The Annals of Applied Probability
2000, Vol. 10, No. 4, 1151–1181

NONUNIFORM RANDOM TRANSFORMATIONS

BY C. A. O'CINNEIDE AND A. V. POKROVSKII

Purdue University and University College, Cork

With a given transformation on a finite domain, we associate a three-dimensional distribution function describing the component size, cycle length and trajectory length of each point in the domain. We then consider a random transformation on the domain, in which images of points are independent and identically distributed. The three-dimensional distribution function associated with this random transformation is itself random. We show that, under a simple homogeneity condition on the distribution of images, and with a suitable scaling, this random distribution function has a limit law as the number of points in the domain tends to ∞. The proof is based on a Poisson approximation technique for matches in an urn model. The result helps to explain the behavior of computer implementations of chaotic dynamical systems.

1. Motivation. In recent years much attention has been given to computer implementations of dynamical systems with chaotic behavior (see [4], [10] and [22]). In contrast to a continuous system, a computer implementation is a transformation on a finite, albeit rather large, set of digital numbers. Such a transformation may be viewed as a digraph on a set $\{1, 2, \ldots, n\}$, with the property that each node is the initial node of exactly one edge. An important combinatorial descriptor of such a transformation is the joint distribution of the cycle lengths, recurrence times and component sizes (see the next section for definitions) of the elements of $\{1, 2, \ldots, n\}$. An exact analysis of quantities such as cycle lengths in discretizations of chaotic dynamical systems is not feasible, as they are typically highly sensitive to small variations in both the dynamical system itself and the discretization scheme. (However, see [16] for a rigorous analysis of a discretization of an interesting planar dynamical system.) Thus, it is natural to consider the statistical characteristics of an ensemble of similar discretizations of a given dynamical system. This approach has been taken in many papers; see, for example, [6], [12], [18] and [21] and the references therein. This literature has given rise to the empirical principle: *ensembles of discretizations of chaotic dynamical systems display regularities which are to some extent insensitive to the details of both the dynamical system and the discretization scheme*. We give an example to illustrate this principle, after introducing some notation.

Let ϕ be a transformation on $\{1, 2, \ldots, n\}$. Then each element $i_0 \in \{1, 2, \ldots, n\}$ generates a trajectory i_0, $i_1 = \phi(i_0)$, $i_2 = \phi(i_1), \ldots$, which is eventually periodic. We denote by $\alpha(\phi)$ the probability that the trajectories

Received December 1998; revised November 1999.
AMS 1991 *subject classifications*. Primary 60F05; secondary 60C05.
Key words and phrases. Random mappings, chaotic dynamical systems, urn models, Poisson approximations.

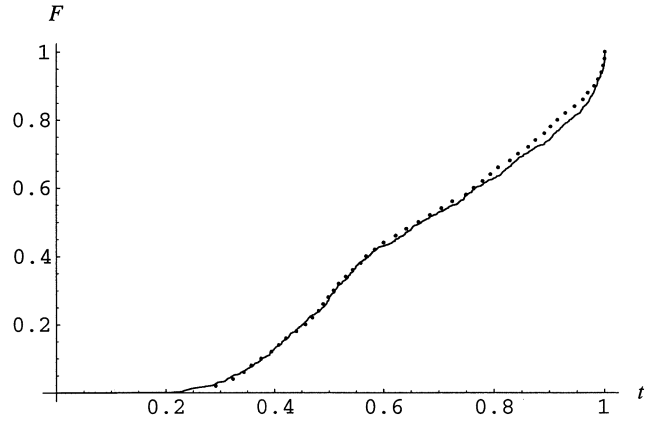

FIG. 1. *The empirical distribution functions* $F_\alpha(\cdot\,|H_1)$ (*dotted line*) *and* $F_\alpha(\cdot\,|H_2)$ (*solid line*).

generated by two independently and uniformly chosen elements intersect. If we have an ensemble $H = \{\phi_1, \phi_2, \ldots, \phi_N\}$ of transformations on $\{1, 2, \ldots, n\}$, we can define the empirical distribution function of $\alpha(\cdot)$ over H by

$$F_\alpha(t|H) = \frac{\#\{\phi \in H: \alpha(\phi) \le t\}}{N}, \qquad t \in \mathbb{R},$$

where # denotes cardinality.

For a positive integer n, the n-discretization of a transformation $f: [0, 1] \mapsto [0, 1]$ is the transformation f_n on $\{1, 2, \ldots, n\}$ defined by $f_n(i) = \lceil nf(i/n) \rceil_+$, where $\lceil \cdot \rceil_+$ is the smallest positive integer not less than the argument. We consider two ensembles of discretizations constructed in different ways from two classical chaotic transformations on the unit interval: the β-mapping $f(x) = \sqrt{2}x \,(\mathrm{mod}\ 1)$ and the logistic mapping $g(x) = 4x(1 - x)$. These are among the best understood examples of chaotic mappings [15].

For the β-mapping f, denote by H_1 the ensemble of discretizations f_n, $n = 10^6 + 1, 10^6 + 2, \ldots, 10^6 + 10^3$. For the logistic mapping g, we discretize in a different way. First define $g(\cdot\,; a) = ag(\cdot)$ for a real. Denote by H_2 the ensemble of discretizations $g_n(\cdot\,; a_k)$ with $n = 10^6$ fixed and with $a_k = 1 - 10^{-6}k$, $k = 1, 2, \ldots, 10^3$. Figure 1 presents the graphs of the empirical distribution functions $F_\alpha(\cdot\,|H_1)$ and $F_\alpha(\cdot\,|H_2)$ of $\alpha(\cdot)$ over these ensembles. The curves are in good agreement. The choice of the ensembles of discretizations H_1 and H_2 of f and g here is quite arbitrary, and the point being made is that the resulting empirical distributions of $\alpha(\cdot)$ are little influenced by the choice. However, we clearly must require that the ensembles be fairly large (each containing 1000 discretizations here), that the discretizations be fairly fine ($n \approx 10^6$ here) and that the discretizations within an ensemble be sufficiently distinct (confirmed here by Figure 1).

Much experimentation of this kind has been done with various mappings (e.g., the Henon mapping), with other graph-theoretical characteristics and with other discretization schemes (especially floating-point implementations [13]), and it supports the italicized observation of the first paragraph of this section.

It is important to understand and to theoretically predict such regularities in the behavior of discretizations of dynamical systems. The obvious approach is a direct investigation of the combinatorics of these computer implementations, but this is not practicable. Thus we consider a phenomenological model of ensembles of discretizations. For such a model to be useful, it should admit rigorous mathematical investigation and produce predictions that are consistent with the results of computer experiments. A direct analog of this approach is to be found in thermodynamics, where, instead of explaining the behavior of every particle of a gas, we use the far less detailed microscopic models of statistical physics, which, despite their simplicity, still account for macroscopic characteristics such as entropy and temperature.

There is a general consensus that random transformations [7] in which images of points are i.i.d. (independent and identically distributed) are the most promising phenomenological models of discretizations of dynamical systems. (Further references on random transformations include [5], [19] and [26].) This consensus is supported, on the one hand, by convincing heuristic arguments (see [6] and [18]) and, on the other hand, by excellent quantitative agreement between the theoretical predictions of these models and the results of extensive computer experiments, especially in situations where the underlying dynamical systems have a Sinai–Ruelle–Bowen invariant measure [23]. We illustrate this with an example. For a completely random mapping Φ on $\{1, 2, \ldots, n\}$ (see the next section for the definition), the distribution function of $\alpha(\Phi)$ has a limit F_α^∞ as $n \to \infty$ which can be calculated theoretically. This is a consequence of Theorem 2.1; it may also be proved by the methods of [26]. Heuristics suggest that the empirical distribution functions $F_\alpha(\cdot\,|H_1)$ and $F_\alpha(\cdot\,|H_2)$ should approximate this limit. Figure 2 presents graphs of these empirical distribution functions along with the prediction F_α^∞. Again, there is good agreement.

The construction of these phenomenological models falls naturally into two stages. At the first stage, one considers random transformations which are nonuniform: each point is assigned a weight reflecting the nature of the original dynamical system—see Section 6.1 for an explanation of the natural choice of weights. However, only broad features of these weights can be convincingly attributed to the underlying continuous transformation whose discretizations we are considering. We hope to identify these features through a "thermodynamic limit" in which the number of points of the discretization increases to ∞. Thus we need a second stage, where the nonuniform random transformations are replaced by suitable limits in which much of the detail of the distribution of images disappears. For example, the limiting behavior of a random transformation may be indistinguishable from that of a completely random transformation [21] or a transformation with an attracting center [14] and

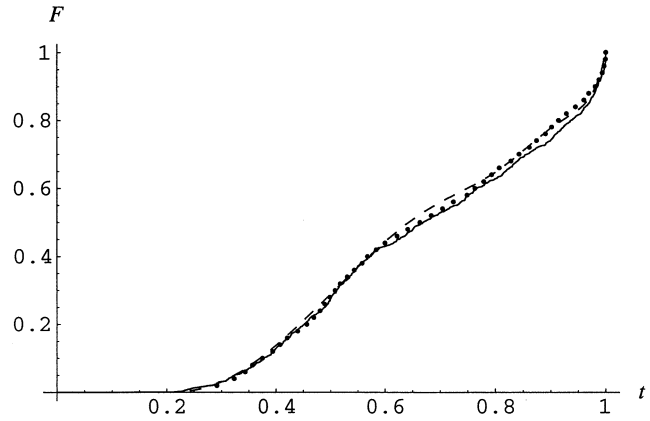

FIG. 2. *The distribution functions* $F_\alpha(\cdot\,|H_1)$ (*dotted line*), $F_\alpha(\cdot\,|H_2)$ (*solid line*) *and* F_α^∞ (*dashed line*).

[27]; see Section 6.2). While the first stage is not amenable to analysis, the second stage is, and this is the focus of the present paper. We derive a simple condition on the distribution of images [see (2.5)], under which the asymptotic behavior of the joint distribution of component size, cycle length and trajectory length of a random mapping is indistinguishable from that of a completely random mapping, apart from a scale factor.

2. Overview.

2.1. *Random transformations.* This paper deals with random transformations on a finite set $\{1, 2, \ldots, n\}$. The elements of this set will be referred to as labels throughout, to distinguish them from other indices that occur. From Section 3 on, they will be the labels of balls in an urn model. Let ϕ be a fixed transformation on $\{1, 2, \ldots, n\}$. In discussing such transformations, we generally follow the notation of the classical paper [19]. We identify ϕ with its digraph: $\phi \equiv \{(i,\ \phi(i))|i = 1, 2, \ldots, n\}$. We define the *trajectory* of a point i_0 under ϕ as the subgraph

$$\mathrm{Tr}(i_0,\ \phi) = \big\{(\phi^k(i_0),\ \phi^{k+1}(i_0))|k = 0,\ 1, 2, \ldots\big\}.$$

All trajectories are eventually cyclic because $\{1, 2, \ldots, n\}$ is finite. We denote by $\mathrm{Cy}(i_0,\ \phi)$ the cycle of this trajectory (also called simply the cycle of i_0). This may be defined formally as follows. Let k_0 be the smallest positive k with $\phi^k(i_0) = \phi^\ell(i_0)$, for some $\ell < k$. Then the cycle of i_0 is the subgraph

$$\mathrm{Cy}(i_0,\ \phi) = \big\{(\phi^k(i_0),\ \phi^{k+1}(i_0))|k = k_0,\ k_0 + 1,\ k_0 + 2, \ldots\big\}.$$

The function ϕ induces an equivalence relation $\sim$ on $\{1, 2, \ldots, n\}$ defined as follows: we write $i \sim j$ if there exists a pair of nonnegative integers $k_1,\ k_2$ such that $\phi^{k_1}(i) = \phi^{k_2}(j)$. This equivalence relation decomposes $\{1, 2, \ldots, n\}$ into equivalence classes which are called the components of ϕ. Each cycle

is wholly contained in one *component*, which, in the language of dynamical systems, is the *basin of attraction* of the cycle. We denote by $Co(i_0, \phi)$ the component containing a particular label $i_0 \in \{1, 2, \ldots, n\}$.

Let p_n be a probability measure on the set $\{1, 2, \ldots, n\}$, and write p_{ni} for the mass $p_n(\{i\})$ that it assigns to i. We place a probability measure π_n on the discrete space Ω_n of transformations $\phi: \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ by setting

$$\pi_n(\{\phi\}) = \prod_{i=1}^{n} p_{n,\,\phi(i)}, \qquad \phi \in \Omega_n. \tag{2.1}$$

This means that the images of the labels $1, 2, \ldots, n$ are chosen independently according to the distribution p_n. We write Φ for the identity transformation on Ω_n, so that, under π_n, Φ is a random transformation on $\{1, 2, \ldots, n\}$. The random transformation Φ is the sole focus of this paper.

If p_n is the uniform distribution on $\{1, 2, \ldots, n\}$, then Φ is called the *completely random transformation* and its law is denoted by π_n^{unif}. The completely random transformation is a "combinatorial" object, in the sense that questions about its distribution are counting problems. The theory of completely random transformations is quite well developed, through combinatorial methods; see [7], [19] and [26] and the bibliographies therein. The related theory of "completely random permutations" was treated in [24]. (For more recent work on random permutations, see [11] and [20] and the references therein.)

We denote by $\#\mathcal{G}$ the order of a graph $\mathcal{G}$, that is, the number of nodes it contains. Thus, for example, $\#\text{Tr}(i_0, \phi)$ is the number of labels in the trajectory of i_0 under ϕ. (We use # to indicate the cardinality of a set also.) Our concern in this paper is with *component size, cycle length and trajectory length* of the labels $\{1, 2, \ldots, n\}$ under the random transformation Φ, which are summarized in the "triplets"

$$(X_i, Y_i, Z_i) \equiv \big(p_n\{Co(i, \Phi)\}, \#\text{Cy}(i, \Phi), \#\text{Tr}(i, \Phi)\big) \quad \text{for } i \in \{1, 2, \ldots, n\}. \tag{2.2}$$

Here component size is measured by p_n. Cycles and trajectories are measured by the number of labels they contain.

The quantity

$$c_n \equiv ||p_n||_2 \equiv \left(\sum_{i=1}^{n} p_i^2\right)^{1/2} \tag{2.3}$$

plays a key role in the following, as a natural scaling parameter for cycle and trajectory length. To explain this, we look ahead for a moment. We shall see in Section 3.1 that, under the homogeneity condition (2.5), the scaled trajectory length $c_n(\#\text{Tr}(L, \phi))$ converges in distribution as $n \to \infty$. See also part E5 of [1]. Similarly, the scaled cycle length $c_n(\#\text{Cy}(L, \phi))$ converges in distribution as $n \to \infty$. This simple fact is not treated directly in the following discussion, but is a consequence of the main result of the paper, Theorem 2.1. We summarize informally by saying that, under (2.5), cycle and trajectory length are

asymptotically of order $1/c_n$. These facts explain the scaling of the triplets $(X_i,\ Y_i,\ Z_i)$ of (2.2) that is introduced at (2.4).

Let L denote a label selected according to the distribution p_n. For $\phi \in \Omega_n$, we denote by F_ϕ the distribution function of the *scaled* triplet

$$\big(p_n\{\mathrm{Co}(L,\ \phi)\},\ c_n(\#\mathrm{Cy}(L,\ \phi)),\ c_n(\#\mathrm{Tr}(L,\ \phi))\big) \tag{2.4}$$

associated with L. This is the joint distribution of component size and scaled cycle and trajectory length of a label under the transformation ϕ, where that label is randomly selected according to the distribution p_n. The support of each F_ϕ lies in the set $[0,\ 1] \times \mathbb{R}^+ \times \mathbb{R}^+$. Were F_ϕ summarizes many of the most important features of ϕ from the viewpoint of dynamical systems and F_Φ is a random distribution function. Our goal is to prove a weak limit theorem for the *law* μ_n of F_Φ as the number of labels n increases to ∞.

2.2. *The main result.* We write $\mathscr{D}_d$ for the space of distribution functions (d.f.'s) on $\mathbb{R}^d$. Thus $F_\phi \in \mathscr{D}_3$ for $\phi \in \Omega_n$. We endow $\mathscr{D}_d$ with the Lévy metric d_L, which is defined for $F_1,\ F_2 \in \mathscr{D}_d$ by

$$\begin{aligned} d_L(F_1,\ F_2) = \inf\big\{&\varepsilon > 0 : F_1(t - \varepsilon e) - \varepsilon \le F_2(t) \\ &\le F_1(t + \varepsilon e) + \varepsilon \text{ for all } t \in \mathbb{R}^d\big\}, \quad \text{where } e = (1,\ 1, \ldots, 1). \end{aligned}$$

We denote the σ-field of Borel sets of $\mathscr{D}_d$ with respect to the Lévy metric by $\mathscr{B}(\mathscr{D}_d)$.

We consider the law μ_n of F_Φ on $\mathscr{D}_3$, defined by $\mu_n(S) = \pi_n(F_\Phi \in S)$, for $S \in \mathscr{B}(\mathscr{D}_3)$. We are interested in the asymptotic behavior as $n \to \infty$ of the measures μ_n, for a sequence p_n, $n = 1, 2, \ldots$, of probability measures, p_n being concentrated on $\{1, 2, \ldots, n\}$. The main result of this paper, stated in Theorem 2.1, is that these measures μ_n converge to a certain limit, provided that p_n is "somewhat uniform" over $\{1, 2, \ldots, n\}$ for large n [see (2.5)].

To treat convergence of the μ_n's, we endow the set $\mathscr{PM}(\mathscr{D}_3)$ of all Borel probability measures over $\mathscr{D}_3$ with the Prohorov metric (Section 3.1 of [17]), defined for $\eta_1,\ \eta_2 \in \mathscr{PM}(\mathscr{D}_3)$ by

$$d_{\mathrm{Pr}}(\eta_1,\ \eta_2) = \inf\big\{\varepsilon > 0 : \eta_1(S) \le \eta_2(S_\varepsilon) + \varepsilon \text{ for all } S \in \mathscr{B}(\mathscr{D}_3)\big\}.$$

Here S_ε is the open ε-neighborhood of the set S.

We remark that we could have worked just as well with the set of Borel measures on $\mathbb{R}^3$ rather than the set of d.f.'s $\mathscr{D}_3$ given, previously and then used the Prohorov metric in place of the Lévy metric. The use of different metrics improves clarity in Section 5.

THEOREM 2.1. *Suppose that for each $n = 1, 2, \ldots$ we have a probability measure p_n on $\{1, 2, \ldots, n\}$. Suppose that this sequence of probability measures satisfies*

$$\lim_{n\to\infty} \varepsilon_n = 0, \ \textit{where} \quad \varepsilon_n \equiv \max_{1\le i\le n} \frac{p_{ni}}{c_n} \ \textit{and} \ c_n \equiv ||p_n||_2. \tag{2.5}$$

Let Φ *be the random transformation with distribution of images* p_n, *and let* μ_n *denote the law of the random d.f.* F_Φ *of scaled triplets of* Φ, *as defined in the paragraph containing* (2.4). *Then the laws* μ_n *converge weakly to a law* μ *on* $\mathscr{D}_3$ *as* $n \to \infty$ *that does not depend on the* p_n*'s.*

Note that

$$(2.6)\quad c_n^2 = \sum_{i=1}^{n} p_{ni}^2 \le \max_{1\le i\le n} p_{ni} = \varepsilon_n c_n, \quad \text{and so} \quad c_n \le \varepsilon_n, \quad n = 1, 2, \ldots .$$

Thus (2.5) also implies that $\lim_{n\to\infty} c_n = 0$.

We call the condition expressed in (2.5) the *homogeneous limit*. Results for a certain *heterogeneous limit*, which explain some empirical observations on discretizations of chaotic transformations with "attracting centers," will be described fully elsewhere; see the discussion in Section 6, in particular, Theorem 6.1.

Theorem 2.1 extends what is known about the random transformation Φ in two ways. First, it extends much of what is known in the uniform case, where $p_{ni} = 1/n, i = 1, 2, \ldots, n,$ to the case of the homogeneous limit. Second, it gives the asymptotic behavior of a fairly comprehensive descriptor of the dynamical system associated with Φ, namely, the joint distribution of component size, cycle length and trajectory length.

In the uniform case, in which Φ is the completely random mapping, we have $c_n = n^{-1/2}$. This means informally that cycles and trajectories are $O(\sqrt{n})$ in length. We denote by μ_n^{unif} the measure μ_n on $\mathscr{D}_3$ in this case, which may also be described as the π_n^{unif}-law of F_Φ. Harris, Stepanov and others have proved many limit theorems for this uniform case. As a simple example, the distribution of the size of the component containing a given label is known to converge to the *Stepanov distribution*, defined by its d.f.

$$F_{\text{Stepanov}}(t) = 1 - \sqrt{1-t}, \qquad 0 \le t \le 1.$$

Theorem 2.1 implies that the distribution of component size has the same limit under (2.5). Even this simple result is new, to the best of our knowledge, for the nonuniform case.

3. Urn models. We first review some basics on Poisson approximation. Then in Section 3.1 we develop a connection between the random transformation Φ and a simple urn model with Poisson drawing times. This connection is extended in Section 3.2 to a coupling of Φ with a many-urn model. The coupling plays a role later in the proof of Theorem 2.1.

The total *variation distance* between two probability measures $\eta_1,\ \eta_2$ on a σ-field $\mathscr{F}$ is defined as

$$d_{\text{TV}}(\eta_1,\ \eta_2) \equiv d_{\text{TV}}(\eta_1,\ \eta_2|\mathscr{F}) \equiv \sup_{G\in\mathscr{F}} |\eta_1(G) - \eta_2(G)|.$$

We denote the law or distribution of a random variable X by $\mathscr{L}(X)$. The Poisson distribution of mean m is denoted by $\text{Po}(m)$. See Barbour, Holst and

Janson [5] for the following well-known results on Poisson approximation, where Theorem 3.1 is (1.23) and Theorem 3.2 is Theorem 1.C.

THEOREM 3.1. *Let* $W = \sum_{i=1}^{n} I_i$, *where the* I_i*'s are independent indicator random variables with* $p_i = \mathbb{E}(I_i) = \mathbb{P}(I_i = 1)$. *Let* $m = \mathbb{E}(W) = \sum_{i=1}^{n} p_i$. *Then*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(m)\big) \le \min(1,\ m^{-1}) \sum_{i=1}^{n} p_i^2 \le \sum_{i=1}^{n} p_i^2.$$

THEOREM 3.2. *Let* M *be a nonnegative random variable and consider the mixture* $\mathrm{Po}(M)$ *of Poisson distributions with mixing measure* $\mathscr{L}(M)$ *over the mean. We have, for all* $m > 0$,

$$d_{\mathrm{TV}}\big(\mathrm{Po}(M),\ \mathrm{Po}(m)\big) \le \min(1, m^{-1/2})\mathbb{E}|M - m| \le \mathbb{E}|M - m|.$$

In each of these theorems, the larger bound is elementary. The refined bounds, involving the "magic factors" m^{-1} and $m^{-1/2}$, are more difficult to prove. The refined bounds are not used in the following discussion.

3.1. *A simple urn model.* Let n be a given positive integer, and suppose we have an urn containing n balls, labeled $\{1, 2, \ldots, n\}$. We draw balls from the urn, independently with replacement. We write $p_{ni} > 0$ for the probability that ball i is selected on a single draw. We first focus on the question: *What is the distribution of the number* D *of draws needed until some ball is drawn twice*? A little reflection shows that, in the notation of Section 2, this is equivalent to the question: *What is the distribution of* $\#\mathrm{Tr}(i_0, \Phi)$*, the number of labels in the trajectory of a given label under the random transformation* Φ? This is answered in a limiting sense, under the homogeneity hypotheses (2.5) on the p_{ni}'s. See Aldous [1], parts E5 and E20, for heuristics and commentary. In the next section, we develop a fuller relationship between Φ and drawing balls from urns.

Our analysis of the italicized questions in the preceding paragraph lays some of the foundation for the proof of Theorem 2.1. The first step is to embed the draws in a Poisson process, as in part E3 of [1]. Let N be a Poisson counting process of rate 1, with jump times $T_1,\ T_2, \ldots$, and suppose the kth draw from the urn takes place at time T_k. This may be described as a *marked Poisson process* [9], the marks being the ball labels. Let $N_i(t)$ denote the number of times that ball i has been drawn by time t. By elementary properties of the Poisson process, the N_i's are independent Poisson processes, N_i having rate p_{ni}. Of course,

$$N = \sum_{i=1}^{n} N_i.$$

Let R denote the earliest instant at which some ball is drawn for the second time. Thus, the number D of draws up to this time satisfies

$$R = T_D \quad \text{and} \quad N(R) = D.$$

We prove a limit theorem for $c_n R$ when the number of balls n increases to ∞, under (2.5). Then we prove that $c_n R$ and $c_n D$ are close and so have the same limit.

We now assume the homogeneous limit condition (2.5). We also introduce the quantities

$$a_{ni} \equiv \frac{p_{ni}}{c_n}, \quad \text{satisfying } \sum a_{ni}^2 = 1. \tag{3.1}$$

Under (2.5), (2.6) implies that $c_n \to 0$, or, equivalently, $\max_i p_{ni} \to 0$. So, for sufficiently large n, each ball has a negligible chance of being chosen on a single draw. In fact, (2.5) also ensures that, for sufficiently large n, each ball has a negligible chance of being the first to be drawn twice; this will be apparent from the results of Section 4.

The following calculation identifies the distribution of R explicitly. For $t \geq 0$,

$$\begin{aligned} \mathbb{P}(R > t) &= \mathbb{P}(\text{no repeat draws of ball } i \text{ up to time } t,\ i = 1, 2, \ldots, n) \\ &= \prod_{i=1}^{n} P\big(N_i(t) < 2\big) = \prod_{i=1}^{n} e^{-p_{ni}t}(1 + p_{ni}t). \end{aligned} \tag{3.2}$$

The Poisson embedding has brought about the independence here. We have from (3.2) that

$$\begin{aligned} \mathbb{P}\big(c_n R > t\big) &= \prod_{i=1}^{n} e^{-p_{ni}t/c_n}(1 + p_{ni}t/c_n) \\ &= \prod_{i=1}^{n} e^{-a_{ni}t}(1 + a_{ni}t), \qquad t \geq 0. \end{aligned}$$

Elementary analysis using (3.1) and the assumption (2.5) establishes from this that, for $t \geq 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(c_n R > t\big) = \exp\{-t^2/2\}. \tag{3.3}$$

(This is also a consequence of Theorem 4.1.) This is the required convergence in distribution for $c_n R$ as the number n of balls goes to ∞.

Let $\varepsilon > 0$ be given. By applying Doob's maximal quadratic inequality to the martingale $N(t) - t$, $t \geq 0$, we see that, for any $t_0 > 0$,

$$\mathbb{P}\Big(\max_{0\leq s\leq t_0} |N(s) - s| > \varepsilon\Big) \leq \frac{4\mathbb{E}\big(N(t_0) - t_0\big)^2}{\varepsilon^2} = \frac{4t_0}{\varepsilon^2}.$$

Applying this with $t_0 = t/c_n$, we get

$$\begin{aligned} \mathbb{P}(c_n|D - R| > \varepsilon) &= \mathbb{P}\left(|N(R) - R| > \frac{\varepsilon}{c_n}\right) \\ &\leq \mathbb{P}\left(\max_{0\leq s\leq t/c_n} |N(s) - s| > \frac{\varepsilon}{c_n}\right) + \mathbb{P}\left(R > \frac{t}{c_n}\right) \\ &\leq \frac{4tc_n}{\varepsilon^2} + \mathbb{P}(c_n R > t). \end{aligned} \tag{3.4}$$

Now choose t with $\exp\{-t^2/2\} = \varepsilon/3$. Using (3.3) and the fact that $c_n \to 0$ as $n \to \infty$, it is clear that for sufficiently large n each term on the right-hand side of (3.4) is no greater than $\varepsilon/2$. For such n, we have $\mathbb{P}(c_n|R - D| > \varepsilon) \leq \varepsilon$. This implies that $c_n D$ has the same limit law as $c_n R$.

3.2. *Coupling drawings from urns and the random transformation.* We consider U urns, each one identical to the one described in Section 3.1. Thus balls are drawn independently at Poisson, rate 1, times from each of the U urns, and the probability that ball i is chosen on any draw is p_{ni}. This construction may be viewed as U independent marked Poisson processes, as before. Let $N_{ui}(t)$ denote the number of balls of label i drawn from urn u by time t. N_{ui} is a Poisson process of rate p_{ni}, and all these Poisson processes are independent of one another.

We write $(\Omega, \mathscr{F}, \mathbb{P})$ for the probability space supporting the N_{ui}'s. The probability is 0 that two balls are ever drawn simultaneously, and we excise this event from Ω as doing so simplifies the description of various constructions to follow. This space will be expanded to support other random variables of interest as we proceed. In this section, we construct on an expanded $(\Omega, \mathscr{F}, \mathbb{P})$ a random transformation Φ with the distribution π_n specified in (2.1).

We define an *expiration time* $\mathscr{E}_u$, $u = 1, 2, \ldots, U$, for each of the U urns, inductively, as follows. Suppose we observe each of the U marked Poisson processes until the first moment that some label is drawn for the second time among all the drawings from all the urns. Let U_1 denote the index of the urn from which this label was drawn for the second time, and let $\mathscr{E}_{U_1}$ be the time of that second drawing. We say that urn U_1 is *expired* from time $\mathscr{E}_{U_1}$ on, and refer to $\mathscr{E}_{U_1}$ as the expiration time of urn U_1. We say that draws of labels from an urn before its expiration are *observed*, and draws made at or after its expiration time are *unobserved*. In particular, the draw from urn U_1 at time $\mathscr{E}_{U_1}$ is unobserved.

We proceed inductively to define expiration times for all U urns. Suppose that we have determined the expiration times $\mathscr{E}_{U_1} < \mathscr{E}_{U_2} < \cdots < \mathscr{E}_{U_k}$ of the urns $U_1, U_2, \ldots, U_k$, where $k < U$. At time $\mathscr{E}_{U_k}$, urns $U_1, U_2, \ldots, U_k$ are expired and all other urns are unexpired. We now continue to observe the drawings from the *unexpired urns* from time $\mathscr{E}_{U_k}$ on until we next make a drawing, from an unexpired urn, of a label that has already been drawn from any of the U urns *during its unexpired period*. In other words, we observe the unexpired urns until the next instant when there is a new match with an earlier *observed* draw. This instant is the expiration time $\mathscr{E}_{U_{k+1}}$ of the urn U_{k+1} from which this label is drawn. This process continues until an expiration time has been determined for each urn. That this process terminates may be seen as follows. Let R_{uu} denote the earliest time at which some label is drawn for the second time from urn u. This is clearly finite with probability 1. If the urn has not expired already by time R_{uu}, then there is a match at time R_{uu} with an earlier observed draw and so in this situation R_{uu} is the expiration time

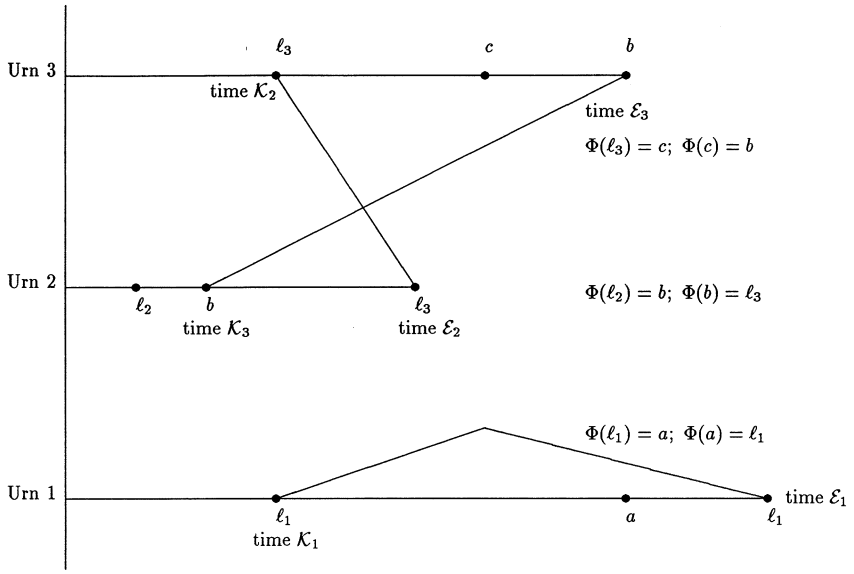

FIG. 3. *Defining* Φ *on the observed labels* $B = \{\ell_1, \ell_2, \ell_3, a, b, c\}$ *from the marked Poisson processes.*

of urn u. We have shown that

$$\mathscr{E}_u \leq R_{uu} < \infty.$$

For future reference, we define $\mathscr{K}_u$ to be the time at which the label drawn at the expiration time $\mathscr{E}_u$ was drawn previously during observed time. We call $\mathscr{K}_u$ the *knot time* associated with urn u, because we think of it as the point at which the time axis of the expired urn u is "tied" to that of another. We denote by ℓ_u the *initial label* of urn u, which is defined as the label of the first ball to be drawn from urn u. Figure 3 illustrates a simple example, in which we have connected each expiration time to the corresponding knot time by a line.

Now consider the set B of all labels drawn during the unexpired periods of the urns. We refer to this as the *set of observed labels*. Each label $i \in B$ has a *unique* successor $i^+ \in B$ defined as follows. Each label $i \in B$ has been drawn *exactly* once during unexpired time. This is the observed draw of that label. We define the successor i^+ of i as the next label drawn after the observed draw of i from the same urn as the observed draw of i was from. That $i^+ \in B$ may be seen as follows. If the next draw after the observed draw of i, from the same urn as i, is not at an expiration time, then clearly the label drawn, namely i^+, is observed and so is in B. On the other hand, if the next label drawn is at an expiration time, then, by definition of expiration times, this label was also drawn at some earlier observed time, and so again this label is in B.

Now we are ready to define the random transformation Φ. First we define Φ on the observed labels B by

$$\Phi(i) = i^+, \qquad i \in B.$$

For $i \notin B$, we choose $\Phi(i)$ at random from $\{1, 2, \ldots, n\}$ according to the distribution p_n, independently of everything else that has been introduced. With this we have extended the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to encompass the random transformation Φ on $\{1, 2, \ldots, n\}$. The construction of Φ on B is illustrated in Figure 3. We now have the following result.

THEOREM 3.3. *The random transformation Φ has the distribution π_n over Ω_n. Moreover, the initial labels $\ell_1, \ell_2, \ldots, \ell_U$ associated with the urns are i.i.d. with distribution p_n and are independent of Φ.*

PROOF. We summarize the construction of Φ and the ℓ_u's in a manner which makes the statement of the theorem obvious. We may view this construction as a process of "visiting" the various urn indices $\{1, 2, \ldots, U\}$ and the various ball labels $\{1, 2, \ldots, n\}$ in a complicated order determined by the marked Poisson processes associated with the urns. When a label or urn index is visited, a label is assigned to it. Here is a precise description of this process of visiting and assigning labels.

We say that urn index u is "visited" the first time a ball is drawn from it. The label assigned to urn index u is the label ℓ_u of the first ball drawn from urn u.

We say that ball label $i \in B$ is "visited" when the next ball is drawn after the unique observed drawing of label i, from the same urn. The label assigned to i is the label $i^+ = \Phi(i)$ of this next ball drawn from the same urn.

When these assignments have been made, the remaining labels $i \notin B$ may be "visited" in an arbitrary order as described in the construction of Φ. They are assigned i.i.d. labels $\Phi(i)$ according to p_n independently of everything else.

We note now that, whenever an urn index or ball label is "visited," a label is assigned to it according to the distribution p_n, independently of all previous assignments of labels. This property is enough to guarantee that all of the assignments are i.i.d. with distribution p_n. This completes the proof. □

In Section 5 we use the coupling described here to get detailed information about the triplets $(X_{\ell_u}, Y_{\ell_u}, Z_{\ell_u})$, $u = 1, 2, \ldots, U$, of (2.2) associated with the ℓ_u's under Φ. This and the Poisson approximation of the next section are the basis of the proof of Theorem 2.1.

4. Poisson process approximation for matches among many urns. In this section we develop a Poisson process approximation for the matches between labels drawn from the U urns of Section 3.2. We derive an explicit total variation norm bound for the approximation. The random transformation Φ plays no role here. The first step is taken in Section 4.1, where we give a Poisson approximation for the number of matches between a pair of urns up

to a fixed time τ. Then in Sections 4.2 and 4.3 we extend this to Poisson approximations for the *processes* of matches between pairs of urns.

4.1. *Poisson approximation for the number of matches among many urns.* We treat point processes somewhat informally for now, postponing some elementary topological and measure-theoretic formalities until the next section.

We consider the U urns of Section 3.2, and the independent Poisson processes N_{ui} of drawing times of label i from urn u. We define

$$N_{\cdot i}(t) \equiv \sum_{u=1}^{U} N_{ui}(t), \qquad t \geq 0,$$

a process which counts the total number of draws of balls of label i from all U urns up to time t. The $N_{\cdot i}$'s are again independent Poisson processes, $N_{\cdot i}$ having rate Up_{ni}. We also define the following independent Poisson processes of rate 1, which track the total number of draws from each urn u up to time t:

$$N_{u\cdot}(t) \equiv \sum_{i=1}^{n} N_{ui}(t), \qquad t \geq 0.$$

The $N_{u\cdot}$'s will not be needed until Section 5.

Our concern is with labels that are drawn more than once, from the totality of all urns. There is said to be a *match between urn u and urn v* when a label is drawn from urn u that has already been drawn from urn v. These are tracked by the planar point processes

$$M_{uv}(A) \equiv \sum_{i=1}^{n} \int\int_{A} dN_{ui}(u_1)\, dN_{vi}(u_2), \qquad A \in \mathscr{B}(\Delta), \quad \text{where } \Delta \equiv \{(t_1, t_2): 0 \leq t_2 \leq t_1\}, \tag{4.1}$$

counting the number of pairs $(t_1,\ t_2)$ in a Borel set A such that the same label was drawn from urn u at time t_1 and from urn v at the earlier time $t_2 \leq t_1$. The letter M here connotes "matches." Here and in the following discussion, a match is characterized by four quantities: the two times $t_1 > t_2$ at which the same label was drawn, the urn u from which that label was drawn at time t_1 and the urn v from which it was drawn at time t_2. The actual label drawn is usually not of interest. In (4.1), M_{uv} is presented as a random measure. It is easily seen to be *simple* [9]: every point of Δ has M_{uv}-mass ≤ 1.

In this section, we fix a time $\tau > 0$ and focus on draws up to time τ from the U urns. We write M_{uv}^{τ} for the "projection" of our point processes onto the triangle

$$\Delta_{\tau} \equiv \{(t_1,\ t_2): 0 \leq t_2 \leq t_1 \leq \tau\}.$$

More formally, for a Borel measure β on Δ, its projection onto Δ_{τ} is the Borel measure $\Pi^{\tau}\beta$ defined by

$$\Pi^{\tau}\beta(A) = \beta(A \cap \Delta_{\tau}) \quad \text{for } A \in \mathscr{B}(\Delta). \tag{4.2}$$

With this notation, $M_{uv}^{\tau} \equiv \Pi^{\tau} M_{uv}$.

Let I_i denote the indicator of the event that label i is drawn *exactly twice* by time τ. Thus $I_i = 1$ iff $N_{\cdot i}(\tau) = 2$. As $N_{\cdot i}$ is a Poisson process of rate Up_{ni}, we have

$$\mathbb{P}(I_i = 1) = \mathbb{P}\big(N_{\cdot i}(\tau) = 2\big) = \frac{(Up_{ni}\tau)^2}{2} e^{-Up_{ni}\tau}. \tag{4.3}$$

Consider the sum

$$J_2 = \sum_{i=1}^{n} I_i,$$

which is the number of labels drawn exactly twice by time τ. The expected value of J_2 is

$$m \equiv \mathbb{E}(J_2) = \sum_{i=1}^{n} \frac{(Up_{ni}\tau)^2}{2} e^{-Up_{ni}\tau}.$$

The next lemma bounds the total variation distance between $\mathscr{L}(J_2)$ and $\mathrm{Po}(m)$.

LEMMA 4.1. *We have*

$$d_{TV}\big(\mathscr{L}(J_2),\, \mathrm{Po}(m)\big) \le \frac{U^4}{4}\tau^4 \sum_{i=1}^{n} p_{ni}^4 = \frac{U^4}{4}\tau^4 ||p_n||_4^4.$$

PROOF. Apply Theorem 3.1, using (4.3) and the fact that the exponential factor there is no greater than 1. □

Using the inequalities $1 - x \le e^{-x} \le 1,\ x \ge 0$, in (4.3) we find that

$$\frac{1}{2}U^2 c_n^2 \tau^2 - \frac{1}{2}U^3 \sum_{i=1}^{n} p_{ni}^3 \tau^3 \le m \le \frac{1}{2}U^2 c_n^2 \tau^2. \tag{4.4}$$

These inequalities allow us to place a bound on the total variation distance between $\mathscr{L}(J_2)$ and $\mathrm{Po}(U^2 c_n^2 \tau^2/2)$, and so allow us to replace the $m = \mathbb{E}(J_2)$ of Lemma 4.1 with something simpler to understand.

LEMMA 4.2. *We have*

$$d_{TV}\big(\mathscr{L}(J_2), \mathrm{Po}(U^2 c_n^2 \tau^2/2)\big) \le \frac{U^3}{2}\tau^3 ||p_n||_3^3 + \frac{U^4}{4}\tau^4 ||p_n||_4^4.$$

PROOF. Using (4.4), Theorem 3.2 gives a bound on the total variation norm distance between the Poisson distributions with means m and $U^2 c_n^2 \tau^2/2$. The triangle inequality allows us to combine this bound with the result of Lemma 4.1 to give the desired conclusion. □

Consider the number J_3 of labels that are drawn *at least three times*. (Recall that J_2 counts only those that have been drawn *exactly twice*.) Boole's inequality gives

$$\begin{aligned}\mathbb{P}(J_3 > 0) &= \mathbb{P}(\text{some label is drawn three times by time } \tau)\\ &\le \sum_{i=1}^{n} \mathbb{P}\big(N_{\cdot i}(\tau) \ge 3\big).\end{aligned}$$

It is elementary that for a Poisson(θ) random variable W, $\mathbb{P}(W \ge 3) \le \theta^3/6$, $\theta \ge 0$. This gives immediately the following result.

LEMMA 4.3. *We have*

$$\mathbb{P}(J_3 > 0) \le \frac{1}{6} U^3 \tau^3 ||p_n||_3^3.$$

Lemmas 4.2 and 4.3 summarize our analysis of the *number* of matches among all draws from the U urns by time τ. When we invoke the homogeneous limit condition (2.5), using the elementary (4.8), Lemma 4.2 tells us that the number of labels drawn exactly twice by time τ is approximately Poisson. Lemma 4.3 tells us, again under (2.5), that it is unlikely that any label is drawn three or more times. Our goal is something larger: a simple total variation approximation for the law of the processes M_{uv}^{τ}, $u,\ v = 1, 2, \ldots, U$. What is missing from Lemmas 4.2 and 4.3 is a description of *between which urns and at what times* the J_2 matches occur, given $J_3 = 0$. This issue is to be addressed fully in the next section, using the following simple lemma.

LEMMA 4.4. *Given that $N_{\cdot i}(\tau) = 2$, that is, that label i was drawn exactly twice among all drawings up to time τ, we have*:

(a) *For any pair $(u,\ v)$, the conditional probability that label i was drawn first from urn v and then (at a later time) from urn u is $1/U^2$.*

(b) *The conditional distribution of the two times $T_1 \ge T_2$ at which label i was drawn is uniform over the triangle $\Delta_\tau = \{(t_1,\ t_2)\colon 0 \le t_2 \le t_1 \le \tau\}$.*

(c) *The pair of urns from which label i was drawn is conditionally independent of the times $(T_1,\ T_2)$ at which it was drawn.*

PROOF. This follows by direct calculation from the premise that the N_{ui}'s, $u = 1, 2, \ldots, U$, are independent and identical Poisson processes. □

4.2. *Poisson process approximation for matches.* In this section and the next, our focus is on matches between urns, ignoring the labels of the balls involved. Our goal is a Poisson process approximation for these matches. In preparation for a precise treatment of certain point processes, we now make some aspects of the underlying probability and metric spaces explicit. We follow [9] for the basic formalities and some notation. Let $\mathscr{X}$ be a complete separable metric space and let $\mathscr{B}(\mathscr{X})$ be its Borel field. Were $\widehat{\mathscr{M}}_{\mathscr{X}}$ denotes the set

of all Borel measures giving finite mass to bounded sets ("boundedly finite") and giving only nonnegative integer masses. There is a natural metric on $\widehat{\mathscr{M}}_{\mathscr{X}}$ (A2.6 of [9]) under which it becomes a complete separable metric space, and we denote its Borel field by $\mathscr{B}(\widehat{\mathscr{M}}_{\mathscr{X}})$. This Borel field is generated by the mappings $m \to m(A)$, $A \in \mathscr{B}(\mathscr{X})$, $m \in \widehat{\mathscr{M}}_{\mathscr{X}}$, from $\widehat{\mathscr{M}}_{\mathscr{X}}$ to $\mathbb{R}$. A *point process* on $\mathscr{X}$ is formally defined in [9] as a Borel-measurable mapping from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ into $\widehat{\mathscr{M}}_{\mathscr{X}}$.

On our probability space $(\Omega, \mathscr{F}, \mathbb{P})$ are defined the Poisson processes N_{ui}, $u = 1, 2, \ldots, U$, $i = 1, 2, \ldots n$, on $[0, \infty)$, from which everything in this section derives. The point processes of matches M_{uv} on $\mathscr{X} = \Delta$ defined in (4.1) comprise a $U \times U$ "matrix"

$$M \equiv (M_{uv}: u,\ v = 1, 2, \ldots, U),$$

which we view as a random element of the U^2-fold product space

$$(\Omega',\ \mathscr{F}') \equiv \left(\widehat{\mathscr{M}}_{\Delta},\ \mathscr{B}(\widehat{\mathscr{M}}_{\Delta})\right)^{U^2}. \tag{4.5}$$

Let $\mathscr{F}_M$ denote the sub-σ-field of $\mathscr{F}$ generated by $M: \Omega \to \Omega'$. It is helpful to think of $\mathscr{F}_M$ as containing only information on the times of matches and the indices of the urns involved, but containing no further information on the labels drawn. We write $\Pi^\tau M$ for the matrix $(M^\tau_{uv}) \equiv (\Pi^\tau M_{uv})$ of point processes of matches up to time τ. We denote by $\mathscr{F}_{M,\tau}$ the sub-σ-field of $\mathscr{F}_M$ generated by $\Pi^\tau M: \Omega \to \Omega'$.

We define a new probability measure ν_n over $\mathscr{F}_M$ by specifying the law of the M_{uv}'s under ν_n, as follows: *under ν_n, the M_{uv}'s are independent Poisson processes of rate c_n^2 on Δ*. Therefore, under ν_n, the M^τ_{uv}'s are independent Poisson processes of rate c_n^2 on the finite triangle Δ_τ. Here is another way to describe ν_n on $\mathscr{F}_{M,\tau}$, which helps to bring out a similarity between ν_n and $\mathbb{P}$ in the light of Lemma 4.4. Under ν_n, the total number of matches by time τ, defined by

$$M_{\text{tot}}(\tau) \equiv \sum_{u,\,v=1}^{U} M_{uv}(\Delta_\tau),$$

has the Poisson distribution of Lemma 4.2, namely, $\text{Po}(U^2 c_n^2 \tau^2/2)$. Furthermore, still under ν_n, given the total number $M_{\text{tot}}(\tau)$ of matches, these matches, taken in random order, are independent of one another and have a common marginal distribution. The marginal distribution may be described as follows. A match is distributed across the urn pairs (u, v) and times $(t_1, t_2) \in \Delta_\tau$ according to the conditional distribution described in (a)–(c) of Lemma 4.4. This similarity between $\mathbb{P}$ and ν_n is the basis of the next lemma, in which we must restrict $\mathbb{P}$ to the event $J_3 = 0$ that no label is drawn more than twice by time τ in order to invoke Lemma 4.4.

LEMMA 4.5. *Let $\tau > 0$ be fixed. The probability measures $\mathbb{P}(\cdot\,|J_2 = m,\ J_3 = 0)$ and $\nu_n(\cdot\,|M_{\text{tot}}(\tau) = m)$ are identical on $\mathscr{F}_{M,\tau}$ for $m = 0, 1, 2, \ldots$. That is, for*

all $G \in \mathscr{F}_{M,\tau}$ *we have*

$$\mathbb{P}(G|J_2 = m, J_3 = 0) = \nu_n\big(G|M_{\text{tot}}(\tau) = m\big), \qquad m = 0, 1, 2 \ldots.$$

PROOF. Note first that the conditional probabilities are well defined: the events on which we condition have positive probability for all $m = 0, 1, 2, \ldots$. Using Lemma 4.4, the independence of the $N_{\cdot i}$'s and the comments immediately preceding the statement of the present lemma, we get the following fact: the conditional $\mathbb{P}$-law of M^τ, given $J_2 = m$ and $J_3 = 0$ and given the random variables $I_1, I_2, \ldots, I_n$ indicating which m of the n labels were drawn twice by time τ [see (4.3)], is identical to the conditional law $\nu_n(\cdot|M_{\text{tot}}(\tau) = m)$. Since this conditional law depends on $I_1, I_2, \ldots, I_n$ only through their sum $J_2 = m$, the lemma follows. □

The next theorem, which is the main result of this section, bounds the total variation distance between the probability measures ν_n and $\mathbb{P}$ over the σ-field $\mathscr{F}_{M,\tau}$. Equivalently, it bounds the total variation distance between the $\mathbb{P}$-law of M^τ and the ν_n-law of M^τ.

THEOREM 4.1. *We have*

$$d_{\text{TV}}(\mathbb{P}, \nu_n|\mathscr{F}_{M,\tau}) \le \frac{4}{3}U^3\tau^3||p_n||_3^3 + \frac{1}{2}U^4\tau^4||p_n||_4^4.$$

PROOF. We write O_3 for the event $J_3 = 0$, and O_3^c for its complement. The fact that $M_{\text{tot}}(\tau) = J_2$ on O_3 is used on the second line of the following calculation. Lemma 4.5 is used on the third. For all $G \in \mathscr{F}_{M,\tau}$ we have

$$\begin{aligned}
\mathbb{P}(G \cap O_3) &= \sum_{m=0}^{\infty} \mathbb{P}\big(G|M_{\text{tot}}(\tau) = m, O_3\big)\mathbb{P}\big(\big\{M_{\text{tot}}(\tau) = m\big\} \cap O_3\big) \\
&= \sum_{m=0}^{\infty} \mathbb{P}(G|J_2 = m, J_3 = 0)\mathbb{P}\big(\big\{M_{\text{tot}}(\tau) = m\big\} \cap O_3\big) \\
&= \sum_{m=0}^{\infty} \nu_n\big(G|M_{\text{tot}}(\tau) = m\big)\mathbb{P}\big(\big\{J_2 = m\big\} \cap O_3\big).
\end{aligned}$$

This is used on the second line of the next calculation. Some caution is needed because $O_3 \notin \mathscr{F}_M$, and so, for example, $\nu_n(O_3)$ is not defined:

$$\begin{aligned}
|\mathbb{P}(G) - \nu_n(G)| &\le \mathbb{P}(O_3^c) + |\mathbb{P}(G \cap O_3) - \nu_n(G)| \\
&\le \mathbb{P}(O_3^c) + \sum_{m=0}^{\infty} \big|\nu_n\big(G|M_{\text{tot}}(\tau) = m\big)\mathbb{P}\big(\{J_2 = m\} \cap O_3\big) \\
&\qquad - \nu_n\big(G|M_{\text{tot}}(\tau) = m\big)\nu_n\big(M_{\text{tot}}(\tau) = m\big)\big|
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}(O_3^c) + \sum_{m=0}^{\infty} \nu_n\big(G | M_{\mathrm{tot}}(\tau) = m\big) \big| \mathbb{P}\big(\{J_2 = m\} \cap O_3\big) \\
&\qquad\qquad - \nu_n\big(M_{\mathrm{tot}}(\tau) = m\big) \big| \\
&\le \mathbb{P}(O_3^c) + \sum_{m=0}^{\infty} \nu_n\big(G | M_{\mathrm{tot}}(\tau) = m\big) \\
&\quad \times \Big\{ \big| \mathbb{P}(J_2 = m) - \nu_n\big(M_{\mathrm{tot}}(\tau) = m\big) \big| + \mathbb{P}\big(\{J_2 = m\} \cap O_3^c\big) \Big\} \\
&\le \mathbb{P}(O_3^c) + \sum_{m=0}^{\infty} \Big\{ \big| \mathbb{P}(J_2 = m) - \nu_n\big(M_{\mathrm{tot}}(\tau) = m\big) \big| \\
&\qquad\qquad + \mathbb{P}\big(\{J_2 = m\} \cap O_3^c\big) \Big\} \\
&= 2\mathbb{P}(J_3 > 0) + 2 d_{\mathrm{TV}}\big(\mathcal{L}(J_2),\, \mathrm{Po}(U^2 c_n^2 \tau^2/2)\big).
\end{aligned}
$$

The two terms in the final expression are bounded by the results of Lemmas 4.3 and 4.2, respectively. □

4.3. *Poisson process approximation up to a random time.* As in Section 3.2, R_{uu} denotes the *time of the first match* for the pair $(u,\, u)$ of urns. Let

$$
R \equiv \max(R_{11}, R_{22}, \ldots, R_{UU}). \tag{4.6}
$$

(This notation is in concert with the use of R for the time of the first match in a single urn in Section 3.1.) We use Theorem 4.1 to deduce a total variation approximation for the law of the M_{uv}'s "up to time R." The new feature is that the "ending time" R, formerly fixed at τ, is now random.

We introduce the processes of matches $M^* = (M^*_{uv})$ observed up to time R, defined by

$$
M^*_{uv}(\omega) \equiv \Pi^{R(\omega)} M_{uv}(\omega) \text{ for } \omega \in \Omega.
$$

We define $\mathcal{F}^*_M$ to be the sub-σ-field of $\mathcal{F}_M$ induced by this mapping. In fact, R is a stopping time with respect to the filtration $\mathcal{F}_{M,\tau}$, $\tau \ge 0$, and $\mathcal{F}^*_M$ is the σ-field $\mathcal{F}_{M,R}$ of this stopping time.

Now we are ready to derive a total variation norm bound on the distance between $\mathbb{P}$ and ν_n on $\mathcal{F}^*_M$. For $G \in \mathcal{F}^*_M$, we have

$$
\begin{aligned}
|\mathbb{P}(G) - \nu_n(G)| &\le |\mathbb{P}\big(G \cap (R \le \tau)\big) - \nu_n\big(G \cap (R \le \tau)\big)| + \mathbb{P}(R > \tau) \\
&\quad + \nu_n(R > \tau) \\
&\le 2 d_{\mathrm{TV}}(\mathbb{P},\, \nu_n \mid \mathcal{F}_{M,\tau}) + 2\nu_n(R > \tau).
\end{aligned}
$$

On the second line, we have used a total variation distance bound twice: first in bounding the difference between $\mathbb{P}(G \cap (R \le \tau))$ and $\nu_n(G \cap (R \le \tau))$, and then in bounding the difference between $\mathbb{P}(R > \tau)$ and $\nu_n(R > \tau)$. These are justified by the fact that the events $R > \tau$ and $G \cap (R \le \tau)$ are in $\mathcal{F}_{M,\tau}$. As the

final bound does not depend on G, it also bounds the total variation distance between $\mathbb{P}$ and ν_n on $\mathscr{T}_M^*$. Thus we continue

$$\begin{aligned} d_{\mathrm{TV}}(\mathbb{P},\, \nu_n \mid \mathscr{T}_M^*) &\le 2\nu_n(R > \tau) + 2d_{\mathrm{TV}}(\mathbb{P},\, \nu_n \mid \mathscr{T}_{M,\tau}) \\ &\le 2U \exp\{-c_n^2\tau^2/2\} + \tfrac{8}{3}U^3\tau^3||p_n||_3^3 + U^4\tau^4||p_n||_4^4. \end{aligned} \tag{4.7}$$

The last step uses Theorem 4.1 and the following bound on the tail probability of R under ν_n. For any urn u, $\nu_n(R_{uu} > \tau) = \nu_n(M_{uu}(\Delta_\tau) = 0) = \exp\{-c_n^2\tau^2/2\}$, the latter following from the fact that M_{uu} is a Poisson process of rate c_n^2 on Δ under ν_n and that the area of Δ_τ is $\tau^2/2$. [This is, in essence, (3.3), which gives the limiting d.f. of $c_n R$ in the case of a single urn.] Now use Boole's inequality to deduce that $\nu_n(R > \tau) = \nu_n(\max_v R_{vv} > \tau) \le U\nu_n(R_{uu} > \tau)$. This explains the first term of (4.7).

The homogeneous limit condition (2.5) states that $\varepsilon_n \to 0$ as $n \to \infty$, where $\varepsilon_n \equiv \max_i p_{ni}/c_n$. We now make a choice of τ such that $c_n\tau \to \infty$ as $n \to \infty$ but also such that the final bound of (4.7) goes to 0 as $n \to \infty$. Set

$$\tau = \frac{\sqrt{-2\log \varepsilon_n^*}}{c_n}, \text{ where } \varepsilon_n^* \equiv \min\Big(\frac{1}{e}, \varepsilon_n\Big).$$

Substituting this choice into (4.7) leads to following result.

THEOREM 4.2. *We have*

$$d_{\mathrm{TV}}(\mathbb{P},\, \nu_n|\mathscr{T}_M^*) \le 14U^4\varepsilon_n|\log \varepsilon_n^*|^{3/2}.$$

PROOF. It follows from definitions that

$$\frac{||p_n||_3^3}{c_n^3} = \frac{||p_n||_3^3}{||p_n||_2^3} \le \varepsilon_n \le 1 \text{ and } \frac{||p_n||_4^4}{c_n^4} \le \varepsilon_n^2. \tag{4.8}$$

Using this and (4.7) with the special choice of τ, we have

$$\begin{aligned} d_{\mathrm{TV}}(\mathbb{P},\, \nu_n \mid \mathscr{T}_M^*) &\le 2U \exp\{-c_n^2\tau^2/2\} + \tfrac{8}{3}U^3\tau^3||p_n||_3^3 + U^4\tau^4||p_n||_4^4 \\ &\le 2U\varepsilon_n^* + \tfrac{8}{3}U^3\varepsilon_n|2\log \varepsilon_n^*|^{3/2} + U^4\varepsilon_n^2|2\log \varepsilon_n^*|^2 \\ &\le 2U\varepsilon_n^* + 8U^3\varepsilon_n|\log \varepsilon_n^*|^{3/2} + 4U^4\varepsilon_n|\log \varepsilon_n^*|^{3/2} \\ &\le 14U^4\varepsilon_n|\log \varepsilon_n^*|^{3/2}. \end{aligned}$$

The next-to-last line is because $3 > 2^{3/2}$ and $\varepsilon_n|\log \varepsilon_n^*|^{1/2} \le 1$. (For the latter, note that $\varepsilon_n^* \le \varepsilon_n \le 1$.) The last line is because $\varepsilon_n^* \le \varepsilon_n$ and $|\log \varepsilon_n^*| \ge 1$. □

It is natural to "speed up time" by a factor of $1/c_n$ in studying the limiting behavior of the counting processes M_{uv}, since they are close to Poisson processes of rate $c_n^2 \to 0$ on Δ. The time scaling produces the point processes on Δ defined by

$$\widetilde{M}_{uv}(A) \equiv M_{uv}\Big(\frac{1}{c_n}A\Big), \qquad A \in \mathscr{B}(\Delta), \quad \text{where } \frac{1}{c_n}A \equiv \{a/c_n \mid a \in A\}.$$

Under ν_n, the $\widetilde{M}_{uv}$'s are independent Poisson processes of rate 1 on Δ.

5. Proof of the main result. In this section we prove Theorem 2.1. The proof is divided into three parts. In the first part, treated in Section 5.1, we present an $\mathscr{F}_M^*$-measurable approximation $F^*_{\Phi,U}$ to the random d.f. F_Φ. In the second part, treated in Section 5.2, we progress in a series of small steps, as measured by the Lévy metric, from F_Φ to $F^*_{\Phi,U}$. In the third part, treated in Section 5.3, the Poisson approximation of Theorem 4.2 is invoked to give detailed information on the law of $F^*_{\Phi,U}$, and the proof of Theorem 2.1 is easily completed.

It is helpful to keep in mind that the word "label" is reserved for the labels $\{1, 2, \ldots, n\}$ of the balls in the urn model. Quantities described as "indices" do not refer to balls.

5.1. *An $\mathscr{F}_M^*$-measurable approximation.* In this section, we make the connection between the random transformation Φ of Section 2 and the Poisson approximation for matches of Section 4, by invoking the coupling of Φ with the urn model described in Section 3.2. Recall that ℓ_u denotes the initial label of urn u, that is, the label of the first ball drawn from urn u. We introduce $\mathscr{F}_M^*$-measurable quantities that approximate the size $X_{\ell_u} = p_n\{\mathrm{Co}(\ell_u, \Phi)\}$ of the component of Φ containing ℓ_u, the length $Y_{\ell_u} \equiv \#\mathrm{Cy}(\ell_u, \Phi)$ of the cycle of ℓ_u under Φ and the length $Z_{\ell_u} \equiv \#\mathrm{Tr}(\ell_u, \Phi)$ of the trajectory of ℓ_u under Φ. Here $\mathscr{F}_M^*$-measurability means precisely that we can construct these approximating random variables by observing the processes M^*_{uv} of matches up to time R. The goodness of these approximations is established in the next Section.

In the course of the construction of the coupling in Section 3.2, we associated with each urn index u, $u = 1, 2, \ldots, U$, an expiration time $\mathscr{E}_u$. We also introduced the time $\mathscr{K}_u$ at which the label drawn at time $\mathscr{E}_u$ was drawn previously during observed time. We called $\mathscr{K}_u$ was called the "knot time," being the point at which the time axis of an expired urn is "tied" to that of another, as indicated by the sloped lines in Figure 3 (or Figure 4). Two knot times may coincide. This happens when the same label is drawn at two different expiration times. The label of the ball drawn at the expiration time $\mathscr{E}_u$ of urn u is called the *knot label* of urn u; this is also the label of the ball drawn at the knot time $\mathscr{K}_u$.

We define $\varphi(u)$ to be the index of the unique urn from which the label drawn at time $\mathscr{E}_u$ from urn u was drawn previously during observed time. Thus φ is a transformation on the urn indices $\{1, 2, \ldots, U\}$. In Figure 4 we identify φ for the situation of Figure 3. We have the following result.

PROPOSITION 5.1. *Two urn indices $u, v \in \{1, 2, \ldots, U\}$ are in the same component of φ if and only if the initial labels ℓ_u and ℓ_v of the corresponding urns are in the same component of Φ.*

PROOF. This result is clear once the definitions are understood. We outline the argument, avoiding some additional notation needed for a formal proof. The key is that the trajectory of an observed label, say i, under Φ, may be described as follows. Let u be the index of the unique urn from which i was

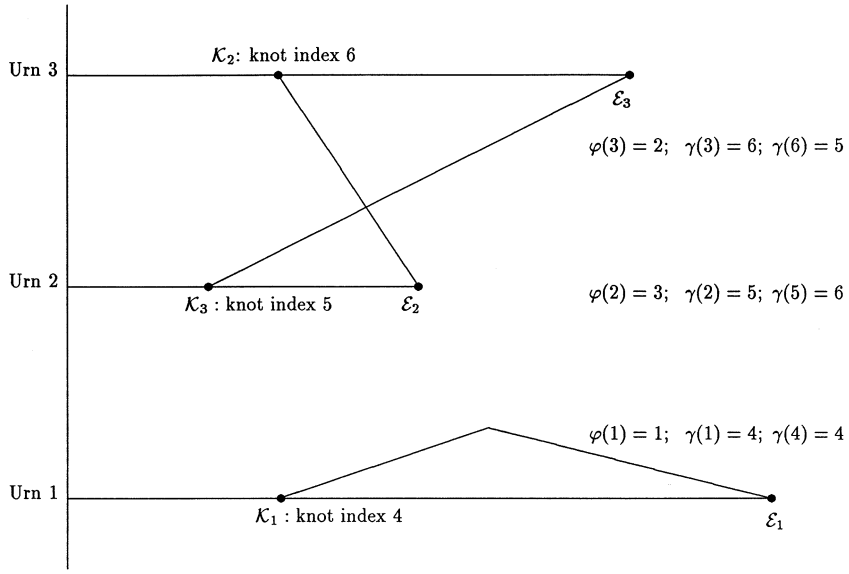

FIG. 4. *Defining φ on the urn indices $\{1, 2, 3\}$ and γ on the urn and knot indices $\{1, 2, 3, 4, 5, 6\}$. (Only the $\mathcal{F}_M^*$-measurable information of Figure 3 is presented here.)*

drawn during observed time. Then $i, \Phi(i), \Phi(\Phi(i)) \cdots$ is the sequence of labels that arise if we start at the time of the observed drawing of label i and record the labels drawn from urn u until just before its expiration time, whereupon we jump to urn $\varphi(u)$ at the knot time $\mathcal{K}_u$ and observe the labels drawn from that point until just before the expiration time of urn $\varphi(u)$, whereupon we jump to urn $\varphi(\varphi(u))$, and so forth.

Suppose that u and v are in the same component of φ. Then their trajectories under φ intersect, at say urn index w. Thus, by the preceding description, the trajectories of ℓ_u and ℓ_v under Φ must intersect at the knot label of urn w. This implies that ℓ_u and ℓ_v are in the same component of Φ.

Conversely, if ℓ_u and ℓ_v are in the same component of Φ, then, by definition, their trajectories under Φ must intersect at some label i. Let w be the index of the urn from which this label was drawn during observed time. Then, by the first paragraph of this proof, it follows that the trajectories of u and v under φ both visit w, and so we conclude that u and v are in the same component of φ. □

We introduce

$$X^u \equiv \frac{\#\mathrm{Co}(u,\ \varphi)}{U}, \qquad u = 1, 2, \ldots, U. \tag{5.1}$$

By Proposition 5.1, *this is the proportion of urns whose initial labels belong to the the same component of* Φ *as* ℓ_u. It is a natural approximation to X_{ℓ_u} based on the urn model. It is $\mathcal{F}_M^*$-measurable because the function φ is defined

entirely in terms of the processes M^*_{uv} of matches up to time R. This is proved to be a good approximation to X_{ℓ_u} in Lemma 5.1. Next we develop similar approximations for Y_{ℓ_u} and Z_{ℓ_u}.

Let V be the number of distinct knot times, and assign "knot indices" $U+1, U+2, \ldots, U+V$ to these knot times in an arbitrary order. We write ℓ_v for the label of the ball drawn at the knot time with index v, $v = U+1, U+2, \ldots, U+V$. Thus the ℓ_v, $v = U+1, U+2, \ldots, U+V$, are the knot labels. We define a transformation γ:

$$\gamma: \{1, 2, \ldots, U, U+1, U+2, \ldots, U+V\} \to \{1, 2, \ldots, U, U+1, U+2, \ldots, U+V\}$$

on the set of urn and knot indices, as follows. First consider an urn index $u = 1, 2, \ldots, U$. As we evaluate the iterates $\Phi^k(\ell_u)$, $k = 0, 1, 2, \ldots$, of Φ in turn, starting with the initial label ℓ_u of urn u, we eventually reach a knot label. We define $\gamma(u)$ to be the index of this first knot label that arises in these iterates. Now consider a knot index $v = U+1, U+2, \ldots, U+V$. We define $\gamma(v)$ to be the index of the first knot label arising in the iterates $\Phi^k(\ell_v)$, $k = 1, 2, \ldots$. (Note that the iterations here start with $k = 1$, not $k = 0$.) The function γ is evaluated in Figure 4 for the situation of Figure 3.

Again we identify the function γ with its digraph. The digraph γ becomes a *weighted graph* when we associate with each edge $(v, \gamma(v))$ the weight $\text{Weight}(v, \gamma(v))$ defined as follows. For urn indices $u = 1, 2, \ldots, U$, $\text{Weight}(u, \gamma(u))$ is the time until the first drawing of a knot label from urn u. For knot indices $v = U+1, U+2, \ldots, U+V$, $\text{Weight}(v, \gamma(v))$ is the time from the observed drawing of the knot label ℓ_v until the next drawing, whether observed or unobserved, of a knot label from the same urn. The weighted digraph of γ is $\mathcal{F}^*_M$-measurable as it depends only on the point processes M^*_{uv} of matches up to time R.

We define the *weight of a subgraph* $\mathcal{G}$ of the weighted graph γ as the sum of the weights of all of its edges:

$$\text{Weight}(\mathcal{G}) \equiv \sum_{(v,\gamma(v)) \in \mathcal{G}} \text{Weight}(v, \gamma(v)).$$

With this we define

$$Y^u \equiv \text{Weight}(\text{Cy}(u, \gamma)) \quad \text{and} \quad Z^u \equiv \text{Weight}(\text{Tr}(u, \gamma)) \quad \text{for } u = 1, 2, \ldots, U.$$

These are our $\mathcal{F}^*_M$-measurable approximations to Y_{ℓ_u} and Z_{ℓ_u}. That they are good approximations will be shown in Lemma 5.3. Informally, the idea is that Y^u and Z^u measure the continuous time taken by the Poisson processes of Section 3.2 to trace out the corresponding cycles and trajectories of Φ, and since the Poisson processes are of rate 1 this continuous time is close to the actual numbers of labels drawn, namely Y_{ℓ_u} and Z_{ℓ_u}. This is analogous to the relationship established in Section 3.1 between the number of draws D from a single urn until the first match and the (continuous) time R until the first

match. We record the important measurability property of the approximations developed in this Section as follows.

PROPOSITION 5.2. *The random variables* $X^u, Y^u, Z^u, u = 1, 2, \ldots, U$, *are* $\mathscr{T}_M^*$*-measurable*.

Now we introduce our $\mathscr{T}_M^*$-measurable approximation to F_Φ. Let $F^*_{\Phi,U}$ denote the empirical d.f. of $(X^u, c_n Y^u, c_n Z^u)$, $u = 1, 2, \ldots, U$. Proposition 5.2 implies that $F^*_{\Phi,U}$ is $\mathscr{T}_M^*$-measurable. Our next goal is to show that $F^*_{\Phi,U}$ is a good approximation to F_Φ in the Lévy metric.

5.2. *Goodness of the* $\mathscr{T}_M^*$*-measurable approximation.* We present a series of approximations in the Lévy metric leading from F_Φ to its $\mathscr{T}_M^*$-measurable approximation $F^*_{\Phi,U}$ defined previously. This section culminates in Lemma 5.4, which describes the accuracy of the approximation $F^*_{\Phi,U}$ of F_Φ.

We make use of the following notation scheme. The notation $\psi(\cdot)$ will represent a function which has 0 as a certain limit; we indicate variables to be held fixed in taking this limit by placing them to the right of a "|" symbol. Thus, for example, we introduce

$$\psi_0(\varepsilon \mid U) \equiv 14U^4\varepsilon|\log \varepsilon^*|^{3/2} \quad \text{where } \varepsilon^* \equiv \min\left(\frac{1}{e}, \varepsilon\right), \tag{5.2}$$

to represent the total-variation norm bound in Theorem 4.2. For fixed U, $\psi_0(\varepsilon \mid U)$ tends to 0 as $\varepsilon \to 0$.

We define $F_{\Phi,U}$ to be the empirical d.f. of the scaled triplets $(X_{\ell_u}, c_n Y_{\ell_u}, c_n Z_{\ell_u})$, $u = 1, 2, \ldots, U$. Conditional on Φ, this is the empirical d.f. of a random sample [25] from F_Φ. This is because the initial labels $\ell_1, \ell_2, \ldots, \ell_U$ form a random sample from p_n that is independent of Φ, by Theorem 3.3. We have the following result.

LEMMA 5.1. *There is a function* $\psi_1(\cdot \mid \cdot)$ *satisfying* $\lim_{U\to\infty} \psi_1(U|\varepsilon) = 0$ *for all* $\varepsilon > 0$ *such that*

$$\mathbb{P}\big(d_L(F_\Phi, F_{\Phi,U}) > \varepsilon\big) \le \psi_1(U|\varepsilon), \qquad \varepsilon > 0.$$

PROOF. The theory of Vapnik–Chervonenkis classes implies that there is a function $\psi_1(U|\varepsilon)$ satisfying the limit condition of the lemma such that, for any d.f. G on $\mathbb{R}^3$, with G_U denoting the empirical d.f. of a random sample of size U from G, we have

$$\mathbb{P}\big(d_L(G, G_U) > \varepsilon\big) \le \psi_1(U|\varepsilon).$$

See, for example, [25], Theorem 1 of Chapter 26. [In fact, we may take $\psi_1(U|\varepsilon) = A/(U\varepsilon^8)$ for a constant A.] Specializing this to $G = F_\phi$ [see (2.4)] and averaging with respect to the distribution $\pi_n(\phi)$, $\phi \in \Omega_n$, of Φ [see (2.1)] gives the desired result. □

At the next step we show that we still get a good approximation to F_Φ using $F^1_{\Phi,U}$, the empirical d.f. of the triplets $(X^u, c_n Y_{\ell_u}, c_n Z_{\ell_u})$, $u = 1, 2, \ldots, U$, in place of $F_{\Phi,U}$. The superscript "1" indicates that the first components X_{ℓ_u} of the random sample have been replaced by the approximations X^u. This lemma is based on the following observation, which is clear when the definition (5.1) of X^u is viewed in the light of Theorem 3.3. Conditional on Φ and ℓ_u, X^u is, in essence, a "sample proportion" estimator of the "population proportion" X_{ℓ_u}. Note that, unlike $F_{\Phi,U}$, $F^1_{\Phi,U}$ is not the empirical d.f. of a random sample given Φ.

LEMMA 5.2. *For all $\varepsilon > 0$ we have*

$$\mathbb{P}\big(d_{\mathrm{L}}(F_{\Phi,U}, F^1_{\Phi,U}) > \varepsilon\big) \le \frac{2}{\varepsilon^2 U}.$$

PROOF. It is elementary that

$$\begin{aligned}
&\mathbb{P}\big(d_L(F_{\Phi,U}, F^1_{\Phi,U}) > \varepsilon\big)\\
&\quad\le \mathbb{P}(\max|X^u - X_{\ell_u}| > \varepsilon)\\
&\quad\le \mathbb{P}\left(\sum_{u=1}^{U}(X^u - X_{\ell_u})^2 > \varepsilon^2\right) \le \frac{1}{\varepsilon^2}\mathbb{E}\left(\sum_{u=1}^{U}(X^u - X_{\ell_u})^2\right)\\
&\quad= \frac{1}{\varepsilon^2}\mathbb{E}\left(\sum_{u=1}^{U}\mathbb{E}\big((X^u - X_{\ell_u})^2 \mid X_{\ell_u}\big)\right)\\
&\quad= \frac{1}{\varepsilon^2}\mathbb{E}\left(\sum_{u=1}^{U}\left\{\frac{X_{\ell_u}(1 - X_{\ell_u})}{U} + \frac{(1 - X_{\ell_u})(1 - 2X_{\ell_u})}{U^2}\right\}\right) \le \frac{2}{\varepsilon^2 U}.
\end{aligned} \tag{5.3}$$

The equality on the last line is because the conditional distribution of $UX^u - 1$ given X_{ℓ_u} is binomial$(U - 1, X_{\ell_u})$. This is a consequence of Theorem 3.3. □

Now we make use of the $\mathscr{T}^*_M$-measurable approximations of cycle and trajectory length given in Section 5.1. Recall from the end of Section 5.1 that $F^*_{\Phi,U}$ denotes the empirical d.f. of $(X^u, c_n Y^u, c_n Z^u)$, $u = 1, 2, \ldots, U$. The difference between this and $F_{\Phi,U}$ is that all three components of the triplets $(X_{\ell_u}, c_n Y_{\ell_u}, c_n Z_{\ell_u})$ in the random sample of size U have been replaced by their $\mathscr{T}^*_M$-measurable approximations $(X^u, c_n Y^u, c_n Z^u)$.

LEMMA 5.3. *There exists a function ψ_2 satisfying $\lim_{x\downarrow 0}\psi_2(x|\varepsilon, U) = 0$ for all $\varepsilon > 0$ and $U = 1, 2, \ldots$, such that*

$$\mathbb{P}\big(d_L(F^1_{\Phi,U}, F^*_{\Phi,U}) > \varepsilon\big) \le \psi_2(\varepsilon_n|\varepsilon, U).$$

PROOF. Recall from Section 4.1 that $N_{u\cdot}$ is the Poisson process of rate 1 describing the times at which balls are drawn from urn u. First we point out

that $N_{u\cdot}(t)$ tends to be close to t for all urns u and all times $t \le R$. We use an argument paralleling that of (3.4). We have, for any $t_0 > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{0\le t\le R} c_n|N_{u\cdot}(t)-t| > \varepsilon\Big)\\
&\le \mathbb{P}\Big(\max_{0\le s\le t_0/c_n} |N_{u\cdot}(s)-s| > \frac{\varepsilon}{c_n}\Big) + \mathbb{P}\Big(R > \frac{t_0}{c_n}\Big)\\
&\le \frac{4t_0c_n}{\varepsilon^2} + \mathbb{P}(c_nR > t_0) \le \frac{4t_0\varepsilon_n}{\varepsilon^2} + Ue^{-t_0^2/2} + \psi_0(\varepsilon_n|U).
\end{aligned} \tag{5.4}$$

The bound on the tail probability of R here is argued as in (4.7). We also used (2.6) at the last step. Now choose t_0 so that $\exp\{-t_0^2/2\} = \varepsilon_n$, and then define the last expression to be $\psi_2'(\varepsilon_n|\varepsilon, U)$. Then ψ_2' satisfies $\lim_{x\downarrow 0}\psi_2'(x|\varepsilon, U) = 0$ and

$$\mathbb{P}\Big(\max_{0\le t\le R} c_n|N_{u\cdot}(t)-t| > \varepsilon\Big) \le \psi_2'(\varepsilon_n|\varepsilon, U).$$

An application of Boole's inequality to bound the union over u of the events on the left here leads readily to a bound showing that the increments in the $N_{u\cdot}$'s up to time $R \equiv \max R_{vv}$ are close to the corresponding time intervals:

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le u\le U} c_n|N_{u\cdot}(t) - N_{u\cdot}(s) - (t-s)| > 2\varepsilon \text{ for some } 0\le s\le t\le R\Big)\\
&\qquad \le U\psi_2'(\varepsilon_n|\varepsilon, U).
\end{aligned} \tag{5.5}$$

The following facts are direct consequences of the definitions in Section 5.1. Each Y_{ℓ_u} is a sum of at most U increments of the form $N_{v\cdot}(\mathscr{E}_v) - N_{v\cdot}(\mathscr{K}_w)$, where $v = \varphi(w)$, and each Y^u is a sum of the corresponding time intervals $\mathscr{E}_v - \mathscr{K}_w$. Similarly, each Z_{ℓ_u} is a sum of $N_{u\cdot}(\mathscr{E}_u)$ and at most $U-1$ increments of the form $N_{v\cdot}(\mathscr{E}_v) - N_{v\cdot}(\mathscr{K}_w)$, where $v = \varphi(w)$, and each Z^u is again a sum of the corresponding time intervals $\mathscr{E}_u$ and $\mathscr{E}_v - \mathscr{K}_w$. Thus (5.5) gives

$$\mathbb{P}\Big(\max_{1\le u\le U}|Y^u - Y_{\ell_u}| > 2U\varepsilon \text{ or } \max_{1\le u\le U}|Z^u - Z_{\ell_u}| > 2U\varepsilon\Big) \le U\psi_2'(\varepsilon_n|\varepsilon, U).$$

The left-hand side here is a bound on $\mathbb{P}(d_L(F^1_{\Phi,U}, F^*_{\Phi,U}) > 2U\varepsilon)$. The conclusion of the lemma follows now with this definition of ψ_2: $\psi_2(x|\varepsilon, U) \equiv U\psi_2'(x|\varepsilon/(2U), U)$. □

By applying the triangle inequality to combine Lemmas 5.1, 5.2 and 5.3, we can summarize this section as *following result*.

LEMMA 5.4. *We have*

$$\mathbb{P}\big(d_L(F_\Phi, F^*_{\Phi,U}) > 3\varepsilon\big) \le \psi_1(U|\varepsilon) + \frac{2}{\varepsilon^2 U} + \psi_2(\varepsilon_n|\varepsilon, U).$$

5.3. *Invoking the Poisson approximation.* First we quote an elementary result bounding the Prohorov distance in terms of a "coupling."

LEMMA 5.5. *Let H_i, $i = 1, 2$, be random variables taking values in a separable metric space $\mathscr{X}$ with metric d, and let η_i, $i = 1, 2$, be their laws. Then for all $\varepsilon > 0$ we have*

$$d_{\mathrm{Pr}}(\eta_1, \eta_2) \le \varepsilon + \mathbb{P}\big(d(H_1, H_2) > \varepsilon\big).$$

PROOF. Let $\varepsilon' \equiv \varepsilon + \mathbb{P}(d(H_1, H_2) > \varepsilon) \ge \varepsilon$. Then we have $\mathbb{P}(d(H_1, H_2) > \varepsilon') \le \mathbb{P}(d(H_1, H_2) > \varepsilon) \le \varepsilon'$, and the conclusion follows from Theorem 3.1.2 of [17]. □

The random d.f. $F^*_{\Phi,U} \in \mathscr{D}_3$ is measurable with respect to the σ-field $\mathscr{F}^*_M$. We write $\mu^*_{n,U}$ for its law on $\mathscr{D}_3$. Since the law ν_n is defined on $\mathscr{F}^*_M$, the measurable mapping $F^*_{\Phi,U}$ from $(\Omega, \mathscr{F}^*_M, \nu_n)$ to $\mathscr{D}_3$ induces a law on $\mathscr{D}_3$ which we denote by $\mu_{\langle U\rangle}$: $\mu_{\langle U\rangle}(G) = \nu_n(F^*_{\Phi,U} \in G)$ for $G \in \mathbb{B}(\mathscr{D}_3)$. This is an approximation to the law $\mu^*_{n,U}$. An important point is that *the law $\mu_{\langle U\rangle}$ does not depend on n*. To see this, first note that working with the *scaled* triplets $(X^u, c_n Y^u, c_n Z^u)$ in Section 5.1 is equivalent to replacing the point processes M_{uv} by their time-scaled versions $\widetilde{M}_{uv}$ defined at the end of section 4.3. But under ν_n the $\widetilde{M}_{uv}$'s are independent Poisson processes of rate 1 on Δ, and this law clearly does not depend on n.

The following result is immediate now from Theorem 4.2.

LEMMA 5.6. *We have*

$$d_{\mathrm{TV}}\big(\mu^*_{n,U}, \mu_{\langle U\rangle}\big) \le \psi_0(\varepsilon_n|U) \equiv 14U^4\varepsilon_n|\log \varepsilon^*_n|^{3/2}.$$

PROOF OF THEOREM 2.1. Using Lemma 5.5 to turn Lemma 5.4 into a statement about the Prohorov distance between laws on $\mathscr{D}_3$, we obtain, for all $\varepsilon > 0$ and $U = 1, 2, \ldots,$

$$d_{\mathrm{Pr}}\big(\mu_n, \mu^*_{n,U}\big) \le 3\varepsilon + \psi_1(U|\varepsilon) + \frac{2}{\varepsilon^2 U} + \psi_2(\varepsilon_n|\varepsilon, U).$$

Using this with Lemma 5.6 and the triangle inequality, we get

$$\begin{aligned} d_{\mathrm{Pr}}\big(\mu_n, \mu_{\langle U\rangle}\big) &\le d_{\mathrm{Pr}}\big(\mu_n, \mu^*_{n,U}\big) + d_{\mathrm{Pr}}\big(\mu^*_{n,U}, \mu_{\langle U\rangle}\big) \\ &\le 3\varepsilon + \psi_1(U|\varepsilon) + \frac{2}{\varepsilon^2 U} + \psi_2(\varepsilon_n|\varepsilon, U) + \psi_0(\varepsilon_n|U). \end{aligned} \tag{5.6}$$

Here we have used the fact that the Prohorov metric is dominated by the total variation distance. It follows that μ_n, $n = 1, 2, \ldots,$ is a Cauchy sequence. This is because, for any given $\delta > 0$, we can make the right-hand side here smaller than this δ by first choosing $\varepsilon = \varepsilon_0 = \delta/30$, and then choosing for U a fixed value U_0 so large that both $\psi_1(U_0 \mid \varepsilon_0) < \delta/10$ and $2/(\varepsilon_0^2 U_0) < \delta/10$ hold. Having chosen $\varepsilon = \varepsilon_0$ and $U = U_0$, we see that *for all sufficiently large n* each of the last two terms on the right-hand side, $\psi_2(\varepsilon_n \mid \varepsilon_0, U_0)$ and $\psi_0(\varepsilon_n \mid U_0)$,

are less than $\delta/10$, by the homogeneous limit condition $\lim_{n\to\infty}\varepsilon_n = 0$, (2.5). It then follows that for all sufficiently large n and n' we have

$$d_{\mathrm{Pr}}(\mu_n, \mu_{n'}) \le d_{\mathrm{Pr}}(\mu_n, \mu_{\langle U_0\rangle}) + d_{\mathrm{Pr}}(\mu_{n'}, \mu_{\langle U_0\rangle}) \le \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Thus μ_n is a Cauchy sequence as claimed. As the space $\mathscr{P}\mathscr{M}(\mathscr{D}_3)$ of probability measures on $\mathscr{D}_3$ is complete under the Prohorov metric, these have a limit μ. This completes the proof. □

6. Concluding remarks.

6.1. *Relationships with analysis of discretizations.* First we explain in a little more detail how Theorem 2.1 relates to the analysis of discretizations of dynamical systems with chaotic behavior. We consider a transformation $f:[0,1] \to [0,1]$ with absolutely continuous Sinai–Ruelle–Bowen invariant measure χ_f. As in Section 1, we write f_n for the discretization $f_n(i) = \lceil nf(i/n)\rceil_+$, which is a transformation on $\{1, 2, \ldots, n\}$. We consider the simplified triplets

$$\big(\#\mathrm{Co}(i, f_n), \#\mathrm{Cy}(i, f_n), \#\mathrm{Tr}(i, f_n)\big), \qquad i \in \{1, 2, \ldots, n\}. \tag{6.1}$$

These differ from the triplets of (2.2) in that component size is now also measured by cardinality (#). We use a superscript # in the following discussion whenever component size is measured by cardinality. For given f and n we consider the distribution function $F^{\#}_{f_n}$ of these triplets when the label i is selected at random from $\{1, 2, \ldots, n\}$. We are interested in the ensemble of d.f.s $F^{\#}_{f_n}$, $n = n_0 + 1, n_0 + 2, \ldots, n_0 + n_1$, for large n_0 and n_1.

The most natural way to approximate an ensemble of discretizations of f by a random mapping Φ is the following. We take the distribution p_n of images of Φ to be a discretization of the invariant measure χ_f of f:

$$p_{ni} \equiv \chi_f\big(((i-1)/n, i/n]\big). \tag{6.2}$$

The following hypothesis expresses the expected relationship between the random mapping Φ and the discretizations of f.

HYPOTHESIS 1. *For large n_0, n_1 the empirical law of $F^{\#}_{f_n}$, $n = n_0 + 1, n_0 + 2, \ldots, n_0 + n_1$, is close to the law of the random d.f. $F^{\#}_{\Phi}$ of the simplified triplets associated with the random mapping Φ with the preceding choice p_n of the distribution of images.*

See [14], [18] and [21] for detailed discussion. We do not know how to prove such a hypothesis, nor can we even test it empirically without detailed knowledge of the measure χ_f. Consider, however, the case where χ_f has a density χ_f' bounded by a constant $K < \infty$. Then the distribution of images p_n satisfies the inequalities $p_{ni} \le K/n$, and so, since we always have $c_n \ge 1/\sqrt{n}$ by Jensen's inequality,

$$\frac{p_{ni}}{c_n} \le \frac{K}{\sqrt{n}}.$$

Thus the homogeneous limit condition (2.5) holds and Theorem 2.1 is applicable to F_Φ. Furthermore, we have here that

$$c_n \approx \frac{\theta}{\sqrt{n}} \quad \text{with} \quad \theta = \int_0^1 \big(\chi_f'(t)\big)^2 dt.$$

The preceding hypothesis in conjunction with Theorem 2.1 then gives rise to the following hypothesis.

HYPOTHESIS 2. *There exists a positive constant θ such that for large n_0 and n_1 the empirical law of the sequence of d.f.'s $F^{\#}_{f_n,\theta}$, $n = n_0+1, n_0+2, \dots, n_0+n_1$, of the scaled triplets*

$$\left(\frac{\#\mathrm{Co}(i, f_n)}{n}, \theta\frac{\#\mathrm{Cy}(i, f_n)}{\sqrt{n}}, \theta\frac{\#\mathrm{Tr}(i, f_n)}{\sqrt{n}}\right), \qquad i \in \{1, 2, \dots, n\},$$

is close to the law μ of Theorem 2.1.

[We ignore the technicality that component size in the simplified triplets (6.1) is measured using the uniform distribution rather than p_n as in Theorem 2.1.] This hypothesis may be viewed as a more precise form of the italicized principle stated in the first paragraph of Section 1. It is much more convenient to work with than Hypothesis 1, because all of the details of f and χ_f have disappeared in the limit, except for the scalar parameter θ. Hypothesis 2 is easily tested in numerical experiments. Figure 2 in Section 1 presents such a test, and lends support to the hypothesis by exhibiting the proximity of $F_\alpha(\cdot\,|H_1)$ and F^∞_α.

6.2. *Remark on mappings with attracting centers.* To understand the role of the homogeneity condition (2.5) of Theorem 2.1, we consider the situation where all weights p_{ni} but one are the same. Specifically, we suppose that there is a constant β such that, for all n,

$$\frac{p_{n1}}{c_n} = \frac{\beta}{\sqrt{\beta^2+1}}, \qquad \frac{p_{ni}}{c_n} = \frac{1}{\sqrt{\beta^2+1}\sqrt{n-1}}, \quad i = 2, 3, \dots, n.$$

Thus, in a sense, the homogeneous limit condition (2.5) holds for all balls but the first. In this case Φ is the classical mapping with a single attracting center [27]. Certain limiting properties of the corresponding random d.f. F_Φ are well known and these are quite different from those of the completely random mapping. The methods of this paper are applicable to random mappings with attracting centers. We will devote a future paper to detailed analysis of this situation. In the meantime, let us quote one result.

THEOREM 6.1. *Suppose that for each $n = 1, 2, \dots$ we have a probability measure p_n on $\{1, 2, \dots, n\}$. Suppose further that*

$$\lim_{n\to\infty} \frac{p_{n1}}{c_n} = a < \infty \quad \textit{and} \quad \lim_{n\to\infty} \max_{i>1} \frac{p_{ni}}{c_n} = 0.$$

Let Φ be the random transformation with distribution of images p_n, and let μ_n denote the law of the random d.f. F_Φ of the scaled triplets defined by (2.4). *Then there exists a probability measure $\mu^{(a)}$ over $\mathscr{D}^3$ such that μ_n converges weakly to $\mu^{(a)}$ as $n \to \infty$. The law $\mu^{(a)}$ depends on the p_n's only through a.*

This may be proved using Theorem 4.2 to approximate the matches involving labels 2, 3, ..., n, and then noting that, in the urn model, draws of label 1 form a Poisson process independent of draws of the other labels.

6.3. *Relationships with analysis of discretizations: continuation.* We now discuss some experimentation showing that the completely random mapping model fails to capture the behavior of discretizations of certain interesting dynamical systems. This is a situation in which results along the lines of Theorem 6.1 could be useful.

Let us consider the transformation $f^{(\ell)}(x) = 1 - |2x-1|^\ell,\ x \in [0,1]$, where $\ell > 2$. (See [13] for a study of discretizations of this. The case $\ell = 2$ is the logistic mapping discussed in Section 1.) In this case there exists an absolutely continuous Sinai–Ruelle–Bowen invariant measure $\chi_{(\ell)}$, but its density is unbounded at the endpoints of the interval. In fact, there is a positive θ_0 for which

$$\frac{\chi_{(\ell)}[0,x]}{x^{1/\ell}} \to \theta_0 \quad \text{and} \quad \frac{\chi_{(\ell)}[1-x,1]}{x^{1/\ell}} \to \theta_0 \quad \text{as} \quad x \to 0.$$

Using (6.2) for the distribution of images p_n of a random transformation approximating the behavior of an ensemble of discretizations of f, we have

$$p_{ni} \approx \frac{\theta_0 i^{1/\ell - 1}}{\ell n^{1/\ell}}, \qquad p_{n,n-i} \approx \frac{\theta_0 i^{1/\ell - 1}}{\ell n^{1/\ell}}, \qquad c_n = \|p_n\|_2 \approx \frac{\theta_1}{n^{1/\ell}}$$

for some $\theta_1 > 0$. Thus in this case $p_{ni} = O(c_n)$ and $p_{n,n-i} = O(c_n)$ as $n \to \infty$, and so, in the terminology of the previous section, we might say loosely that every ball is an attracting center. Numerical experiments show again that the statistical behavior of the d.f.'s of scaled triplets over ensembles of discretizations of $f^{(\ell)}$ is inconsistent with predictions based on the theory of completely random mappings. Let us denote by H_ℓ the ensemble of discretizations $f_n^{(\ell)}$, $n = 10^6+1, 10^6+2, \ldots, 10^6+10^3$. Figure 5 graphs the distribution functions $F_\alpha(\cdot|H_\ell)$ [of the same parameter $\alpha(\cdot)$ treated in Section 1] for a selection of values of $\ell > 2$ against the prediction F_α^∞ based on a completely random mapping. In sharp contrast to Figure 2, all d.f.'s are quite distinct. [The behavior of $F_\alpha(\cdot\,|H_\ell)$ as ℓ increases is a little complex. At most argument values, if not all, these d.f.'s increase as ℓ increases from 2 to about 2.5. Thereafter they typically decrease in ℓ. These observations are based on much careful experimentation.]

6.4. *Relationship with the continuum random tree.* While the triplets defined in Section 2 are indeed the most important descriptors of a random transformation when we are not interested in the individual identities of the

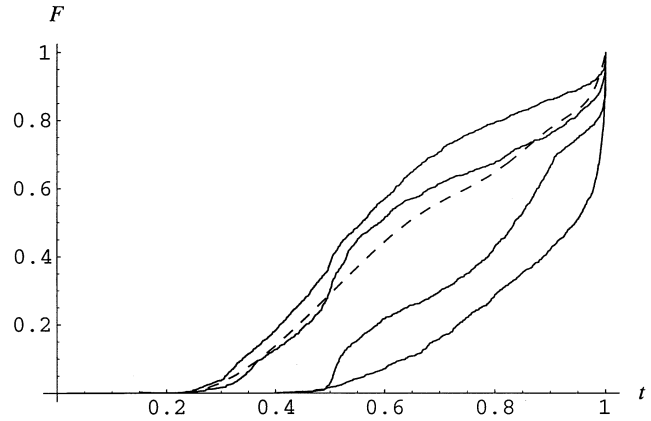

FIG. 5. *The distribution function* F_α^∞ (*dashed line*) *and the empirical distribution functions* $F_\alpha(\cdot\,|H_{2.5})$, $F_\alpha(\cdot\,|H_3)$, $F_\alpha(\cdot\,|H_4)$ *and* $F_\alpha(\cdot\,|H_5)$, *which appear from top to bottom.*

elements of the domain, there are many other quantities of interest. For example, when the cycles are removed, the graph Φ becomes a forest [26], and we may ask about the distribution of the sizes of the trees in this forest. A complete picture of what can be achieved in this direction requires an embedding of the entire graph Φ in a large space in such a way that the induced laws are tight. Aldous has shown how to do this for a natural family of random trees. The connection with Aldous's *compact continuum random tree* (see [2] and [3], and further references in the latter) is very close. Indeed, we believe that the trees in the forest described previously converge to the compact continuum random tree under the homogeneous limit condition (2.5). To indicate the connection, the Poisson processes on the triangle Δ, which were so essential in Sections 4.2 and 4.3, capture the same "dynamics" as Aldous's inhomogeneous Poisson process with rate t and uniformly distributed marks (see Section 4 of [2]). The random mapping is not a tree, but there is an elegant way to construct a sequence of approximations to the limit process using the Poisson law ν_n of Section 4.2 that is analogous to Aldous's presentation of the compact continuum tree. This is to be the subject of a future paper.

Acknowledgments. This work was begun while the authors were visiting University College, Cork. They express their gratitude to their host departments at UCC, Statistics and Physics. The references to Boole's inequality in the text are another salute to UCC, where George Boole was the first Professor of Mathematics. The authors thank S. Lalley for several useful comments and references.

REFERENCES

[1] ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic.* Springer, New York.

[2] ALDOUS, D. (1991). The continuum random tree I. *Ann. Probab.* **19** 1–28.

[3] ALDOUS, D. (1993). The continuum random tree III. *Ann Probab.* **21** 248–289.
[4] ALLIGOOD, K. T., SAUER, T. D. and YORKE, J. A. (1997). *Chaos: An Introduction to Dynamical Systems.* Springer, New York.
[5] BARBOUR, A. D., HOLST, L. and Janson, S. (1992). *Poisson Approximation.* Clarendon, Oxford.
[6] BECK, C. (1989). Scaling behavior of random maps. *Phys. Lett. A* **136** 121–125l.
[7] BOLLOBAS, B. (1985). *Random Graphs.* Academic Press, London.
[8] BURTIN, Y. D. (1980). On a simple formula for random mappings and its applications. *J. Appl. Probab.* **17** 403–414.
[9] DALEY, D. J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes.* Springer, New York.
[10] DEVANEY, R. L. (1990). Chaos, Fractals and Dynamics: *Computer Experiments in Mathematics.* Addison-Wesley, Menlo Park, CA.
[11] DIACONIS, P. MCGRATH, M. and PITMAN, J. (1995). Riffle shuffles, cycles, and descents. *Combinatorica* **15** 11–20.
[12] DIAMOND, P., KLEMM, A., Kloeden, P. and POKROVSKII, A. (1996). Basin of attraction of cycles of discretizations of dynamical systems with SRB invariant measures. *J. Statist. Phys.* **84** 713–733.
[13] DIAMOND, P., KLOEDEN, P., POKROVSKII, A. and SUZUKI, M. (1996). Statistical properties of discretizations of a class of chaotic dynamical systems. *Comput. Math. Appl.* **31** 83–95.
[14] DIAMOND, P., KLOEDEN, P., POKROVSKII, A. and VLADIMIROV, A. (1995). Collapsing effects in numerical simulation of a class of chaotic dynamical systems and random mappings with a single attracting centre. *Phys. D* **86** 559–571.
[15] DI MELO, W. and VAN STRIEN, S. (1993). *One-Dimensional Dynamics.* Springer, Berlin.
[16] DYSON, F. J. and FALK, H. (1992). Period of a discrete cat mapping. *Amer. Math. Monthly* **99** 603–614.
[17] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.
[18] GREBOGI, C., OTT, E. and YORKE, J. A. (1988). Roundoff-induced periodicity and the correlation dimension of chaotic attractors. *Phys. Rev. A* **34** 3688–3692.
[19] HARRIS, B. (1960). Probability distributions related to random mappings. *Ann. Math. Statist* **31** 1045–1062.
[20] LALLEY, S. (1996). Cycle structure of riffle shuffles. *Ann. Probab.* **24** 49–73.
[21] LEVY, Y. E. (1982). Some remarks about computer studies of dynamical systems. *Phys. Lett. A* **88** 1–3.
[22] OTT, E. (1993). *Chaos in Dynamical Systems.* Cambridge Univ. Press.
[23] RUELLE, D. (1989). *Elements of Differentiable Dynamics and Bifurcation Theory.* Academic Press, Boston.
[24] SHEPP, L. A. and LLOYD, S. P. (1966). Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.* **121** 340–357.
[25] SHORAK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications.* Wiley, New York.
[26] STEPANOV, V. E. (1969). Limit distributions of certain characteristics of random mappings. *Theory Probab. Appl.* **14** 612–626.
[27] STEPANOV, V. E. (1971). Random mappings with a single attracting centre. *Theory Probab. Appl.* **16** 155–161.

SCHOOL OF INDUSTRIAL ENGINEERING
GRISSOM HALL 1287
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907-1287
E-MAIL: colm@ecn.purdue.edu

INSTITUTE FOR TRANSMISSION PROBLEMS
RUSSIAN ACADEMY OF SCIENCE
19 BOLSHOI KARETNY LANE
MOSCOW 101447
RUSSIA