Adv Data Anal Classif (2009) 3:281–290
DOI 10.1007/s11634-009-0048-5

REGULAR ARTICLE

The value of the last digit: statistical fraud detection with digit analysis

Stephan Dlugosz · Ulrich Müller-Funk

Received: 15 May 2009 / Revised: 17 October 2009 / Accepted: 28 October 2009 /
Published online: 13 November 2009

Abstract Digit distributions are a popular tool for the detection of tax payers' noncompliance and other fraud. In the early stage of digital analysis, Nigrini and Mittermaier (A J Pract Theory 16(2):52–67, 1997) made use of Benford's Law (Benford in Am Philos Soc 78:551–572, 1938) as a natural reference distribution. A justification of that hypothesis is only known for multiplicative sequences (Schatte in J Inf Process Cyber EIK 24:443–455, 1988). In applications, most of the number generating processes are of an additive nature and no single choice of 'an universal first-digit law' seems to be plausible (Scott and Fasli in Benford's law: an empirical investigation and a novel explanation. CSM Technical Report 349, Department of Computer Science, University of Essex, http://cswww.essex.ac.uk/technical-reports/2001/CSM-349.pdf, 2001). In that situation, some practioneers (e.g. financial authorities) take recourse to a last digit analysis based on the hypothesis of a Laplace distribution. We prove that last digits are approximately uniform for distributions with an absolutely continuous distribution function. From a practical perspective, that result, of course, is only moderately interesting. For that reason, we derive a result for 'certain' sums of lattice-variables as well. That justification is provided in terms of stationary distributions.

Keywords Fraud detection · Last digits · Digit analysis · Benford's law

Mathematics Subject Classification (2000) 60B10 · 62P20 · 91B99

S. Dlugosz (✉)
ZEW, L7, 1, 68161 Mannheim, Germany
e-mail: dlugosz@zew.de

U. Müller-Funk
ERCIS, Leornado-Campus 3, 48149 Münster, Germany
e-mail: ulrich.mueller-funk@ercis.de

1 Introduction

By fraud detection we mean techniques that help to single out suspects in a non-technical way, i.e., to preclassify people on the basis of operative data (e.g., lists of prices, taxable amounts, …) without entering into a detailed technical investigation. The idea behind is that fraudsters might inadvertently violate some formal aspects. One such aspect concerns the distribution of numbers. More precisely, those techniques are not intended to identify a single irregular entry—as it is intended in credit card fraud detection (Bolton and Hand 2002)—instead they are used to decide, whether a whole dataset is contaminated by abnormal numbers. An approach to that effects dates back to M. Nigrini who presumed that mantissas resp. first digits had to obey Benford's law and that a deviation from that law in a table of figures hints at an abnormal behavior (Nigrini and Mittermaier 1997; Benford 1938).

Benford's law, however, is closely related to the multiplicative structure of numbers and can only be supported in rather special cases (Schatte 1988). Most tables, however, result from adding up costs, turnovers etc., and show a completely different distribution of numbers. Some tax offices surmised that the last digits should follow a uniform distribution, a hypothesis that could be checked with the help of the χ^2 test. The present paper supports that point of view. For some readers the whole approach might be rather doubtful as it is completely based on formal peculiarities. Alternative methods relying on pairwise comparisons, however, are hard to implement in a feasible way and, moreover, are not legally accepted in some countries (like Germany). For that reason, authorities are thrown back to the detection of deviations from some sort of 'normality'—to be justified in a mathematical way.

After introducing the necessary notation in Sect. 2, we present two theorems for the uniform distribution of two different concepts of 'last' digits in Sect. 3. The first theorem investigates more precisely the intuition of having a more and more uniformly distributed 'lower end' of mantissas. In the second case, we look at numbers from the integer range that are generated by some additive process like it is common in bookkeeping, especially for calculating transaction volumes. The other theorems are surely the more interesting results for investigators. The proofs of theorems and lemmas is deferred to Sect. 5.

2 Mantissas and last digits

Let $2 \leq b \in \mathbb{N}$ be a 'base' for the expansion of numbers and $M = [1, b[$ ('set of mantissas'). The most prominent mantissa distribution, Benford's law, is defined according to

$$P_H([c, d[) = \log_b(d) - \log_b(c), \quad 1 \leq c < d < b.$$

All $x \in \mathbb{R}^{>0}$ allow for a unique representation

$$x = m(x)b^{e(x)}$$

with $m(x) = m_1(x)\,m_2(x)\,m_3(x)\ldots \in M$ ('mantissa') and $e(x) \in \mathbb{Z}$ ('exponent'). For example, the number $x = 12.345$ in the usual decimal system (i.e., $b = 10$) has the representation 1.2345×10^1.

The probability law on $\{1, \ldots, b-1\}$ resp. $\{0, \ldots, b-1\}$ induced by the kth digit m_k is called the kth Benford distribution, e.g.,

$$
\begin{aligned}
P_H\,(m_1(X) = j) &= \log_b(1 + j^{-1}) \\
P_H\,(m_2(X) = j) &= \sum_{i=1}^{b-1} \log_b\left(1 + (ib + j)^{-1}\right) \\
P_H\,(m_k(X) = j) &= \sum_{i_1=1}^{b-1} \sum_{i_2=0}^{b-1} \cdots \sum_{i_{k-1}=0}^{b-1} \log_b\left(1 + (i_1 b^{k-1} + \cdots + i_{k-1} b + j)^{-1}\right).
\end{aligned}
$$

There are similar expressions for the common probabilities of two or more digits.

Remark $M = [1, b[$ is equipped with multiplication $\mod b$ and the quotient topology. Those specifications turn M into a compact Abelian group with the normalized Haar measure P_H. Accordingly, P_H can be perceived as some sort of 'uniform distribution' as well.

Now, let X be a positive random variable expressing some real-world phenomenon, e.g., the amount of yearly taxes paid by some person. Digit analysis for fraud detection is based on specifications of the probability law $\mathscr{L}(m_k(X))$. Nigrini and Mittermaier (1997) originally claimed that $\mathscr{L}(m_1(X))$ 'typically' obeys Benford's law. Various studies shed doubt on the universality of that distribution, in particular with data sets related to additive—and not multiplicative operations (Scott and Fasli 2001; Bolton and Hand 2002). For that reason, we study the last digits and show that they are approximatively uniformly distributed.

For these (quasi) integer-valued data, we need some further notation:[1]

$$
\begin{aligned}
d_k^{(b)}(x) : \mathbb{N}_0 &\to B = \{0, \ldots, b-1\} \\
x &\mapsto \left\lfloor \frac{x}{b^{k-1}} \right\rfloor - \left\lfloor \frac{x}{b^k} \right\rfloor
\end{aligned}
$$

denote the kth last digit of a number $x \in \mathbb{N}$, which consists of more than k digits, in a number system with b different digits ranging from zero to $b-1$. For example, the last digit of the number $x = 12{,}345$ is 5 and the 4th last digit is 2.

For simplicity of notation, we will use $d_k(x)$ instead of $d_k^{(b)}(x)$ for the kth last digit and $d_{2,1}(x)$ instead of $d_1^{(b \cdot b)}(x)$ for the combination of the second last and the last digit throughout this paper.

[1] For the purpose of generality, B denotes a finite set of numbers that, with some addition operation $+$, forms a circulant finite group.

3 Uniform distribution of the last digits

At first, we show that a block of digits becomes independent and uniformly distributed in the limit.

Theorem 1 *Let $X > 0$ be a random attribute with distribution function F and continuous density $f = F'$. Then, for all $\ell > 1$ and $1 \le k_r < b$ $(1 \le r \le \ell)$:*

$$P\left(m_n(X) = k_1, \ldots, m_{n+l-1}(X) = k_\ell\right) \to_{n\to\infty} b^{-l}$$

The result can be supplemented in various ways. Assuming that f is Lipschitzian, one can establish the rate of convergence $O(n^{-1})$, for instance.

In real world problems, the convergence shown in Theorem 1 is appropriate only for very special cases as the number of digits, i.e., the 'size' of the number has to be very large (or very precise). Therefore, for numbers with a limited number of digit positions another set of criteria is needed to decide whether the last digits obey the discrete uniform distribution. Many financial figures are results of a 'summation' process, i.e., they are given by $\sum_i x_i$ with $x_i \in \mathbb{N}$, e.g., turnover. This basic idea leads to another theorem that states that the last digit is uniformly distributed on B with the help of Markov chains (Kemeny and Snell 1976). We will derive a set of criteria that justifies the uniform distribution for the last digits and can be checked easily.

In order to outline the idea behind this approach, let X_i be a sequence of i.i.d. random variables on $\mathbb{N}$. Interpreted in an economic context, these random variables represent selling positions of a certain period of time. With respect to addition, the last digits form a finite Abelian group on B (think of $B = \{0, \ldots, 9\}$ in most cases).

Each addition operation can be seen as a transition on a finite Markov chain with the following transition matrix: Let vector $\mathbf{v} \in [0, 1]^z$ describe the probabilities $v_z = P\left(d_1(X_i) = z\right)$. The vector $\mathbf{v}^{(n)} \in [0, 1]^b$ is given by $v_z^{(n)} = P\left(d_1(\sum_{i=1}^n X_i) = z\right)$. Let

$$A = \begin{pmatrix} v_0 & v_{b-1} & v_{b-2} & \cdots & v_1 \\ v_1 & v_0 & v_{b-1} & \cdots & v_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ v_{b-2} & \cdots & v_1 & v_0 & v_{b-1} \\ v_{b-1} & v_{b-2} & \cdots & v_1 & v_0 \end{pmatrix} \tag{1}$$

be the transition probability matrix describing the transition of the last digit performed by a single addition. A is double-stochastic, i.e., $v_z \ge 0\ \forall z$ and $\sum_z v_z = 1$.

As the X_i are i.i.d., the Markov property is given. Now, the probabilities

$$P\left(d_1\left(\sum_{i=1}^{n-1} X_i + X_n\right) = z\right)$$

can be calculated via:

$$\mathbf{v}^{(n)} = A\mathbf{v}^{(n-1)}$$

Furthermore, matrix A is a double-stochastic circulant matrix, i.e., a special form of a Toeplitz matrix (Gray 2006).

Obviously, the last digits of many numbers in financial application, which are results of a longer summation process, are a result of this finite homogeneous Markov chain. To show the ergodicity of this Markov chain, we have to prove that it is irreducible and aperiodic (Kemeny and Snell 1976). The stationary distribution of this Markov chain would be the discrete uniform distribution, because A is double-stochastic.

Lemma 1 *Let A be as in (1). Furthermore, put $I = \{i \in B : v_i > 0\}$. A is irreducible iff there exists $i \in I$ with $\gcd(i, b) = 1$ or (alternatively) $i, j \in I, i \neq j$ with $\gcd(i + j \mod b, b) = 1$.*

Aperiodicity cannot be shown that easy and we have to do some more work.

Lemma 2 *The structure of A is given by (1).*

a) *A is aperiodic if $v_0 > 0$.*
b) *Let $I = \{i \in B : v_i > 0\}$. If $\exists i, j \in I, 1 \neq j$, $\gcd(|i - j|, b) = 1$, then A is aperiodic.*

Unfortunately, Lemma 2 does not give us any information on aperiodicity in other cases than mentioned. To turn argumentation around, the following Lemma 3 shows us some cases, where the Markov chain with transition matrix A is periodic.

Lemma 3 *Let A be as in (1). Furthermore, put $I = \{i \in B : v_i > 0\}$. Let $P = \{p_1, \ldots, p_{|P|}\}$ be (the pairwise different) prime divisors of b. If $\forall i, j \in I : \prod_{\ell=1}^{P} p_\ell^{f_\ell} \mid |j - i|$ with $p_\ell \in P$ and $\mathbb{N} \ni f_\ell \leq e_\ell \quad \forall \ell$ and if $\prod_{\ell=1}^{P} p_\ell^{f_\ell} \nmid i$ and $\prod_{\ell=1}^{P} p_\ell^{f_\ell} \nmid j$, then A is periodic.*

The following fact follows directly from Lemmas 1, 2 and 3.

Theorem 2 *Let X_i be a sequence of i.i.d. random variables on $\mathbb{N}$ and $i \in \{0, \ldots, n\}$ with $n \in \mathbb{N}$.*

If

(1) *we have*
   - *$P(d_1(X_i) = z) \neq 0$ for any $z \in B$ with $\gcd(z, b) = 1$ or*
   - *$P(d_1(X_i) = z_1) \neq 0 \neq P(d_1(X_i) = z_2)$ for some $z_1, z_2 \in B$ and $z_1 \neq z_2$ with $\gcd(z_1 + z_2 \mod b, b) = 1$*

(2) *and*
   - *$P(d_1(X_i) = 0) \neq 0$ or*
   - *$P(d_1(X_i) = z_1) \neq 0$ and $P(d_1(X_i) = z_2) \neq 0$ for $z_1, z_2 \in B$ with $z_1 \neq z_2$ and $\gcd(|z_1 - z_2|, b) = 1$,*

then:

$$P\left(d_1\left(\sum_{i=1}^{n} X_i\right) = z\right) \to_{n\to\infty} b^{-1}$$

From a practical perspective, the number of transitions needed to come close to the stationary distribution is very important (e.g., if there are only a few sale positions for some days of the year). As for that aspect we state:

Theorem 3 *Let A be ergodic and let*

$$(\lambda_0, \lambda_1, \ldots, \lambda_{b-1}) = \left(1, \sum_z v_z e^{-\frac{2\pi i z}{b}}, \ldots, \sum_z v_z e^{-\frac{2(b-1)\pi i z}{b}}\right)$$

denote the eigenvalues of A and $p^n(z)$ the zth entry of the vector $p^n = A^n(1, 0, \ldots, 0)^t \in \mathbb{R}^b$.

Then, it is true that:

$$\frac{1}{b}\sum_z \left(p^n(z) - \frac{1}{b}\right)^2 = \frac{1}{b^2}\sum_{m=1}^{b-1} |\lambda_m|^{2n} \leq \frac{b-1}{b^2}(\lambda_*)^n$$

where $\lambda_* < 1$ denotes the second largest eigenvalue of A (the largest is $\lambda_0 = 1$).

Note, that for the convergence result, A has not necessarily to be known exactly. All we have to ensure is that some entries of A are positive. Regarding the speed of convergence, however, we have to calculate the eigenvalues of A and thus have to know the size of its entries. If these data are missing, we can often estimate them using some additional data (e.g., using the data from competitors on fractions of certain products).

In a more general setting, we might not only be interested in the last digit, but also in the second, third, … last one. Especially in retail, we often have psychologically motivated prices ending with 9. This results in a periodic Markov chain. On the other hand, there might be (e.g., for simplicity of cashing) a tendency to use last digits like 0 and 5. This causes a reducible Markov chain. In these cases the Markov chain describing the last digits is not ergodic. Therefore, the second or even third last digit is used for analysis. This is why we extend the analysis to the second last digit. Results are also true for third, forth etc., last digits and the necessary proofs follow the same ideas.

In general, ergodicity of the Markov chain for the last digits implies ergodicity of the Markov chain for the two last digits:

Theorem 4 *Assume the validity of Theorem 2, for the last digits, i.e., on B. Then it also holds for the last two digits (combined), i.e., on $B \times B$.*

Let us first turn to the case of a reducible Markov chain on B, e.g., if there are only 'psychologically-motivated' prices.

Theorem 5 *The structure of A_1 is given by (1) for B and the structure of A_2 is given by (1) for B^2. If A_1 is reducible and if there is a $z \in B\backslash\{0\}$ with $P(d_1(X) = z) > 0$ and if A_2 is aperiodic, then it holds for the second last digit of the sum $\sum_i X_i$:*

$$P\left(d_2\left(\sum_{i=1}^n X_i\right) = z\right) \to_{n\to\infty} b^{-1}.$$

In the case of a periodic Markov chain on B we cannot use such a simple argument due to the fact that A_1 is periodic iff A_2 is periodic. Nevertheless, we can show the following result for a periodic Markov chain on A_1:

Theorem 6 *The structure of A_1 is given by (1) for B and the structure of A_2 is given by (1) for $B \times B$. If A_1 is periodic with period $p \in P = \{p : p|b\}$ (cf. 3) and if A_2 is irreducible, we get*

$$P\left(d_2\left(\sum_{i=1}^{n} X_i\right) = z\right) \to_{n\to\infty} b^{-1}.$$

Theorems 5 and 6 can be combined to show that the second last digits are uniformly distributed in more complex cases.

Furthermore, we are in need of a theorem analogue to Theorem 3 that complements Theorems 4, 5 and 6. First we define:

Definition 1 (Projection matrix from B^2 to B (second last digit)) Let

$$P_b = \operatorname{diag}(\underbrace{\mathbf{1}_b, \ldots, \mathbf{1}_b}_{b \text{ times}}) \in \mathbb{R}^{b\times b^2} \text{ with } \mathbf{1}_b = \underbrace{(1, \ldots, 1)}_{b \text{ times}}.$$

P_b is the projection matrix, which maps each stochastic vector from $\mathbb{R}^{b^2}$ to a stochastic vector on $\mathbb{R}^b$, whereby b entries are aggregated to a single entry.

Now, we can state:

Theorem 7 *Let $A_2 \in \mathbb{R}^{b^2\times b^2}$ be a double-stochastic, ergodic, circulant transition matrix and $z \in B$.*

With $P = P_b$ as in Definition 1 holds:

$$b \cdot \left|\sum_z (Pp^n)(z) - \frac{1}{b}\right|^2 \le \sum_z \left(\frac{1}{b^2}\sum_{m=1}^{b^2-1} |(Pv_m)(z)||(\lambda_m)^n|\right)^2,$$

with $\lambda_0, \ldots, \lambda_{b^2-1}$ denote the b^2 eigenvalues of A_2.

4 Conclusion

Assume that we have information on prices and daily turnovers of a restaurant. Let $X_{ij} \in \mathbb{N}$ be the turnover associated with the i-th customer on day j (measured in the smallest unit available, i.e., pence, cent etc.). Then $X_j = \sum_i X_{ij}$ represents the daily turnover. We are interested in the distribution of the last digits of X_j and we want to test whether this distribution is irregular or not. Under the conditions of Theorem 2, the 'regular' last distribution of X_j is the discrete uniform distribution in the limit. Theorem 3 may help us to decide, whether X_j is close enough to this asymptotic result in order to apply the χ^2-test on uniform distribution of the last digits. In the case of psychologically motivated prices, we can make use of the Theorems 5, 6 and 7 to obtain similar results for the second, third etc. last digit. This approach has successfully been applied to data from ice cream parlors.

We conclude that fraud detection is very well possible on the basis of formal peculiarities, provided the structure of the data is appropriately taken into account.

5 Technicalities

Proof of Theorem 1 For the sake of (notational) convenience we only prove the case $\ell = 1$—the argument carries over to higher dimensions. For $i \in \mathbb{Z}$, $1 \le j_1 < b$, $0 \le k$, $j_r < b$ $(r > 2)$ put

- $t_n(i, j) = \sum_{r=1}^{n} j_r b^{i-r+1}$
- $I_n(i, j, k) = [t_n(i, j) + kb^{i-n}, t_n(i, j) + (k+1)b^{i-n}[$

Accordingly,

$$P\,(m_{n+1}(X) = k) = \sum_{i \in Z} \sum_{j_1=1}^{b-1} \sum_{j_2=0}^{b-1} \cdots \sum_{j_n=0}^{b-1} \int_{I_n(i,j,k)} f(t)dt$$

Choose $\varepsilon > 0$. For some $q > 1$:

- $P\,(e(X) > q) = P(X \ge b^{q+1}) < \varepsilon$
- $P\,(e(X) < -q) = P(X \le b^{-q}) < \varepsilon$

For n sufficiently large, moreover,

$$\left| f(t) - f\left(t_n(i, j) + kb^{i-n}\right) \right| < \frac{\varepsilon}{2q}$$

independently of i, j. Note, that the points $t_n(i, j)$ form a mesh of size b^{i-n+1} and that $t_n(i, j) + kb^{i-n}$ can be perceived as a set of supporting points for a Riemannian sum—for every fixed $-q \le i \le q$. Now

- $\left| \sum_{j_1=1}^{b-1} \sum_{j_2=0}^{b-1} \cdots \sum_{j_n=0}^{b-1} \int_{I_n(i,j,k)} - b^{i-n} \sum_{j_1=1}^{b-1} \sum_{j_2=0}^{b-1} \cdots \sum_{j_n=0}^{b-1} f\left(t_n(i, j) + kb^{i-n}\right) \right| < \varepsilon$
- $b^{i-n+1} \sum_{j_1=1}^{b-1} \sum_{j_2=0}^{b-1} \cdots \sum_{j_n=0}^{b-1} f\left(t_n(i, j) + kb^{i-n}\right) \to \int_{b^i}^{b^{i+1}} f(t)dt$

As a consequence,

$$\overline{\lim}_n \left| P\,(m_{n+1}(X) = k) - \frac{1}{b} \int_{b^{-q}}^{b^{q+1}} f(t)dt \right| < 3\varepsilon$$

The assertion follows from $q \to \infty$. □

Proof of Lemma 1 $I \ne \emptyset$ as A is double-stochastic. The second case is covered by the first using A^2 instead of A, and the proof of the first case is as follows:

'$\Rightarrow$' Let $I = \{i\}$ and $\gcd(i, b) = k > 1$. Then we have $i > 1$ and only the states $n \cdot k \mod b$ with $n \in \{0, \ldots, \frac{b}{k} - 1\}$ are reachable. The state 1 could be reduced from the Markov chain.

'$\Leftarrow$' $\exists i \in I : \gcd(i, b) = 1$. Then we have $n \cdot i \mod b \neq 0$ for all $n \in B \backslash \{0\}$ and for A^2 we have $v_{2i \mod b} > 0$, i.e., one more reachable state. Mathematical induction ending at $b \cdot i \mod b = 0$ (after b steps) supports the statement. □

Proof of Lemma 2 W.l.o.g. let $i - j = k$. As $b \cdot j \mod b = 0$ we have $(b-i) \cdot j + j \cdot i \mod b = 0$. This implies that the minimal period is equal or smaller than $\gcd(b, b-k)$. With $\gcd(k, b) = 1$ we get $\gcd(b, b-k) = 1$. If there is no common divisor of b and k, it follows that b and $b - k$, $0 < k < b$ cannot have a common divisor. □

Proof of Lemma 3 W.l.o.g. let $j > i$ and f_ℓ such that $j = i + \prod_{\ell=1}^{P} p_\ell^{f_\ell}$. Obviously it is true that $b = \prod_{\ell=1}^{P} p_\ell^{f_\ell} \cdot x$ and let $k, \ell \in \mathbb{N}_0$ such that:

$$k \cdot i + l \cdot \left(i + \prod_{\ell=1}^{P} p_\ell^{f_\ell} \right) = m \cdot b$$

$$\Leftrightarrow (k + l) \cdot i = m \cdot \prod_{\ell=1}^{P} p_\ell^{f_\ell} \cdot x - l \cdot \prod_{\ell=1}^{P} p_\ell^{f_\ell} = (m \cdot x - l) \cdot \prod_{\ell=1}^{P} p_\ell^{f_\ell}$$

The right hand side $\prod_{\ell=1}^{P} p_\ell^{f_\ell}$ is divisible, and thus $\prod_{\ell=1}^{P} p_\ell^{f_\ell} \mid (k + l)$ holds, because of assuming $\prod_{\ell=1}^{P} p_\ell^{f_\ell} \nmid i$. Therefore, we have for all $d_o = \prod_{\ell=1}^{P} p_\ell^{f_\ell}$ for all 0. □

Proof of Theorem 3 The eigenvalues are defined due to the fact, that A is a Toeplitz matrix (Gray 2006). With Rosenthal (1995) (Fact 3, p. 391) and that in his notation $a_m = \frac{1}{b}$ and because the eigenvectors form an orthonormal basis. □

Proof of Theorem 4 Using the notations from Theorem 2 applied to $B \times B$, it is trivial, that

- $\gcd(z, b) = 1 \rightarrow \gcd(z, b^2) = 1$
  because $|B \times B| = b^2$ and the divisors of b and b^2 are the same.
- $\gcd(z_1 + z_2 \mod b, b) = 1 \rightarrow \gcd(z_1 + z_2 \mod b^2, b^2) = 1$
  for the same reason.
- $\gcd(|z_1 - z_2|, b) = 1 \rightarrow \gcd(|z_1 - z_2|, b^2) = 1$
  ditto.

□

Proof of Theorem 5 The Markov chain based on matrix A_2 is not ergodic, because it is reducible. Nevertheless, the state 00 can be reached because there is a $z \in B_1 \backslash \{0\}$ with $P(d_1(X) = z) > 0$ and $\nexists_{i \in I} \gcd(i, b) = 1$, i.e., $\exists_{j \in \mathbb{N}} i \cdot j \mod b = 0$ and the starting state in the considered case here is 00, obviously. We delete the unreachable states from the Markov chain and this reduced Markov chain is ergodic. $P(d_1(X) | d_2(X) = z) = P(d_1(X))$ gives the result. □

Proof of Theorem 6 Periodicity of A_2 follows from the periodicity of A_1. Now, we can divide the state space $B \times B$ into subspaces S_0 to S_{p-1}, with the property that

$P(S_i|S_j) = 1$ if $i = (j+1) \mod (m+1)$ and zero otherwise. A Markov chain living on one of the state spaces S_0 to S_{p-1} is ergodic and the corresponding transition matrices are double-stochastic. Furthermore, these subspaces S_0, S_{p-1} are of the same size and thus, the stationary distributions are the same. As the elements of B are also equally 'distributed' over the subspaces, $P(S_i) = P(S_j)$ for all i, j and we have for any $z \in S_j$ for all j:

$$P\left(d_{2,1}\left(\sum_{i=1}^{n} X_i\right) = z\right) \to_{n\to\infty} pb^{-1}.$$

□

Proof of Theorem 7, c.f. Fact 3, p. 391f. of Rosenthal (1995) Because eigenvalues and -vectors exist for circulant matrices and because A is ergodic and doubly stochastic, we can state for the initial distribution of the Markov chain π_0:

$$\pi_0 = \frac{1}{b^2} \sum_{m=0}^{b^2-1} v_m$$

With $Pp^n = P\pi_0 A_2^n$, $v_m A_2 = \lambda v_m$ and $\lambda_0 = 1$ we have:

$$Pp^n = \frac{1}{b^2}\left(Pv_0 + \sum_{m=1}^{b^2-1} Pv_m(\lambda_m)^n\right)$$

Because $\lambda_* < 1$ and with a similar argument to Fact 3 of Rosenthal (1995) follows:

$$(Pp^n)(z) - \frac{1}{b} = \frac{1}{b^2} \sum_{m=1}^{b^2-1} (Pv_m)(z)(\lambda_m)^n$$

Apply the Triangle inequality and sum over z. □

References

Benford F (1938) The law of anomalous numbers. Proc Am Philos Soc 78:551–572

Bolton RJ, Hand DJ (2002) Statistical fraud detection: a review (with discussion). Stat Sci 17(3):235–255

Gray RM (2006) Toeplitz and circulant matrices: a review. Found Trends Commun Inf Theory 2(3):155–239

Kemeny JG, Snell JL (1976) Finite Markov Chains. Springer, NewYork

Nigrini MJ, Mittermaier LJ (1997) The use of Benford's law as an aid in analytical procedures: audit. A J Pract Theory 16(2):52–67

Rosenthal JS (1995) Convergence rates for Markov chains. SIAM Rev 37(3):387–405

Schatte P (1988) On mantisse distributions in computing and Benford's law. J Inf Process Cyber EIK 24:443–455

Scott PD, Fasli M (2001) Benford's law: an empirical investigation and a novel explanation. Technical report. CSM Technical Report 349, Department of Computer Science, University of Essex, http://cswww.essex.ac.uk/technical-reports/2001/CSM-349.pdf