RESEARCH

How Many Digits are Needed?

Ira W. Herbst^1 · Jesper Møller^2 · Anne Marie Svane^2

Abstract

Let $X_1, X_2, ...$ be the digits in the base-q expansion of a random variable X defined on $[0, 1)$ where $q \geq 2$ is an integer. For $n = 1, 2, ...$, we study the probability distribution P_n of the (scaled) remainder $T^n(X) = \sum_{k=n+1}^{\infty} X_k q^{n-k}$: If X has an absolutely continuous CDF then P_n converges in the total variation metric to the Lebesgue measure μ on the unit interval. Under weak smoothness conditions we establish first a coupling between X and a non-negative integer valued random variable N so that $T^N(X)$ follows μ and is independent of $(X_1, ..., X_N)$, and second exponentially fast convergence of P_n and its PDF f_n. We discuss how many digits are needed and show examples of our results.

Keywords Asymptotic distribution · Coupling · Exponential convergence rate · Extended Newcomb-Benford law · Remainder of a digit expansion · Total variation distance

Mathematics Subject Classification 60F25 · 62E17 · 37A50

1 Introduction

Let X be a random variable so that $0 \leq X < 1$, and for $x \in \mathbb{R}$, let $F(x) = \mathrm{P}(X \leq x)$ be the cumulative distribution function (CDF) of X. For a given integer $q \geq 2$, we consider the base-q transformation $T : [0, 1) \mapsto [0, 1)$ given by

$$T(x) = xq - \lfloor xq \rfloor \tag{1}$$

✉ Jesper Møller
  jm@math.aau.dk

  Ira W. Herbst
  iwh@virginia.edu

  Anne Marie Svane
  annemarie@math.aau.dk

^{1} Department of Mathematics, University of Virginia, 141 Cabell Drive, Kerchof Hall, Charlottesville VA 22903, Virginia, USA

^{2} Department of Mathematical Sciences, Aalborg University, Skjernvej 4A, Aalborg Ø DK-9220, Denmark

where $\lfloor \cdot \rfloor$ is the floor function (so $\lfloor xq \rfloor$ is the integer part of xq). For $n = 1, 2, ...$, let $T^n = T \circ \cdots \circ T$ denote the composition of T with itself n times and define

$$X_n = \lfloor T^{n-1}(X)q \rfloor \tag{2}$$

where $T^0(X) = X$. Then

$$X = \sum_{n=1}^{\infty} X_n q^{-n} \tag{3}$$

is the base-q expansion of X with digits $X_1, X_2, \ldots$. Note that X is in a one-to-one correspondence to the first n digits $(X_1, ..., X_n)$ together with $T^n(X) = \sum_{k=n+1}^{\infty} X_k q^{n-k}$, which is the remainder multiplied by q^n. Let μ denote Lebesgue measure on $[0, 1)$, P_n the probability distribution of $T^n(X)$ and F_n its CDF, so X follows P_0 and has CDF $F_0 = F$. The following facts are well-known (see Cornean et al. 2022 and the references therein):

(a) $P_0 = P_1$ (i.e., invariance in distribution under T) is equivalent to stationarity of the process $X_1, X_2, \ldots$.
(b) $P_0 = P_1$ and F is absolutely continuous if and only if $P_0 = \mu$.
(c) $P_0 = \mu$ if and only if $X_1, X_2, \ldots$ are independent and uniformly distributed on $\{0, 1, ..., q - 1\}$.

Items (a)–(c) together with the fact that T is ergodic with respect to μ are used in metric number theory (see Dajani and Kalle 2021; Schweiger 1995 and the references therein) to establish properties such as 'for Lebesgue almost all numbers between 0 and 1, the relative frequency of any finite combination of digits of a given length n and which occurs among the first $m > n$ digits converges to q^{-n} as $m \to \infty$' (which is basically the definition of a normal number in base-q, cf. Borel (1909)). To the best of our knowledge, less (or perhaps no) attention has been paid to the asymptotic behaviour of the (scaled) remainder $T^n(X)$ as $n \to \infty$. This paper fills this gap.

Assuming F is absolutely continuous with a probability density function (PDF) f we establish the following. We start in Section 2 to consider a special case of f where $T^n(X)$ follows exactly μ when n is sufficiently large. Then in Section 3, under a weak assumption on f, we specify an interesting coupling construction involving a non-negative integer-valued random variable N so that $T^N(X)$ follows exactly μ and is independent of $(X_1, ..., X_N)$. Moreover, in Section 4, we show that $\lim_{n \to \infty} d_{\text{TV}}(P_n, \mu) = 0$ where d_{TV} is the total variation metric (as given later in (12)). Because of these results, if in an experiment a realization of X is observed and the first n digits are kept, and if (so far) the only model assumption is absolute continuity of F, then the remainder rescaled by q^n is at least approximately uniformly distributed when n is large. Since we interpret the uniform distribution as the case of complete randomness, no essential information about the distribution is lost. On the other hand, if the distribution of the remainder is far from uniform, this may indicate that the distribution one is trying to find has finer structure that one is missing by looking only at the first n digits. We return to this issue in Section 5 when discussing sufficiency and ancillarity. Furthermore, in Section 4 we study the convergence rate of $d_{\text{TV}}(P_n, \mu)$ and other related properties. In Section 5, we illustrate our results from Sections 3 and 4 when F follows the extended Newcomb-Benford law (Example 1).

The present paper is a shorter version of Herbst et al. (2023) to which we refer for further results, comments, and examples. In particular, Herbst et al. (2023) gives an extension of our convergence results in Section 4 to the case where X is a multivariate random variable with values in the k-dimensional unit cube $[0, 1)^k$ and each of the k coordinates of X is transformed

by T. We plan in a future paper to study the asymptotic behaviour of the remainder in other expansions, including a certain base-β expansion of a random variable, namely when q is replaced by $\beta = (1 + \sqrt{5})/2$ (the golden ratio) in all places above.

2 Preliminaries

Let again the situation be as in (1)–(3). The following lemma is true in general (i.e., without assuming F is absolutely continuous). As in Cornean et al. (2022), we define a base-q fraction in $[0, 1)$ to be a number of the form $\sum_{k=1}^{n} j_k q^{-k}$ with $(j_1, ..., j_n) \in \{0, 1, ..., q - 1\}^n$ and $n \in \mathbb{N}$.

Lemma 2.1 *If F has no jump at any base-q fraction in $[0, 1)$ then for every $x \in [0, 1]$,*

$$F_n(x) = \sum_{j=0}^{q^n - 1} F(q^{-n}(j + x)) - F(q^{-n} j). \tag{4}$$

Proof Clearly, (4) holds for $x = 1$, so let $0 \le x < 1$. For $j_1, ..., j_n \in \{0, 1, ..., q - 1\}$ and $j = \sum_{i=1}^{n} j_i q^{n-i}$, the event that $X_1 = j_1, ..., X_n = j_n$, and $T^n(X) \le x$ is the same as the event that $q^{-n} j \le X < q^{-n}(j + 1)$ and $X \le q^{-n}(j + x)$. Hence, since $0 \le x < 1$,

$$F_n(x) = \sum_{j=0}^{q^n - 1} P(q^{-n} j \le X \le q^{-n}(j + x))$$

whereby (4) follows since $F(x)$ has no jumps at the base-q fractions. □

The property that F has no jump at any base-q fraction is of course satisfied when F is continuous.

For the remainder of this section and the following Sections 3, 4 and 5 we assume that X has a probability density function (PDF) f concentrated on $(0, 1)$, meaning that F is absolutely continuous with $F(x) = \int_{-\infty}^{x} f(t) \, dt$ for all $x \in \mathbb{R}$. Then, by (4), F_n is absolutely continuous with PDF

$$f_n(x) = q^{-n} \sum_{j=0}^{q^n - 1} f(q^{-n}(j + x)) \tag{5}$$

for $0 < x < 1$.

In the following special case of f, convergence of P_n is obtained within a finite number of steps.

Proposition 2.2 *Let $m \ge 1$ be an integer. Then $P_m = \mu$ (and hence $P_n = \mu$ for $n = m$, $m + 1, ...$) if and only if for all $k \in \{0, 1, ..., q^m - 1\}$ and Lebesgue almost every $u \in [0, 1)$,*

$$f((k + u)q^{-m}) = q^m P\left(\sum_{i=1}^{m} X_i q^{m-i} = k \,\middle|\, T^m(X) = u\right) f_m(u). \tag{6}$$

In particular, if f is constant Lebesgue almost everywhere on each of the intervals $[jq^{-m}, (j + 1)q^{-m})$, $j = 0, 1, ..., q^m - 1$, then for $n = m, m + 1, ...$, $P_n = \mu$ and $(X_1, ..., X_n)$ is independent of $T^n(X)$.

Proof If $P_m = \mu$ then by invariance of μ under T, $P_n = \mu$ for $n = m, m+1, ...$ Let $K = \sum_{i=1}^{m} X_i q^{m-i}$ and $U = T^m(X)$, so $X = (K + U)q^{-m}$. For Lebesgue almost every $t \in [0, 1)$,

$$f(t) = q^m P(K = \lfloor q^m t \rfloor \mid U = q^m t - \lfloor q^m t \rfloor) f_m(q^m t - \lfloor q^m t \rfloor)$$

since

$$F(t) = P((K + U)q^{-m} \leq t)$$
$$= F(q^{-m} \lfloor q^m t \rfloor) + \int_0^{q^m t - \lfloor q^m t \rfloor} P(K = \lfloor q^m t \rfloor \mid U = u) f_m(u) \, du.$$

Thereby the first assertion follows.

Suppose that c_j is a constant and $f = c_j$ Lebesgue almost everywhere on $[jq^{-m}, (j+1)q^{-m})$ for $j = 0, 1, ..., q^m - 1$. Then

$$\sum_{j=0}^{q^m - 1} c_j q^{-m} = \sum_{j=0}^{q^m - 1} \int_{jq^{-m}}^{(j+1)q^{-m}} c_j = \int_0^1 f = 1,$$

and so for Lebesgue almost all $x \in [0, 1)$, (5) gives that $f_m(x) = 1$. Therefore, $P_m = \mu$, and hence $P_n = \mu$ for $n = m, m+1, ...$ Consequently, the last assertion follows from (6), using that $\sum_{i=1}^{m} X_i q^{m-i}$ and $(X_1, ..., X_m)$ are in a one-to-one correspondence.   □

3 Couplings

Let f be a PDF on $[0, 1)$. We introduce the following notation. Let $I_\emptyset = I_{1;0} = [0, 1)$ and $c_\emptyset = c_{1;0} = \inf_{I_\emptyset} f$. For $n = 1, 2, ...$ and $x_1, x_2, ... \in \{0, 1, ..., q-1\}$, let $k = 1 + \sum_{i=1}^{n} x_i q^{n-i}$ and

$$I_{x_1,...,x_n} = I_{k;n} = [(k-1)q^{-n}, kq^{-n})$$

and

$$c_{x_1,...,x_n} = c_{k;n} = \inf_{I_{x_1,...,x_n}} f - \inf_{I_{x_1,...,x_{n-1}}} f.$$

Recall that f is lower semi-continuous at a point x if for any sequence $y_n \to x$, it holds that $\liminf_n f(y_n) \geq f(x)$. Note that if $x = \sum_{n=1}^{\infty} x_n q^{-n} \in [0, 1)$ is not a base-q fraction, then lower semi-continuity at x is equivalent to

$$f(x) = \lim_{n \to \infty} \inf_{y \in I_{x_1,...,x_n}} f(y). \tag{7}$$

Commonly used PDFs are lower semi-continuous almost everywhere. For an example where this condition does not hold, see Remark 1 in Herbst et al. (2023).

Write $U \sim \mu$ if U is a uniformly distributed random variable on $[0, 1)$.

Theorem 3.1 *Suppose f is lower semi-continuous at Lebesgue almost all points in $[0, 1)$. Then there is a coupling between $X \sim f$ and a non-negative integer-valued random variable N such that $T^N(X) \sim \mu$ is independent of $(X_1, ..., X_N)$.*

Proof For Lebesgue almost all $x = \sum_{n=1}^{\infty} x_n q^{-n} \in [0, 1)$ with $x_n = \lfloor T^{n-1}(x)q \rfloor$, assuming x is not a base-q fraction (recalling that the set of base-q fractions is a Lebesgue nullset), (7) gives

$$f(x) = \inf_{I_\emptyset} f + \left(\inf_{I_{x_1}} f - \inf_{I_\emptyset} f \right) + \left(\inf_{I_{x_1,x_2}} f - \inf_{I_{x_1}} f \right) + \dots$$
$$= \sum_{n=0}^{\infty} c_{x_1,\dots,x_n}. \tag{8}$$

Let N be a random variable such that for $f(x) > 0$, conditionally on $X = x$,

$$P(N = n \mid X = x) = c_{x_1,\dots,x_n}/f(x), \quad n = 0, 1, \dots$$

By (8) and since $c_{x_1,\dots,x_n} \geq 0$, this is a well-defined conditional distribution.

By Bayes theorem, conditioned on $N = n$ with $P(N = n) > 0$, X follows an absolutely continuous distribution with PDF

$$f(x \mid n) = c_{x_1,\dots,x_n}/P(N = n).$$

Therefore, since $f(x|n)$ is constant on each of the intervals $I_{k;n}$, conditioned on $N = n$ we immediately see that $(X_1, \dots, X_n)$ (interpreted as nothing if $n = 0$) and $T^n(X)$ are independent and that $T^n(X) \sim \mu$. The latter implies that $T^N(X) \sim \mu$ is independent of N. Consequently, if we do not condition on N, we have that $(X_1, \dots, X_N)$ and $T^N(X) \sim \mu$ are independent. $\qquad\square$

Corollary 3.2 *For the coupling construction in the proof of Theorem 3.1, conditioned on $X = x$ with $f(x) > 0$, we have*

$$P(N \leq n \mid X = x) = \sum_{k=0}^{n} c_{x_1,\dots,x_k}/f(x), \quad n = 0, 1, \dots, \tag{9}$$

where $x_k = \lfloor T^{k-1}(x)q \rfloor$ for $1 \leq k \leq n$. Moreover,

$$P(N \leq n) = q^{-n} \sum_{k=1}^{q^n} \inf_{I_{k;n}} f, \quad n = 0, 1, \dots \tag{10}$$

Remark 1 Corollary 3.2 is used in Section 5 to quantify how many digits are needed to make the remainder uniformly distributed with sufficiently high probability. By (10), in order to make N as small as possible, we prefer a version of f which is as large as possible, cf. Herbst et al. (2023).

Proof The proof of Theorem 3.1 gives immediately (9). Thus, for $n = 0, 1, \dots,$

$$P(N \leq n) = \int_0^1 P(N \leq n \mid X = x) f(x) \, dx = \sum_{k=0}^{n} \sum_{j=1}^{q^k} c_{j;k} q^{-k}.$$

So $P(N = 0) = c_\emptyset$ in agreement with (10). For $n = 1, 2, ...$, we have

$$\sum_{k=0}^{n} \sum_{j=1}^{q^k} c_{j;k} q^{-k} = c_\emptyset + \sum_{k=1}^{n} \sum_{(x_1,...,x_k) \in \{0,1,...,q-1\}^k} c_{x_1,...,x_k} q^{-k}$$

$$= \inf_{I_\emptyset} f + \sum_{x_1 \in \{0,1,...,q-1\}} \left(\inf_{I_{x_1}} f - \inf_{I_\emptyset} f \right) q^{-1} + ...$$

$$+ \sum_{(x_1,...,x_n) \in \{0,1,...,q-1\}^n} \left(\inf_{I_{x_1,...,x_n}} f - \inf_{I_{x_1,...,x_{n-1}}} f \right) q^{-n}$$

$$= q^{-n} \sum_{(x_1,...,x_n) \in \{0,1,...,q-1\}^n} \inf_{I_{x_1,...,x_n}} f$$

$$= q^{-n} \sum_{j=1}^{q^n} \inf_{I_{j;n}} f.$$

Thereby (10) follows. □

Corollary 3.3 *Let the situation be as in Theorem 3.1. The output of the following simulation algorithm is distributed as $X \sim f$:*

(a) Draw N from (10).
(b) Conditionally on N, generate a discrete random variable K with

$$P(K = k - 1 \mid N = n) \propto c_{k;n}, \quad k = 1, ..., q^n, \, n = 0, 1, ... \tag{11}$$

(c) Independently of (N, K) pick a random variable $U \sim \mu$.
(d) Output $(K + U)q^{-N}$.

Proof Let $a_n = \sum_{k=1}^{q^n} c_{k;n}$ be the normalizing constant in (11). Conditioned on $N = n$ with $P(N = n) > 0$, steps (b) and (c) give that $U \sim \mu$ and K are independent, so the conditional distribution of $(K + U)q^{-N}$ is absolutely continuous with a conditional PDF given by

$$f(x \mid n) = q^n c_{k;n}/a_n \quad \text{if } x \in I_{k;n}.$$

Moreover, we get from (10) that $P(N = 0) = c_\emptyset$ and

$$P(N = n) = P(N \leq n) - P(N < n) = a_n q^{-n}, \quad n = 1, 2, ...$$

Therefore, the (unconditional) distribution of $(K + U)q^{-N}$ is absolutely continuous with a PDF which at each point $x = \sum_{n=1}^{\infty} x_n q^n \in [0, 1)$ with $x_n = \lfloor T^{n-1}(x)q \rfloor$ is given by

$$\sum_{n=0}^{\infty} f(x \mid n) P(N = n) = \sum_{n=0}^{\infty} q^n \left(c_{x_1,...,x_n}/a_n \right) a_n q^{-n} = \sum_{n=0}^{\infty} c_{x_1,...,x_n}.$$

This PDF agrees with (8), so $(K + U)q^{-N} \sim f$. □

Denote by $\mathcal{B}$ the class of Borel subsets of $[0, 1)$. The total variation distance between two probability measures ν_1 and ν_2 defined on $\mathcal{B}$ and with PDFs g_1 and g_2, respectively, is given by

$$d_{\text{TV}}(\nu_1, \nu_2) = \sup_{A \in \mathcal{B}} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \|g_1 - g_2\|_1, \tag{12}$$

(see Tsybakov 2009, Lemma 2.1). Then Theorem 3.1 shows the following.

Corollary 3.4 *Let the situation be as in Theorem 3.1. Then*

$$d_{\mathrm{TV}}(P_n, \mu) \leq \mathrm{P}(N > n), \quad n = 0, 1, \ldots \tag{13}$$

Remark 2 In general the coupling inequality (13) is sharp, see Herbst et al. (2023). It follows from Corollary 3.2 and 3.4 that (7) implies $\lim_{n \to \infty} d_{\mathrm{TV}}(P_n, \mu) = 0$. In Theorem 4.1 below we show that (7) is not needed for this convergence result.

Proof Using Corollary 3.3, let $X = (K + U)q^{-N}$. For $n = 0, 1, \ldots$, if Q_n denotes the probability distribution of $T^n(U)$, then $Q_n = \mu$, and so

$$d_{\mathrm{TV}}(P_n, \mu) = d_{\mathrm{TV}}(P_n, Q_n) \leq \mathrm{P}(T^n(X) \neq T^n(U)) \leq \mathrm{P}(N > n),$$

where the first inequality is the standard coupling inequality for the coupled random variables $T^n(X)$ and $T^n(U)$, and the last inequality follows since $N \leq n$ implies $T^n(X) = T^n(U)$. Thereby (13) is verified.   □

Remark 3 By the Kantorovich-Rubinstein theorem, the Wasserstein distance between two probability measures ν_1 and ν_2 on [0, 1] is given by

$$W_1(\nu_1, \nu_2) = \inf_{\gamma \in \Gamma(\nu_1, \nu_2)} \{\mathrm{E}|Y_1 - Y_2| \mid (Y_1, Y_2) \sim \gamma\},$$

where $\Gamma(\nu_1, \nu_2)$ consists of all couplings of ν_1 and ν_2. By (Gibbs and Su 2002, Thm 4),

$$W_1(\nu_1, \nu_2) \leq d_{\mathrm{TV}}(\nu_1, \nu_2),$$

so by Remark 2, Corollary 3.4 implies

$$W_1(P_n, \mu) \leq \mathrm{P}(N > n) \to 0.$$

The latter bound can be improved by using the coupling between $T^n(X)$ and $T^N(X) \sim \mu$ to obtain

$$\begin{aligned} W_1(P_n, \mu) &\leq \mathrm{E}|(T^n(X) - T^N(X))1_{N>n}| \\ &\leq \int_0^1 \max\{|x|, |x - 1|\}\mathrm{d}x \, \mathrm{P}(N > n) \\ &= \frac{3}{4}\mathrm{P}(N > n). \end{aligned}$$

See also Gibbs and Su (2002) for an overview of the relation between the total variation distance and other measures of distance between probability measures.

4 Asymptotic Results

We need some notation for the following theorem. For a real, measurable function g defined on $(0, 1)$, denote its L_1- and supremum-norm by $\|g\|_1 = \int_0^1 |g(t)| \, \mathrm{d}t$ and $\|g\|_\infty = \sup_{x \in (0,1)} |g(x)|$, respectively, and denote the corresponding L_1-space by $L_1(0, 1) = \{g \mid \|g\|_1 < \infty\}$ (here, $\|g\|_\infty$ may be infinite when there are no further assumptions on g). Let $\bar{L}_1(0, 1) = \{g \mid \int_0^1 g(t)dt = 1, \|g\|_1 < \infty\}$ be the subset of functions with finite L_1-norm and integral over [0, 1] equal one, and $\bar{L}_1'(0, 1) \subset \bar{L}_1(0, 1)$ its subset of differentiable functions g such that $\|g'\|_\infty < \infty$. For $g \in \bar{L}_1'(0, 1)$, $n \in \mathbb{N}$, $j = 0, 1, \ldots, q^n - 1$, and

$0 < x < 1$, define $g'_{n,j}(x) = g'(x)$ if $q^{-n}j < x < q^{-n}(j+1)$ and $g'_{n,j}(x) = 0$ otherwise, and define

$$g_n(x) = q^{-n} \sum_{j=0}^{q^n-1} g(q^{-n}(j+x)). \tag{14}$$

Henceforth, we also think of f as an element of $\bar{L}_1(0, 1)$.

Theorem 4.1 *If $f \in \bar{L}_1(0, 1)$ then for every $g \in \bar{L}'_1(0, 1)$,*

$$d_{\text{TV}}(P_n, \mu) \le \frac{1}{2}\|f - g\|_1 + \frac{1}{6}q^{-2n} \sum_{j=0}^{q^n-1} \|g'_{n,j}\|_\infty \le \frac{1}{2}\|f - g\|_1 + \frac{1}{6}q^{-n}\|g'\|_\infty. \tag{15}$$

In particular,

$$\lim_{n \to \infty} d_{\text{TV}}(P_n, \mu) = 0 \tag{16}$$

and we have the following sharper convergence results. If $f \in \bar{L}'_1(0, 1)$ then P_n converges exponentially fast:

$$d_{\text{TV}}(P_n, \mu) \le \frac{1}{6}q^{-2n} \sum_{j=0}^{q^n-1} \|f'_{n,j}\|_\infty \le \frac{1}{6}q^{-n}\|f'\|_\infty. \tag{17}$$

If $\|f\|_\infty < \infty$ and f is continuous except for finitely many points, then

$$|f_n(x) - 1| \to 0 \quad \text{uniformly for } x \in (0, 1). \tag{18}$$

If f is twice differentiable with $\|f''\|_\infty < \infty$ then we have the following improvement of (17):

$$d_{\text{TV}}(P_n, \mu) = \frac{1}{8}q^{-2n} \left| \sum_{j=0}^{q^n-1} f'(\xi_{nj}) \right| + O(\|f''\|_\infty q^{-2n}) \tag{19}$$

$$\le \frac{1}{8}q^{-n}\|f'\|_\infty + O(\|f''\|_\infty q^{-2n})$$

where $\xi_{n,j} \in (q^{-n}j, q^{-n}(j+1))$ is arbitrary.

Before proving this theorem we need the following lemma.

Lemma 4.2 *Let $f \in \bar{L}_1(0, 1)$, $g \in \bar{L}'_1(0, 1)$. For every $x \in (0, 1)$,*

$$|g_n(x) - 1| \le q^{-2n}\left(x^2 - x + \frac{1}{2}\right) \sum_{j=0}^{q^n-1} \|g'_{n,j}\|_\infty \le \frac{1}{2}q^{-n}\|g'\|_\infty, \tag{20}$$

and

$$\int_0^1 |f_n(x) - g_n(x)|\, dx \le \|f - g\|_1. \tag{21}$$

If g is twice differentiable on $(0, 1)$ with $\|g''\|_\infty < \infty$ then for every $x \in (0, 1)$,

$$g_n(x) - 1 = q^{-2n}\left(x - \frac{1}{2}\right) \sum_{j=0}^{q^n-1} g'(\xi_{n,j}) + O(\|g''\|_\infty q^{-2n}), \tag{22}$$

where each $\xi_{n,j} \in (q^{-n}j, q^{-n}(j+1))$ is arbitrary.

Proof Let $x \in (0, 1)$. From (14) we have

$$
\begin{aligned}
g_n(x) - 1 &= \sum_{j=0}^{q^n-1} \int_{q^{-n}j}^{q^{-n}(j+1)} [g(q^{-n}(j+x)) - g(t)] \, dt \\
&= \sum_{j=0}^{q^n-1} \int_0^1 q^{-n} [g(q^{-n}(j+x)) - g(q^{-n}(j+y))] \, dy.
\end{aligned}
\tag{23}
$$

If g is differentiable on $(0, 1)$ with $\|g'\|_\infty < \infty$, we get by the mean value theorem,

$$
|g(q^{-n}(j+x)) - g(q^{-n}(j+y))| \le \|g'_{n,j}\|_\infty q^{-n} |x - y|,
$$

which yields the bound

$$
|g_n(x) - 1| \le q^{-2n} \int_0^1 |x - y| \, dy \sum_{j=0}^{q^n-1} \|g'_{n,j}\|_\infty = q^{-2n} \left(x^2 - x + \frac{1}{2} \right) \sum_{j=0}^{q^n-1} \|g'_{n,j}\|_\infty.
\tag{24}
$$

Thereby (20) follows. Moreover,

$$
\begin{aligned}
\int_0^1 |f_n(x) - g_n(x)| \, dx &\le \sum_{j=0}^{q^n-1} \int_0^1 q^{-n} |f(q^{-n}(j+x)) - g(q^{-n}(j+x))| \, dx \\
&= \|f - g\|_1
\end{aligned}
\tag{25}
$$

whereby (21) follows. If g is twice differentiable on $(0, 1)$ with $\|g''\|_\infty < \infty$, the mean value theorem gives

$$
\begin{aligned}
g(q^{-n}(j+x)) - g(q^{-n}(j+y)) &= g'(\xi_{x,y}) q^{-n} (x - y) \\
&= g'(\xi_{n,j}) q^{-n} (x - y) + O(\|g''\|_\infty q^{-2n}),
\end{aligned}
$$

where $\xi_{x,y} \in (q^{-n}j, q^{-n}(j+1))$ depends on x and y and $\xi_{n,j} \in (q^{-n}j, q^{-n}(j+1))$ is arbitrary. The second equality was obtained by applying the mean value theorem to $g'(\xi_{x,y}) - g'(\xi_{n,j})$. Inserting this into (23) yields

$$
g_n(x) - 1 = q^{-2n} \sum_{j=0}^{q^n-1} g'(\xi_{n,j}) \int_0^1 (x - y) \, dy + O(\|g''\|_\infty q^{-2n}),
$$

which reduces to (22). □

We are now ready for the proof of Theorem 4.1.

Proof We have

$$
\begin{aligned}
d_{\mathrm{TV}}(P_n, \mu) &= \frac{1}{2} \int_0^1 |f_n(x) - 1| \, dx \\
&\le \frac{1}{2} \int_0^1 |f_n(x) - g_n(x)| \, dx + \frac{1}{2} \int_0^1 |g_n(x) - 1| \, dx \\
&\le \frac{1}{2} \|f - g\|_1 + \frac{1}{6} q^{-2n} \sum_{j=0}^{q^n-1} \|g'_{nj}\|_\infty \\
&\le \frac{1}{2} \|f - g\|_1 + \frac{1}{6} q^{-n} \|g'\|_\infty
\end{aligned}
\tag{26}
$$

where we get the equality from (12) and the second inequality from (20), (21), and since $\int_0^1 (x^2 - x + 1/2)\, dx = 1/3$. Thereby (15) is verified. Taking $n \to \infty$ in (15) and using that $\bar{L}'_1(0,1)$ is dense in $\bar{L}_1(0,1)$, we get (16). Equation (17) follows from (15) by setting $g = f$.

For the proof of (18) we suppose f is continuous except at $x_1, \ldots, x_m \in (0,1)$ and set $x_0 = 0$ and $x_{m+1} = 1$. Let $\delta > 0$ and

$$I_n = \{ j \in \{0, 1, \ldots, q^n - 1\} \mid \exists i \in \{0, 1, \ldots, m+1\} : |q^{-n} j - x_i| < \delta \},$$
$$J_n = \{0, 1, \ldots, q^n - 1\} \backslash I_n.$$

By (23),

$$
f_n(x) - 1 = \sum_{j \in I_n} \int_{q^{-n} j}^{q^{-n}(j+1)} (f(q^{-n}(j+x)) - f(t))\, dt
$$
$$
+ \sum_{j \in J_n} \int_{q^{-n} j}^{q^{-n}(j+1)} (f(q^{-n}(j+x)) - f(t))\, dt. \tag{27}
$$

Given $\varepsilon > 0$, we choose δ so that $\delta < \varepsilon / (6(m+2)\|f\|_\infty)$. Then, since the cardinality of I_n is at most $(m+2)(2q^n\delta + 1)$, the first sum in (27) is bounded by

$$
\left(\frac{2q^n \varepsilon}{6\|f\|_\infty} + m + 2 \right) q^{-n} \|f\|_\infty = \frac{\varepsilon}{3} + (m+2)q^{-n}\|f\|_\infty < \frac{\varepsilon}{2}
$$

for n sufficiently large. Moreover, for n large enough, the second sum in (27) is bounded by $\varepsilon/2$ since f is uniformly continuous on $(0,1) \backslash \bigcup_{i=0}^{m+1} (x_i - \delta/2, x_i + \delta/2)$, which is a closed set. Thus, for large enough n, $|f_n(x) - 1| < \varepsilon$ which gives (18) since $\varepsilon > 0$ is arbitrary.

To prove (19) we use (22) with g replaced by f. Then, for every $A \in \mathcal{B}$,

$$
\int_A (f_n(t) - 1)\, dt = q^{-2n} \int_A \left(t - \frac{1}{2} \right) dt \sum_{j=0}^{q^n - 1} f'(\xi_{nj}) + O(\|f''\|_\infty q^{-2n}).
$$

We have

$$
\sup_{A \in \mathcal{B}} \left| \int_A \left(t - \frac{1}{2} \right) dt \right| = \frac{1}{2} \sup_{A \in \mathcal{B}} \left| \int_A (2t - 1)\, dt \right| = \frac{1}{4} \int_0^1 |2t - 1|\, dt = \frac{1}{8}
$$

where the second identity follows from (12). This gives (19). □

Remark 4 In continuation of Remark 2, by Theorem 4.1, $d_{\text{TV}(P_n, \mu)} \to 0$ and under weak conditions the convergence is exponentially fast. Lemma 4.2 provides estimates of $g_n(x) - 1$.

5 So How Many Digits are Needed?

This section starts with some theoretical statistical considerations and continues then with Example 1 (Example 2 in Herbst et al. (2023) illustrates how the convergence rate in Theorem 4.1 depends on the smoothness of f). For another paper it would be interesting to study the practical applications of our results, including how to apply a missing data approach which treats N as a hidden variable, cf. Herbst et al. (2023).

Consider a parametric model for the probability distribution of X given by a parametric class of lower semi-continuous densities f_θ where θ is an unknown parameter. By Theorem 3.1 this specifies a parametric model for $(X_1, ..., X_N)$ which is independent of $T^N(X) \sim \mu$. In practice we cannot expect N to be observable, but let us imagine it is. Then, according to general statistical principles (see e.g. Barndorff-Nielsen 1978), statistical inference for θ should be based on the sufficient statistic $(X_1, ..., X_N)$, whilst $T^N(X)$ is an ancillary statistic and hence contains no information about θ. Moreover, Theorem 4.1 ensures (without assuming that the densities are lower semi-continuous) that $T^n(X)$ is approximately uniformly distributed. Hence, if n is 'large enough', nearly all information about θ is contained in $(X_1, ..., X_n)$.

The following Example 1 demonstrates how the results in Sections 3 and 4 can be used to quantify the number n of digits needed in order that $N > n$ (conditioned or not on $X = x$) with a small probability or that $d_{\mathrm{TV}}(P_n, \mu)$ is small. See also Example 2 in Herbst et al. (2023).

Example 1 Any number $y \neq 0$ can uniquely be written as $y = sq^k(y_0 + y_f)$ where $s = s(y) \in \{\pm 1\}$ is the sign of y, $k = k(y) \in \mathbb{Z}$ determines the decimal point of y in base-q, $y_0 = y_0(y) \in \{1, ..., q - 1\}$ is the leading digit of y in base-q, and $y_0 + y_f$ is the so-called significand of y in base-q, where $y_f = y_f(y) \in [0, 1)$ is the fractional part of $y_0 + y_f$ in base-q. Correspondingly, consider any real-valued random variable $Y \neq 0$ (or just $\mathrm{P}(Y = 0) = 0$), so (almost surely) $Y = Sq^K(X_0 + X)$ where $S = s(Y)$, $K = k(Y)$, $X_0 = y_0(Y)$, and $X = y_f(Y)$ are random variables. Let $X_1, X_2, ...$ be the digits of X in the base-q expansion, cf. (3). We call $X_0, X_1, X_2, ...$ the significant digits of Y in base-q. Now, suppose Y satisfies the extended Newcomb-Benford law, that is, the log-significand of Y in base-q, $\log_q(X_0 + X)$, is uniformly distributed on $[0, 1)$ (Berger and Hill 2015, Theorem 4.2):

$$F(x) = \sum_{j=1}^{q-1} \left(\log_q(j + x) - \log_q j \right), \quad f(x) = \sum_{j=1}^{q-1} \frac{1}{\ln q} \frac{1}{j + x}, \tag{28}$$

for $0 \leq x \leq 1$. This law applies to a wide variety of real datasets and a number of appealing properties are then satisfied by Y, see Hill (1998) and Berger and Hill (2015) and the references therein.

Combining (10) and (28) gives for $n = 0, 1, ...$ that

$$\mathrm{P}(N \leq n) = \frac{q^{-n}}{\ln q} \sum_{j=1}^{q-1} \sum_{k=1}^{q^n} \frac{1}{j + kq^{-n}}.$$

The tail probabilities $\mathrm{P}(N > n)$ decrease quickly as n and q increase, see the left panel in Fig. 1 for plots of $\mathrm{P}(N > n)$ against n for $q = 2, 3, 5, 10$. The middle panel of Fig. 1 shows $\mathrm{P}(N > 1 \mid X = x)$ as a function of x for $q = 10$. We see large fluctuations, with probabilities dropping to zero when approaching the right limit of the intervals $I_{k;1}$, where $\inf_{I_{k;1}} f$ is attained. To avoid these fluctuations, the right panel of Fig. 1 shows an upper bound on $\mathrm{P}(N > n \mid X = x)$ as a function of x for $q = 10$ and $n = 0, 1, 2, 3$. The upper bound is found by noting that on each $I_{k;n}$, $\mathrm{P}(N > n \mid X = x)$ is convex decreasing towards zero. Hence an upper bound is given by evaluating at the left end points and interpolating linearly. The plot shows that $\mathrm{P}(N > n \mid X = x)$ is very close to zero for all x already for $n = 2$.
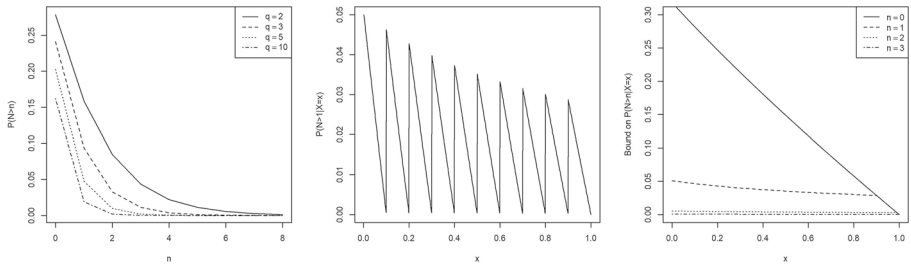
Fig. 1 Left panel: $P(N > n)$ as a function of n for $q = 2, 3, 5, 10$. Middle panel: $P(N > 1 \mid X = x)$ as a function of x for $q = 10$. Right panel: An upper bound for $P(N > n \mid X = x)$ as a function of x for $n = 0, 1, 2, 3$ and $q = 10$

This is also in accordance with Theorem 4.1 stating that $T^n(X)$ converges to a uniform distribution on $[0, 1)$ and hence the first digit X_n of $T^n(X)$ is approximately uniformly distributed on $\{0, 1, ..., q - 1\}$ when n is large. For $n = 1, 2, ...$ and $x_n \in \{0, 1, ..., q - 1\}$, we have

$$P(X_n = x_n) = \log_q \left(\prod_{j=1}^{q-1} \prod_{i=1}^{q^{n-1}} \left(1 + \frac{1}{jq^n + (i-1)q + x_n} \right) \right)$$

where $P(X_n = x_n)$ is a decreasing function of x_n. The left part of Fig. 2 shows plots of $P(X_n = 0) - P(X_n = q - 1)$ versus n for $q = 2, 3, 5, 10$ indicating fast convergence to uniformity and that the convergence speed increases with q. The right part of Fig. 2 illustrates the stronger statement in (18) that the PDF f_n of $T^n(X)$ converges uniformly to the uniform PDF.

Remark 5 In conclusion, Example 1 demonstrates that the answer to the title of our paper ('How many digits are needed?') of course depends much on q (the higher q is, the fewer digits are needed). Example 2 in Herbst et al. (2023) shows how the answer depends on how much f deviates from the uniform PDF on $[0, 1)$ (the more skew f is, the more digits are needed).
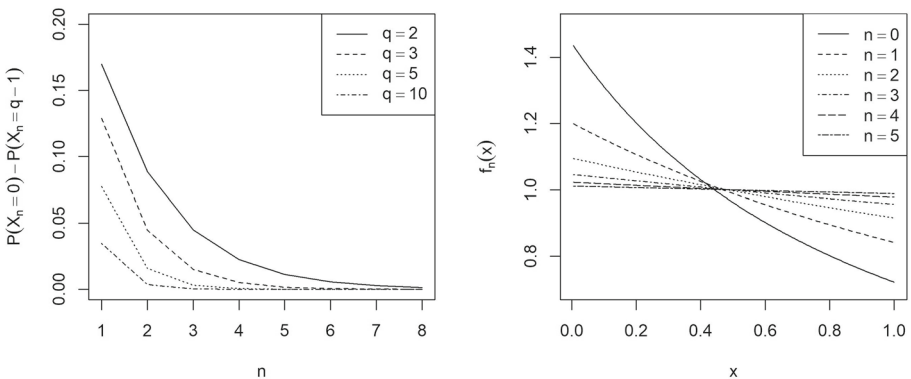

Fig. 2 Left panel: $P(X_n = 0) - P(X_n = q - 1)$ as a function of n for various values of q when f is as in (28). Right panel: f_n when $q = 2$ and $n = 0, \ldots, 5$

Declarations

Competing Interests The authors declare no competing interests.

References

Barndorff-Nielsen OE (1978) Information and exponential families in statistical theory. Wiley, Chichester

Berger A, Hill TP (2015) An introduction to Benford's Law. Princeton University Press, Princeton, NJ

Borel E (1909) Les probabilités dénombrables et leurs applications arithmétiques. Rendiconti del Circolo Matematico di Palermo 27:247–271

Cornean H, Herbst I, Møller J, Støttrup BB, Studsgaard KS (2022) Characterization of random variables with stationary digits. J Appl Prob 59:931–947

Dajani K, Kalle C (2021) A first course in Ergodic theory. Chapman and Hall/CRC

Gibbs AL, Su FE (2002) On choosing and bounding probability metrics. Int Stat Rev 70:419–435

Herbst I, Møller J, Svane AM (2023) How many digits are needed? Preprint at http://arxiv.org/abs/2307.06685

Hill TP (1998) The first digit phenomenon. American Scientist 86:358–363

Schweiger F (1995) Ergodic theory of fibered systems and metric number theory. Clarendon Press, Oxford

Tsybakov AB (2009) Introduction to nonparametric estimation. Springer Series in Statistics, Springer - Verlag, New York

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.