\cdot\frac{k^2}{2N}\right)^r$$

Hence, using $\mathbb{E}Y^p \le \sum_{r\ge 1} pr^{p-1}\mathbb{P}(Y\ge r)$ we get

$$\begin{aligned} E_{k,p} &\le \left(\frac{e^2k^2}{2N}\right)^p + \sum_{r\ge\frac{e^2k^2}{2N}} pr^{p-1}\mathbb{P}(\sum_{j=2}^{k} I_j \ge r) \le \left(\frac{e^2k^2}{2N}\right)^p + \sum_{r\ge\frac{e^2k^2}{2N}} pr^{p-1}\left(\frac{e}{r}\cdot\frac{k^2}{2N}\right)^r \\ &\le \left(\frac{e^2k^2}{2N}\right)^p + \sum_{r\ge\frac{e^2k^2}{2N}} pr^{p-1}e^{-r} \le C_p\left(\frac{k^2}{2N}\right)^p, \end{aligned}$$

which proves (iii). □

We now return to the proof of Theorem 2. Part (i) of Lemma 3 gives us an expected size of a jump from a state k but with a possibility of remaining in k. To find the expected size of a jump without that possibility we need to condition on the fact that we do leave the state k. That is, if E_k^*, var_k^* and $E_{k,p}^*$ denote the same quantities for the process (X_j) as those without stars for (Y_j), then

$$E_{k,p}^* = \frac{E_{k,p}}{\mathbb{P}_k(Y_1\ne k)}, \quad \text{for} \quad p>0 \tag{11}$$

where $\mathbb{P}_k$ is the conditional probability, given $Y_0 = k$. In particular,

$$\mathrm{var}_k^* = E_{k,2}^* - (E_k^*)^2 = \frac{E_{k,2}}{\mathbb{P}_k(Y_1\ne k)} - \left(\frac{E_k}{\mathbb{P}(Y_1\ne k)}\right)^2. \tag{12}$$

Since

$$\mathbb{P}_k(Y_1 = k) = \prod_{j=1}^{k-1}\frac{N-j}{N} \le \exp\left(\sum_{j=1}^{k-1}\log(1-\frac{j}{N})\right) \le \exp\left(-\sum_{j=1}^{k-1}\frac{j}{N}\right) \le \exp\left(-\frac{k^2}{2N}\right),$$

we obtain from (11)

$$E_k^* \le \frac{E_k}{1-\exp(-k^2/2N)}.$$

On the other hand,

$$\begin{aligned} E_k^* &= \frac{E_k}{1-\exp(-k^2/2N)}\cdot\frac{1-\exp(-k^2/2N)}{\mathbb{P}_k(Y_1\ne k)} \\ &= \frac{E_k}{1-\exp(-k^2/2N)}\left(1-\frac{\exp(-k^2/2N)-\mathbb{P}_k(Y_1=k)}{\mathbb{P}_k(Y_1\ne k)}\right). \end{aligned}$$

For $k\le\sqrt{N\log N}$,

$$\begin{aligned} \left|\exp(-k^2/2N)-\mathbb{P}_k(Y_1=k)\right| &\le \prod_{j=1}^{k-1}\exp(-\frac{j}{N}) - \prod_{j=1}^{k-1}\left(1-\frac{j}{N}\right) \le \sum_{j=1}^{k-1}\left(\exp(-\frac{j}{N})-(1-\frac{j}{N})\right) \\ &\le \sum_{j=1}^{k-1}\frac{j^2}{N^2} = O(\frac{k^3}{N^2}) = o(\mathbb{P}_k(Y_1\ne k)), \end{aligned}$$

and for $k \geq \sqrt{N \log N}$ each of the two terms on the left hand side is $O(1/\sqrt{N})$. Thus we have

$$\frac{E_k}{1-\exp(-k^2/2N)}(1-o(1)) \leq E_k^* \leq \frac{E_k}{1-\exp(-k^2/2N)},$$

which means that we may take

$$f(N) = \sqrt{N} \quad \text{and} \quad \mu(x) = \frac{x^2/2}{1-\exp(-x^2/2)}.$$

Similarly, using the above, (12) and Lemma 3(ii) we see that $V(x)$ may be taken to be

$$V(x) = \frac{(x^2/2)^2 + x^2/2}{1-\exp(-x^2/2)} - \left(\frac{x^2/2}{1-\exp(-x^2/2)}\right)^2 = \frac{x^2}{2} \cdot \frac{1-\exp(-x^2/2)-(x^2/2)\exp(-x^2/2)}{(1-\exp(-x^2/2))^2}.$$

Finally, part (iii) gives a control of a growth of higher moments, namely we can take

$$\nu_p(x) = c_p \frac{x^{2p}/2}{1-\exp(-x^2/2)}.$$

Since the process starts at N and moves down until it reaches 1, the range, after rescaling, is between 0 and $\sqrt{N}$ and thus, formally at least, does not satisfy the assumptions of Theorem 1 (of course, the fact that the process decreases rather than increases is inessential). Yet, if we were able to show that for these particular functions $\mu(x)$ and $V(x)$ the expected error between the Riemann sum and the integral $\int_0^N \frac{dx}{\mu(x/\sqrt{N})}$ is $o(\sqrt{N})$ then we would have known that the number of states visited has expected value

$$\int_0^N \frac{1-\exp(-x^2/(2N))}{x^2/(2N)} dx \sim \sqrt{N} \int_0^\infty \frac{1-\exp(-x^2/2)}{x^2/2} dx = \sqrt{2\pi N},$$

the variance asymptotic to

$$\begin{aligned} \sqrt{N} \int_0^\infty \frac{V(x)}{\mu^3(x)} dx &= \sqrt{N} \int_0^\infty \frac{(1-\exp(-x^2/2)-(x^2/2)\exp(-x^2/2))(1-\exp(-x^2/2))}{(x^2/2)^2} dx \\ &= \frac{2-\sqrt{2}}{3}\sqrt{\pi N}, \end{aligned}$$

and that it satisfies the CLT. But this is not difficult; with $\gamma_N \to \infty$, $\gamma_N = o(\sqrt{N})$ let $\tau = \inf\{k \geq 0 : \ X_k \leq \gamma_N \sqrt{N}\}$ (recall that X_k decreases from N to 1) and consider the sum

$$\left| \sum_j \left\{ \int_{X_{j+1}}^{X_j} \frac{dx}{\mu(x/\sqrt{N})} - \frac{X_j - X_{j+1}}{\mu(X_j/\sqrt{N})} \right\} \right|.$$

We will treat the cases $j < \tau - 1$, $j \geq \tau$, and $j = \tau - 1$ separately. If $j < \tau - 1$, then

$$\sum_{j=0}^{\tau-2} \int_{X_{j+1}}^{X_j} \frac{dx}{\mu(x/\sqrt{N})} \leq \int_{X_{\tau-1}}^\infty \frac{dx}{\mu(x/\sqrt{N})} \leq \sqrt{N} \int_{\gamma_N}^\infty \frac{dx}{\mu(x)} \leq C\frac{\sqrt{N}}{\gamma_N},$$

and since $1/\mu(x)$ is decreasing in that range, the Riemann sum underestimates the integral, so that

$$\mathbb{E}\left|\sum_{j<\tau-1}\left\{\int_{X_{j+1}}^{X_j}\frac{dx}{\mu(x/\sqrt{N})}-\frac{X_j-X_{j+1}}{\mu(X_j/\sqrt{N})}\right\}\right| = O(\sqrt{N}/\gamma_N) = o(\sqrt{N}).$$

For $j \geq \tau$ the jth term is bounded by

$$(X_j - X_{j+1})\sup_{X_{j+1}\leq x\leq X_j}\left|\frac{1}{\mu(x/\sqrt{N})}-\frac{1}{\mu(X_j/\sqrt{N})}\right| \leq \frac{(X_j-X_{j+1})^2}{\sqrt{N}}\sup_{X_{j+1}/\sqrt{N}\leq x\leq X_j/\sqrt{N}}\left|\frac{\mu'(x)}{\mu^2(x)}\right|.$$

Let $b_N \to \infty$, $b_N = o(\sqrt{N})$; on the set $\{X_j - X_{j+1} \leq b_N\mu(X_j/\sqrt{N})\}$, the right-hand side is bounded by

$$\frac{X_j-X_{j+1}}{\sqrt{N}}b_N \sup\left\{\left|\frac{\mu'(x)}{\mu(x)}\right| : \frac{X_j - b_N\mu(X_j/\sqrt{N})}{\sqrt{N}} \leq x \leq \frac{X_j}{\sqrt{N}}\right\}$$

If $X_j/\sqrt{N}$ is bounded away from zero the supremum is bounded by twice $|\mu'(X_j/\sqrt{N})/\mu(X_j/\sqrt{N})|$, otherwise is bounded by 2. Hence,

$$\begin{aligned}
&\sum_j I(\tau\leq j)I(X_j - X_{j+1}\leq b_N\mu(X_j/\sqrt{N}))\frac{(X_j-X_{j+1})^2}{\sqrt{N}}\sup_{X_{j+1}/\sqrt{N}\leq x\leq X_j/\sqrt{N}}\left|\frac{\mu'(x)}{\mu^2(x)}\right|\\
&\quad\leq 2\sum_j I(\tau\leq j)\frac{X_j-X_{j+1}}{\sqrt{N}}\left(1\wedge\left|\frac{\mu'(X_j/\sqrt{N})}{\mu(X_j/\sqrt{N})}\right|\right) \leq 2\frac{\gamma_N\sqrt{N}}{\sqrt{N}} + 2\int_1^{\gamma_N}\frac{|\mu'(x)|}{\mu(x)}dx = o(\sqrt{N}).
\end{aligned}$$

On the complementary set $\{X_j - X_{j+1} > b_N\mu(X_j/\sqrt{N})\}$, since $|\mu'(x)/\mu(x)|$ is bounded we get, for $p \geq 1$

$$\begin{aligned}
&\mathbb{E}_{\mathcal{F}_j}\left(I(X_j - X_{j+1} > b_N\mu(X_j/\sqrt{N}))\frac{(X_j-X_{j+1})^2}{\sqrt{N}}\sup_x\left|\frac{\mu'(x)}{\mu(x)}\right|\right)\\
&\quad\leq \frac{K}{\sqrt{N}b_N^p\mu^p(X_j/\sqrt{N})}\cdot\mathbb{E}_{\mathcal{F}_j}(X_j - X_{j+1})^{p+2}\\
&\quad\leq \frac{K(1-\exp(-X_j^2/2N))^p}{\sqrt{N}b_N^p(X_j^2/(2N))^p}\cdot\frac{(X_j^2/(2N))^{p+2}}{1-\exp(-X_j^2/(2N))} \leq \frac{KX_j^4}{\sqrt{N}b_N^p(2N)^2}.
\end{aligned}$$

Since there are no more than $\gamma_N\sqrt{N}$ terms and for $j\geq\tau$, $X_j \leq \gamma_N\sqrt{N}$ we obtain that the entire sum is bounded by

$$\frac{(\gamma_N\sqrt{N})^5}{\sqrt{N}b_N^p(2N)^2} = O(\gamma_N^5/b_N^p),$$

which can be made $o(\sqrt{N})$ by an appropriate choice of γ_N and b_N. The argument for the function $V(x)/\mu^3(x)$ is essentially the same. It remains to show that both

$$\mathbb{E}\left|\int_{X_\tau}^{X_{\tau-1}}\frac{dx}{\mu(x)}-\frac{X_{\tau-1}-X_\tau}{\mu(X_\tau/\sqrt{N})}\right| \quad\text{and}\quad \mathbb{E}\left|\int_{X_\tau}^{X_{\tau-1}}\frac{V(x/\sqrt{N})}{\mu^3(x/\sqrt{N})}dx - (X_{\tau-1}-X_\tau)\frac{V(X_{\tau-1}/\sqrt{N})}{\mu^3(X_{\tau-1}/\sqrt{N})}\right|,$$

are $o(\sqrt{N})$. We have

$$\left|\int_{X_\tau}^{X_{\tau-1}} \frac{dx}{\mu(x/\sqrt{N})} - \frac{X_{\tau-1}-X_\tau}{\mu(X_\tau/\sqrt{N})}\right| \le (X_{\tau-1}-X_\tau) \sup_{X_\tau \le x \le X_{\tau-1}} \left|\frac{1}{\mu(x/\sqrt{N})} - \frac{1}{\mu(X_{\tau-1}/\sqrt{N})}\right|$$
$$\le (X_{\tau-1}-X_\tau) \sup_{X_\tau \le x \le X_{\tau-1}} \left|\frac{\mu(X_{\tau-1}/\sqrt{N}) - \mu(x/\sqrt{N})}{\mu(x/\sqrt{N})\cdot\mu(X_{\tau-1}/\sqrt{N})}\right|$$
$$\le C\frac{(X_{\tau-1}-X_\tau)^2}{\sqrt{N}\mu(X_\tau/\sqrt{N})}\cdot\frac{\mu'(X_{\tau-1}/\sqrt{N})}{\mu(X_{\tau-1}/\sqrt{N})} \le C\frac{(X_{\tau-1}-X_\tau)^2}{X_{\tau-1}},$$

where we have used the fact that $\mu'(x)/\mu(x)$ behaves like $2/x$ for large x. Taking expectations yields

$$\mathbb{E}\frac{(X_{\tau-1}-X_\tau)^2}{X_{\tau-1}} = \sum_j \mathbb{E}\frac{(X_{\tau-1}-X_\tau)^2}{X_{\tau-1}} I(\tau=j).$$

Since

$$\{\tau=j\} = \bigcup_{m>\gamma_N\sqrt{N}} \{\tau=j\}\cap B_{j,m},$$

where $B_{j,m} = \{X_0 > m, \ldots, X_{j-2} > m, X_{j-1} = m\} \in \mathcal{F}_{j-1}$, denoting by $\mathbb{E}_m$ and $\mathbb{P}_m$ the conditional expectation, given that $X_0 = m$, our expectation is further equal to

$$\sum_j \sum_{m>\gamma_N\sqrt{N}} \mathbb{E}\frac{(m-X_j)^2}{m} I_{B_{j,m}} I(X_j \le \gamma_N\sqrt{N})$$
$$= \sum_j \sum_{m>\gamma_N\sqrt{N}} \frac{1}{m}\mathbb{E}I_{B_{j,m}}\mathbb{E}_{\mathcal{F}_{j-1}}(m-X_j)^2 I(X_j \le \gamma_N\sqrt{N})$$
$$= \sum_j \sum_{m>\gamma_N\sqrt{N}} \mathbb{E}\frac{I_{B_{j,m}}}{m}\mathbb{E}_m(m-X_1)^2 I(X_1 \le \gamma_N\sqrt{N})$$
$$\le \sum_{m>\gamma_N\sqrt{N}} \mathbb{E}\frac{I(m\in R)}{m}(\mathbb{E}_m(m-X_1)^4)^{1/2}\mathbb{P}_m^{1/2}(X_1 \le \gamma_N\sqrt{N})$$
$$\le \sum_{m>\gamma_N\sqrt{N}} \frac{1}{m}\left(\frac{m^2}{2N}\right)^2 \mathbb{P}_m^{1/2}(m-X_1 \ge m-\gamma_N\sqrt{N}).$$

We split the sum into two pieces, according to whether $m \le 3\gamma_N\sqrt{N}$ or not. In the first case, bounding the sum by the number of terms times the largest one, we see that this part is no more than

$$\frac{2\gamma_N\sqrt{N}}{(2N)^2}(3\gamma_N\sqrt{N})^3 = O(\gamma_N^4) = o(\sqrt{N}),$$

with the appropriate choice of γ_N. In order to bound the second sum, note that $m \ge 3\gamma_N\sqrt{N}$ implies that $m - \gamma_N\sqrt{N} \ge 2m$. Hence, by the computations used in the proof of Lemma (iii) we get

$$\mathbb{P}_m^{1/2}(m-X_1 \ge m-\gamma_N\sqrt{N}) \le \mathbb{P}_m^{1/2}(m-X_1 \ge 2m) \le \left(\frac{e}{2m}\cdot\frac{m^2}{2N}\right)^{2m/2} = \left(\frac{em}{4N}\right)^m \le \left(\frac{e}{4}\right)^m,$$

for $m \le N$. Therefore, the entire sum is bounded above by

$$\frac{1}{4N^2}\sum_{m>3\gamma_N\sqrt{N}} m^3\left(\frac{e}{4}\right)^m = o(\sqrt{N}).$$

The argument for the function $V(x)/\mu^3(x)$is virtually the same and is omitted. □

4 Further Remarks

The martingale array defined by (2) can be modified to study other characteristics of the process (Y_j). For example, there has been quite a bit of work concerning the total time T until the iteration becomes a constant function. This is just the sum, over all visited states, of holding times T_j,

$$T = \sum_{j\in\mathcal{T}} T_j.$$

Thus, the martingale array for for this problem is defined by

$$d_{N,k} = T_{X_{k-1}} - \frac{X_{k-1}-X_k}{\mu_N(X_{k-1})}\mathbb{E}_{\mathcal{F}_{k-1}}T_{X_{k-1}}, \quad \mathcal{F}_k = \sigma\{X_0,\ldots,X_k,T_{X_0},\ldots,T_{X_{k-1}}\}.$$

Given that the state j is visited, T_j is geometric distribution with parameter $1-P_{j,j}$, where $P_{i,j}$ are transition probabilities for the chain (X_k). Hence $\mathbb{E}_{\mathcal{F}_j}T_j = 1/\mathbb{P}_j(X_1 \ne j)$ which is exactly a correction term between E_j and E_j^* (see (11)). Thus, reasoning as before we get

$$\mathbb{E}T = \mathbb{E}\sum_k (X_{k-1}-X_k)\frac{\mathbb{E}_{\mathcal{F}_{k-1}}T_{X_{k-1}}}{\mu_N(X_{k-1})} \sim \mathbb{E}\sum_k \frac{X_{k-1}-X_k}{X_{k-1}^2/(2N)} \sim 2N\int_1^\infty \frac{dx}{x^2} = 2N,$$

a long known result. The CLT does not hold since $d_{N,k}$'s fail to satisfy the neglibility condition (8). In fact (see e.g. [12, 6, 17, 8, 9]), $T/\mathbb{E}T$ converges in distribution to a random variable whose density is

$$f(x) = \sum_{k\ge 2}(-1)^k\binom{k}{2}(2k-1)e^{-\binom{k}{2}x}, \quad x>0.$$

Acknowledgment Part of the research of the first author was carried out while he was visiting the University of Jyväskylä, Finland in the Fall of 2002. He would like to thank the Department of Mathematics and Statistics for the invitation, and Christel and Stefan Geiss for their hospitality. Also, he would like to thank his colleagues Bill Goh and Eric Schmutz for many helpful conversations on this topic, especially during their work on [9].

References

[1] D. Aldous. Deterministic and stochastic models for coalescence (aggregation, coagulation): a review of mean field theory for probabilists. *Bernoulli*, 5:3–48, 1999.

[2] S. Asmussen. *Applied Probability and Queues.* Springer, 2nd edition, 2003.

[3] P. Billingsley. *Probability and Measure.* Wiley, 3rd edition, 1995.

[4] C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach, i. haploid models. *Adv. Appl. Probab.*, 6:260–290, 1974.

[5] A. Dalal and E. Schmutz. Compositions of random functions on a finite set. *Electron. J. Combin.*, 9:R26, 2002.

[6] P. Donnelly. Weak convergence to a Markov chain with an entrance boundary: ancestral processes in population genetics. *Ann. Probab.*, 19:1102–1117, 1991.

[7] W. Ewens. *Mathematical Population Genetics*, volume 27 of *Interdisciplinary Mathematical Series.* Springer, 2nd edition, 2004.

[8] J. Fill. On compositions of random functions on a finite set. unpublished manuscript.

[9] W. M. Y. Goh P. Hitczenko and E. Schmutz. Iterating random functions on a finite set, 2002. manuscript available at `http://www.arxiv.org`, paper math.CO/0207276.

[10] J. F. C. Kingman. *Mathematics of genetic diversity*, volume 34 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics, 3rd edition, 1980.

[11] J. F. C. Kingman. The coalescent. *Stochastic Proc. Appl.*, 13:235–248, 1982.

[12] J. F. C. Kingman. Exchangeability and the evolution of large populations. In *Exchangeability in probability and statistics*, pages 97–112. North-Holland, 1982.

[13] J. F. C. Kingman. On the genealogy of large populations. Essays in statistical science. *J. Appl. Probab.*, 19A:27–43, 1982.

[14] M. Möhle. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.*, 32:983–993, 2000.

[15] M. Möhle. The time back to the most recent common ancestor in exchangeable population models. *Adv. Appl. Probab.*, 36:78–97, 2004.

[16] J. Propp and D. Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Struct. Algorithms*, 9:223–252, 1996.

[17] S. Tavare. Line-of-descent and geneological processes and their applications in population genetics models. *Theoretical Population Biology*, 26:119–164, 1984.

[18] G. A. Watterson. On the number of segregating sites in genetic models without recombination. *Theoretical Population Biology*, 7:256–276, 1975.