Check for updates

J. Appl. Probab. **62,** 1340–1359 (2025)
*doi:*10.1017/jpr.2025.3

COUPLING RESULTS AND MARKOVIAN STRUCTURES FOR NUMBER REPRESENTATIONS OF CONTINUOUS RANDOM VARIABLES

JESPER MØLLER,* *Aalborg University*

Abstract

Consider nested subdivisions of a bounded real set into intervals defining the digits $X_1, X_2, \ldots$ of a random variable X with a probability density function f. If f is almost everywhere lower semi-continuous, there is a non-negative integer-valued random variable N such that the distribution of $R = (X_{N+1}, X_{N+2}, \ldots)$ conditioned on $S = (X_1, \ldots, X_N)$ does not depend on f. If also the lengths of the intervals exhibit a Markovian structure, $R \mid S$ becomes a Markov chain of a certain order $s \geq 0$. If $s = 0$ then $X_{N+1}, X_{N+2}, \ldots$ are independent and identically distributed with a known distribution. When $s > 0$ and the Markov chain is uniformly geometric ergodic, there is a random time M such that the chain after time $\max\{N, s\} + M - s$ is stationary and M follows a simple known distribution.

Keywords: Approximation; beta-expansion; continued fraction; dynamical system; Lüroth series; Markov chain; nested partitions; pseudo golden ratio; read-once algorithm; sufficient digits

2010 Mathematics Subject Classification: Primary 60G07
Secondary 60G99

1. Introduction

Let X be a continuous random variable with digit expansion $X = 0.X_1X_2\ldots$ in the usual decimal numeral system (base-10) or another numeral system as described below. This paper shows that under very mild conditions, a non-negative integer-valued random variable N exists such that $S = (X_1, \ldots, X_N)$ specifies the 'sufficient digits' in the sense that the distribution of $R = (X_{N+1}, X_{N+2}, \ldots)$ conditioned on S does not depend on the distribution of X. Under a further condition ensuring a Markovian structure of the numeral system, which is satisfied for base-10 and many other numeral systems, it is also shown that $R \mid S$ is a Markov chain of a certain order $s \geq 0$. When $s = 0$, which, e.g., is satisfied for base-10, $X_{N+1}, X_{N+2}, \ldots$ are shown to be independent and identically distributed (i.i.d.) with a known distribution. When $s > 0$ and the Markov chain is uniformly geometric ergodic, a coupling is established between (X, N) and a random time M such that the chain after time $\max\{N, s\} + M - s$ is stationary and M follows a simple known distribution. Before further presenting these results and the outline of this paper, the background material on numeral systems presented in Section 1.1 is needed. The importance of the results is wide ranging, as discussed in Section 5.

Received 30 April 2024; accepted 9 January 2025.
* Postal address: Department of Mathematical Sciences, Aalborg University, Skjernvej 4A, 9220 Aalborg, Denmark. Email: jm@math.aau.dk

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

1.1. Nested subdivisions

The general setting of nested subdivisions described in this section is considered throughout the text. Specific constructions and examples are given in Section 2.

Let $\mathcal{X}$ be a finite or countable set, and denote by $\mathbb{R}$, $\mathbb{N}$, and $\mathbb{N}_0$ the sets of real numbers, positive integers, and non-negative integers, respectively, and by $\mathcal{X}^{\mathbb{N}}$ the set of all sequences $(x_1, x_2, \dots)$ with elements in $\mathcal{X}$.

Suppose $I = I_\emptyset \subset \mathbb{R}$ is a Borel set of Lebesgue measure $\ell_\emptyset = |I| \in (0, \infty)$ and for every $n \in \mathbb{N}$ there is a subdivision (at level n)

$$I = \bigcup_{(x_1,\dots,x_n)\in\mathcal{X}^n} I_{x_1,\dots,x_n} \bigcup D_n \tag{1.1}$$

such that the following properties are satisfied. The $I_{x_1,\dots,x_n}$s and D_n are pairwise disjoint sets. Each I_k with $k \in \mathcal{X}$ is an open non-empty interval, and for $n > 1$, $I_{x_1,\dots,x_n}$ is an open and possibly empty interval such that $I_{x_1,\dots,x_n} \subseteq I_{x_1,\dots,x_{n-1}}$. Each D_n is a subset of the endpoints of the non-empty $I_{x_1,\dots,x_n}$s, and so $D_n \subseteq D_{n+1}$.

Some comments, further notation, and one more assumption are in order.

For illustrative purposes, in the usual decimal numeral system ('base-10 case'), $\mathcal{X} = \{0, 1, \dots, 9\}$, $I = [0, 1)$, and for every $n \in \mathbb{N}$ and $(x_1, \dots, x_n) \in \{0, 1, \dots, 9\}^n$, $I_{x_1,\dots,x_n}$ has left and right endpoints $\sum_{i=1}^n x_i 10^{-i}$ and $10^{-n} + \sum_{i=1}^n x_i 10^{-i}$, respectively, and D_n is the set of all such left endpoints. This example can be much generalized, as discussed in Examples 2.1 and 2.2. Continuous fractions (see Example 2.3) provide an example where $\mathcal{X}$ is countably infinite and non-equidistant subdivisions are used. Furthermore, considering non-equidistant subdivisions will be of importance for statistical models, as discussed in Section 5.1.

When $n \geq 2$, $I_{x_1,\dots,x_n}$ is allowed to be empty, since restrictions on the digits may be needed in some cases (e.g. Example 2.2).

Let $D = \cup_{n=1}^\infty D_n$ be the set of all endpoints. Since each D_n is finite or countable, D is finite or countable, and so $|D| = 0$.

It is only for convenience that each interval $I_{x_1,\dots,x_n}$ is assumed to be open. The assumption is of no importance for the probabilistic results in this paper, since $|D| = 0$ and all considered probability distributions on I will be assumed to be absolutely continuous. Therefore, ignoring Lebesgue nullsets, the partitions given by (1.1) for $n = 1, 2, \dots$ are said to be nested subdivisions of I.

Let $J = I \setminus D$. Then $|I| = |J|$ and

$$J = \bigcap_{n=1}^\infty \bigcup_{(x_1,\dots,x_n)\in\mathcal{X}^n} I_{x_1,\dots,x_n} \setminus D. \tag{1.2}$$

For every $(x_1, x_2, \dots) \in \mathcal{X}^{\mathbb{N}}$ and $n \in \mathbb{N}$, let $\ell_{x_1,\dots,x_n} = |I_{x_1,\dots,x_n}|$ be the length of $I_{x_1,\dots,x_n}$ and assume that

$$\ell_{x_1,\dots,x_n} \to 0 \quad \text{as } n \to \infty. \tag{1.3}$$

Denote by $a_{x_1,\dots,x_n}$ the left endpoint of $I_{x_1,\dots,x_n}$. For $x \in I_{x_1,\dots,x_n}$, the error $e_n(x)$ and the relative error $u_n(x)$ of the approximation $x \approx a_{x_1,\dots,x_n}$ as compared to $\ell_{x_1,\dots,x_n}$ are defined by $e_n(x) = x - a_{x_1,\dots,x_n}$, and $u_n(x) = e_n(x)/\ell_{x_1,\dots,x_n}$. By (1.3), $e_n(x) \to 0$ as $n \to \infty$. If $x \in J$ then $0 < u_n(x) < 1$.

Define the digits (or coefficients or symbols) of every $x \in J$ as the unique sequence $(x_1, x_2, \dots) \in \mathcal{X}^{\mathbb{N}}$ such that $x \in I_{x_1,\dots,x_n}$ for $n = 1, 2, \dots$ The assumption in (1.3) means that

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

$x = \lim_{n\to\infty} a_{x_1,\dots,x_n}$ is determined by its digits. Express this one-to-one correspondence between $x \in J$ and its digits by writing

$$x = .x_1x_2\dots \quad \text{for } x \in J. \tag{1.4}$$

This notation agrees with the usual digit representations used for the example of the base-10 case.

For $n \in \mathbb{N}_0$ and $x \in J$, call $x^{[n]} = .x_{n+1}x_{n+2}\dots$ and $x_{>n} = (x_{n+1}, x_{n+2}, \dots)$ the nth (scaled) remainder of x and the nth tail of x, respectively, where $x^{[0]} = x$. Let $J_{>0} = \{x_{>0} \mid x \in J\}$. Then, for any fixed $n \in \mathbb{N}_0$, there is a one-to-one correspondence between $x^{[n]} \in J$ and $x_{>n} \in J_{>0}$.

1.2. Objective and main results

1.2.1. *Overview.* Let X be a random variable with values in J and a probability density function (PDF) f concentrated on J. Denote its digits by $X = 0.X_1X_2\dots$, cf. (1.4). Then $(X_1, X_2, \dots)$ is a stochastic process with state space $\mathcal{X}$. For $n \in \mathbb{N}$, writing $A_n = a_{X_1,\dots,X_n}$ and $L_n = \ell_{X_1,\dots,X_n}$, by (1.2) and (1.3), $A_n < X < A_n + L_n$ and $L_n \to 0$ as $n \to \infty$. The aim of this paper is to show how coupling methods can be used for establishing a number of interesting results for the distribution of

$$X^{[n]} = .X_{n+1}X_{n+2}\dots \quad \text{or, equivalently,} \quad X_{>n} = (X_{n+1}, X_{n+2}, \dots); \tag{1.5}$$

the distributions of

$$E_n = e_n(X) = X - A_n \quad \text{and} \quad U_n = u_n(X) = E_n/L_n; \tag{1.6}$$

and the distributions of further stochastic variables/processes obtained when $n \in \mathbb{N}_0$ in (1.5) and (1.6) is replaced by a random time obtained from one of the various coupling constructions studied in detail in Sections 3 and 4. These distributions are of interest when considering number representations of X as in Section 2.3, as well as many other examples (see [6, 7, 12, 15, 24] and the references therein). The main results in this paper extend those included in [15] (in which only base-q expansions with $q \in \{2, 3, \dots\}$ are considered), however the proofs are shorter and different. In particular, under weak conditions the existence of a random variable $N \in \mathbb{N}_0$ will be established so that the distribution of the residual process

$$R = (R_1, R_2, \dots) = (X_{N+1}, X_{N+2}, \dots) = X_{>N} \tag{1.7}$$

conditioned on

$$S = (X_1, \dots, X_N) \tag{1.8}$$

does not depend on the distribution of X. Surprisingly, this distribution depends only on ratios of the lengths of intervals in the nested subdivisions, and often a Markovian structure appears as discussed in Section 1.2.2. Moreover, the results have important applications for statistical methodology and metric number theory, as discussed in Section 5.

1.2.2. *Details.* The weak condition we give in (1.9) is used when establishing any result in Sections 3 and 4. Recall that lower semi-continuity (LSC) of f at a point $x \in I$ means $\liminf_{x_n\to x} f(x_n) \geq f(x)$ for any sequence of points x_n converging to x. Commonly used PDFs are LSC almost everywhere. Consider the condition that

$$f \text{ is LSC on a Borel set } H \subseteq J, \text{ where } |J \setminus H| = 0 \tag{1.9}$$

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

(hence $|I \setminus H| = 0$). This is indeed a very mild condition; for an example where it is violated, see [13, Remark 1]. The set H is just introduced for generality since a PDF is only unique up to a Lebesgue nullset. The choice of H is discussed in Remark 3.1.

Assuming (1.9), there is a useful coupling between X and a non-negative integer-valued random variable N. The coupling construction is detailed in Theorem 3.1, which is summarized below using the following notation. If $n = 0$, interpret $(x_1, \dots, x_n)$ as the 'empty sequence of digits', denoted $\emptyset$, and set $\mathcal{X}^0 = \{\emptyset\}$ (this notation is similar to the use of $\emptyset$ for the empty point configuration in connection with finite point processes [21]). Then S defined by (1.8) becomes a discrete random variable with state space $\cup_{n=0}^{\infty} \mathcal{X}^n$, the set of all finite sequences of digits including the case of no digits. For $\ell > 0$, denote by Unif$(0, \ell)$ and Unif(I) the uniform distribution on $(0, \ell)$ and I, respectively. For mathematical convenience, introduce two random variables E_0 and U_0 such that, conditioned on $N = 0$, $S \perp\!\!\!\perp E_0 \sim \text{Unif}(0, \ell_\emptyset)$ and $S \perp\!\!\!\perp U_0 \sim \text{Unif}(0, 1)$, where $\perp\!\!\!\perp$ means independence. Define

$$E = E_N, \quad L = \begin{cases} \ell_N & \text{if } N > 0, \\ \ell_\emptyset & \text{if } N = 0, \end{cases} \quad \text{and} \quad U = U_N = E/L. \tag{1.10}$$

Then (X, N) and (S, U) are in a one-to-one correspondence.

Among other things, Theorem 3.1 states that $S \perp\!\!\!\perp U \sim \text{Unif}(0, 1)$; conditioned on L, we have $S \perp\!\!\!\perp E \sim \text{Unif}(0, L)$; the conditional distribution of $R \mid S$ does not depend on f but is expressible in terms of ratios between the lengths of the intervals in the nested subdivisions. These results extend those in [15], which only considered the special case of base-q expansions (with $q > 1$ an integer, cf. Example 2.1; see also [4, 5]).

The results in Section 3 are expanded in Section 4, assuming the lengths of the intervals in the nested subdivisions of I exhibit the following structure. Set $0/0 = 0$ and, for a given $s \in \mathbb{N}_0$, assume that, for any integer $n \geq s + 2$ and every $(x_1, \dots, x_n) \in \mathcal{X}^n$, the ratio $\ell_{x_1,\dots,x_n}/\ell_{x_1,\dots,x_{n-1}}$ depends only on $(x_{n-s}, \dots, x_n)$ and is written as

$$\ell_{x_1,\dots,x_n}/\ell_{x_1,\dots,x_{n-1}} = p_{x_{n-s},\dots,x_n}. \tag{1.11}$$

Then the lengths of the intervals are said to exhibit a Markovian structure of order s because $R \mid S$ turns out to be a Markov chain of order s, as explained below.

The special case $s = 0$ is treated in Section 3.2, where it is shown that $X_1, X_2, \dots$ are i.i.d. with a known distribution. Suppose $s > 0$. Define

$$\Omega = \{(x_1, \dots, x_s) \in \mathcal{X}^s \mid \ell_{x_1,\dots,x_s} > 0\} \tag{1.12}$$

and a matrix P with entries indexed by $((x_1, \dots, x_s), (y_1, \dots, y_s)) \in \Omega \times \Omega$ such that

$$P_{(x_1,\dots,x_s),(y_1,\dots,y_s)} = \begin{cases} p_{x_1,\dots,x_s,y_s} & \text{if } y_i = x_{i+1} \text{ for } 1 \leq i < s, \\ 0 & \text{otherwise.} \end{cases} \tag{1.13}$$

For $i \in \mathbb{N}$, let $X_i^{(s)} = (X_i, X_{i+1}, \dots, X_{i+s-1})$ and $R_i^{(s)} = X_{N+i}^{(s)}$, and for $m \in \mathbb{N}_0$ let

$$X^{(s)} = \big(X_1^{(s)}, X_2^{(s)}, \dots\big), \quad X_{>m}^{(s)} = \big(X_{m+1}^{(s)}, X_{m+1}^{(s)}, \dots\big), \tag{1.14}$$

$$R^{(s)} = \big(R_1^{(s)}, R_2^{(s)}, \dots\big), \quad R_{>m}^{(s)} = \big(R_{m+1}^{(s)}, R_{m+2}^{(s)}, \dots\big), \tag{1.15}$$

so $R_{>m}^{(s)} = X_{>N+m}^{(s)}$. It is shown in Section 4.1 that $R^{(s)} \mid S$ is a Markov chain with transition matrix P. In particular, if the chain $R^{(s)} \mid S$ is uniformly geometric ergodic, a coupling of (X, N)

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

with a random time M is established in Section 4.2 such that $X^{(s)}_{>\max\{N,s\}+M-s}$ is stationary, and M is independent of the chain and follows a simple known distribution. This coupling construction is based on the read-once algorithm for perfect simulation developed by Wilson [26].

1.3. Outline

Section 2 discusses various procedures for constructing nested subdivisions and presents examples used later on. The coupling construction between X and N and related results are given in Section 3. Under various conditions, Section 4 shows that $R^{(s)} \mid S$ is a Markov chain, specifies the coupling between (X, N) and M, and establishes related results. Section 5 concludes with a brief discussion of the importance of the results and some open problems.

2. Procedures for nested subdivisions

Assuming the first subdivision of I into the intervals $I_k = (a_k, a_k + \ell_k)$ with $k \in \mathcal{X}$ is given, we explain in Section 2.1 how probability distributions and in Section 2.2 how dynamical systems may be used to generate nested subdivisions of I. Various illustrative examples are presented in Section 2.3, including generalized Lüroth series, β-expansions, and continued-fraction representations.

2.1. Markovian constructions

Given the subdivision of I into the intervals $I_k = (a_k, a_k + \ell_k)$ with $k \in \mathcal{X}$, an integer $s \geq 0$, and a probability mass function (PMF) on $\mathcal{X}^m$ with $m = s + 1$, the following procedure is in accordance with the Markovian structure in (1.11). Consider first the case $s = 0$ and define a PMF by $p_k = \ell_k/\ell_\emptyset$ for all $k \in \mathcal{X}$. In accordance with (1.11), for each $(x_1, \ldots, x_n) \in \mathcal{X}^n$ with $n \in \mathbb{N}$, let $\ell_{x_1,\ldots,x_n} = p_{x_1} \cdots p_{x_n} \ell_\emptyset$. Recursively, for $n = 1, 2, \ldots$, assuming $a_{x_1,\ldots,x_n}$ has been specified, let the subdivision of $I_{x_1,\ldots,x_n}$ into the intervals $I_{x_1,\ldots,x_n,k}$ with $k \in \mathcal{X}$ be such that $a_{x_1,\ldots,x_n,j} < a_{x_1,\ldots,x_n,k}$ if $a_j < a_k$. This ordering is such that the corresponding $\ell_{x_1,\ldots,x_n,k}$s are proportional to the ℓ_ks (the lengths of the intervals in the first subdivision of I). Of course, another ordering could have been used as well.

Consider next the case $s \geq 1$. Let p_k be as before and, for $(x_1, \ldots, x_{n+1}) \in \mathcal{X}^{n+1}$ with $1 \leq n \leq s$, suppose $p_{x_1,\ldots,x_n,x_{n+1}}$ is a PMF with respect to $x_{n+1} \in \mathcal{X}$ such that $p_{x_1} \cdots p_{x_1,\ldots,x_m}$ becomes a PMF with respect to $(x_1, \ldots, x_m) \in \mathcal{X}^m$. For $(x_1, \ldots, x_{n+1}) \in \mathcal{X}^{n+1}$ with $n \in \mathbb{N}$, let

$$\ell_{x_1,\ldots,x_{n+1}} = \ell_{x_1,\ldots,x_n} \times \begin{cases} p_{x_1,\ldots,x_n,x_{n+1}} & \text{if } n \leq s, \\ p_{x_{n-s+1},\ldots,x_n,x_{n+1}} & \text{if } n > s. \end{cases}$$

If $\ell_{x_1,\ldots,x_n} > 0$, specify the left endpoints of the intervals in the subdivision of $I_{x_1,\ldots,x_n}$ as before, i.e. $a_{x_1,\ldots,x_n,j} < a_{x_1,\ldots,x_n,k}$ if $a_j < a_k$. Thus (1.11) is satisfied and the nested subdivisions of I are specified such that the ordering of the intervals at the first level is reused for the intervals in the subdivision of any interval at a later level.

2.2. Dynamical systems

A common way to construct nested partitions is by using a dynamical system as follows (see also [6, 7, 24). Let $T\colon I \mapsto I$ be a function and $I = \cup_{k\in\mathcal{X}} \tilde{I}_k$ a countable partition such that, for each $k \in \mathcal{X}$, $\tilde{I}_k$ is a bounded interval of length $\ell_k > 0$ and the restriction of T to $\tilde{I}_k$ is a strictly monotone function, denoted T_k. For $n = 1, 2, \ldots$, let $T^n(x) = T(T^{n-1}(x))$ be the state

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

of the dynamical system at time n when it is started at $T^0(x) = x$, and let the nth digit of x be $x_n = k$ if $T^n(x) \in \tilde{I}_k$. For every $(x_1, \dots, x_n) \in \mathcal{X}^n$ with $n \in \mathbb{N}$, define

$$\tilde{I}_{x_1,\dots,x_n} = \{x \in I \mid T^{i-1}(x) \in \tilde{I}_{x_i},\ i = 1, \dots, n\},$$

which by induction is seen to be a bounded interval, and let

$$I_{x_1,\dots,x_n} \text{ agree with } \tilde{I}_{x_1,\dots,x_n} \text{ if the endpoints of } \tilde{I}_{x_1,\dots,x_n} \text{ are excluded.} \tag{2.1}$$

As mentioned in Section 1.1, it is of no importance for the results in this paper but only for convenience that the endpoints are excluded. In the dynamical system literature, $\tilde{I}_{x_1,\dots,x_n}$ is called a cylinder.

Henceforth, 'a setting for a dynamical system' refers to a situation where a dynamical system as above is given such that (1.4) holds and each $I_{x_1,\dots,x_n}$ is specified by (2.1). By induction, if $x = .x_1x_2 \dots \in J$ then $x^{[n]} = T^n(x) \in J$ for $n = 0, 1, \dots$ Thus, since $X \in J$, $X^{[n]} = T^n(X) \in J$ for $n = 0, 1, \dots$ Moreover, if each T_k is a diffeomorphism on I_k except for a Lebesgue nullset then, for every $n \in \mathbb{N}_0$, the distribution of $X^{[n]}$ is absolutely continuous and has a PDF on J which, for almost all $x = .x_1x_2 \dots \in J$, is given by

$$f^{[n]}(x) = \sum_{(k_1,\dots,k_n)\in\mathcal{X}^n:\, \ell_{k_1,\dots,k_n}>0} f(.k_1 \dots k_n x_1 x_2 \dots)\left|\left(T_{k_1}^{-1} \circ \cdots \circ T_{k_n}^{-1}\right)'(x)\right|. \tag{2.2}$$

This follows from the transformation theorem of densities by considering each case of $X \in I_{k_1,\dots,k_n}$, and since $T_{k_1}^{-1} \circ \cdots \circ T_{k_n}^{-1}(x) = .k_1 \dots k_n x_1 x_2 \dots$

By definition, the process $X_{>0} = (X_1, X_2, \dots)$ is stationary if and only if $X_{>0}$ and $X_{>1}$ are identically distributed, which in the present setting can be reformulated: stationarity is equivalent to $X = 0.X_1X_2 \dots \sim f$ implying $X^{[1]} = 0.X_2X_3 \dots \sim f$, and for short f is then said to be shift invariant. In the metric number theory literature (e.g. [6, 7, 24]), a distribution μ on $(I, \mathcal{B})$ is said to be T-invariant if $\mu(A) = \mu(T^{-1}(A))$ for all $A \in \mathcal{B}$. Then μ being absolutely continuous and T-invariant corresponds to f being shift invariant.

2.3. Examples

Using the approach in Section 2.1, it may not always be obvious to see whether the nested subdivisions could also be constructed using a setting for a dynamical system as in Section 2.2; an exception is the case $s = 0$, which is treated in Example 2.1. On the other hand, many examples of important number representations are generated by a dynamical system, and they are not all of the 'Markovian type'. A few examples are summarized in this section and used for illustrative purposes later on in the text. Further examples can be found, e.g., in [6, 7, 24].

In Examples 2.1–2.3, first a dynamical system is specified and second it is understood that nested partitions are constructed as in Section 2.2.

Example 2.1. Suppose that I agrees with $[0, 1]$ perhaps except at 0 or 1, a dynamical system is given such that (1.4) holds, and, for every $k \in \mathcal{X}$, the map T_k restricted to I_k is a linear bijection into $(0, 1)$ with slope t_k/ℓ_k, where $t_k \in \{\pm 1\}$. This is the setting for a generalized Lüroth series (GLS), cf. [2]. Since the signs t_k can vary, it is convenient that it is not necessary in the present paper to specify whether I and $\tilde{I}_k$ include or not each of their endpoints. Note that $T_{x_n} \circ \cdots \circ T_{x_1}$ restricted to $I_{x_1,\dots,x_n}$ is a linear bijection $I_{x_1,\dots,x_n} \mapsto (0, 1)$ with slope $(t_{x_1}/\ell_{x_1}) \cdots (t_{x_n}/\ell_{x_n})$. Thus, $\ell_{x_1,\dots,x_n} = \ell_{x_1} \cdots \ell_{x_n}$ for all $(x_1, \dots, x_n) \in \mathcal{X}^n$, and so the Markovian structure in (1.11) is satisfied when $s = 0$. Moreover, the left endpoints of the intervals in the subdivision of $I_{x_1,\dots,x_n}$ are ordered in the same way as in Section 2.1 if $t_{x_1} \cdots t_{x_n} = 1$,

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

and in the reverse order if $t_{x_1}\cdots t_{x_n}=-1$ (i.e. $a_{x_1,\dots,x_n,j}>a_{x_1,\dots,x_n,k}$ if $a_j<a_k$). Consequently, if $t_k=1$ for all $k\in\mathcal{X}$, the construction from Section 2.1 results in the same nested partitions, and the construction can be modified so that it works in any case. For the subdivision of $I_{x_1,\dots,x_n}$, use the same procedure as in Section 2.1 except if $t_{x_1}\cdots t_{x_n}=-1$, in which case the left endpoints should appear in the reverse order. One important and well-known case occurs when $q\geq 2$ is an integer, $I=[0,1)$, $\mathcal{X}=\{0,1,\dots,q-1\}$, and, for each $k\in\mathcal{X}$, $\tilde{I}_k=[k/q,(k+1)/q)$ and $T_k(x)=qx-k$. Then every $x\in[0,1)$ has a base-q expansion

$$x=\sum_{i=1}^{\infty}x_iq^{-i}\quad\text{with } x_i=\lfloor qT^{i-1}(x)\rfloor,\ i=1,2,\dots, \tag{2.3}$$

and J is the set of those x where the sum is not finite and there exists no $n\in\mathbb{N}$ such that $x_n=x_{n+1}=\cdots=q-1$. Here, $\lfloor a\rfloor$ is the integer part of $a\geq 0$. Moreover, $a_{x_1,\dots,x_n}=\sum_{i=1}^{n}x_iq^{-i}$, $\ell_{x_1,\dots,x_n}=q^{-n}$, and $T^n(x)=\sum_{i=1}^{\infty}x_{i+n}q^{-i}$.

Another special case is an 'original' Lüroth series, where $I=(0,1)$ and $\mathcal{X}=\{2,3,\dots\}$, and where Lüroth [19] showed that any $x\in(0,1)$ has a series

$$x=\frac{1}{x_1}+\frac{1}{x_1(x_1-1)x_2}+\cdots+\frac{1}{x_1(x_1-1)\cdots x_{n-1}(x_{n-1}-1)x_n}+\cdots \tag{2.4}$$

with $(x_1,x_2,\dots)\in\mathcal{X}^{\mathbb{N}}$ (sometimes in the literature, e.g. in [7], 0 is included in I, but again this is of no importance for the present paper). Defining, for each $k\in\mathcal{X}$, $\tilde{I}_k=[1/k,1/(k-1))$ and $T_k(x)=(k-1)kx-k+1$, it turns out that each digit x_n is uniquely given by $x_n=k$ if $1/k<T^n(x)<1/(k-1)$ (expressions for $T^n(x)$ and technical details for the case $T^n(x)=1/k$ are given in [7]; fortunately this case can be ignored in the present paper). It follows that J is the set of those x where the sum in (2.4) is not finite.

Example 2.2. Let $\beta>1$, $I=[0,1)$, and $T(x)=\beta x-\lfloor\beta x\rfloor$. If $\beta=q\in\{2,3,\dots\}$ then define $\mathcal{X}$ as in the case of the base-q expansion previously described, otherwise let $\mathcal{X}=\{0,1,\dots,\lfloor\beta\rfloor\}$, $\tilde{I}_k=[k/\beta,(k+1)/\beta)$ if $0\leq k<\lfloor\beta\rfloor$, and $\tilde{I}_{\lfloor\beta\rfloor}=[\lfloor\beta\rfloor/\beta,1)$. Rényi [23] showed that every $x\in[0,1)$ has a so-called β-expansion

$$x=\sum_{i=1}^{\infty}x_i\beta^{-i}\quad\text{with } x_i=\lfloor\beta T^{i-1}(x)\rfloor,\ i=1,2,\dots, \tag{2.5}$$

which of course agrees with (2.3) when $\beta\in\{2,3,\dots\}$. The convergence of the series in (2.5) implies that (1.4) holds.

Denote the set of all admissible sequences by $\mathcal{A}_\beta=\{(\lfloor\beta T^0(x)\rfloor,\lfloor\beta T^1(x)\rfloor,\dots)\mid 0\leq x<1\}$, referring to [7, 22] for a detailed description of $\mathcal{A}_\beta$. Clearly, $J_{>0}\subset\mathcal{A}_\beta$. For instance, if $\beta=q\in\{2,3,\dots\}$ then $A_\beta=\mathcal{X}^{\mathbb{N}}$ and $\mathcal{A}_\beta\setminus J_{>0}$ is the set of base-q fractions, i.e. numbers x where the sum in (2.3) is finite.

Another tractable case is the following. Let $s\in\mathbb{N}$, $m=s+1$, and $\beta=\beta_m>1$ be the pseudo golden mean of order m, that is, the positive root of $z^m-z^{m-1}-\cdots-z-1=0$. Then $\beta_2<\beta_3<\ldots<\lim_{n\to\infty}\beta_n=2$, where $\beta_2=(1+\sqrt{5})/2$ is the golden mean. So $\mathcal{X}=\{0,1\}$. Using (2.5) and induction for $n=1,2,\dots$, it is straightforwardly verified that, for any $(x_1,\dots,x_n)\in\mathcal{X}^n$, $\ell_{x_1,\dots,x_n}$ is equal to

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

$$\beta^{-n} \times \begin{cases} 0 & \text{if } n > s,\ x_i = \cdots = x_{i+s} = 1,\ 1 \le i \le n-s, \\ 1 & \text{if } x_n = 0, \\ \beta^i - \sum_{j=0}^{i-1} \beta^j & \text{if } x_{n-i} = 0,\ x_{n-i+1} = \cdots = x_n = 1,\ 1 \le i \le s,\ i < n, \\ \beta^n - \sum_{j=0}^{n-1} \beta^j & \text{if } x_1 = \cdots = x_n = 1,\ s \ge n. \end{cases}$$

Thus, the admissible sequences of 0s and 1s are those with at most s consecutive 1s (in accordance with [22]). It is easily seen that (1.11) holds with $p_{x_1,\dots,x_s,1} = 1 - p_{x_1,\dots,x_s,0}$ and

$$p_{x_1,\dots,x_s,0} = \begin{cases} 1/\beta & \text{if } x_s = 0, \\ 1/(\beta^{i+1} - \sum_{j=1}^{i} \beta^j) & \text{if } x_{s-i} = 0,\ x_{s-i+1} = \cdots = x_s = 1,\ 1 \le i < s, \\ 1 & \text{if } x_1 = \cdots = x_s = 1. \end{cases} \tag{2.6}$$

Finally, $\Omega = \mathcal{X}^s$.

Example 2.3. Recall the following about continued fractions [16]. Let $I = (0, 1]$, $\mathcal{X} = \mathbb{N}$, $\tilde{I}_k = (1/(k+1), 1/k]$ for $k \in \mathcal{X}$, and $T(x) = 1/x - \lfloor 1/x \rfloor$ be the Gauss map [10]. A rational number in I has a (regular) finite continued-fraction representation

$$[0;x_1, \dots, x_n] = \cfrac{1}{x_1 + \cfrac{1}{x_2 + \cfrac{1}{\ddots \cfrac{1}{x_n}}}} \in (0, 1],$$

and each irrational number $x \in (0, 1)$ has a unique infinite continued-fraction representation $x = [0;x_1, x_2, \dots] = \lim_{n\to\infty} [0;x_1, \dots, x_n]$, with $x_1, x_2, \dots \in \mathbb{N}$. So J consists of all irrational numbers between 0 and 1, and

$$\tilde{I}_{x_1,\dots,x_n} = \begin{cases} \big[[0;x_1, \dots, x_n],\ [0;x_1, \dots, x_n + 1]\big) & \text{for } n = 2, 4, 6, \dots, \\ \big([0;x_1, \dots, x_n + 1],\ [0;x_1, \dots, x_n]\big] & \text{for } n = 1, 3, 5, \dots. \end{cases}$$

For any $x = [0;x_1, x_2, \dots] \in J$, the continued-fraction remainder is given by $T^n(x) = [0;x_{n+1}, x_{n+2}, \dots] \in J$. Defining $q_n = x_n q_{n-1} + q_{n-2}$ for $n \in \mathbb{N}$ with $q_{-1} = 0$ and $q_0 = 1$, then q_n is a strictly increasing function of n and $\ell^{-1}_{x_1,\dots,x_n} = q_n(q_n + q_{n-1})$, so $\ell_{x_1,\dots,x_{n+1}}/\ell_{x_1,\dots,x_n} = q_n(q_n + q_{n-1})/[q_{n+1}(q_{n+1} + q_n)]$ depends on every digit $x_1, \dots, x_{n+1}$. Thus, (1.11) cannot be satisfied for any $s \in \mathbb{N}_0$.

3. The sufficient digits

Theorem 3.1 establishes the coupling between X and a non-negative integer-valued random variable N under the weak assumption that f is almost everywhere LSC, cf. (1.9). As noted after Theorem 3.1, $S = (X_1, \dots, X_N)$ becomes a sufficient statistic (if $N = 0$ then S is interpreted as the empty set). Therefore, $X_1, \dots, X_N$ are called the sufficient digits (though it is of course not meant that the approximation $X \approx A_n$ does not improve when $n > N$).

3.1. Coupling construction between X and N

Some notation is needed. Assume (1.9) and recall the definition of H. For $n \in \mathbb{N}$ and $(x_1, \dots, x_n) \in \mathcal{X}^n$, let $c_\emptyset = i_\emptyset = \inf_H f$ and $c_{x_1,\dots,x_n} = i_{x_1,\dots,x_n} - i_{x_1,\dots,x_{n-1}}$, where

$$i_{x_1,\dots,x_n} = \begin{cases} \inf_{H \cap I_{x_1,\dots,x_n}} f & \text{if } H \cap I_{x_1,\dots,x_n} \ne \emptyset, \\ 0 & \text{if } H \cap I_{x_1,\dots,x_n} = \emptyset. \end{cases}$$

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

If $n = 0$, interpret $(x_1, \dots, x_n) \in \mathcal{X}^n$ as $\emptyset \in \{\emptyset\}$, $c_{x_1,\dots,x_n}$ as $c_\emptyset$, and $\ell_{x_1,\dots,x_n}$ as $\ell_\emptyset$. Let ν be the product measure on $H \times \mathbb{N}_0$ given by the product of Lebesgue measure on H and the counting measure on $\mathbb{N}_0$.

Theorem 3.1. *Assuming (1.9) and recalling the notation in (1.6)–(1.8) and (1.10), the following hold.*

(i) *There is a coupling of X with a non-negative integer-valued random variable N such that the distribution of (X,N) is absolutely continuous with respect to ν, with density for any $(x, n) \in H \times \mathbb{N}_0$ given by*

$$f(x, n) = c_{x_1,\dots,x_n} \quad \text{if } x = .x_1x_2\dots \tag{3.1}$$

(ii) *For any $n \in \mathbb{N}_0$, $(x_1, \dots, x_n) \in \mathcal{X}^n$, and $x \in H$ with $f(x) > 0$,*

$$\mathbb{P}(S = (x_1, \dots, x_n)) = c_{x_1,\dots,x_n}\ell_{x_1,\dots,x_n}, \tag{3.2}$$
$$\mathbb{P}(N \le n \mid X = x) = i_{x_1,\dots,x_n}/f(x), \tag{3.3}$$
$$\mathbb{P}(N \le n) = \textstyle\sum_{(x_1,\dots,x_n)\in\mathcal{X}^n} i_{x_1,\dots,x_n}\ell_{x_1,\dots,x_n}. \tag{3.4}$$

(iii) *We have*

$$S \perp\!\!\!\perp U \sim \mathrm{Unif}(0, 1), \tag{3.5}$$

and R conditioned on $S = (x_1, \dots, x_n)$ has finite-dimensional distributions

$$\mathbb{P}(R_1 = x_{n+1}, \dots, R_m = x_{n+m} \mid S = (x_1, \dots, x_n)) = \ell_{x_1,\dots,x_{n+m}}/\ell_{x_1,\dots,x_n} \tag{3.6}$$

when $(x_1, \dots, x_{n+m}) \in \mathcal{X}^{n+m}$ with $m \in \mathbb{N}$, $n \in \mathbb{N}_0$, and $\mathbb{P}(S = (x_1, \dots, x_n)) > 0$, i.e. provided $c_{x_1,\dots,x_n}\ell_{x_1,\dots,x_n} > 0$.

(iv) *Conditioned on L, $S \perp\!\!\!\perp E \sim \mathrm{Unif}(0, L)$.*

Proof. The theorem is an extension of results in [15] for the special case of base-q expansions (with $q > 1$ an integer, cf. Example 2.1). A general, different, and shorter proof is given here.

Let $x = .x_1x_2\dots \in H$. Then

$$f(x) = \lim_{n\to\infty} i_{x_1,\dots,x_n} = \lim_{n\to\infty}\left(i_\emptyset + \sum_{i=1}^{n} c_{x_1,\dots,x_i}\right) = \sum_{n=0}^{\infty} c_{x_1,\dots,x_n},$$

where the first equality follows since f is LSC at $x \in H$ and x is in the interior of $I_{x_1,\dots,x_n}$, and the second equality follows from the definition of $i_{x_1,\dots,x_n}$. Clearly, $c_{x_1,\dots,x_n} \ge 0$ and

$$\int \sum_{n=0}^{\infty} c_{x_1,\dots,x_n}\,\mathrm{d}x = \int f(x)\,\mathrm{d}x = 1,$$

so (3.1) specifies a well-defined distribution on $H \times \mathbb{N}_0$. Thus the existence of the coupling with density (3.1) follows, and so (i) is verified.

Equation (3.2) follows immediately from (3.1). A straightforward calculation first gives (3.3) using (3.1), and second (3.4) using (3.3). This proves (ii).

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

Since (X, N) and (S, U) are in a one-to-one correspondence, combining (3.1) and (3.2) shows that, for $u \in (0, 1)$ and $\mathbb{P}(S = (x_1, \ldots, x_n)) > 0$,

$$\mathbb{P}(U \le u \mid S = (x_1, \ldots, x_n)) = \frac{\mathbb{P}(a_{x_1,\ldots,x_n} < X \le a_{x_1,\ldots,x_n} + \ell_{x_1,\ldots,x_n} u, N = n)}{\mathbb{P}(S = (x_1, \ldots, x_n))} = \frac{\int_{a_{x_1,\ldots,x_n}}^{a_{x_1,\ldots,x_n} + u\ell_{x_1,\ldots,x_n}} c_{x_1,\ldots,x_n} \, \mathrm{d}x}{c_{x_1,\ldots,x_n} \ell_{x_1,\ldots,x_n}} = u.$$

Thus, $U|S \sim \mathrm{Unif}(0, 1)$ and hence (3.5) follows. Furthermore, (3.1) and (3.2) give

$$\begin{aligned}\mathbb{P}(R_1 = x_{n+1}, \ldots, R_m = x_{n+m} \mid S = (x_1, \ldots, x_n)) &= \frac{\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n, X_{n+1} = x_{n+1}, \ldots, X_{n+m} = x_{n+m}, N = n)}{\mathbb{P}(S = (x_1, \ldots, x_n))} \\ &= \frac{\int_{I_{x_1,\ldots,x_{n+m}}} c_{x_1,\ldots,x_n} \, \mathrm{d}x}{c_{x_1,\ldots,x_n} \ell_{x_1,\ldots,x_n}} = \ell_{x_1,\ldots,x_{n+m}} / \ell_{x_1,\ldots,x_n},\end{aligned}$$

whereby (3.6) is proven. Finally, (iv) is immediately obtained from (3.5), which completes the proof. □

Remark 3.1. Various comments on Theorem 3.1 are in order.

- The joint distribution of (X, N) is specified by $X \sim f$. Since N is not a function of X, it may be treated as a latent variable. Suppose a statistical model class for the PDF f is given (so that f is LSC almost everywhere on J). Since (X, N) is in a one-to-one correspondence with (S, U), where $S \perp\!\!\!\perp U \sim \mathrm{Unif}(0, 1)$, it follows that S is a sufficient statistic and U is an ancillary statistic when imagining N is observable, cf. [1]. Some suggestions on how to construct models for S appear in Section 5.
- The distribution of N depends on the choice of H, that is, the version of the PDF f. In brief, to make N small, pick H in (1.9) as small as possible, cf. (3.3).
- By (3.4), $\mathbb{P}(N \le n)$ is large if f is approximately constant on the intervals $I_{x_1,\ldots,x_n}$ where $\ell_{x_1,\ldots,x_n}$ is not small.
- It is remarkable that the conditional distribution of R given S depends not on f but only on the ratios between the lengths of the intervals, cf. (3.6). Therefore, for any $n \in \mathbb{N}_0$, conditioned on $N \le n$, the distribution of $X_{>n}$ (or equivalently $X^{[n]}$) depends only on these ratios. This property is exploited further in Section 4.

3.2. Corollaries

From (3.4) we can easily get the following corollary.

Corollary 3.1. *Assume (1.9). Then, for any fixed $n \in \mathbb{N}_0$, $\mathbb{P}(N \le n) = 1$ if and only if, for every $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $f(x)$ is constant for almost all $x \in I_{x_1,\ldots,x_n}$. In particular, $N = 0$ almost surely if and only if $X \sim \mathrm{Unif}(I)$.*

Example 3.1. For continued-fraction representations, where T is the Gauss map given in Example 2.3, it is well known that

$$f_{\mathrm{G}}(x) = \frac{1}{\ln 2} \frac{1}{1 + x}, \quad 0 < x \le 1, \tag{3.7}$$

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

is the PDF of the unique T-invariant distribution; see, e.g., [6]. By Corollary 3.1, N is unbounded when $X \sim f_{\mathrm{G}}$.

The simplest case of (1.11) happens when $s = 0$. Defining the PMF $\pi_{\mathrm{inv}}(k) = \ell_k/\ell_\emptyset$ for $k \in \mathcal{X}$, then (1.11) with $s = 0$ is equivalent to

$$\ell_{x_1,\ldots,x_n}/\ell_\emptyset = \pi_{\mathrm{inv}}(x_1)\cdots\pi_{\mathrm{inv}}(x_n) \quad \text{whenever } n \geq 2 \text{ and } (x_1,\ldots,x_n) \in \mathcal{X}^n. \tag{3.8}$$

Corollary 3.2. *Assuming (1.9) and (3.8), the following items hold.*

(i). *The following statements are equivalent:*

(a) $X_1, X_2, \ldots$ *are i.i.d. with PMF* π_{inv}.

(b) $X \sim \mathrm{Unif}(I)$.

(c) $N = 0$ *almost surely.*

(ii). $S \perp\!\!\!\perp R$ *and* $R_1, R_2, \ldots$ *are i.i.d. with PMF* π_{inv}.

(iii). $S \perp\!\!\!\perp X^{[N]} \sim \mathrm{Unif}(I)$.

(iv). *For every* $n \in \mathbb{N}_0$, *conditioned on* $N \leq n$, $X^{[n]} \sim \mathrm{Unif}(I)$ *and* $X_{n+1}, X_{n+2}, \ldots$ *are i.i.d. with PMF* π_{inv}.

Proof. Let $(x_1,\ldots,x_{n+m}) \in \mathcal{X}^{n+m}$, $m \in \mathbb{N}$, and $n \in \mathbb{N}_0$. For $\mathbb{P}(S = (x_1,\ldots,x_n)) > 0$, or equivalently $c_{x_1,\ldots,x_n}\ell_{x_1,\ldots,x_n} > 0$, cf. (3.2), it follows from (3.6) and (3.8) that

$$\mathbb{P}(R_1 = x_{n+1},\ldots,R_m = x_{n+m} \mid S = (x_1,\ldots,x_n)) = \pi_{\mathrm{inv}}(x_{n+1})\cdots\pi_{\mathrm{inv}}(x_{n+m}).$$

Thus (ii) is verified. It is straightforwardly seen that (i) holds because of the following equivalences: (a) follows from (ii) and (c); (b) implies (c); and (a) and (b) are equivalent. Finally, since $X^{[N]} = .R_1R_2\ldots$, both (iii) and (iv) follow from (i) and (ii). □

For instance, (3.8) is satisfied in the setting for a GLS, including e.g. base-q expansions, cf. Example 2.1. Conversely, given a PMF π_{inv} on $\mathcal{X}$, (3.8) may be used for constructing nested subdivisions of I where each non-empty interval $I_{x_1,\ldots,x_n}$ (interpreted as I if $n = 0$) is divided into intervals $I_{x_1,\ldots,x_n,k}$ with proportions $\pi_{\mathrm{inv}}(k)$ for $k \in \mathcal{X}$. For the GLS setting, the fact that (a) implies (b) in Corollary 3.2 (i) also follows from [2, Lemma 1].

Let $Q^{[n]}$ be the probability distribution of $X^{[n]}$, μ be a probability measure on the measure space $(I, \mathcal{B})$, where $\mathcal{B}$ is the class of Borel sets included in I, and $\|Q^{[n]} - \mu\|_{\mathrm{TV}} = \sup_{A\in\mathcal{B}} |Q^{[n]}(A) - \mu(A)|$ be the total variation distance of the signed measure $Q^{[n]} - \mu$. It is well known that if μ has PDF h and $X^{[n]}$ has PDF $f^{[n]}$, then

$$\|Q^{[n]} - \mu\|_{\mathrm{TV}} = \frac{1}{2}\int_I |f^{[n]}(x) - h(x)|\,\mathrm{d}x.$$

For instance, $f^{[n]}$ may be given by (2.2), and in the following corollary, $h = 1/\ell_\emptyset$.

Corollary 3.3. *Assume (1.9) and (3.8), and let* $\mu_{\mathrm{inv}} = \mathrm{Unif}(I)$. *Then there is a coupling between (X,N) and a random variable* $Y \sim \mu_{\mathrm{inv}}$ *such that* $X^{[n]} = Y^{[n]} \sim \mu_{\mathrm{inv}}$ *when* $N \leq n$. *In particular,*

$$\|Q^{[n]} - \mu_{\mathrm{inv}}\|_{\mathrm{TV}} \leq \mathbb{P}(N > n) \to 0 \quad \text{as } n \to \infty. \tag{3.9}$$

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

Proof. Consistently with (i) and (ii) in Corollary 3.2, let $S \perp\!\!\!\perp Y = .Y_1 Y_2 \ldots \sim \mu_{\text{inv}}$ and $R_1 = Y_{N+1}, R_2 = Y_{N+2}, \ldots$. Conditioned on $N \le n$, we have $X^{[n]} = Y^{[n]} \sim \mu_{\text{inv}}$, cf. Corollary 3.2(iv). Since $X^{[n]} \sim Q^{[n]}$ and $Y^{[n]} \sim \mu_{\text{inv}}$, the coupling inequality (see, e.g., [25]) gives $\|Q^{[n]} - \mu_{\text{inv}}\|_{\text{TV}} \le \mathbb{P}(X^{[n]} \ne Y^{[n]})$, and so the inequality in (3.9) follows. Since N is finite, the convergence to 0 in (3.9) is obvious. □

4. Markov chains

This section treats the case of the Markovian structure in (1.11). The special case $s = 0$ has been treated in Section 3.2. Throughout this section, let $s \in \mathbb{N}$ be given, let $X \sim f$, where f satisfies (1.9), and let S, N, R be as in Theorem 3.1. Of course, it is natural to pick the value of s as small as possible such that (1.11) holds.

4.1. When the residual process is Markov

Recall the notation in (1.12)–(1.15). Observe that (1.11) with $n = s + 1$ gives

$$p_{x_1,\ldots,x_s,x_{s+1}} = \ell_{x_1,\ldots,x_{s+1}} / \ell_{x_1,\ldots,x_s} \quad \text{for } (x_1, \ldots, x_s), (x_2, \ldots, x_{s+1}) \in \Omega. \tag{4.1}$$

It follows by induction that (1.11) is equivalent to

$$\ell_{x_1,\ldots,x_n} = \ell_{x_1,\ldots,x_s}\, p_{x_1,\ldots,x_{s+1}} \cdots p_{x_{n-s},\ldots,x_n} \tag{4.2}$$

whenever $n \in \{s+2, s+3, \ldots\}$ and $(x_1, \ldots, x_s), \ldots, (x_{n-s+1}, \ldots, x_n) \in \Omega$. Conversely, given $s \in \mathbb{N}$, nested subdivisions of I at levels $n = 1, \ldots, s$, and transition probabilities $p_{x_1,\ldots,x_{s+1}} \ge 0$ for $(x_1, \ldots, x_s) \in \Omega$ and $x_{s+1} \in \mathcal{X}$ such that $\sum_{k\in\mathcal{X}} p_{x_1,\ldots,x_s,k} = 1$ and $p_{x_1,\ldots,x_s,k} = 0$ if $(x_2, \ldots, x_s, k) \notin \Omega$, then (4.2) may be used for constructing nested subdivisions of I, where each non-empty interval $I_{x_1,\ldots,x_n}$ with $n \ge s$ is divided into intervals $I_{x_1,\ldots,x_n,k}$ with proportions $p_{x_{n-s+1},\ldots,x_n,k}$ for $k \in \mathcal{X}$.

Theorem 4.1. *Assume (1.9) and (1.11). Consider any $n \in \mathbb{N}_0$ and $(x_1, \ldots, x_n) \in \mathcal{X}^n$ with $\mathbb{P}(S = (x_1, \ldots, x_n)) > 0$, that is, $c_{x_1,\ldots,x_n} \ell_{x_1,\ldots,x_n} > 0$, cf. (3.2). Conditioned on $S = (x_1, \ldots, x_n)$, $R^{(s)}$ is a time-homogeneous Markov chain with transition matrix P and initial distribution*

$$\begin{aligned} &\mathbb{P}(R_1 = x_{n+1}, \ldots, R_s = x_{n+s} \mid S = (x_1, \ldots, x_n)) \\ &\qquad = \begin{cases} \prod_{j=n+1-s}^{n} p_{x_j,\ldots,x_{j+s}} & \text{if } n \ge s, \\ (\ell_{x_1,\ldots,x_s} / \ell_{x_1,\ldots,x_n}) \prod_{j=1}^{n} p_{x_j,\ldots,x_{j+s}} & \text{if } n < s, \end{cases} \end{aligned} \tag{4.3}$$

where $\prod_{j=1}^{0} \cdots$ is interpreted as 1.

Proof. By (1.13) and (4.1), P is a transition matrix since its entries are non-negative and each row sums to 1. To simplify the notation, write $\ell_{i:j} = \ell_{x_i,\ldots,x_j}$, $p_{i:j} = p_{x_i,\ldots,x_j}$, $c_{i:j} = c_{x_i,\ldots,x_j}$, and $P_{i:j} = \prod_{k=i}^{j} p_{k:k+s}$ when $i \le j$. Then, for any $m \in \mathbb{N}_0$ and $(x_1, \ldots, x_{n+m+s}) \in \mathcal{X}^{n+m+s}$ with $c_{1:n}\ell_{1:n} > 0$,

$$\begin{aligned} &\mathbb{P}(R_1 = x_{n+1}, \ldots, R_{m+s} = x_{n+m+s} \mid S = (x_1, \ldots, x_n)) \\ &\qquad = \ell_{1:n+m+s} / \ell_{1:n} \end{aligned} \tag{4.4}$$

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

$$= \begin{cases} P_{n+1-s:m+n} & \text{if } n \geq s \\ (\ell_{1:s}/\ell_{1:n})P_{1:n+m} & \text{if } n < s \end{cases}$$
$$= \begin{cases} P_{n+1-s:n}P_{n+1:n+m} & \text{if } n \geq s \\ (\ell_{1:s}/\ell_{1:n})P_{1:n}P_{n+1:n+m} & \text{if } n < s \end{cases} \tag{4.5}$$

where (4.4) is obtained from (3.6), $0/0 = 0$, and (4.5) follows from (1.11) and (4.1). From (4.5) with $m = 0$ follows (4.3). If $m > 0$ then

$$\mathbb{P}(R_{m+s} = x_{n+m+s} \mid S = (x_1, \ldots, x_n), R_1 = x_{n+1}, \ldots, R_{m+s-1} = x_{n+m+s-1})$$
$$= \frac{\mathbb{P}(R_1 = x_{n+1}, \ldots, R_{m+s} = x_{n+m+s} \mid S = (x_1, \ldots, x_n))}{\mathbb{P}(R_1 = x_{n+1}, \ldots, R_{m+s-1} = x_{n+m+s-1} \mid S = (x_1, \ldots, x_n))} = p_{n+m:n+m+s},$$

where (4.5) is used in the last equality. So $R^{(s)} \mid S = (x_1, \ldots, x_n)$ is a time-homogeneous Markov chain with transition matrix P. □

The following will be needed in Corollary 4.1.

To any PMF $\pi = (\pi_{x_1,\ldots,x_s})_{(x_1,\ldots,x_s)\in\Omega}$ corresponds a PDF f_π on H that is defined, for every $x = .x_1x_2 \ldots \in H$, by

$$f_\pi(x) = \begin{cases} \pi_{x_1,\ldots,x_s}/\ell_{x_1,\ldots,x_s} & \text{if } (x_1, \ldots, x_s) \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

Clearly, f_π is LSC on H. Conversely, from any PDF f on H which is constant almost everywhere on each interval $I_{x_1,\ldots,x_s}$ with $(x_1, \ldots, x_s) \in \Omega$, we obtain a PMF π on Ω such that (4.6) is satisfied with $f = f_\pi$.

Assume (1.11) and $X \sim f_\pi$. Then f_π is shift invariant if and only if π is invariant with respect to P (or, for short, P-invariant), which means that for every $(x_1, \ldots, x_s) \in \Omega$,

$$\sum_{k:\,(k,x_1,\ldots,x_{s-1})\in\Omega} \pi_{k,x_1,\ldots,x_{s-1}} p_{k,x_1,\ldots,x_s} = \pi_{x_1,\ldots,x_s}. \tag{4.7}$$

For $s = 1$, this is interpreted as $\sum_{k\in\mathcal{X}} \pi_k p_{k,x_1} = \pi_{x_1}$.

Corollary 4.1. *Assume (1.11) and $X \sim f_\pi$, where π is P-invariant. Then $N \leq s$ almost surely and $X^{(s)}$ is a stationary Markov chain with initial distribution $(X_1, \ldots, X_s) \sim \pi$ and transition matrix P. Conditioned on $(X_1, \ldots, X_s)$, we have $N \perp\!\!\!\perp X_{>s}$.*

Proof. By (4.6), from $X \sim f_\pi$ follows $(X_1, \ldots, X_s) \sim \pi$, and $N \leq s$ almost surely, cf. Corollary 3.1. By Theorem 4.1, conditioned on both N and $(X_1, \ldots, X_s)$, or equivalently, conditioned on S and on those (if any) R_i with $N < i \leq s - N$, $X^{(s)}$ is a Markov chain with transition matrix P and the distribution of $X_{>s}$ depends only on $(X_1, \ldots, X_s)$ (and not on N). Therefore, conditioned on $(X_1, \ldots, X_s)$, we have $N \perp\!\!\!\perp X_{>s}$, and since $(X_1, \ldots, X_s) \sim \pi$, the distribution of the chain $X^{(s)} = \{X_i^{(s)}\}_{i=1,2,\ldots}$ is invariant under left shifts. The proof is complete. □

It is well known that there exists a unique P-invariant PMF if and only if P is irreducible. The notation π_{inv} will be used to indicate that this is the unique P-invariant PMF. Letting $f_{\text{inv}} = f_{\pi_{\text{inv}}}$, we get the following result.

Corollary 4.2. *Assume (1.11) and that P is irreducible. Then f_{inv} is the unique shift-invariant PDF (up to equivalence of PDFs).*

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

The condition in Corollary 4.2 that P is irreducible translates by (1.11) into the condition that, for any distinct $(x_1, \dots, x_s), (y_1, \dots, y_s) \in \Omega$, there exist $s + k - 1 \geq 1$ elements $x_{s+1}, \dots, x_{2s+k-1} \in \mathcal{X}$ with $x_{s+k} = y_1, \dots, x_{2s+k-1} = y_s$ and

$$\ell_{x_1,\dots,x_{s+1}} > 0, \dots, \ell_{x_{s+k-1},\dots,x_{2s+k-1}} > 0.$$

In the setting of a dynamical system (Section 2.2), Corollary 4.2 is in line with the well-known result that if T is ergodic with respect to Lebesgue measure and f specifies a T-invariant distribution, then f specifies the unique T-invariant distribution (in fact, in all specific examples of T given in this paper, T is ergodic with respect to Lebesgue measure; see, e.g., [6] and [24]).

Example 4.1. In continuation of Example 2.2 when β is the pseudo golden mean of order $s + 1$, it follows by Theorem 4.1 that $R^{(s)} \mid S$ is a Markov chain. Since the transition probabilities are given by (2.6), the chain is seen to be irreducible. Thus, by Corollary 4.2, f_{inv} specifies the unique shift-invariant PDF, or equivalently T-invariant probability measure, where $T(x) = \beta x - \lfloor \beta x \rfloor$. In fact, for any value of $\beta > 1$, [11, 22] independently obtained the T-invariant distribution such that the corresponding PDF can be derived in closed form.

4.2. Coupling construction between (X, N) and (M_1, M_2)

Some notation is needed for the following theorem. For every $z \in \Omega$ and $i \in \mathbb{N}$, let $Y_i^{(s)}(z) = (Y_i(z), \dots, Y_{i+s-1}(z))$ and suppose $Y^{(s)}(z) = \{Y_i^{(s)}(z)\}_{i=1,2,\dots}$ is a Markov chain on Ω with transition matrix P and $Y_1^{(s)}(z)$ following the PMF given by row z in P. For $i \in \mathbb{N}$, denote by $P_i^{(s)}(z)$ the distribution of $Y_i^{(s)}(z)$, and by P^i the ith-step transition matrix of $Y^{(s)}(z)$, i.e. P^i is the product of P with itself i times. Let $P_{z,w}^i$ be entry (z, w) of P^i, and $\|\cdot\|_{\text{TV}}$ the total variation distance for signed finite measures on Ω. Recall that uniform geometric ergodicity of a Markov chain with transition matrix P means that there is a unique invariant PMF π_{inv} with respect to P and, for some constants α and $\beta < 1$ and for all $(z, i) \in \Omega \times \mathbb{N}$,

$$\left\| P_i^{(s)}(z) - \Pi_{\text{inv}} \right\|_{\text{TV}} = \frac{1}{2} \sum_{w \in \Omega} \left| P_{z,w}^i - \pi_{\text{inv}}(w) \right| \leq \alpha \beta^i, \tag{4.8}$$

where Π_{inv} is the probability measure on Ω with PMF π_{inv}. If Ω is finite then uniform geometric ergodicity is equivalent to irreducibility and aperiodicity of the chain. When Ω is countable, see [20, Section 16.2.1] for conditions of uniform geometric ergodicity. These criteria of uniform geometric ergodicity translate into conditions on the lengths of the intervals at level $s + 1$. This was illustrated in the remark after the proof of Corollary 4.2 when discussing irreducibility. Aperiodicity holds if, e.g., $\ell_{x_1,\dots,x_{2s}} > 0$ and $\ell_{x_1,\dots,x_{2s+1}} > 0$ for all $(x_1, \dots, x_{2s+1}) \in \mathcal{X}^{2s+1}$.

Theorem 4.2. *Assume (1.9), (1.11), and (4.8), and let π_{inv} be the unique P-invariant PMF and f_{inv} the corresponding shift-invariant PDF. Then there is a time $t \in \mathbb{N}$, a positive probability $\varepsilon_t > 0$ (given by (4.11)), and a coupling of (X,N) with two independent geometrically distributed random variables M_1 and M_2 with the same mean $1/\varepsilon_t$ and support $\mathbb{N}$ such that the following items hold. (Let $b = t + s - 1$, $M = b(M_1 + M_2 - 1)$, and $K = \max\{N, s\} + M - s$.)*

(i) *$(X_1, \dots, X_{\max\{N,s\}}, N)$ is independent of both $X^{[K]} \sim f_{\text{inv}}$ and $X_{>K}^{(s)}$, which is a stationary Markov chain with initial PMF π_{inv} and transition matrix P.*

(ii) *For any $n \in \mathbb{N}_0$, conditioned on $K \leq n$, $X^{[n]} \sim f_{\text{inv}}$.*

(iii) $(X_1, \dots, X_{\max\{N,s\}}, N) \perp\!\!\!\perp (M_1, M_2)$.

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

Theorem 4.2 and the following corollary are proved and exemplified in Section 4.4.

Corollary 4.3. *Assume (1.9), (1.11), and (4.8), let f_{inv} and K be as in Theorem 4.2, and let μ_{inv} be the probability measure with PDF f_{inv}. Then there is a coupling between (X,K) and a random variable $Z \sim \mu_{\mathrm{inv}}$ such that $X^{[n]} = Z^{[n]}$ when $K + s \le n$, and so*

$$\|Q^{[n]} - \mu_{\mathrm{inv}}\|_{\mathrm{TV}} \le \mathbb{P}(K > n) \to 0 \quad \text{as } n \to \infty.$$

Theorem 4.2 and Corollary 4.3 are discussed together with Corollaries 3.2 and 3.3 in Section 5.2.

4.3. Read-once algorithm

The proof in Section 4.4 of Theorem 4.2 is based on the read-once algorithm developed by Wilson [26]. The algorithm is stated below for the setting used in the present paper.

The Markov chain $Y^{(s)}(z)$ can be specified as a stochastic recursive sequence (SRS) so that

$$Y_i(z) = g(Y_{i-s}(z), \dots, Y_{i-1}(z), V_i), \quad i = 1, 2, \dots, \tag{4.9}$$

where $V_1, V_2, \dots$ are i.i.d. random variables, $(Y_{1-s}(z), \dots, Y_0(z)) = z$, and there are many possible choices of the updating function g and the distribution of V_1. Without loss of generality, assume $\mathcal{X} \subset \mathbb{R}$ is countable, and for specificity, let $V_1 \sim \mathrm{Unif}(0, 1)$; for $(x_1, \dots, x_s) \in \Omega$, let

$$(0, 1) \ni v \mapsto g(x_1, \dots, x_s, v) = F^-(v \mid x_1, \dots, x_s) \tag{4.10}$$

be the generalized inverse of the CDF $F(x \mid x_1, \dots, x_s) = \sum_{k:\, k \le x,\, (x_2,\dots,x_s,k) \in \Omega} p_{x_1,\dots,x_s,k}$, $x \in \mathbb{R}$, where $(x_2, \dots, x_s, k)$ is interpreted as k if $s = 1$. Note that all the chains $Y^{(s)}(z)$ with $z \in \Omega$ are coupled via $(V_1, V_2, \dots)$

For an integer $t \ge s$, letting $b = t + s - 1$, $B = B(V_1, \dots, V_b) = \{(Y_1(z), \dots, Y_b(z)) \mid z \in \Omega\}$ is called a block of length b. The block agrees with $\{(Y_1^{(s)}(z), \dots, Y_t^{(s)}(z)) \mid z \in \Omega\}$ and the notation $B(V_1, \dots, V_b)$ is used to stress that the block depends only on $V_1, \dots, V_b$. Let $O(z, V_1, \dots, V_b) = Y_t^{(s)}(z)$, which is interpreted as the output of B when z is the input in B. Call B a coalescent block with common output $O(V_1, \dots, V_b) = O(z, V_1, \dots, V_b)$ if this is the same for all $z \in \Omega$. Let

$$\varepsilon_t = \mathbb{P}(B \text{ is a coalescent block}), \tag{4.11}$$

which is an increasing function of t. Suppose $\varepsilon_t > 0$, and consider the sequence of blocks $B_i = B_i(V_{(i-1)b+1}, \dots, V_{ib})$, $i = 1, 2, \dots$, which are i.i.d. copies of B. Now, Wilson's read-once algorithm consists of three steps:

(i) Draw $B_1, \dots, B_{M_1}, B_{M_1+1}, \dots, B_{M_1+M_2}$, where B_{M_1} and B_{M_2} are the first and second coalescent blocks. Let $O_{M_1 b} = O(V_{(M_1-1)b+1}, \dots, V_{M_1 b})$ be the common output of B_{M_1}.

(ii) If $M_2 \ge 2$, for $i = 0, \dots, M_2 - 2$ use the output $O_{(M_1+i)b}$ of B_{M_1+i} as the input in the following block, which gives the output $O_{(M_1+i+1)b} = O(O_{(M_1+i)b}, V_{(M_1+i)b+1}, \dots, V_{(M_1+i)b+b})$ in B_{M_1+i+1}.

(iii) Return $O = O_M$, where $M = (M_1 + M_2 - 1)b$.

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

It follows from [9] and [26] that the existence of a $t \in \mathbb{N}$ such that $\varepsilon_t > 0$ is equivalent to (4.8), and it has already been noted that (4.8) implies the unique existence of the invariant PMF π_{inv}. Thus, the following lemma follows from [26].

Lemma 4.1. *Assume (1.11) and (4.8), and let π_{inv} be the unique P-invariant PMF. Then $O \sim \pi_{\text{inv}}$.*

4.4. Proofs and example

Proof of Theorem 4.2. Assume (4.8), and let π_{inv} be the unique P-invariant PMF and f_{inv} the corresponding shift-invariant PDF. Using Theorem 3.1, $X = 0.X_1X_2 \ldots \sim f$ and N can be generated as follows. First, generate $(X_1, \ldots, X_N)$ from the PMF (3.2). If $N < s$ then generate $(X_{N+1}, \ldots, X_s)$ such that $(X_{N+1}, \ldots, X_s)$ conditioned on $(X_1, \ldots, X_N) = (x_1, \ldots, x_n)$ has PMF $p(x_{n+1}, \ldots, x_s \mid x_1, \ldots, x_n) = \ell_{x_1,\ldots,x_s}/\ell_{x_1,\ldots,x_n}$ whenever $\ell_{x_1,\ldots,x_n} > 0$. This is in accordance with (3.6). Second, independent of $(X_1, \ldots, X_{\max\{N,s\}}, N)$, generate i.i.d. $V_1, V_2, \ldots \sim \text{Unif}(0, 1)$ and use the SRS (4.9) when generating

$$X_{>\max\{N,s\}} = Y^{(s)}(X_{\max\{N,s\}-s+1}, \ldots, X_{\max\{N,s\}}).$$

Thus, $(X_1, \ldots, X_{\max\{N,s\}}, N)$ is independent of the output O of the read-once algorithm. By Lemma 4.1, $X^{(s)}_{\max\{N,s\}+M-s+1} = (X_{\max\{N,s\}+M-s+1}, \ldots, X_{\max\{N,s\}+M}) = O \sim \pi_{\text{inv}}$. Combining these facts with Corollaries 4.1 and 4.2, Theorem 4.2(i) follows; (ii) follows directly from (i). Finally, since $(X_1, \ldots, X_{\max\{N,s\}}, N) \perp\!\!\!\perp (V_1, V_2, \ldots)$, (iii) follows from the read-once algorithm. Hence, the proof of Theorem 4.2 is complete. □

Proof of Corollary 4.3. Imagine that the first block starts just after time $\max\{N, s\}$. Then $\max\{N, s\} + M$ is the time at which the output of the read-once algorithm was produced. Hence, using Theorem 4.2(i), along similar lines to the proof of Corollary 3.3, Corollary 4.3 is obtained. □

Example 4.2. Consider again Example 2.2 when β is the pseudo golden mean of order $s + 1$. Since Ω is finite and a Markov chain with transition matrix P given by (2.6) is irreducible and aperiodic, $\varepsilon_t > 0$ for all sufficiently large t. So, for large t, the assumptions in Theorem 4.2 are satisfied, and hence the results in Theorem 4.2 and Corollary 4.3 apply.

Consider the case $s = 1$. Solving (4.7) to obtain π_{inv} and then using (4.6) together with the fact that $\beta^2 = \beta + 1$, we get

$$\pi_{\text{inv}} = \left(\frac{\beta^2}{1+\beta^2}, \frac{1}{1+\beta^2}\right), \qquad f_{\text{inv}}(x) = \begin{cases} (1+2\beta)/(2+\beta) & \text{if } 0 < x < 1/\beta, \\ (1+\beta)/(2+\beta) & \text{if } 1/\beta < x < 1. \end{cases}$$

Choosing the updating function g as in (4.10) and letting $V \sim \text{Unif}(0, 1)$, then

$$\varepsilon_1 = \mathbb{P}(F^-(V \mid 0) = F^-(V \mid 1)) = \mathbb{P}(V \le \beta^{-1}) = \beta^{-1} = 0.6180\ldots$$

This is for the case $t = 1$, so the block length is just $b = 1$ and the expected running time of the read-once algorithm is just $\mathbb{E}(M_1 + M_2) = 2/\varepsilon_1 = \sqrt{5} + 1 = 3.2361\ldots$ So, it may be expected that $R_n \sim f_{\text{inv}}$ for small values of n.

5. Concluding remarks

This section concludes the paper with a discussion of the main results, including some open problems for future research.

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

Assuming only almost-everywhere lower semi-continuity of f, Theorem 3.1 establishes the general results of this paper. In my opinion, the two most important general results are that $S \perp\!\!\!\perp U \sim \text{Unif}(0, 1)$ and $R \mid S$ follows a distribution that does not depend on f but is specified by ratios of the lengths of intervals in the nested subdivisions, see (3.6). The first result leads to a discussion in Section 5.1 of statistical modelling, and the importance of the latter result is discussed in Section 5.2.

5.1. Statistical models

After Theorem 3.1 it was noted that $S = (X_1, \ldots, X_N)$ is a sufficient statistic when N is treated as a latent variable. Hence, for likelihood-based statistical inference, no matter whether it is in a classical/frequentist or Bayesian setting, the likelihood function based on (X, N) is equivalent to the likelihood based on the discrete random variable S [1]. So a natural question is how to specify the distribution of S.

Statistical models where N is expected to be small would possibly suffice in many applications. For the case of base-q expansions, specific examples of PDFs f in [13] show that N will be small when q is not small and f does not deviate much from $f_{\text{inv}} = 1$.

Since S is a (discrete) random variable of varying dimension, one modelling approach could be inspired by the construction of point process models [8, 21]. For example, in analogy with the use of Janossy densities when specifying finite point process densities, by specifying the distribution of first N and next $S \mid N$. Or, in the same spirit as when Gibbs distribution models for finite point process are defined, specifying the distribution of S with respect to some natural reference measure under which, e.g., $X_1, \ldots, X_N \mid N$ are i.i.d. and independent of N.

In connection to Bayesian inference, another question is how to construct a random PDF $f(\cdot \mid Y)$, assuming Y is a stochastic process such that $\mathbb{P}(X \in B \mid Y) = \int_B f(x \mid Y)\,\mathrm{d}x$ for any $B \in \mathcal{B}$ (the class of Borel sets included in I). One possibility is the mixture of finite Polya tree distributions given by (5.1), which is obtained as follows. Let I be the unit interval and $\mathcal{X} = \{0, 1\}$. Denote by $B(\alpha_0, \alpha_1)$ the beta-distribution with shape parameters $\alpha_0 \geq 0$ and $\alpha_1 \geq 1$ such that $\alpha_0 + \alpha_1 > 0$, where if $k \in \{0, 1\}$ and $\alpha_k = 0$ then $B(\alpha_0, \alpha_1)$ is concentrated at the point $k - 1$ and hence the value of $\alpha_{1-k} > 0$ plays no role. Suppose $Y = \{Y_{x_1,\ldots,x_n} \mid (x_1, \ldots, x_n) \in \{0, 1\}^n,\ n \in \mathbb{N}\}$, where the $Y_{x_1,\ldots,x_{n-1},0}$s (interpreted as Y_0 if $n = 0$) are independent $B(\alpha_{x_1,\ldots,x_{n-1},0}, \alpha_{x_1,\ldots,x_{n-1},1})$-distributed, $Y_{x_1,\ldots,x_{n-1},1} = 1 - Y_{x_1,\ldots,x_{n-1},0}$, and $\alpha_{x_1,\ldots,x_n} = 0$ if $\ell_{x_1,\ldots,x_{n-1}} > 0$ and $\ell_{x_1,\ldots,x_n} = 0$. Let N be a non-negative integer-valued random variable that is independent of Y. Conditioned on (Y, N), for $n = 0, \ldots, N$, let the X_ns be independent and let $Y_{X_1,\ldots,X_{n-1},0}$ be the probability that $X_n = 0$. Thereby, conditioned on Y, $S = (X_1, \ldots, X_N)$ has been generated. In accordance with Theorem 3.1, conditioned on S, independent of Y, if $N = 0$ then generate $X \sim \text{Unif}(I)$, otherwise generate $E \sim \text{Unif}(0, \ell_{X_1,\ldots,X_N})$ and set $X = a_{X_1,\ldots,X_N} + E$. Thus, conditioned on Y and $N = n$ with $\mathbb{P}(N = n) > 0$, X follows a random PDF $f(\cdot \mid Y, N = n)$, which for $x \in I_{x_1,\ldots,x_n}$ is given by

$$f(x \mid Y, N = n) = \prod_{j=1}^{n} Y_{x_1,\ldots,x_{j-1},0}^{1-x_j} Y_{x_1,\ldots,x_{j-1},1}^{x_j},$$

where $\prod_{j=1}^{0} \cdots$ is interpreted as 1. This random PDF, where $N = n$ is fixed, specifies a finite Polya tree distribution in the sense of [18]; if N was replaced by ∞ when generating $X_1, X_2, \ldots$, the random probability distribution given by $\mathbb{P}(0.X_1X_2 \ldots \in B \mid Y)$ with $B \in \mathcal{B}$ would be a Polya tree distribution as defined in [17]. Arguments for considering instead a finite Polya tree distribution can be found in [18, 27]. For instance, a Polya tree distribution (in contrast to the finite Polya tree distribution) is not always absolutely continuous (unless the αs

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

increase sufficiently rapidly, which leads to problems of robustness [18]), and if it is absolutely continuous then the random PDF is almost surely discontinuous almost everywhere. Note that $X \mid Y$ specifies a random probability distribution $\mathcal{P}$ on $\mathcal{B}$ that is absolutely continuous with random PDF

$$f(x \mid Y) = \sum_{n=0}^{\infty} f(x \mid Y, N = n)\mathbb{P}(N = n) \tag{5.1}$$

(setting $f(x \mid Y, N = n) = 0$ if $\mathbb{P}(N = n) = 0$). The use of $\mathcal{P}$ for statistical modelling and a comparison with other ways of imposing random stopping rules in a Polya tree distribution (including allowing optional stopping as in [27]) will be investigated in future work [3].

Yet another interesting question is how to model random nested partitions. For instance, if I is a unit interval and $\mathcal{X} = \{0, 1\}$ is binary, we may consider a distribution of the lengths $\{\ell_{x_1,\dots,x_n} \mid (x_1, \dots, x_n) \in \{0, 1\}, n \in \mathbb{N}\}$, which is of a similar type to Y above and is independent of (N, Y), and then proceed in a similar way when specifying a model for S conditioned on both Y and the random nested partitions (this will also be investigated in [3]). How should this be extended to multivariate cases (where the ideas behind randomized partitioning schemes in [27] may be a source of inspiration)? How should one create random models for the intervals used to define the Markovian constructions in Section 2.1, including the case of generalized Lüroth series in Example 2.1? These questions are left for future research.

5.2. How important is the Markovian structure?

Birkhoff's ergodic theorem is a major tool in metric number theory when studying ergodic dynamical systems in connection with stationary digit systems $X_1, X_2, \dots$; see, e.g., [6, 7, 24]. Birkhoff's theorem is in close agreement with Corollaries 3.2 and 3.3, Theorem 4.2, and Corollary 4.3, although there are differences. To see this, suppose $X_{>0}$ is stationary and ergodic. Here, ergodicity means that $\mathbb{P}(X \in A) \in \{0, 1\}$ whenever $A \subseteq I$ is a Borel set such that $.x_1x_2 \dots \in A$ implies $.x_2x_3 \dots \in A$. Let $h\colon I \mapsto \mathbb{R}$ be a Borel function, where the mean $\mathbb{E}h(X)$ exists. Then Birkhoff's theorem states that 'the time average converges to the space average': $(1/m)\sum_{n=1}^{m} h(X^{[n]}) \to \mathbb{E}h(X)$ as $m \to \infty$. On the other hand, the results in the present paper state that under certain regularity conditions, but without assuming $X_{>0}$ is stationary, it is not possible to distinguish between the distributions of the remainders $X^{[n]}, X^{[n+1]}, \dots$ when n is larger than a certain random time: at the random time $N + 1$ (under the conditions given in Corollary 3.2) or $\max\{N, s\} + M$ (under the conditions given in Theorem 4.2) and further on in time, the remainders exactly follow the shift-invariant PDF f_{inv}. Moreover, under the conditions given in Corollaries 3.3 and 4.3, the distribution of $X^{[n]}$ converges towards the invariant distribution with respect to the total variation distance.

For this convergence result the assumption that f is LSC almost everywhere may be replaced by other assumptions. For a general PDF f, if there is only one PDF invariant under the shift, namely f_{inv}, and if $\int_I |g^{[n]} - f_{\text{inv}}| \to 0$ for a PDF $g^{[n]}$ corresponding to a smooth PDF g (using a similar notation to $f^{[n]}$ and f in Section 3.2), the triangle inequality and the inequality $\int_I |f^{[n]} - g^{[n]}| \le \int_I |f - g|$ give the convergence result. Moreover, bounds for the rate of convergence are given in [15] for the case of base-q expansions (Example 2.1), in [12] for β-expansions when β is the golden mean (Example 2.2), and in [14] when β is the pseudo golden mean (Example 2.2). For these cases the assumption that f is LSC almost everywhere is indeed not needed.

For cases where the Markovian structure (1.11) is not satisfied for any $s \in \mathbb{N}_0$, it would of course also be interesting to establish similar convergence results, to study the distribution of $R \mid S$ given by (3.6) (which do not depend on f), and if possible to establish a coupling between

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

X and a random time K (depending on f) so that $X^{[K]} \sim f_{\text{inv}}$. Here, one particular interesting case is continued-fraction representations, cf. Example 2.3, where $f_{\text{inv}} = f_{\text{G}}$ is given by (3.7).

Acknowledgements

I am grateful to Anne Marie Svane and Ira Herbst for a careful reading of a draft of this paper and their useful comments. Discussions with Charlene Kalle and Simon Kristensen on various details, including continued fractions, are appreciated. I thank Mario Beraha for pointing my attention to Polya tree distributions.

Funding information

Supported by The Danish Council for Independent Research—Natural Sciences, grant DFF – 10.46540/2032-00005B.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. John Wiley, Chichester.
[2] Barrionuevo, J., Burton, R., Dajani, K. and Kraiikamp, C. (1996). Ergodic properties of generalized Lüroth series. *Acta Arith.* **74**, 311–327.
[3] Beraha, M. and Møller, J. (2025). Generalized finite Polya tree distributions for statistical modelling. In preparation.
[4] Cornean, H., Herbst, I., Møller, J., Støttrup, B. and Studsgaard, K. (2022). Characterization of random variables with stationary digits. *J. Appl. Prob.* **59**, 931–947.
[5] Cornean, H., Herbst, I., Møller, J., Støttrup, B. and Studsgaard, K. (2023). Singular distribution functions for random variables with stationary digits. *Methodol. Comput. Appl. Prob.* **25**, 31.
[6] Dajani, K. and Kalle, C. (2021). *A First Course in Ergodic Theory*. Chapman and Hall/CRC, Boca Raton, FL.
[7] Dajani, K. and Kraaikamp, C. (2002). *Ergodic Theory of Numbers*. Mathematical Association of America, Washington, DC.
[8] Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. I, *Elementary Theory and Methods*. Springer, New York.
[9] Foss, S. and Tweedie, R. (1998). Perfect simulation and backward coupling. *Stoch. Models* **14**, 187–203.
[10] Gauss, C. (1976). *Mathematisches Tagebuch 1796–1814*. Akademische Verlagsgesellschaft, Geest & Portig K.G., Leipzig.
[11] Gelfond, A. (1959). A common property of number systems. *Izv. Akad. Nauk. SSSR. Ser. Mat.* **23**, 809–814.
[12] Herbst, I., Møller, J. and Svane, A. (2023). The asymptotic distribution of the remainder in a certain base-β expansion. Preprint, arXiv:2312.09652.
[13] Herbst, I., Møller, J. and Svane, A. (2023). How many digits are needed? Preprint, arXiv:2307.06685.
[14] Herbst, I., Møller, J. and Svane, A. (2025). The asymptotic distribution of the scaled remainder for pseudo golden ratio expansions of a continuous random variable. To appear in *Methodol. Comput. Appl. Prob.* **27**.
[15] Herbst, I. W., Møller, J. and Svane, A. M. (2024). How many digits are needed? *Methodol. Comput. Appl. Prob.* **26**, 5.
[16] Khintchine, A. (1963). *Continued Fractions*. Noordhoff, Groningen.
[17] Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *Ann. Statist.* **20**, 1222–1235.
[18] Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *Ann. Statist.* **22**, 1161–1176.
[19] Lüroth, J. (1883). Ueber eine eindeutige Entwickelung von Zahlen in eine unendliche Reihe. *Math. Annalen* **21**, 411–423.

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press

[20] MEYN, S. AND TWEEDIE, R. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
[21] MØLLER, J. AND WAAGEPETERSEN, R. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, FL.
[22] PARRY, W. (1960). On the β-expansions of real numbers. *Acta Math. Acad. Sci. Hungar.* **11**, 401–416.
[23] RÉNYI, A. (1957). Representations for real numbers and their ergodic properties. *Acta Math. Acad. Sci. Hungar.* **8**, 477–493.
[24] SCHWEIGER, F. (1995). *Ergodic Theory of Fibered Systems and Metric Number Theory*. Clarendon Press, Oxford.
[25] THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer, New York.
[26] WILSON, D. (2000). How to couple from the past using a read-once source of randomness. *Random Structures Algorithms* **16**, 85–113.
[27] WONG, W. AND MA, L. (2010). Optional Pólya tree and Bayesian inference. *Ann. Statist.* **38**, 1433–1459.

https://doi.org/10.1017/jpr.2025.3 Published online by Cambridge University Press