Probability Surveys
Vol. 11 (2014) 177–236
ISSN: 1549-5787
DOI: 10.1214/12-PS213

Distribution of the sum-of-digits function of random integers: A survey

Louis H. Y. Chen

Department of Mathematics
National University of Singapore
10 Lower Kent Ridge Road
Singapore 119076
e-mail: matchyl@nus.edu.sg

Hsien-Kuei Hwang*

Institute of Statistical Science
Institute of Information Science
Academia Sinica
Taipei 115
Taiwan
e-mail: hkhwang@stat.sinica.edu.tw

and

Vytas Zacharovas†

Department of Mathematics and Informatics
Vilnius University
Naugarduko 24, Vilnius
Lithuania
e-mail: vytas.zacharovas@mif.vu.lt

Abstract: We review some probabilistic properties of the sum-of-digits function of random integers. New asymptotic approximations to the total variation distance and its refinements are also derived. Four different approaches are used: a classical probability approach, Stein's method, an analytic approach and a new approach based on Krawtchouk polynomials and the Parseval identity. We also extend the study to a simple, general numeration system for which similar approximation theorems are derived.

AMS 2000 subject classifications: Primary 60F05, 60C05; secondary 62E17, 11N37, 11K16.
Keywords and phrases: Sum-of-digits function, Stein's method, Gray codes, total variation distance, numeration systems, Krawtchouk polynomials, digital sums, asymptotic normality.

Received December 2012.

*Part of this author's work was done while visiting the Institute for Mathematical Sciences, National University of Singapore, supported by Grant C-389-000-012-101; he thanks the Institute for its support and hospitality.

†Part of the work of this author was done during his visit at Institute of Statistical Science, Academia Sinica, Taipei, and the Institute for Mathematical Sciences, National University of Singapore. He gratefully acknowledges the support provided by both institutes.

Contents

1. Introduction

Positional numeral systems have long been used in the history of human civilizations, and the sum-of-digits function of an integer, which equals the sum of all its digits in some given base, appeared naturally in multitudinous applications such as divisibility check or check-sum algorithms. Early publications dealing with divisibility of integers using digit sums date back to at least Blaise Pascal's *Œuvres*[1] in the mid-1650's; see Glaser's interesting account [56]. Numerous properties of the sum-of-digits function have been extensively studied in the literature since then; see Chapter XX of Dickson's *History of Number Theory* [34], which contains a detailed annotated bibliography for publications up to the early 20th century dealing with the digits of an integer, and properties discussed include relations between the digit structures between n and n^2, iterated sum-of-digits function, general numeral bases, etc. Modern reviews on digital sums and number systems can be found in Stolarsky's paper [131], and the books by Knuth [78, §4.1], Ifrah [69], Allouche and Shallit [3, Ch. 3], Sándor and Crstici [118, § 4.3], Berthé and Rigo [12]. See also the two papers by Barat and Grabner [4] and by Mauduit and Rivat [99] for more useful pointers to several directions relevant to the sum-of-digits function. We are concerned in this paper with the distributional aspect of the sum-of-digits function of random integers. Many other types of results have been investigated in the literature and will not be reviewed here; most of these results deal with dynamical properties, exponential sums, Dirichlet series, block occurrences, Thue-Morse sequence, congruential properties, connections to other structures, additivity, uniform distribution and discrepancy, sum-of-digits under special subsequences, etc.

More precisely, let $q \geqslant 2$ be a fixed integer and $n = \sum_{0\leqslant j\leqslant\lambda} \varepsilon_j q^j$, where $\varepsilon_j \in \{0,\dots,q-1\}$ and $\lambda = \lfloor \log_q n \rfloor$. Then the sum-of-digits function $\nu_q(n)$ of n in base q is defined as $\sum_{0\leqslant j\leqslant\lambda} \varepsilon_j$. When $q = 2$, we write $\nu(n) = \nu_2(n)$, which is the number of ones in the binary representation of n.

Since the distribution of $\nu_q(n)$ is irregular in the sense that its values can vary between $(q-1)\lfloor \log_q n \rfloor$ and 1 (see Figure 1 for $q = 2$), we consider $X_n = X_n(q)$, which denotes the random variable equal to $\nu_q(U_n)$, where U_n assumes each of the values $\{0,\dots,n-1\}$ with equal probability $1/n$. The behavior of X_n is then more smooth.

Obviously, when $n = q^k - 1$, the distribution of X_n is exactly multinomial with parameters k and q identical probabilities $1/q$. The general difficulty then lies in estimating the closeness between the distribution of X_n and a suitably chosen multinomial distribution. Periodicities are then ubiquitous in the study of most asymptotic problems involved.

We will mainly review known results for the mean, the variance, the higher moments and the limit distribution of X_n, as well as related asymptotic approximations. It turns out that many of such results have been derived independently in the literature, and rediscoveries are not uncommon. As Stolarsky [131] puts it

[1] Pascal's *Œuvres* is freely available on Wikisource.

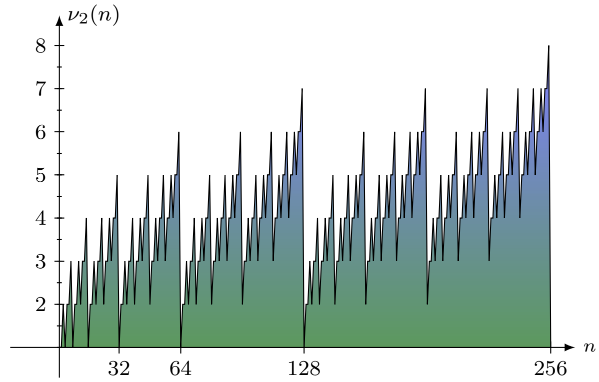


FIG 1. $\nu_2(n)$, $n = 1, \ldots, 256$.

> *"Whatever its mathematical virtues, the literature on sums of digital sums reflects a lack of communication between researchers."*

In view of the large number of independent discoveries it is likely that we missed some papers in our attempt to give a more complete collection of relevant known results.

In addition to reviewing known stochastic properties of X_n, we will present new approximations to the distribution of X_n. For simplicity, we focus on the binary case $q = 2$, leaving the straightforward extension to other numeration systems to the interested reader. In particular, our results imply that the total variation distance between the distribution of X_n and a binomial random variable Y_λ of parameters $\lambda := \lfloor \log_2 n \rfloor$ and $\frac{1}{2}$ is asymptotic to (see Figure 2)

$$\begin{aligned} d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) &= \frac{1}{2} \sum_{k \geqslant 0} \left| \mathbb{P}(X_n = k) - 2^{-\lambda} \binom{\lambda}{k} \right| \\ &= \frac{\sqrt{2}\,|F(\log_2 n)|}{\sqrt{\pi \lambda}} + O\left(\lambda^{-1}\right), \end{aligned} \tag{1}$$

where, interestingly,

$$F(\log_2 n) = \mathbb{E}(X_n) - \frac{\lambda}{2}.$$

The function F is a bounded, periodic function (namely, $F(x) = F(x+1)$) with discontinuities at integers; see (19) for the definition of F for arbitrary x, and Figure 3 for a graphical rendering.

We see that, up to an error of order $(\log_2 n)^{-1/2}$, the total variation distance is essentially asymptotic to the absolute difference between the mean and $\lambda/2$. Finer approximations will also be derived.

Four different proofs will be given for clarifying the total variation distance, and each has its own generality; these include an elementary probability approach, Stein's method, Fourier analysis, and a new Krawtchouk-Parseval approach. Indeed, these approaches easily extend to the consideration of more

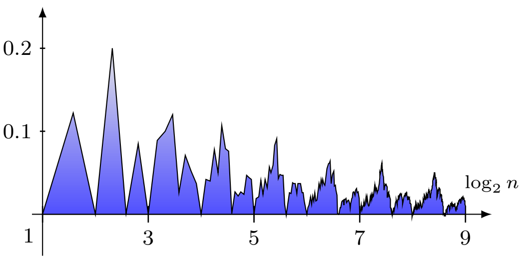


FIG 2. $d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) - \frac{\sqrt{2}\,|F(\log_2 n)|}{\sqrt{\pi\lambda}}$.

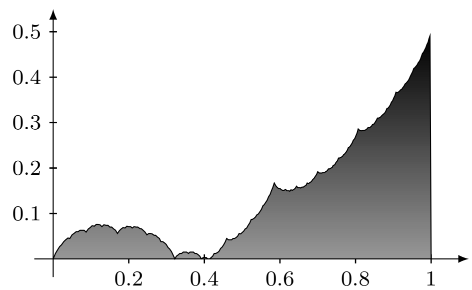


FIG 3. $|F(x)|$.

general frameworks, a simple one being briefly considered that applies in particular to the number of ones in the binary-reflected Gray codes.

1.1. First moment of X_n

The mean of X_n is essentially the partial sum of $\nu_q(j)$

$$S_q(n) := n\mathbb{E}(X_n) = \sum_{0\leqslant j<n} \nu_q(j),$$

which, by the relation $\nu_q(qj+r) = \nu_q(j)+r$ for $0 \leqslant r < q$, satisfies the following recurrence

$$S_q(n) = \sum_{1\leqslant r\leqslant q} S_q\left(\left\lfloor \frac{n+r-1}{q} \right\rfloor\right) + \sum_{1\leqslant r\leqslant q} (q-r)\left\lfloor \frac{n+r-1}{q} \right\rfloor \qquad (n \geqslant 2),$$

with $S_q(n) = 0$ for $n \leqslant 1$. In particular, when $q = 2$, this recurrence has the form

$$S_2(n) = S_2\left(\left\lfloor \frac{n}{2} \right\rfloor\right) + S_2\left(\left\lceil \frac{n}{2} \right\rceil\right) + \left\lfloor \frac{n}{2} \right\rfloor.$$

For many other recurrences for $S_2(n)$, see [100]. Interestingly, the quantity $S_2(n)$ appeared naturally in a large number of concrete applications and is given as

A000788 in Sloane's *Encyclopedia of Integer Sequences*. A partial list when $q = 2$ is given as follows.

- The number of bisecting strategies in certain games [53];
- Linear forms in number theory [87];
- Determinant of some matrix of order n [23]; see also [71] for an extension to $q \geqslant 2$;
- Bounds for the number of edges in certain class of graphs [60, 63, 101];
- The solution to the recurrence $f(n) = \max_k\{f(k) + f(n-k) + \min\{k, n-k\}\}$ with $f(1) = 0$ is exactly $S_2(n)$; concrete instances where this recurrence arise can be found in [61, §2.2.1] and [63, 100]; see also [1, 101];
- The number of comparators used by Batcher's bitonic sorting network [68];
- External left length of some binary trees [86];
- The minimum number of comparisons used by
  - top-down recursive mergesort [46];
  - bottom-up mergesort [111];
  - queue-mergesort [21];
- The number of runs for the output sequence or recursive mergesort with high erroneous comparisons; see [62].

This list of concrete examples, albeit nonrandom in nature, shows the richness and diversity of the sum-of-digits function.

Legendre, in his *Théorie des nombres* whose first edition was published in 1798, derived the relation

$$\nu_q(n) = n - (q-1)\sum_{j\geqslant 1}\left\lfloor \frac{n}{q^j}\right\rfloor;$$

see [85, Tome I, p. 12]. This relation has proved useful in establishing many properties connected to $\nu_q(n)$, including notably the identity (5) below. On the other hand, since $\sum_{j\geqslant 1}\lfloor n/q^j\rfloor$ equals the q-adic valuation of $n!$ (namely, the largest power of q that divides $n!$), the above relation has also been widely used in the q-adic valuations of many famous numbers. For an extension of the right-hand side, see [112].

About nine decades later, d'Ocagne [35] proved in 1886 an identity for $S_q(n)$ for $q = 10$ (see also [34, p. 457]); his identity easily extends to any base $q \geqslant 2$ and can be rewritten as follows. Write $n = \sum_{0\leqslant j\leqslant k}\varepsilon_j q^j$, where $\varepsilon_j = \varepsilon_j(n) \in \{0, 1, \ldots, q-1\}$. Then d'Ocagne's expression is identical to

$$\sum_{0\leqslant j<n}\nu_q(j) = \sum_{0\leqslant j\leqslant k}\varepsilon_j q^j\left(\frac{\varepsilon_j - 1 + (q-1)j}{2} + \sum_{j<\ell\leqslant k}\varepsilon_\ell\right). \tag{2}$$

In particular, when $q = 2$, we can write $n = \sum_{1\leqslant j\leqslant s}2^{\lambda_j}$, where $\lambda_1 > \cdots > \lambda_s \geqslant 0$, and (2) has the alternative form

$$\sum_{0\leqslant j<n}\nu_2(j) = \sum_{1\leqslant j\leqslant s}2^{\lambda_j}\left(\frac{\lambda_j}{2} + j - 1\right), \tag{3}$$

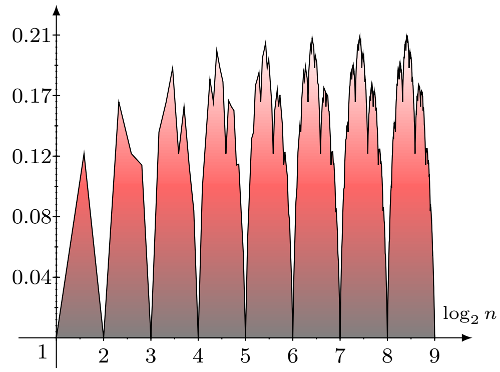


FIG 4. $\frac{1}{2}\log_2 n - \mathbb{E}(X_n)$.

where $s = \nu_2(n)$. An extension of this expression can be found in [114,138]. Since the proof of d'Ocagne's expression is very simple (summing over all coefficients block by block), it has remained almost unnoticed in the literature. Similar expressions appeared and used in several later publications; see, for example, [11, 22, 49, 62, 83, 86, 123, 138].

The first asymptotic result for $\mathbb{E}(X_n)$ was derived by Bush [15] about half a century after d'Ocagne's 1886 paper [35], and he proved that

$$\mathbb{E}(X_n) \sim \frac{q-1}{2}\log_q n,$$

as $n \to \infty$, inspired by an expression derived earlier in Bowden's book [13][2]. Note that, by (2),

$$\mathbb{E}(X_{aq^k}) = \frac{(q-1)k + a - 1}{2} \qquad (a = 1, \ldots, q-1).$$

Bush proved his formula by providing upper and lower bounds for the sum $\sum_{m<n} \varepsilon_j(m)$ using the periodicity of ε_j: $\varepsilon_j(m + q^{j+1}) = \varepsilon_j(m)$. In particular, when $q = 2$, $\varepsilon_j(m)$ is a sequence starting with a series of 2^j zeros followed by 2^j ones. His estimates imply indeed a more precise result (see Figure 4 for $q = 2$)

$$\mathbb{E}(X_n) = \frac{q-1}{2}\log_q n + O(1),$$

where the O-term is optimal. Note that this estimate can also be derived easily from d'Ocagne's expression (2) by observing that the sum $\frac{1}{2}(q-1)\sum_{0\leqslant j\leqslant\lambda} \varepsilon_j j q^j$ provides the major contribution, the others being of order $O(n)$.

Bellman and Shapiro [11] were primarily concerned with the binary representation $q = 2$ and provided an independent proof of Bush's result

$$\mathbb{E}(X_n) = \frac{1}{2}\log_2 n + O(\log\log n).$$

[2]We were unable to find a copy of this book.

They use two different proofs (one by generating functions and Tauberian theorems and the other by recurrence) and briefly mention in a footnote that the remainder can be improved to $O(1)$.

The same paper also initiated a very important notion called "dyadically additive", which has later on been fruitfully extended and explored mostly under the name of q-additivity (and its multiplicative counterpart q-multiplicativity); see [29, 52, 124] for the early publications and [97] and the papers cited there for more recent developments.

Mirsky [102], following [11], proved that

$$\mathbb{E}(X_n) = \frac{q-1}{2} \log_q n + O(1). \tag{4}$$

His simple, half-page proof is based on the decompositions

$$\mathbb{E}(X_n) = \frac{1}{n} \sum_{0 \leqslant j < n} \nu_q(j) = \frac{1}{n} \sum_{0 \leqslant j < n} \sum_{\ell \geqslant 0} \varepsilon_\ell(j) = \frac{1}{n} \sum_{0 \leqslant r < q} r \sum_{\ell \geqslant 1} f(n, \ell, r),$$

where $f(n, \ell, r)$ denotes number of integers $0 \leqslant j < n$ such that $\varepsilon_\ell(j) = r$. Then (4) follows from the simple estimate $f(n, \ell, r) = n/q + O(q^\ell)$.

Mirsky's result was independently re-derived by Cheo and Yien [22] and Tang [134] (judged to be virtually identical to [22] in MathSciNet), and referred to as Cheo and Yien's theorem in [25, 74]. Cheo and Yien proved additionally in [22] a theorem for the density of X_n of the form

$$\mathbb{P}\big(X_n = m\big) \sim \frac{1}{n} \cdot \frac{(\log_q n)^m}{m!},$$

for each finite $m \geqslant 0$.

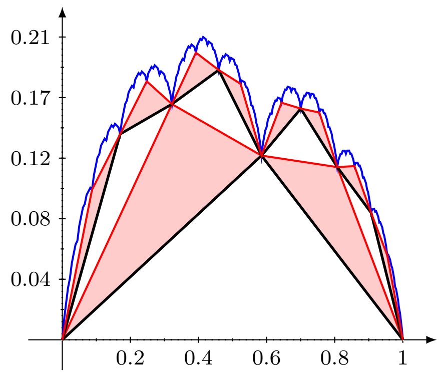
Drazin and Griffith [37] studied the sum of integer powers of the digits and derived estimates similar to (4). They also commenced the study of more precise numerical bounds for the $O(1)$-term in (4), which was followed later in [23, 45, 49–51, 100, 123, 139]. In particular, no mention is made in [23, 45, 100, 123] of known results for the $O(1)$-term in (4), and in particular the bounds derived in [45] about half a century later are weaker than those in [37].

The next stage of refinement was accomplished by Trollope in 1968 where he showed that the $O(1)$-term in (4) is indeed a periodic function when $q = 2$ for which an explicit expression is also given. His proof is based on d'Ocagne's formula (3), which he derived in [138] in a more general setting.

Delange [30] made an important step towards the ultimate understanding of the underlying periodic function. He extended Trollope's result to any base $q \geqslant 2$ and showed, by a very simple, elegant, elementary proof, that (see Figure 4 for $q = 2$):

$$\mathbb{E}(X_n) - \frac{q-1}{2} \log_q n = F_1(\log_q n), \tag{5}$$

where $F_1(x) = F_1(x+1)$ is a continuous, periodic, and *nowhere differentiable function* (see also [135, 137]). His expression for F_1 is as follows; see Figure 5 for



,

FIG 5. $-F_1(x)$: $q = 2$.

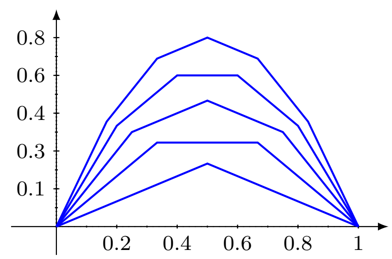


FIG 6. $h(x) : q = \begin{matrix} 6 \\ \vdots \\ 2 \end{matrix}$.

a plot of $-F_1(x)$ and its first few approximations by $\frac{1}{2}\log_2 n - \mathbb{E}(X_n)$.

$$F_1(x) = \frac{q-1}{2}\,(1 - \{x\}) + q^{1-\{x\}} g(q^{-1+\{x\}}),$$

where $\{x\}$ denotes the fractional part of x and $g(x)$ is a Takagi function [64, 84]

$$g(x) = \sum_{j \geqslant 0} q^{-j} h(q^j x),$$

with the 1-periodic function h defined by (see Figure 6)

$$h(x) = \int_0^x \left(q\{t\} - \{qt\} - \frac{q-1}{2} \right) \mathrm{d}t.$$

Furthermore, the Fourier series expansion of F is also computed; see also [47] for a systematic approach by analytic means. Delange's proof is based on the simple observation that

$$\varepsilon_j(n) = \left\lfloor \frac{n}{q^j} \right\rfloor - q \left\lfloor \frac{n}{q^{j+1}} \right\rfloor = \int_n^{n+1} \left(\left\lfloor \frac{t}{q^j} \right\rfloor - q \left\lfloor \frac{t}{q^{j+1}} \right\rfloor \right) \mathrm{d}t. \tag{6}$$

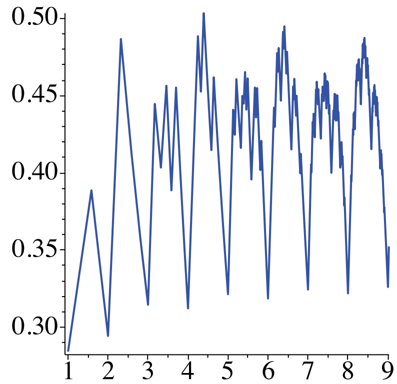


Fig 7. *(Kolmogorov dist.)* $\sqrt{\log n}$.

His paper [30] has since become a classic and has stimulated much recent research on various themes related to digital sums and different numeration systems; also different asymptotic tools have been developed.

In particular, the Trollope-Delange formula (5) for $\mathbb{E}(X_n)$, which is not only an asymptotic expansion but also an identity for all $n \geqslant 1$, is not exceptional but a distinguishing feature of many digital sums; see below and [47,58,135] for more examples.

1.2. Beyond the mean: Variance, higher moments and limit distribution of X_n

The first paper dealing with the distribution of X_n beyond the mean value is by Kátai and Mogyoródi [73] in 1968. They derived the asymptotic normality of X_n with a rate of the form

$$\sup_x \left| \mathbb{P}\left(\frac{X_n - \frac{1}{2}(q-1)\log_q n}{\sqrt{\frac{1}{12}(q^2-1)\log_q n}} < x \right) - \Phi(x) \right| = O\left(\frac{\log\log n}{\sqrt{\log n}} \right), \tag{7}$$

where Φ denotes the standard normal distribution function and the variance is implicit in their proof, namely,

$$\mathbb{V}(X_n) \sim \frac{q^2-1}{12} \log_q n.$$

Their approach consists in decomposing X_n into sums of suitable number of independent random variables, each assuming the values $\{0, 1, \ldots, q-1\}$ with equal probability. See (30) below for the binary case.

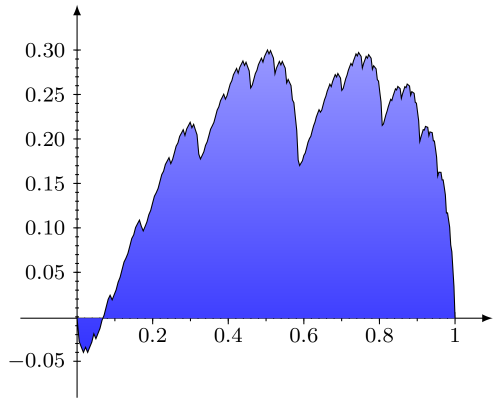


FIG 8. $-F_2(x)$.

About a decade later, Diaconis [31] obtained, by Stein's method, an optimal Berry-Esseen bound for $q = 2$ of the form

$$\sup_x \left| \mathbb{P}\left(\frac{X_n - \frac{1}{2}\log_2 n}{\sqrt{\frac{1}{4}\log_2 n}} < x \right) - \Phi(x) \right| = O\left(\frac{1}{\sqrt{\log n}} \right); \tag{8}$$

(see Figure 7) he also proved that ($q = 2$)

$$\mathbb{V}(X_n) = \frac{\log_2 n}{4} + O\left(\sqrt{\log n}\right).$$

Moments of X_n. Stolarsky [131], in addition to giving a wide list of references, carried out a systematic study of the asymptotics of the moments of X_n when $q = 2$; in particular, he proved that

$$\mathbb{E}(X_n^m) = \frac{1}{n} \sum_{0 \leqslant j < n} \nu_2(j)^m = \left(\frac{\log_2 n}{2} \right)^m + O(\log_2^{m-1} n), \tag{9}$$

for any positive integer m. The O-term is however too weak to obtain a more precise asymptotic approximation to the central moments of X_n of order $\geqslant 2$.

Later Coquet [26] in 1986 improved Stolarsky's result by providing a formula of the Trollope-Delange type ($q = 2$)

$$\mathbb{E}(X_n)^m = \left(\frac{\log_2 n}{2} \right)^m + \sum_{0 \leqslant j < m} (\log_2 n)^j F_{m,j}(\log_2 n); \tag{10}$$

here $F_{m,j}$ are bounded, continuous, 1-periodic functions. Coquet's method of proof starts from defining

$$S_2^{[m]}(n) := \sum_{0 \leqslant j < n} \nu_2(j)^m,$$

and shows that the quantity $S_2^{[m]}(2n) - 2S_2^{[m]}(n)$ is expressible in terms of a sum of $S_2^{[j]}(n)$ with $j < m$; then an induction is used. In particular, his result for the second moment implies the identity

$$\mathbb{V}(X_n) = \frac{\log_2 n}{4} + F_2(\log_2 n),$$

where F_2 is bounded, continuous and periodic of period 1; see Figure 8. Coquet [26] mentioned that the function F_2 is nowhere differentiable and his proofs extend to any q-ary base. An independent proof of the above identity à la Delange was given later by Kirschenhofer [75]; see also Osbaldestin [110] for an interesting discussion of several digital sums, as well as an alternative expression for F_2.

On the other hand, Coquet's expressions for the $F_{m,j}$'s (except for F_2) are nonconstructive; see [59] for the third moment. Dumont and Thomas [42] studied the moments of X_n in a general framework and derived more explicit expressions for $F_{m,j}$, as well as properties such as continuity and nowhere differentiability. Their approach relies on substitutions on finite alphabet and matrix analysis; see [41]. In addition to the moments, they also considered in the same paper [42] the moments of $X_n - \frac{1}{2}(q-1)\log_q n$ and showed that

$$\begin{aligned}\mathbb{E}\left(X_n - \frac{q-1}{2}\log_q n\right)^m = \frac{1+(-1)^m}{2}\cdot\frac{m!}{(m/2)!2^{m/2}}\left(\frac{q^2-1}{12}\log_q n\right)^{m/2}\\ + \sum_{0\leqslant j<m/2}(\log_q n)^j\tilde{F}_{m,j}(\log_q n) + o(1),\end{aligned} \tag{11}$$

where the $\tilde{F}_{m,j}$'s are continuous, 1-periodic, and nowhere differentiable functions. These estimates imply of course the asymptotic normality of X_n by the method of moments, which Dumont and Thomas later established in [43] (in a more general framework).

Other constructive expressions, together with interesting functional properties, are derived by Okada et al. [107], based on binomial measures; see also [104] and the recent paper [81]. Extensions of the same approach to cover the moments of X_n for any $q \geqslant 2$ were carried out in [104,105], the required tools being developed in [109].

Unaware of Stolarsky's and Coquet's results, Kennedy and Cooper considered the cases when $q = 10$: $m = 2$ in [74] and any positive integer m in [24] but with a non-optimal error term in the corresponding expression of (9) for $q = 10$; see also [14]. The optimal error term follows indeed from Dumont and Thomas's result in [42] (see also [104]) and was later re-proved by Yu in [142] (see also [16] for an extension).

A general procedure, based on the classical approach of Dirichlet series and Mellin-Perron integral formula (fully discussed in [47]), was developed in [58] and leads to absolutely convergent Fourier series expansions for $G_{m,j}$. The approach there can be easily extended to q-ary case.

Probability generating function of X_n. By definition, the probability generating function of X_n is given by

$$\mathbb{E}\left(y^{X_n}\right) = \frac{1}{n} \sum_{0 \leqslant j < n} y^{\nu_q(j)}.$$

The special cases when $q = y = 2$ appeared as the total number of odd numbers of $\binom{j}{i}$ for $0 \leqslant i \leqslant j < n$, a result derived by Glaisher [55] in 1899; earlier results of similar character can be found in the papers by Kummer [82] and by Lucas [89]. For another interesting occurrence in cellular automata, see [44, 140, 141].

The distribution of X_n is closely connected to the notion of q-additive and q-multiplicative functions, first introduced by Bellman and Shapiro [11], and later systematically investigated by Gel'fond [52] and Delange [29]; see also [4, 97] and the references cited there. We did not find a more complete survey on q-additive or q-multiplicative functions but a simple search on MathSciNet resulted in more than 152 papers (as of July 1, 2014); see [12, Ch. 9] and [4].

A function $f : \mathbb{N} \to \mathbb{C}$ is said to be *q-multiplicative* if

$$f\left(aq^r + b\right) = f(aq^r) f(b),$$

for $1 \leqslant a \leqslant q-1$ and $0 \leqslant b < q^r$, $r \geqslant 1$. This implies that $f(0) = 1$. Similarly, one defines q-additive functions by $f\left(aq^r + b\right) = f(aq^r) + f(b)$. By definition, one then obtains, for a q-multiplicative function f (see [29, 52])

$$\sum_{0 \leqslant j < n} f(j) = \sum_{0 \leqslant j \leqslant \lambda} \left(\prod_{0 \leqslant r < j} \left(1 + \sum_{1 \leqslant \ell < q} f(\ell q^r) \right) \right) \left(\prod_{j < r \leqslant \lambda} f(\varepsilon_r q^r) \right) \sum_{0 \leqslant \ell < \varepsilon_j} f(\ell q^j),$$

where $n = \sum_{0 \leqslant j \leqslant \lambda} \varepsilon_j q^j$. Now taking $f(n) = y^{\nu_q(n)}$, which is obviously a q-multiplicative function, we obtain, by re-grouping nonzero summands,

$$\mathbb{E}\left(y^{X_n}\right) = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} y^{c_1 + \cdots + c_{j-1}} \left(1 + y + \cdots + y^{c_j - 1}\right) \left(1 + y + \cdots + y^{q-1}\right)^{\lambda_j}, \tag{12}$$

where

$$n = c_1 q^{\lambda_1} + c_2 q^{\lambda_2} + \cdots + c_s q^{\lambda_s},$$

with $\lambda_1 > \cdots > \lambda_s \geqslant 0$ and $c_j \in \{1, \ldots, q-1\}$. The closed-form expression (12) was later derived and stated explicitly by Stein [127].

Special cases of (12) appeared in Roberts [116] for $q = y = 2$ (later re-derived in [131]), and in Stein [126] for $q = 2$, which has the form

$$\mathbb{E}(y^{X_n}) = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} y^{j-1} (1+y)^{\lambda_j}, \tag{13}$$

when $n = 2^{\lambda_1} + 2^{\lambda_2} + \cdots + 2^{\lambda_s}$, where $\lambda_1 > \lambda_2 > \cdots > \lambda_s$.

In the same paper [127], Stein also obtained many bounds for the exponential sum (12); in particular, the function

$$G(\log_q n; y) := \frac{\mathbb{E}(y^{X_n})}{n^{\log_q(1+y+\cdots+y^{q-1})}} \qquad (y > 0)$$

is bounded and periodic ($G(x; y) = G(x + 1; y)$).

Okada et al. [104, 108] later gave more explicit expressions for the periodic function G by multinomial measures. A different approach was provided in [81]. A Fourier expansion for $q = 2$ was given in [58], which is absolutely convergent when $\sqrt{2} - 1 < y < \sqrt{2} + 1$.

The closed-form expression (12) contains much information; for example, the d'Ocagne's formula (2) follows from (12) by taking derivative with respect to $y = 1$ and then substituting $y = 1$. We will see later that (12) is also helpful in proving effective approximations for distances between X_n and some binomials.

For other approaches to q-additive and q-multiplicative functions, see [2, 38, 57, 93, 95, 96, 106].

1.3. Asymptotic distribution of sum-of-digits function

We mentioned Kátai and Mogyoródi's [73] and Diaconis's [31] Berry-Esseen bounds for X_n. We group here known results concerning limit and approximation theorems for X_n according to the major approach used, focusing mostly on the case $q = 2$ for simplicity of presentation and comparison. See Table 1 for a summary.

Table 1

A summary of known approaches leading to the asymptotic normality of X_n; here CLT denotes "central limit theorem" and LLT "local limit theorem"

Authors & Papers	Year	Results	Approach	Notes
Kátai & Mogyoródi [73]	1968	CLT+rate	Elementary	q-ary
Diaconis [31]	1977	CLT+rate	Stein's method	binary
Schmidt [121]	1983	Multivariate CLT	Probabilistic	binary
Schmid [120]	1984	Multivariate LLT +rate	Matrix GF Markov chain	binary
Stein [129]	1986	Binomial approximation	Stein's method	binary
Dumont & Thomas [42]	1992	CLT	Method of moments	general
Loh [88]	1992	Multinomial approximation	Stein's method	q-ary
Barbour & Chen [7]	1992	Approximation by a mixture of binomial	Stein's method	binary
Grabner [57]	1993	(implicit)	Mellin transform	q-additive
Bassily & Kátai [9]	1995	CLT	Method of moments	q-additive
Manstavičius [93]	1997	Functional CLT	Probabilistic	q-additive
Dumont & Thomas [43]	1997	LLT+rate	Markov chain	general
Drmota & Gajdosik [40]	1998	LLT+ rate	Generating function	general
Drmota et al. [39]	2003	Functional CLT	Probabilistic	q-ary

1.3.1. *Classical probabilistic approach*

Kátai and Mogyoródi's approach uses elementary probability tools and relies their Berry-Esseen bound (7) on the following decomposition (for $q = 2$)

$$\mathbb{P}\big(X_n = \ell\big) = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j} \mathbb{P}(Y_{\lambda_j} = \ell - j + 1), \tag{14}$$

which follows immediately from (13). Here Y_j denotes the sum of j independent Bernoulli indicators, each assuming 0 and 1 with equal probability 1/2. The identity (14) implies that the random variable X_n is itself a mixture of independent binomial distributions. The remaining proof then proceeds along standard classical lines (by using estimates for sums of independent random variables).

Heppner [65] later proved, in the same spirit, a simple Chernoff-type inequality for X_n when $q = 2$ ($\lambda = \lfloor \log_2 n \rfloor$)

$$\mathbb{P}(|X_n - (\lambda + 1)/2| > C) \leqslant 2\mathbb{P}\big(|Y_{\lambda+1} - (\lambda + 1)/2| > C\big);$$

since the right-hand side of this inequality decreases exponentially as C grows, one concludes that $\nu_2(m)$ is close to $(\lambda+1)/2$ for most $m < n$. This observation is useful in establishing precise estimates for sums of the form $\sum_{m \in B_n} \nu_2(m)$, where B_n is an arbitrary subset of nonnegative integers $< n$. For results concerning the distribution of $\nu_q(n)$ for given subsequences of integers (such as prime numbers and squares), see [98, 99] and the references therein. Similar estimates will be used below.

A central limit theorem for the distribution of the values assumed by the sequence $\nu_2(3n) - \nu_2(n)$ was derived by Kátai [72], while the corresponding local limit theorem was given independently by Stolarsky [132]. The proof of Stolarsky's local limit theorem starts from matrix generating functions, obtaining a closed-form expression, and then applies the saddle-point method for the corresponding sum. Schmidt [121] then proved, motivated by Stolarsky's [132] result, a multidimensional central limit theorem (the joint distribution of the values of $\nu_2(K_1 n), \ldots, \nu_2(K_d n)$ for odd numbers $K_1, \ldots, K_d$) using tools from Markov chains. The intuition behind such a limit law is that the two events $\nu_2(K_1 U_n)$ and $\nu_2(K_2 U_n)$ are more or less independent, where $U_n \sim \text{Uniform}[0, n-1]$ and K_1, K_2 are odd numbers.

Dumont and Thomas [43] use again Markov chains and large deviations to characterize the asymptotic distribution of a class of digital sums (covering in particular X_n) associated with substitutions, a Berry-Esseen bound being also derived.

1.3.2. *q-multiplicative functions*

Since $\nu_q(n)$ is q-additive, the function $e^{it\nu_q(n)}$ is q-multiplicative. The distribution of the values of q-additive functions has been widely studied in the number-theoretic literature. We mention briefly an early result. Delange [29] showed that

$$\frac{1}{n}\sum_{0\leqslant m<n} f(m) = \prod_{1\leqslant r\leqslant \lfloor \log_q n\rfloor} \frac{1+f(q^r)+\cdots+f((q-1)q^r)}{q} + o(1),$$

for any q-multiplicative function f with $|f| \leqslant 1$ and (see [94])

$$\lim_{k\to\infty} \prod_{r_0\leqslant r\leqslant k} \left|\frac{1+f(q^r)+\cdots+f((q-1)q^r)}{q}\right| > 0.$$

This result roughly says that *the mean value of q-multiplicative functions with bounded modulus is close to some multinomial distribution.*

In particular, if one applies *formally* this result to $f(n) = e^{it\nu_q(n)}$, then the left-hand side corresponds to the characteristic function of X_n, while the dominant term on the right-hand side to a multinomial distribution. We cannot however conclude directly from this result that X_n is asymptotically multinomially distributed due to lack of uniformity in t. For asymptotic normality and related results for q-additive functions, see [9, 10, 93, 130] and [12, Ch. 9].

1.3.3. Stein's method

Stein's method is a method of probability approximation invented by Charles Stein in 1972 [128]. It does not involve Fourier analysis but hinges on the solution of a functional equation. In a nutshell, Stein's method can be described as follows. Let W and Z be random variables. In approximating the distribution $\mathscr{L}(W)$ of W by the distribution $\mathscr{L}(Z)$ of Z, the difference between $\mathbb{E}(h(W))$ and $\mathbb{E}(h(Z))$ for a class of functions h is expressed as

$$\mathbb{E}(h(W)) - \mathbb{E}(h(Z)) = \mathbb{E}\left(L[f_h](W)\right),$$

where L is a linear operator and f_h a bounded solution of the equation

$$L[f] = h - \mathbb{E}(h(Z)).$$

The error $\mathbb{E}(L[f_h](W))$ is then bounded by studying the solution f_h and exploiting the probabilistic properties of W. The operator L has the property that $\mathbb{E}(L[f](Z)) = 0$ for a sufficiently large class of f and therefore characterizes $\mathscr{L}(Z)$. Examples of L are (i) $L[f](w) = f'(w) - wf(w)$ for normal approximation, that is, if Z is the standard normal distribution [128], and (ii) $L[f](w) = \lambda f(w+1) - wf(w)$ for Poisson approximation, that is, if Z has the Poisson distribution with mean λ [17]. The operator L is not unique. It can be chosen to be the generator of a Markov process whose stationary distribution is the approximating distribution $\mathscr{L}(Z)$. This generator approach to Stein's method is due to Barbour [5, 6].

Using Stein's method with $L[f](w) = f'(w) - wf(w)$, Diaconis [31] proved that

$$\sup_x \left|\mathbb{P}\left(\frac{X_{n+1}-(\lambda+1)/2}{\sqrt{(\lambda+1)/4}} \leqslant x\right) - \Phi(x)\right| \leqslant \frac{c_1}{\sqrt{\lambda}},$$

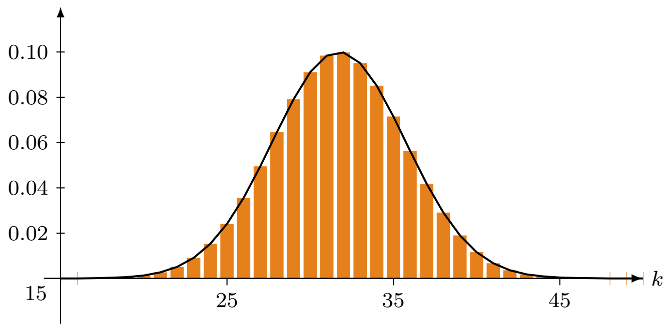


FIG 9. *The histogram of* $X_{31415926535897932384}$*, where the black smooth curve represents the density with the same mean and the same variance.*

which implies (8) since $\lambda = \lfloor \log_2 n \rfloor$. Chen and Shao [19] refined Diaconis's proof to obtain

$$\sup_x \left| \mathbb{P}\left(\frac{X_n - \lambda_0/2}{\sqrt{\lambda_0/4}} \leqslant x \right) - \Phi(x) \right| \leqslant \frac{6.2}{\sqrt{\lambda_0}},$$

where $\lambda_0 := \lceil \log_2 n \rceil$.

In his book [129], Stein considered binomial approximation for X_n, and using the equation

$$L[f](w) = (k - w)f(w) - wf(w-1) \tag{15}$$

for f defined on $\{0, 1, \ldots, k\}$, he obtained

$$\max_\ell \left| \mathbb{P}\big(X_n = \ell\big) - \frac{1}{2^\lambda}\binom{\lambda}{\ell} \right| \leqslant \frac{4}{\lambda};$$

see also [67]. By using the generator approach, Loh [88] extended the binary expansion for X_n to q-ary expansion for any base $q \geqslant 2$, and proved that

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Z)) \leqslant \frac{3.3q^{3/2}(q-1)}{\sqrt{\lceil \log_q n \rceil}},$$

where X_n denotes the q-dimensional random vector whose i-th component is the number of the i-th digit in the q-ary expansion and

$$Z \sim \text{Multinomial}(\lceil \log_q n \rceil, 1/q, \ldots, 1/q).$$

Barbour and Chen [7] also used the generator approach to improve the error bound in the binary expansion case to $1/\lambda$ if the approximating binomial distribution Y_{λ_0} is replaced by $\text{Binom}(2e(n), \frac{1}{2})$, where $e(n)$ is the mean of X_n or by a mixture of $Y_{\lambda_0 - 1}$ with either Y_{λ_0} or $Y_{\lambda_0 - 2}$ chosen to have mean $e(n)$.

1.3.4. Generating functions and analytic approach

Schmid [120] derived, improving earlier results by Stolarsky [132] and by Schmidt [121], a very precise multidimensional local limit theorem of the form

$$\frac{1}{n}\#\{m : 0 \leqslant m < n, \nu_2(K_j m) = k_j, j = 1, \dots, d\}$$
$$= \frac{\exp\left(-\frac{1}{2\log_2 n}\left(\mathbf{k} - \frac{1}{2}\log_2 n\right)\mathbf{V}^{-1}\left(\mathbf{k} - \frac{1}{2}\log_2 n\right)^{\mathrm{tr}}\right)}{(2\pi\log_2 n)^{d/2}\det(\mathbf{V})^{1/2}} + O\left((\log n)^{-(d+1)/2}\right), \tag{16}$$

where $d \geqslant 1$, the K_j's are odd integers > 1, $\mathbf{k} = (k_1, \dots, k_d)$ and $\mathbf{V}$ is the positive-definite $d \times d$ matrix with entries

$$v_{j,\ell} := \frac{\gcd(K_j, K_\ell)^2}{4K_j K_\ell} \qquad (1 \leqslant j, \ell \leqslant d).$$

His proof builds on matrix generating functions and uses tools from Markov chains, following Stolarsky and Schmidt. In addition to providing optimal convergence rate for the corresponding multidimensional central limit theorem (derived in [121]), his result implies very tight estimates for the distribution of $\nu_2(kn) - \nu_2(n)$, a problem receiving much attention in the literature; see the recent paper [27], the Ph.D. Dissertation [130] and the references therein.

In particular, (16) also leads to a local limit theorem for X_n with optimal rate when $q = 2$.

Drmota and Gajdosik [40] use generating functions and complex-analytic method to prove a local limit theorem for the sum-of-digits function in more general numeration systems; see the paper by Madritsch [92] and the references cited there for more recent developments.

1.3.5. Other approaches

We mentioned the result (11) by Dumont and Thomas [42] for the central moments of X_n, which implies the asymptotic normality of X_n by the Frechet-Shohat moment convergence theorem.

The same method of moments was later applied by Bassily and Kátai [9] to derive the asymptotic normality of q-additive functionals; see also [54, 91, 92].

Manstavičius [93], Drmota et al. [39] obtained a functional limit theorem for $\nu_q(n)$.

2. New results

We will derive a few approximation theorems for the distribution of X_n; different approaches will be developed, each having its own advantages and constraints. In particular, an expansion for a refined version of the total variation distance will be given, which will cover (1) as a special case.

Here and throughout this paper, we consider only the case $q = 2$ for simplicity. The following notations will be consistently used. Let $\lambda = \lambda_1 = \lfloor \log_2 n \rfloor$. We then write $n = \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j}$ with $\lambda_1 > \cdots > \lambda_s \geqslant 0$. Let $Y_\lambda \sim \mathrm{Binom}(\lambda, 1/2)$.

Let $H_m(x)$ denote the Hermite polynomials

$$H_m(x) = (-1)^m e^{x^2/2} \frac{\mathrm{d}^m}{\mathrm{d}x^m} e^{-x^2/2} \qquad (m = 0, 1, \dots).$$

Define a sequence $\{h_m\}$ by

$$h_m := \frac{2^{m/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |H_m(x)| \, e^{-x^2/2} \, \mathrm{d}x \qquad (m = 0, 1, \dots). \tag{17}$$

Theorem 2.1. *Let X_n denote the number of 1s in the binary representation of a random integer, where each of the integers $\{0, 1, \dots, n-1\}$ is chosen with equal probability $\mathbb{P}(X_n = m) = 1/n$. Then*

$$\begin{aligned} &\sum_{0 \leqslant k \leqslant \lambda} \left| \mathbb{P}\big(X_n = k\big) - \sum_{0 \leqslant r < m} (-1)^r a_r(n) 2^{-\lambda} \Delta^r \binom{\lambda}{k} \right| \\ &\qquad = \frac{h_m |a_m(n)|}{(\log_2 n)^{m/2}} + O\left((\log n)^{-(m+1)/2} \right), \end{aligned}$$

for $m = 1, 2, \dots,$ where the sequence $a_r(n) = a_r(2n)$ is defined by (see (13))

$$\mathbb{E}(y^{X_n}) \left(\frac{1+y}{2} \right)^{-\lambda} = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j} y^{j-1} \left(\frac{1+y}{2} \right)^{\lambda_j - \lambda} = \sum_{r \geqslant 0} a_r(n) (y-1)^r, \tag{18}$$

and Δ denotes the difference operator

$$\Delta^r \binom{\lambda}{k} = \sum_{0 \leqslant \ell \leqslant r} \binom{r}{\ell} (-1)^\ell \binom{\lambda}{k - \ell} \qquad (r = 0, 1, \dots).$$

Two explicit expressions for $a_r(n)$ are as follows.

$$\begin{aligned} a_r(n) &= \sum_{0 \leqslant \ell \leqslant r} \binom{\lambda + r - \ell - 1}{r - \ell} \frac{(-1)^{r-\ell}}{\ell! 2^{r-\ell}} \mathbb{E}(X_n \cdots (X_n - \ell + 1)) \\ &= \frac{2^\lambda}{n} \sum_{1 \leqslant j \leqslant s} 2^{-(\lambda - \lambda_j)} \sum_{0 \leqslant \ell \leqslant r} \binom{\lambda - \lambda_j + \ell - 1}{\ell} \binom{j-1}{r-\ell} (-1)^\ell 2^{-\ell}, \end{aligned}$$

for $r = 0, 1, \dots$.

Taking $m = 1$ and dividing the left-hand side by 2, we obtain an asymptotic approximation to the total variation distance; see [20, 125] for similar results.

Corollary 2.2. *The total variation distance between the distribution of X_n and that of the binomial random variable Y_λ satisfies*

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) = \frac{\sqrt{2}\,|F(\log_2 n)|}{\sqrt{\pi \log_2 n}} + O\left((\log n)^{-1}\right),$$

where the bounded periodic function F is defined by

$$F(x) = 2^{-x} \sum_{j\geqslant 0} 2^{-d_j}\left(j - 1 - \frac{d_j}{2}\right), \tag{19}$$

for $x \in (0,1]$, when writing $2^x = \sum_{j\geqslant 0} 2^{-d_j} \in (1,2]$ with $0 = d_0 < d_1 < \cdots$, and $F(x+1) = F(x)$ for other values of x.

Proof. Observe first that $a_0(n) = 1$ and by the definition of F

$$F(\log_2 n) = a_1(n) = \mathbb{E}(X_n) - \frac{\lambda}{2},$$

which has the form

$$F(\log_2 n) = \frac{2^\lambda}{n} \sum_{1\leqslant j\leqslant s} 2^{-(\lambda-\lambda_j)}\left(j - 1 - \frac{\lambda - \lambda_j}{2}\right). \tag{20}$$

On the other hand, since $H_1(x) = x$, we then get $h_1 = \sqrt{2/\pi}$. By considering the values of

$$2^x = \sum_{0\leqslant j\leqslant k} 2^{-d_j} = \sum_{0\leqslant j<k} 2^{-d_j} + \sum_{j>d_k} 2^{-j},$$

we see that F is continuous except at the end points (integers). □

By (20), we see that if $\lambda_j = \lambda - 2(j-1)$ for $j = 1, \ldots, s$, then $F(\log_2 n) = 0$. This yields the sequence $\{\log_2(\sum_{0\leqslant j\leqslant k} 4^{-j})\}_{k=0,1,\ldots}$ for the locations of the zeros of $|F(x)|$; see Figure 10.

Taking $m = 2$, we obtain a refined estimate with smaller errors.

Corollary 2.3.

$$\begin{aligned}\sum_{0\leqslant k\leqslant\lambda} &\left|\mathbb{P}(X_n = k) - 2^{-\lambda}\binom{\lambda}{k} + \left(\mathbb{E}(X_n) - \frac{\lambda}{2}\right) 2^{-\lambda}\binom{\lambda}{k}\frac{\lambda+1-2k}{\lambda+1-k}\right| \\ &= \frac{16|F_2(\log_2 n)|}{\sqrt{2\pi e}\,\log_2 n} + O\left((\log n)^{-3/2}\right),\end{aligned} \tag{21}$$

where $F_2(x)$ is defined for $x \in (0,1]$ by

$$F_2(x) = 2^{-x} \sum_{j\geqslant 0} 2^{-d_j}\left(\frac{d_j(d_j+5)}{8} - \frac{jd_j}{2} + \frac{j(j-3)}{2} + 1\right),$$

by writing $2^x = \sum_{j\geqslant 0} 2^{-d_j}$ as above, and $F_2(x+1) = F_2(x)$ for other values of x (see Figure 11 for a plot).

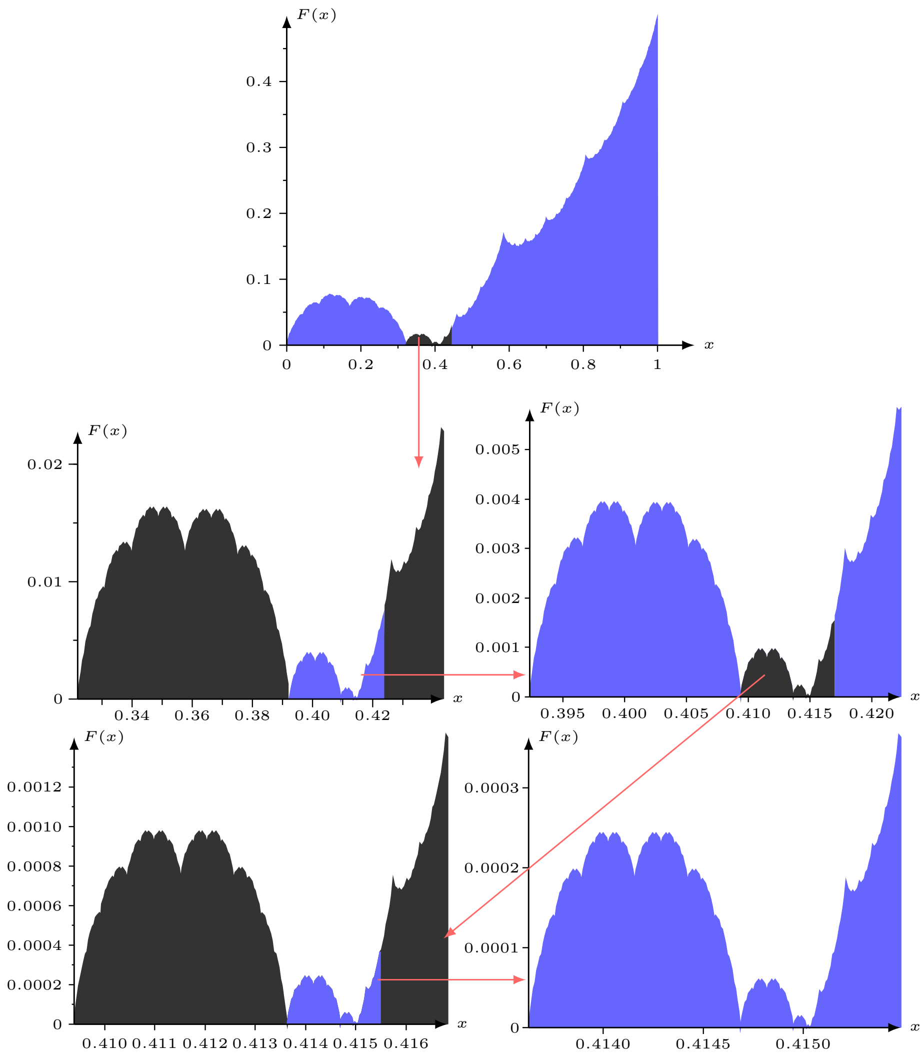


FIG 10. *The fractal nature of the function* $|F(x)|$.

The two functions $F(x)$ and $F_2(x)$ may assume the value zero when x is not an integer; see Figures 10 and 11. This means that in such cases the error term is of a smaller order, and the right-hand side of our result gives simply an O-estimate. One naturally wonders if there are other simple uniform approximants for the total variation distance. We propose a simple one in the following.

Theorem 2.4. *The total variation distance between the distribution of* X_n *and the binomial distribution of* Y_λ *satisfies*

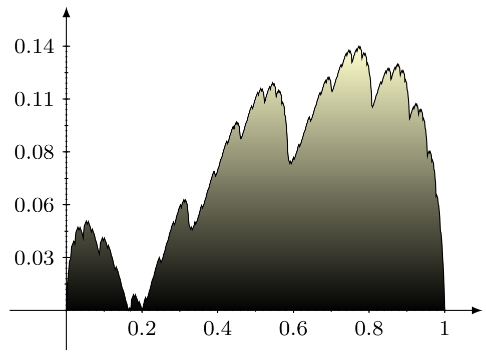


FIG 11. $|F_2(x)|$.

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) \asymp \frac{1}{2^{\lambda-\lambda_2}} \min\left\{1, \frac{\lambda-\lambda_2}{\sqrt{\lambda}}\right\},$$

whenever $\lambda - \lambda_2 \geqslant c$*, where* $c > 0$ *is sufficiently large.*

This result is similar to the estimate proved by Soon [125] (see also [20]), where he considered the distance $d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_{\lambda+1}))$ instead of $d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda))$ by using Stein's method.

We see roughly that the wider the gap between λ and λ_2, the smaller the total variation distance is.

On the other hand, the theorem fails when $c = 2$. In this case, $n = 2^\lambda + 2^{\lambda-2}$ and by (21) or by a direct calculation,

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) \asymp \lambda^{-1}. \tag{22}$$

More generally, if

$$n = 2^\lambda + 2^{\lambda-2} + \cdots + 2^{\lambda-2d} + n_0,$$

where $d \geqslant 1$ and $n_0 = O(n/\lambda^{3/2})$, then $F(\log_2 n) = O(\lambda^{-1/2})$, and (22) holds.

All these results can be extended to $\nu_q(n)$. The major difference is to use generating function (12) instead of (13).

3. Proofs

We first prove Theorem 2.1 by a direct analytic approach based on Fourier analysis. A closely connected semigroup approach (first developed by Deheuvels and Pfeifer for Poisson distribution; see [28]), but relies on more algebraic formulations and manipulations, can also be used for the same purpose; see [117]. Then Theorem 2.4 is proved by two different approaches: one by Stein's method, and the other by a standard probability argument, which starts from decomposing the distribution of X_n into a sum of binomial distributions. Our adaptation

TABLE 2
A summary of approaches used and results proved in this section

Approach	Result	Section
Analytic	Thm 2.1 (for extended d_{TV})	3.1
Elementary Probability	Thm 2.4 (for d_{TV})	3.2
Stein's method	Thm 2.4 (for d_{TV})	3.3
Krawtchouk-Parseval	χ^2-distance	3.4

of Stein's method indeed leads to a refinement of Theorem 2.4, which will be given in Section 3.3. For more methodological interests, we also include another approach using the Krawtchouk polynomials and the Parseval identity, which is the binomial analogue of the Charlier-Parseval approach developed earlier in detail in [143].

3.1. Analytic approach: Proof of Theorem 2.1

We now prove Theorem 2.1 and write the proof in a more general way that can be readily amended for dealing with other cases such as Gray codes; see Section 4.2 below.

3.1.1. Probability generating function

Let

$$P_n(y) := \mathbb{E}\left(y^{X_n}\right) = \frac{1}{n} \sum_{0 \leqslant j < n} y^{\nu_2(j)}. \tag{23}$$

Then

$$P_{2n}(y) = \frac{1+y}{2} P_n(y), \tag{24}$$

and $P_{2^k}(y) = (1+y)^k/2^k$. Note that $\nu_2(n) \leqslant \lambda + 1$.

For convenience, let

$$Q_n(y) = nP_n(y).$$

In terms of Q_n, the relation (24) has the form

$$Q_{2n}(y) = (1+y)Q_n(y).$$

For odd numbers, we have

$$Q_{2n+1}(y) = (1+y)Q_n(y) + y^{\nu_2(n)}.$$

These two recurrences can be written as

$$Q_n(y) = (1+y)Q_{\lfloor n/2 \rfloor}(y) + \delta_n y^{\nu_2(\lfloor n/2 \rfloor)},$$

for all $n \geqslant 0$, where

$$\delta_n = \frac{1-(-1)^n}{2}.$$

By iteration, we then get

$$\begin{aligned} Q_n(y) &= \sum_{0 \leqslant j \leqslant \lambda} \delta_{\lfloor n/2^j \rfloor} y^{\nu_2(\lfloor n/2^{j+1} \rfloor)} (1+y)^j \\ &= (1+y)^\lambda \sum_{0 \leqslant j \leqslant \lambda} \delta_{\lfloor 2^{\{\log_2 n\}+j} \rfloor} \frac{y^{\nu_2(\lfloor 2^{\{\log_2 n\}+j-1} \rfloor)}}{(1+y)^j}, \end{aligned} \tag{25}$$

for any $n \geqslant 1$; compare (13). This means that P_n has the form

$$P_n(y) = \left(\frac{1+y}{2}\right)^\lambda \phi_n(y),$$

where

$$\phi_n(y) = \frac{2^\lambda}{n} \sum_{0 \leqslant j \leqslant \lambda} \delta_j \frac{y^{\rho_j}}{(1+y)^j},$$

and ρ_j are nonnegative integers such that $\rho_j \leqslant j$ and $|\delta_j| \leqslant 1$.

3.1.2. Local expansion of $\phi_n(y)$

The approach we use here relies on the intuition that if ϕ_n is sufficiently "smooth" then X_n is close to the binomial distribution Y_λ. More precisely, let

$$\phi_n(y) = \sum_{j \geqslant 0} a_j(n)(y-1)^j;$$

cf. (18).

Lemma 3.1. *For each $m \geqslant 1$, we have*

$$\left| \phi_n(y) - \sum_{0 \leqslant r < m} a_r(n)(y-1)^r \right| \leqslant \frac{3}{2} \cdot \frac{2^\lambda (2|y-1|)^m}{n(1-2|y-1|)}, \tag{26}$$

if $|y-1| \leqslant 1/2 - \varepsilon$, $\varepsilon > 0$ being an arbitrarily small number.

Proof. We indeed prove a stronger estimate

$$\frac{n}{2^\lambda}|a_r(n)| \leqslant 3 \cdot 2^{r-1},$$

for all $r \geqslant 1$, which then implies (26).

Let $[y^r]f(y)$ denote the coefficient of y^r in the Taylor expansion of f. Since $\rho_j \leqslant j$, we have, for $r \geqslant 1$,

$$\frac{n}{2^\lambda}|a_r(n)| = \left| [w^r] \sum_{0 \leqslant j \leqslant \lambda} \delta_j \frac{(1+w)^{\rho_j}}{(2+w)^j} \right|$$

$$\leqslant [w^r]\sum_{j\geqslant 1}\left(\frac{1+w}{2-w}\right)^j$$
$$= [w^r]\frac{1+w}{1-2w}$$
$$= 3\cdot 2^{r-1},$$

as required. □

3.1.3. An asymptotic expansion for $\mathbb{P}(X_n = k)$

Proposition 3.2. *For all integer $0 \leqslant r \leqslant \lambda$ and each $m \geqslant 1$, we have*

$$\mathbb{P}(X_n = k) = \frac{1}{2^\lambda}\sum_{0\leqslant r<m}(-1)^r a_r(n)\Delta^r\binom{\lambda}{k} + O\left(\frac{2^{3m/2}\Gamma((m+1)/2)}{\lambda^{(m+1)/2}}\right),$$

uniformly in k.

Proof. By Cauchy's integral formula for the coefficient of an analytic function, we have

$$\begin{aligned}\mathbb{P}(X_n = k) &= \frac{1}{2\pi i}\oint_{|y|=1} y^{-n-1}P_n(y)\,\mathrm{d}y\\ &= \frac{1}{2\pi}\left(\int_{-1/2}^{1/2} + \int_{1/2\leqslant|t|\leqslant\pi}\right)e^{-kit}\left(\frac{1+e^{it}}{2}\right)^\lambda \phi_n(e^{it})\,\mathrm{d}t\\ &=: I_1 + I_2.\end{aligned}$$

Since $e^{i/2}$ lies inside the circle $|y-1| = 1/2$, we evaluate I_1 by applying Lemma 3.1 and obtain

$$\begin{aligned}I_1 = \frac{1}{2\pi}\int_{-1/2}^{1/2}\left(\frac{1+e^{it}}{2}\right)^\lambda \sum_{0\leqslant r<m} a_r(n)e^{-kit}(e^{it}-1)^r\,\mathrm{d}t\\ + O\left(\int_{-1/2}^{1/2}\left|\frac{1+e^{it}}{2}\right|^\lambda |1-e^{it}|^m\,\mathrm{d}t\right).\end{aligned}$$

The integral in the O-term is then estimated as follows.

$$\begin{aligned}\int_{-1/2}^{1/2}\left|\frac{1+e^{it}}{2}\right|^\lambda |1-e^{it}|^m\,\mathrm{d}t &= 2^m\int_{-1/2}^{1/2}\left(\cos\tfrac{t}{2}\right)^\lambda\left|\sin\tfrac{t}{2}\right|^m\,\mathrm{d}t\\ &\leqslant 2^{m+1}\int_0^1(1-t)^{(\lambda-1)/2}t^{(m-1)/2}\,\mathrm{d}t \qquad (27)\\ &= O\left(\frac{2^{3m/2}\Gamma((m+1)/2)}{\lambda^{(m+1)/2}}\right).\end{aligned}$$

Now substituting this estimate into the expression of I_1 and using the relation

$$\Delta^r \binom{\lambda}{k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-kit}(1+e^{it})^{\lambda}(1-e^{it})^r \, \mathrm{d}t, \tag{28}$$

we obtain

$$\begin{aligned}
I_1 &= \sum_{0\leqslant r<m} \frac{a_r(n)}{2\pi} \int_{-1/2}^{1/2} \left(\frac{1+e^{it}}{2}\right)^{\lambda} (e^{it}-1)^r e^{-kit} \, \mathrm{d}t \\
&\quad + O\left(\frac{2^{3m/2}\Gamma((m+1)/2)}{\lambda^{(m+1)/2}}\right) \\
&= \frac{1}{2^{\lambda}} \sum_{0\leqslant r<m} (-1)^r a_r(n) \Delta^r \binom{\lambda}{k} \\
&\quad + \sum_{0\leqslant r<k} \frac{|a_r(n)|}{2\pi} \int_{1/2\leqslant|t|\leqslant\pi} \left|\frac{1+e^{it}}{2}\right|^{\lambda} |e^{it}-1|^r \, \mathrm{d}t \\
&\quad + O\left(\frac{2^{3m/2}\Gamma((m+1)/2)}{\lambda^{(m+1)/2}}\right) \\
&= \frac{1}{2^{\lambda}} \sum_{0\leqslant r<m} (-1)^r a_r(n) \Delta^r \binom{\lambda}{k} + O\left(4^m \left(\cos\tfrac{1}{4}\right)^{\lambda}\right) \\
&\quad + O\left(\frac{2^{3m}\Gamma((m+1)/2)}{\lambda^{(m+1)/2}}\right).
\end{aligned}$$

On the other hand, since (by (13))

$$\begin{aligned}
\max_{1/2\leqslant|t|\leqslant\pi} |P_n(e^{it})| &\leqslant \frac{1}{n} \sum_{1\leqslant j\leqslant s} \left|1+e^{i/2}\right|^{\lambda_j} \\
&\leqslant \frac{1}{n} \sum_{1\leqslant j\leqslant s} 2^{\lambda_j} \exp\left(-\frac{\lambda_j}{8\pi^2}\right) \\
&\leqslant \frac{1}{n} \sum_{\lambda_j\geqslant c_0\lambda} 2^{\lambda_j} \exp\left(-\frac{c_0\lambda}{8\pi^2}\right) + \frac{1}{n} \sum_{\lambda_j<c_0\lambda} 2^{c_0\lambda} \\
&= O\left(\exp\left(-\frac{c_0\lambda}{8\pi^2}\right) + \lambda 2^{-(1-c_0)\lambda}\right).
\end{aligned}$$

Choosing

$$c_0 = \frac{\log 2}{\log 2 + 1/(8\pi^2)},$$

so as to balance the two terms in the O-symbol, we obtain

$$\max_{1/2\leqslant|t|\leqslant\pi} |P_n(e^{it})| = O\left(\lambda e^{-c_0'\lambda}\right),$$

where

$$c_0' = \frac{\log 2}{1 + 8\pi^2 \log 2}.$$

Thus

$$I_2 = O\left(\lambda e^{-c_0'\lambda}\right).$$

This proves the proposition. □

3.1.4. Estimates for the differences of binomial coefficients

Lemma 3.3. *For $r \geqslant 0$, we have*

$$2^{-\lambda} \max_{0\leqslant k\leqslant\lambda} \left|\Delta^r \binom{\lambda}{k}\right| = O\left(\frac{2^{3r/2}\Gamma((r+1)/2)}{\lambda^{(r+1)/2}}\right),$$

$$2^{-\lambda} \sum_{0\leqslant k\leqslant\lambda} \left|\Delta^r \binom{\lambda}{k}\right| = \frac{h_r}{\lambda^{r/2}} \left(1 + O\left(\lambda^{-1}\right)\right), \tag{29}$$

where h_r is defined in (17).

Proof. By (28) and an analysis similar to that used in (27), we obtain

$$\begin{aligned}
\max_{0\leqslant k\leqslant\lambda} \left|\Delta^r \binom{\lambda}{k}\right| &\leqslant \frac{2^{\lambda+r}}{2\pi} \int_{-\pi}^{\pi} \left(\cos \tfrac{t}{2}\right)^\lambda \left|\sin \tfrac{t}{2}\right|^r \,\mathrm{d}t, \\
&= \frac{2^{\lambda+r}}{\pi} \int_0^1 (1-t)^{(\lambda-1)/2} t^{(r-1)/2} \,\mathrm{d}t \\
&= O\left(\frac{2^{\lambda+3r/2}\Gamma((r+1)/2)}{\lambda^{(r+1)/2}}\right).
\end{aligned}$$

For the proof of (29), we apply the standard saddle-point method and obtain

$$\begin{aligned}
&2^{-\lambda} \sum_{0\leqslant k\leqslant\lambda} \left|\Delta^r \binom{\lambda}{k}\right| \\
&= \sum_{0\leqslant k\leqslant\lambda} \left|\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{1+e^{it}}{2}\right)^\lambda (1-e^{it})^r e^{-kit} \,\mathrm{d}t\right| \\
&= \sum_{\substack{k=\lambda/2+x\sqrt{\lambda}/2 \\ x=o(\lambda^{1/6})}} \left|\frac{2^{r+1}}{2\pi\lambda^{(r+1)/2}} \int_{-\infty}^{\infty} (-it)^r e^{-t^2/2-xit} \,\mathrm{d}t\right| \left(1 + O\left(\lambda^{-1}\right)\right) \\
&= \frac{2^{r/2}}{\sqrt{2\pi}\,\lambda^{r/2}} \int_{-\infty}^{\infty} |H_r(x)|\, e^{-x^2/2} \,\mathrm{d}x \left(1 + O\left(\lambda^{-1}\right)\right),
\end{aligned}$$

proving (29) by (17). □

Note that when $r = 1$, we have the closed form expression

$$\sum_{0\leqslant k\leqslant \lambda} \left|\binom{\lambda}{k} - \binom{\lambda}{k-1}\right| = 2\binom{\lambda}{\lfloor \lambda/2 \rfloor} - 1.$$

For higher values of r, a closed-form expression can be derived for $\sum_{0\leqslant k\leqslant\lambda} |\Delta^r \binom{\lambda}{k}|$ in terms of the zeros of Krawtchouk polynomials; see [117] for $r = 2$.

3.1.5. *Proof of Theorem 2.1*

Proof. Applying Proposition 3.2, we get

$$\sum_{0\leqslant k\leqslant\lambda} \left| \mathbb{P}(X_n = k) - 2^{-\lambda} \sum_{0\leqslant r\leqslant m+1} (-1)^r a_r(n) \Delta_r \binom{\lambda}{k} \right| = O\left(\frac{2^m \Gamma((m+3)/2)}{\lambda^{(m+1)/2}}\right).$$

Note that the sum over all k for the terms corresponding to $r = m+1$ is of order $\lambda^{-(m+1)/2}$. Theorem 2.1 then follows from (29). $\square$

We will later formulate a simple framework of numeration systems for which the same type of results as X_n hold, using the same method of proofs.

3.2. Elementary probability approach: Proof of Theorem 2.4

A crucial observation that will be elaborated here is the fact that X_n is itself a mixture of binomial distributions. More precisely, by the decomposition of Kátai and Mogyoródi (14),

$$\mathbb{P}\big(X_n = k\big) = \frac{1}{n} \sum_{1\leqslant j\leqslant s} 2^{\lambda_j} \mathbb{P}(Y_{\lambda_j} = k - j + 1). \tag{30}$$

A direct probabilistic proof of the above relation is as follows. Suppose U_n is a uniformly distributed of $[0, n-1]$, then by definition $X_n = \nu_2(U_n)$. First, we have the relation

$$X_{2^k} \stackrel{d}{=} Y_k \qquad (k = 1, 2, \ldots).$$

On the other hand, since $\nu_2(2^r + j) = 1 + \nu_2(j)$ if $0 \leqslant j < 2^r$, we also have

$$j + X_{2^{\lambda_{j+1}}} \stackrel{d}{=} j + Y_{\lambda_{j+1}} \qquad (0 \leqslant j < s).$$

We can now split the interval $\{0, 1, \ldots, n-1\} = \bigcup_{j=0}^{s-1} A_s$, where $A_0 = [0, 2^\lambda)$ and

$$A_j = \left[\sum_{1\leqslant r\leqslant j} 2^{\lambda_r}, \sum_{1\leqslant r\leqslant j+1} 2^{\lambda_r} \right), \tag{31}$$

for $1 \leqslant j \leqslant s-1$. Clearly, $\mathbb{P}(U_n \in A_j) = 2^{\lambda_{j+1}}/n$. We then obtain (30).

We group in the following lemma a few simple properties of the total variation distances involving Y_k, which will be needed later.

Lemma 3.4. *Let Y_k be a binomial random variable with mean parameters k and $1/2$. Then*

$$\begin{aligned} d_{\mathrm{TV}}(\mathscr{L}(Y_k), \mathscr{L}(Y_k+1)) &= O\left(k^{-1/2}\right), \\ d_{\mathrm{TV}}(\mathscr{L}(Y_k), \mathscr{L}(Y_{k+1})) &= \tfrac{1}{2} d_{\mathrm{TV}}(\mathscr{L}(Y_k), \mathscr{L}(Y_k+1)) = O\left(k^{-1/2}\right), \\ d_{\mathrm{TV}}(\mathscr{L}(Y_k), \mathscr{L}(Y_{k+j}+\ell)) &= O\left((j+\ell)k^{-1/2}\right). \end{aligned}$$

Proof. Since $\mathbb{P}(Y_k = j)$ increases monotonically in the interval from $\big[0, \lfloor k/2 \rfloor\big)$ and decreases monotonically in the interval $\big(\lfloor k/2 \rfloor, k\big]$, we have

$$d_{\mathrm{TV}}(\mathscr{L}(Y_k), \mathscr{L}(Y_k+1)) = 2\mathbb{P}\big(Y_k = \lfloor k/2 \rfloor\big) = O\left(k^{-1/2}\right).$$

In a similar way, since $\mathbb{P}(Y_{k+1} = j) = \mathbb{P}(Y_k + I = j) = \big(\mathbb{P}(Y_k = j-1) + \mathbb{P}(Y_k = j)\big)/2$, where I is Bernoulli with parameter $1/2$, we get

$$\begin{aligned} d_{\mathrm{TV}}(\mathscr{L}(Y_k), \mathscr{L}(Y_{k+1})) &= \tfrac{1}{4} \sum_{j \geqslant 0} \big|\mathbb{P}(Y_k = j) - \mathbb{P}(Y_k = j-1)\big| \\ &= \tfrac{1}{2} d_{\mathrm{TV}}(\mathscr{L}(Y_k), \mathscr{L}(Y_k+1)). \end{aligned}$$

This proves the lemma. □

3.2.1. Proof of Theorem 2.4 when $\lambda - \lambda_2 \leqslant \sqrt{\lambda}$

Consider first the case when $\lambda - \lambda_2 \leqslant \sqrt{\lambda}$. By (30) and Lemma 3.4, we have

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) \\ &\quad = \frac{1}{2n} \sum_{\ell \geqslant 0} \left| \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j} \left(\mathbb{P}(Y_{\lambda_j} = \ell - j + 1) - \mathbb{P}(Y_\lambda = \ell)\right) \right| \qquad (32) \\ &\quad \leqslant \frac{1}{n} \sum_{2 \leqslant j \leqslant s} 2^{\lambda_j} d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_j} + j - 1), \mathscr{L}(Y_\lambda)) \\ &\quad \leqslant 2 \sum_{2 \leqslant j \leqslant s} \frac{\lambda - \lambda_j}{2^{\lambda - \lambda_j} \sqrt{\lambda_j}} \\ &\quad \leqslant 2 \sum_{\lambda - \lambda_2 \leqslant k \leqslant \lambda - \lambda_s} \frac{k}{2^k \sqrt{\lambda - k}} \\ &\quad = O\left(\frac{\lambda - \lambda_2}{2^{\lambda - \lambda_2} \sqrt{\lambda_2}}\right), \end{aligned}$$

giving an upper bound for the total variation distance.

To obtain a lower bound, we apply again (32) and Lemma 3.4.

$$\begin{aligned}
d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) \geqslant & \frac{n-2^\lambda}{2n} d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) \\
& - \frac{1}{2n} \sum_{3 \leqslant j \leqslant s} 2^{\lambda_j} d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_j+j-1}), \mathscr{L}(Y_{\lambda_2+1})) \\
\geqslant & \frac{n-2^\lambda}{2n} d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) \\
& + O\left(\frac{1}{n} \sum_{3 \leqslant j \leqslant s} \frac{2^{\lambda_j}}{\sqrt{\lambda_j}} (\lambda_2 - \lambda_j + j)\right) \\
\geqslant & \frac{n-2^\lambda}{2n} d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) \\
& + O\left(\frac{2^{\lambda_3}(\lambda_2-\lambda_3)}{2^\lambda \sqrt{\lambda_3}}\right).
\end{aligned}$$

Now, by Lemma 3.4,

$$d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) = d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}), \mathscr{L}(Y_\lambda)) + O(\lambda^{-1/2}).$$

3.2.2. Proof of Theorem 2.4 when $c \leqslant \lambda - \lambda_2 \leqslant \sqrt{\lambda}$

For $c \leqslant \lambda - \lambda_2 \leqslant \sqrt{\lambda}$, where $c > 0$ is sufficiently large, we have

$$\begin{aligned}
d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}), \mathscr{L}(Y_\lambda)) &\geqslant \mathbb{P}\left(Y_{\lambda_2} \leqslant \sqrt{\lambda\lambda_2}/2\right) - \mathbb{P}\left(Y_\lambda \leqslant \sqrt{\lambda\lambda_2}/2\right) \\
&= \Phi\left(\frac{\sqrt{\lambda\lambda_2}/2 - \lambda_2/2}{\sqrt{\lambda_2}/2}\right) - \Phi\left(\frac{\sqrt{\lambda\lambda_2}/2 - \lambda/2}{\sqrt{\lambda}/2}\right) \\
&\quad + O(\lambda^{-1/2}) \\
&= \Phi\left(\sqrt{\lambda} - \sqrt{\lambda_2}\right) - \Phi\left(\sqrt{\lambda_2} - \sqrt{\lambda}\right) + O\left(\lambda^{-1/2}\right) \\
&\geqslant \varepsilon \frac{\lambda - \lambda_2}{\sqrt{\lambda}},
\end{aligned}$$

for $\varepsilon > 0$, by the central limit theorem of the binomial distribution (with rate). Combining the upper- and the lower-bounds, we get

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) \asymp \frac{\lambda - \lambda_2}{2^{\lambda-\lambda_2}\sqrt{\lambda}},$$

if $c \leqslant \lambda - \lambda_2 \leqslant \sqrt{\lambda}$ and c is sufficiently large.

3.2.3. Proof of Theorem 2.4 when $(\lambda - \lambda_2)/\sqrt{\lambda} \to \infty$

The lower bound becomes less precise if $(\lambda - \lambda_2)/\sqrt{\lambda} \to \infty$. In this case, we first observe that the total variation does not exceed 1; thus

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) \leqslant \frac{n - 2^\lambda}{n} \leqslant \frac{2}{2^{\lambda-\lambda_2}}.$$

Take $C = (\lambda - \lambda_2)/\sqrt{\lambda}$. We have

$$\begin{aligned}
&d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda))\\
&\quad\geqslant \mathbb{P}\left(Y_\lambda \geqslant \frac{\lambda}{2} - \frac{C}{4}\sqrt{\lambda}\right) - \mathbb{P}\left(Y_{\lambda_2}+1 \geqslant \frac{\lambda}{2} - \frac{C}{4}\sqrt{\lambda}\right)\\
&\quad\geqslant \mathbb{P}\left(\left|Y_\lambda - \frac{\lambda}{2}\right| \leqslant \frac{C}{4}\sqrt{\lambda}\right) - \mathbb{P}\left(\left|Y_{\lambda_2} - \frac{\lambda_2}{2}\right| \geqslant \frac{\lambda-\lambda_2}{2} - 1 - \frac{C}{4}\sqrt{\lambda}\right)\\
&\quad= \mathbb{P}\left(\left|Y_\lambda - \frac{\lambda}{2}\right| \leqslant \frac{C}{4}\sqrt{\lambda}\right) - \mathbb{P}\left(\left|Y_{\lambda_2} - \frac{\lambda_2}{2}\right| \geqslant \frac{C}{4}\sqrt{\lambda} - 1\right).
\end{aligned}$$

Applying Chebyshev's inequality, we get

$$1 \geqslant d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) \geqslant 1 + O\left(C^{-1}\right),$$

if $C \geqslant 8$.

When $\lambda_3 < \lambda_2 - 1$, we have the lower bound

$$\begin{aligned}
d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) &\geqslant \frac{2^{\lambda_2} d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) - 2^{\lambda_3} - \cdots - 2^{\lambda_s}}{n}\\
&\geqslant \frac{d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) - 1/2}{2^{\lambda-\lambda_2}}\\
&\geqslant \frac{1}{2^{\lambda-\lambda_2+1}}\left(1 + O\left(C^{-1}\right)\right).
\end{aligned}$$

On the other hand, when $\lambda_3 = \lambda_2 - 1$, we use (32) and get

$$\begin{aligned}
&2d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda))\\
&\quad\geqslant \frac{1}{n}\sum_{\ell\geqslant 0}\big|2^{\lambda_2}\big(\mathbb{P}(Y_{\lambda_2} = \ell-1) - \mathbb{P}(Y_\lambda = \ell)\big)\\
&\qquad\quad + 2^{\lambda_3}\big(\mathbb{P}(Y_{\lambda_3} = \ell-2) - \mathbb{P}(Y_\lambda = \ell)\big)\big| - \frac{2^{\lambda_4} + \cdots + 2^{\lambda_s}}{n}\\
&\quad\geqslant \frac{1}{n}\sum_{\ell\geqslant 0}\big|(2^{\lambda_2} + 2^{\lambda_3})\big(\mathbb{P}(Y_{\lambda_2} = \ell-1) - \mathbb{P}(Y_\lambda = \ell)\big)\\
&\qquad\qquad + 2^{\lambda_3}\big(\mathbb{P}(Y_{\lambda_3} = \ell-2) - \mathbb{P}(Y_{\lambda_2} = \ell-1)\big)\big| - \frac{2^{\lambda_4+1}}{n}\\
&\quad\geqslant \frac{1}{n}\left((2^{\lambda_2} + 2^{\lambda_3}) d_{\mathrm{TV}}(\mathscr{L}(Y_{\lambda_2}+1), \mathscr{L}(Y_\lambda)) - 2^{\lambda_3+1}\right)\\
&\quad\geqslant \frac{1}{2^{\lambda-\lambda_2+1}}\left(1 + O\left(C^{-1}\right)\right).
\end{aligned}$$

This completes the proof of Theorem 2.4.

3.3. Stein's method: An alternative proof of Theorem 2.4

The sum-of-digits function was among one of the first instances used to demonstrate the effectiveness of Stein's method (see [31,129]) with an optimal approximation rate. This method centers on exploiting an equation that characterizes the limiting measure, which, in the case of binomial distribution, is given by (15) and can be derived in the following way.

3.3.1. Stein's equation for binomial distribution

Since Y_k is binomially distributed with parameters k and $p \in (0,1)$, we see that the probabilities $\mathbb{P}(Y_k = j)$ satisfy the difference equation

$$q\big((j+1)\mathbb{P}(Y_k = j+1) - j\mathbb{P}(Y_k = j)\big) = (p(k-j) - jq)\mathbb{P}(Y_k = j). \tag{33}$$

Following Stein's idea [128] for deriving the characteristic equation for the normal distribution

$$\mathbb{E}\left(f'(\mathscr{N})\right) = \mathbb{E}\left(\mathscr{N} f(\mathscr{N})\right) \qquad (f \in C^1(\mathbb{R})),$$

by using integration by parts, we consider the average

$$q\mathbb{E}\big(g(Y_k) - g(Y_k - 1)\big)Y_k = q \sum_{0 \leqslant j \leqslant k} \big(g(j) - g(j-1)\big)\mathbb{P}(Y_k = j),$$

and apply summation by parts, which yields, by (33),

$$\begin{aligned} &q\mathbb{E}\big(g(Y_k) - g(Y_k - 1)\big)Y_k \\ &\quad = q \sum_{0 \leqslant j \leqslant k} g(j) q\big(j\mathbb{P}(Y_k = j) - (j+1)\mathbb{P}(Y_k = j)\big) \\ &\quad = \sum_{0 \leqslant j \leqslant k} g(j)\mathbb{P}(Y_k = j)(jq - p(k-j))\mathbb{P}(Y_k = j) \\ &\quad = q\mathbb{E}\left(Y_k g(Y_k)\right) - p\mathbb{E}\left((k - Y_k)g(Y_k)\right). \end{aligned}$$

Thus the identity

$$q\mathbb{E}\left(Y_k g(Y_k - 1)\right) = p\mathbb{E}\left((k - Y_k)g(Y_k)\right) \tag{34}$$

holds for any function $g : \{0, 1, 2, \ldots, k\} \to \mathbb{R}$.

A simpler proof of (34) starts with the relation

$$q(j+1)\mathbb{P}(Y_k = j+1) = p(k-j)\mathbb{P}(Y_k = j) \qquad (0 \leqslant j < k), \tag{35}$$

multiply both sides by $g(j)$, and sum over all indices j, giving rise to

$$q \sum_{0 \leqslant j \leqslant k} (j+1)g(j)\mathbb{P}(Y_k = j+1) = p \sum_{0 \leqslant j \leqslant k} (k-j)g(j)\mathbb{P}(Y_k = j),$$

which is nothing but (34).

Conversely, if for some discrete random variable Z the identity

$$\mathbb{E}\big(qZg(Z-1) + p(Z-k)g(Z)\big) = 0$$

holds for any function $g(j)$, then the probabilities $\mathbb{P}(Z = j)$ satisfy the equation (35) as $\mathbb{P}(Y_k = j)$. Thus

$$\mathbb{P}(Z = j) = \mathbb{P}(Y_k = j).$$

3.3.2. Binomial approximation

In the special case when $p = q = 1/2$, we have

$$\mathbb{E}\big((Y_k - k)g(Y_k) + Y_k g(Y_k - 1)\big) = 0. \tag{36}$$

Thus we expect that the above quantity will be small for any random variable whose distribution is close to $\mathrm{Binom}(k, 1/2)$. Assume $h : \{0, 1, \ldots, \lambda\} \to \mathbb{C}$ is an arbitrary function. Let g be a solution to the recurrence relation

$$(x - \lambda)g(x) + xg(x - 1) = h(x) - \mathbb{E}(h(Y_\lambda)). \tag{37}$$

Then we can represent the difference of means as

$$\mathbb{E}(h(X_n)) - \mathbb{E}(h(Y_\lambda)) = \mathbb{E}\big((X_n - \lambda)g(X_n) + X_n g(X_n - 1)\big).$$

By Stein's equation (37), the expectation on the right-hand side of the above identity will be zero if X_n were distributed according to binomial distribution $B(\lambda, 1/2)$. Thus we expect that this quantity will be small if the distribution of X_n is close to $B(\lambda, 1/2)$.

Recalling that A_j is defined in (31), we see that $\mathbb{P}(X_n < x | U_n \in A_j) = \mathbb{P}(Y_{\lambda_j} + j - 1 < x)$. It follows that

$$\begin{aligned}
&\mathbb{E}\left(h\big(X_n\big)\right) - \mathbb{E}(h(Y_\lambda)) \\
&\quad = \mathbb{E}\big((X_n - \lambda)g(X_n) + X_n g(X_n - 1)\big) \\
&\quad = \sum_{0 \leqslant j < s} \mathbb{P}(U_n \in A_j)\mathbb{E}\big((X_n - \lambda)g(X_n) + X_n g(X_n - 1) | U_n \in A_j\big) \\
&\quad = \sum_{1 \leqslant j \leqslant s} \frac{2^{\lambda_j}}{n}\mathbb{E}\big((Y_{\lambda_j} + j - 1) - \lambda)g(Y_{\lambda_j} + j - 1) \\
&\qquad\qquad + (Y_{\lambda_j} + j - 1)g(Y_{\lambda_j} + j - 2)\big).
\end{aligned}$$

The term with $j = 1$ in the last sum is zero since Y_λ is binomially distributed. Hence, with $g_j(x) := g(x + j - 1)$, we then obtain

$$\begin{aligned}
&\mathbb{E}\left(h\big(X_n\big)\right) - \mathbb{E}(h(Y_\lambda)) \\
&\quad = \sum_{2 \leqslant j \leqslant s} \frac{2^{\lambda_j}}{n}\mathbb{E}\big((\lambda_j - \lambda + j - 1)g_j(Y_{\lambda_j}) + (j - 1)g_j(Y_{\lambda_j} - 1)\big) \\
&\qquad + \sum_{2 \leqslant j \leqslant s} \frac{2^{\lambda_j}}{n}\mathbb{E}\big((Y_{\lambda_j} - \lambda_j)g_j(Y_{\lambda_j}) + Y_{\lambda_j} g_j(Y_{\lambda_j} - 1)\big).
\end{aligned}$$

But the second sum is identically zero by (36). It follows that

$$\begin{aligned}
&\mathbb{E}\big(h(X_n)\big) - \mathbb{E}(h(Y_\lambda)) \\
&\quad = \sum_{2 \leqslant j \leqslant s} \frac{2^{\lambda_j}}{n}\mathbb{E}\big((\lambda_j - \lambda + j - 1)g(Y_{\lambda_j} + j - 1) + (j - 1)g(Y_{\lambda_j} + j - 2)\big),
\end{aligned}$$

which can be alternatively rewritten as

$$\mathbb{E}\big(h(X_n)\big) - \mathbb{E}(h(Y_\lambda)) = \mathbb{E}\big(Q_1 g(X_n)\big) + \mathbb{E}\big(Q_2 g(X_n - 1)\big), \tag{38}$$

where Q_1 is a random variable taking value $\lambda_j - \lambda + j - 1$ if $U_n \in A_j$ and Q_2 takes value $j-1$ if $U_n \in A_j$ for $1 \leqslant j \leqslant s$. Note that

$$\begin{aligned}
\mathbb{E}(Q_1) &= \sum_{1\leqslant j\leqslant s} \mathbb{P}(U_n \in A_j)(\lambda_j - \lambda + j - 1) \\
&= \sum_{1\leqslant j\leqslant s} \frac{2^{\lambda_j}}{n}(\lambda - \lambda_j + j - 1) \\
&= O\left(\frac{\lambda - \lambda_2}{2^{\lambda-\lambda_j}}\right),
\end{aligned}$$

and, similarly, $\mathbb{E}(Q_2) = O\left(1/2^{\lambda-\lambda_j}\right)$.

3.3.3. Solving the equation $(x-\lambda)g(x) + xg(x-1) = h(x) - \mathbb{E}(h(Y_\lambda))$

Solving the equation (37) is equivalent to finding the solution x_m of the difference equation

$$(k-m)x_m - mx_{m-1} = \delta_m,$$

for $1 \leqslant m \leqslant k$ (note that x_{-1} and x_n do not affect the solution of this equation and therefore can be assumed to be equal to zero), where the δ_m's are given and satisfy the condition

$$\delta_0\binom{k}{0} + \delta_1\binom{k}{1} + \cdots + \delta_k\binom{k}{k} = 0.$$

The solution is obtained by introducing new variables $z_m = (k-m)\binom{k}{m}x_m$ for which our difference equation takes form

$$z_m - z_{m-1} = \binom{k}{m}\delta_m,$$

Iterating this, we obtain the following solution to Stein's equation (37).

Lemma 3.5 ([129])**.** *Let $Y_\lambda \sim \mathrm{Binom}(\lambda, 1/2)$. Define the function $g : \{0, 1, \ldots, \lambda - 1\} \to \mathbb{C}$ by*

$$g(m) = \frac{1}{(\lambda - m)\binom{\lambda}{m}} \sum_{0\leqslant r\leqslant m} \binom{\lambda}{r}\big(h(r) - \mathbb{E}(h(Y_\lambda))\big).$$

Then g is the solution to the recurrence equation

$$(x-\lambda)g(x) + xg(x-1) = h(x) - \mathbb{E}(h(Y_\lambda)),$$

for all $x \in \{1, \ldots, \lambda - 1\}$.

Note that

$$g(m) = -\frac{1}{(\lambda-m)\binom{\lambda}{m}} \sum_{m<r\leqslant\lambda} \binom{\lambda}{r}\big(h(r) - \mathbb{E}(h(Y_\lambda))\big).$$

Lemma 3.6. *The sequence*

$$y_m := \frac{1}{\binom{\lambda}{m}} \sum_{0\leqslant r\leqslant m} \binom{\lambda}{r}$$

is monotonically increasing in m.

Proof. By induction using the recurrence relation

$$y_{m+1} = \frac{m+1}{\lambda-m} y_m + 1,$$

and the monotonicity of $\frac{m+1}{\lambda-m}$. □

Lemma 3.7 ([88], [125])**.** *If $0 \leqslant h(m) \leqslant 1$, then the solution of Stein's equation provided by Lemma 3.5 satisfies the uniform estimate*

$$|g(m)| = O\left(\lambda^{-1/2}\right) \qquad (m = 0, 1, \dots).$$

Proof. If $m \leqslant \lambda/2$, then, by the monotonicity of the sequence y_m, we obtain

$$\begin{aligned}
|g(m)| &\leqslant \frac{1}{(\lambda-m)\binom{\lambda}{m}} \sum_{0\leqslant r\leqslant m} \binom{\lambda}{r} = \frac{y_m}{\lambda-m} \leqslant \frac{y_{\lfloor\lambda/2\rfloor}}{\lambda-\lfloor\lambda/2\rfloor} \\
&= \frac{1}{(\lambda-\lfloor\lambda/2\rfloor)\binom{\lambda}{\lfloor\lambda/2\rfloor}} \sum_{0\leqslant r\leqslant\lfloor\lambda/2\rfloor} \binom{\lambda}{r} \\
&\leqslant \frac{2^\lambda}{\lfloor\lambda/2\rfloor\binom{\lambda}{\lfloor\lambda/2\rfloor}} = O\left(\lambda^{-1/2}\right).
\end{aligned}$$

The case when $m > \lfloor\lambda/2\rfloor$ is treated similarly. Indeed, if $m > \lfloor\lambda/2\rfloor$, then, using the identity

$$(k-m)\binom{k}{m} = (m+1)\binom{k}{k-m-1},$$

we have

$$\begin{aligned}
|g(m)| &\leqslant \frac{1}{(\lambda-m)\binom{\lambda}{m}} \sum_{m<r\leqslant\lambda} \binom{\lambda}{r} \\
&= \frac{1}{(m+1)\binom{\lambda}{\lambda-m-1}} \sum_{0\leqslant r<\lambda-m} \binom{\lambda}{r} = \frac{y_{\lambda-m-1}}{m+1} \leqslant \frac{y_{\lambda-\lfloor\lambda/2\rfloor}}{m+1} \\
&\leqslant \frac{2^\lambda}{\lfloor\lambda/2\rfloor\binom{k}{\lfloor\lambda/2\rfloor}} = O\left(\lambda^{-1/2}\right).
\end{aligned}$$

□

3.3.4. Proof of Theorem 2.4 by Stein's method

Assume now that $A \subset \mathbb{R}$ is an arbitrary set. Define

$$h(m) := I_A(m) = \begin{cases} 1, & \text{if } m \in A; \\ 0, & \text{if } m \notin A. \end{cases}$$

Then

$$\begin{aligned} P\big(X_n \in A\big) - P\big(Y_\lambda \in A\big) &= \mathbb{E}\big(Q_1 g(X_n)\big) + \mathbb{E}\big(Q_2 g(X_n - 1)\big) \\ &= O\left(\frac{\mathbb{E}(Q_1) + \mathbb{E}(Q_2)}{\sqrt{\lambda}}\right) \\ &= O\left(\frac{\lambda - \lambda_2}{2^{\lambda-\lambda_2}\sqrt{\lambda}}\right). \end{aligned}$$

Thus

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) = O\left(\frac{\lambda - \lambda_2}{2^{\lambda-\lambda_2}\sqrt{\lambda}}\right).$$

3.3.5. A refinement of Theorem 2.4

A finer result can be obtained by using the following lemma.

Lemma 3.8 ([8]). *If $0 \leqslant h(x) \leqslant 1$ and g is defined in (3.5), then*

$$\max_{1 \leqslant j \leqslant \lambda - 1} |g(j) - g(j-1)| \leqslant 2 \min\left\{\frac{1}{j}, \frac{1}{\lambda - j}\right\} \leqslant \frac{4}{\lambda}.$$

Proof. If $m \leqslant \lambda/2$, then

$$\begin{aligned} &g(m) - g(m-1) \\ &\quad = \left(\frac{1}{(\lambda - m)\binom{\lambda}{m}} - \frac{1}{(\lambda - m + 1)\binom{\lambda}{m-1}}\right) \sum_{0 \leqslant r < m} \binom{\lambda}{r}\big(h(r) - \mathbb{E}(h(Y_\lambda))\big) \\ &\qquad + \frac{h(m) - \mathbb{E}(h(Y_\lambda))}{\lambda - m}. \end{aligned}$$

By the elementary inequality

$$\binom{\lambda}{m - r} \leqslant \left(\frac{m}{\lambda - m + 1}\right)^r \binom{\lambda}{m},$$

we see that

$$\begin{aligned} |g(m) - g(m-1)| &\leqslant \left(\frac{1}{m} - \frac{1}{\lambda - m}\right) \sum_{0 \leqslant r \leqslant m-1} \frac{\binom{\lambda}{r}}{\binom{\lambda}{m}} + \frac{1}{\lambda - m} \\ &\leqslant \frac{\lambda - 2m}{m(\lambda - m)} \sum_{1 \leqslant r \leqslant m} \left(\frac{m}{\lambda - m + 1}\right)^r + \frac{1}{\lambda - m} \end{aligned}$$

$$
\begin{aligned}
&\leqslant \frac{\lambda-2m}{m(\lambda-m)}\cdot\frac{\frac{m}{\lambda-m+1}}{1-\frac{m}{\lambda-m+1}}+\frac{1}{\lambda-m}\\
&=\frac{\lambda-2m}{(\lambda-m)(\lambda-2m+1)}+\frac{1}{\lambda-m}\\
&\leqslant\frac{2}{\lambda-m}.
\end{aligned}
$$

In a similar way we obtain the estimate

$$|g(m)-g(m-1)|\leqslant\frac{2}{m},$$

in the case when $m>\lambda/2$. □

The following result is similar in nature to that obtained by Soon [125] for unbounded function $h(x)$ that he later applied to derive several large and moderate deviations results for X_n.

Proposition 3.9. *Assume h is any real function such that $0\leqslant h(x)\leqslant 1$. Then*

$$\mathbb{E}(h(X_n))-\mathbb{E}(h(Y_\lambda))=4a_1(n)\mathbb{E}\left(h(Y_\lambda)\frac{\lambda/2-Y_\lambda}{\lambda}\right)+O\left(\frac{(\lambda-\lambda_2)^2}{2^{\lambda-\lambda_2}\lambda}\right),$$

where $a_1(n)=F(\log_2 n)$ is defined in (20).

Proof. The lemma implies that $g(x+j-1)=g(x)+O(j/k)$. Since Y_{k+s} has the same distribution as Y_k+W_s, where W_s is independently and binomially distributed $W_s\sim B(s,1/2)$, we can replace the mean $\mathbb{E}(g(Y_{k+s}))$ by $\mathbb{E}(g(Y_k+W_s))$, the error so introduced being bounded above by

$$\mathbb{E}(g(Y_{k+s}))-\mathbb{E}(g(Y_k))=\mathbb{E}\big(g(Y_k+W_s)-\mathbb{E}(g(Y_k))\big)=O\left(\frac{s}{k}\right),$$

where we used the estimate $|W_s|\leqslant s$. Thus

$$
\begin{aligned}
&\mathbb{E}(h(X_n))-\mathbb{E}(h(Y_\lambda))\\
&\quad=\sum_{2\leqslant j\leqslant s}\frac{2^{\lambda_j}}{n}\mathbb{E}\big((\lambda_j-\lambda+j-1)g(Y_{\lambda_j}+j-1)\\
&\qquad+(j-1)g(Y_{\lambda_j}+j-2)\big)\\
&\quad=\sum_{2\leqslant j\leqslant s}\frac{2^{\lambda_j}}{n}\big(\lambda_j-\lambda+2(j-1)\big)\mathbb{E}\left(g(Y_{\lambda_j}+j-1)\right)\\
&\qquad+O\left(\frac{1}{2^{\lambda-\lambda_2}\lambda}\right)\\
&\quad=\mathbb{E}(g(Y_\lambda))\sum_{2\leqslant j\leqslant s}\frac{2^{\lambda_j}}{n}\big(\lambda_j-\lambda+2(j-1)\big)+O\left(\frac{(\lambda-\lambda_2)^2}{2^{\lambda-\lambda_2}\lambda}\right)\\
&\quad=2a_1(n)\mathbb{E}(g(Y_\lambda))+O\left(\frac{(\lambda-\lambda_2)^2}{2^{\lambda-\lambda_2}\lambda}\right).
\end{aligned}
$$

We now evaluate the quantity $\mathbb{E}(g(Y_\lambda))$ appearing in the last expression

$$\begin{aligned}\mathbb{E}(g(Y_\lambda)) &= \frac{1}{2^\lambda}\sum_{0\leqslant m<\lambda}\binom{\lambda}{m}\frac{1}{(\lambda-m)\binom{\lambda}{m}}\sum_{0\leqslant r\leqslant m}\binom{\lambda}{r}\big(h(r)-\mathbb{E}(h(Y_\lambda))\big)\\ &= \frac{1}{2^\lambda}\sum_{0\leqslant r<\lambda}\binom{\lambda}{r}\big(h(r)-\mathbb{E}(h(Y_\lambda))\big)\sum_{r\leqslant m<\lambda}\frac{1}{\lambda-m}\\ &= \mathbb{E}\left(\big(h(Y_\lambda)-\mathbb{E}(h(Y_\lambda))\big)\sum_{Y_\lambda\leqslant m<\lambda}\frac{1}{\lambda-m}\right)\\ &= \mathbb{E}\left(\big(h(Y_\lambda)-\mathbb{E}(h(Y_\lambda))\big)\left(\sum_{Y_\lambda\leqslant m<\lambda}\frac{1}{\lambda-m}-\sum_{\lambda/2\leqslant m<\lambda}\frac{1}{\lambda-m}\right)\right),\end{aligned}$$

since $\mathbb{E}\big(h(Y_\lambda)-\mathbb{E}h(Y_\lambda)\big)=0$ and the sum $\sum_{\lambda/2\leqslant m<\lambda}(\lambda-m)^{-1}$ is a constant independent of Y_λ. Then

$$\mathbb{E}(g(Y_\lambda)) = -\sum_{0\leqslant r<\lambda}\mathbb{P}(Y_\lambda=r)\big(h(r)-\mathbb{E}(h(Y_\lambda))\big)\sum_{m=\lambda/2}^{r}\frac{1}{\lambda-m},$$

where we use the convention that $\sum_{m=a}^{b} = -\sum_{m=b}^{a}$. We then split the sum into two parts and obtain

$$\begin{aligned}\mathbb{E}(g(Y_\lambda)) = -&\sum_{|r-\lambda/2|\leqslant\lambda^{3/4}}\mathbb{P}(Y_\lambda=r)\big(h(r)-\mathbb{E}(h(Y_\lambda))\big)\sum_{m=\lambda/2}^{r}\frac{1}{\lambda-m}\\ &+O(\mathbb{P}(|Y_\lambda-\lambda/2|>\lambda^{3/4})).\end{aligned}$$

If $\lambda/2\leqslant r\leqslant\lambda/2+\lambda^{3/4}$, then

$$\begin{aligned}\sum_{m=\lambda/2}^{r}\frac{1}{\lambda-m} &= \sum_{m=\lambda/2}^{r}\left(\frac{1}{\lambda-m}-\frac{1}{\lambda/2}\right)+\frac{r-\lceil\lambda/2\rceil+1}{\lambda/2}\\ &= \frac{r-\lambda/2}{\lambda/2}+\sum_{m=\lambda/2}^{r}\frac{m-\lambda/2}{(\lambda-m)\lambda/2}+O(\lambda^{-1})\\ &= \frac{r-\lambda/2}{\lambda/2}+O\left(\frac{(r-\lambda/2)^2}{\lambda^2}+\lambda^{-1}\right).\end{aligned}$$

The same estimate holds when r lies in the range $\lambda/2-\lambda^{3/4}\leqslant r\leqslant\lambda/2$. Thus

$$\begin{aligned}\mathbb{E}(g(Y_\lambda)) = 2\mathbb{E}&\left(\big(h(Y_\lambda)-\mathbb{E}(h(Y_\lambda))\big)\frac{\lambda/2-Y_\lambda}{\lambda}\right)\\ &+O\left(\frac{1}{\lambda^2}\mathbb{E}\big|h(Y_\lambda)-\mathbb{E}(h(Y_\lambda))\big|(Y_\lambda-\lambda/2)^2\right)\\ &+O\left(\lambda^{-1}+\mathbb{P}(|Y_\lambda-\lambda/2|>\lambda^{3/4})\right)\end{aligned}$$

$$= 2\mathbb{E}\left(\big(h(Y_\lambda) - \mathbb{E}(h(Y_\lambda))\big)\frac{\lambda/2 - Y_\lambda}{\lambda}\right)$$
$$+ O\left(\frac{1}{\lambda^2}\mathbb{E}(Y_\lambda - \lambda/2)^2\right) + O(\lambda^{-1})$$
$$= 2\mathbb{E}\left(\big(h(Y_\lambda) - \mathbb{E}(h(Y_\lambda))\big)\frac{\lambda/2 - Y_\lambda}{\lambda}\right) + O(\lambda^{-1})$$
$$= 2\mathbb{E}\left(h(Y_\lambda)\frac{\lambda/2 - Y_\lambda}{\lambda}\right) + O(\lambda^{-1}),$$

since $\mathbb{E}\,(\lambda/2 - Y_\lambda) = 0$. This proves the proposition. □

3.3.6. Corollaries of Proposition 3.9

Corollary 3.10. *We have*

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) = |a_1(n)|\mathbb{E}\left|\frac{Y_\lambda - \lambda/2}{\lambda/2}\right| + O\left(\frac{(\lambda-\lambda_2)^2}{2^{\lambda-\lambda_2}\lambda}\right). \tag{39}$$

Proof. By the definition of the total variation distance

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) = \sup_h \big|\mathbb{E}(h(X_n)) - \mathbb{E}(h(Y_\lambda))\big|,$$

where the supremum is taken over all functions h assuming only binary values $\{0, 1\}$. It is easy to see that the supremum of the average containing h in the above relation is reached by the function

$$h(x) = \begin{cases} 1, & \text{if} \quad x \leqslant \lambda/2, \\ 0, & \text{if} \quad x > \lambda/2, \end{cases}$$

and we thus get, by Proposition 3.9, the estimate

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) = |a_1(n)|\mathbb{E}\left|\frac{Y_\lambda - \lambda/2}{\lambda/2}\right| + O\left(\frac{(\lambda-\lambda_2)^2}{2^{\lambda-\lambda_2}\lambda}\right).$$ □

Corollary 3.11. *For all $c \leqslant \lambda - \lambda_2$ with c large enough, we have*

$$d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) = \frac{|a_1(n)|}{\sqrt{2\pi\lambda}}\left(1 + O\left(\frac{\lambda-\lambda_2}{\sqrt{\lambda}}\right)\right).$$

Proof. This follows from the estimate (39) because

$$\mathbb{E}\left|\frac{Y_\lambda - \lambda/2}{\sqrt{\lambda}/2}\right| = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}|x|e^{-x^2/2}\,dx + O(\lambda^{-1/2}) = \sqrt{\frac{2}{\pi}} + O(\lambda^{-1/2}),$$

and the quantity $|a_1(n)|$ can be bounded from bellow by

$$\frac{\lambda-\lambda_2}{2^{\lambda-\lambda_2}} = O(a_1(n)),$$

if $\lambda - \lambda_2 > c$ with $c > 0$ large enough. □

Remark. The Stein method we adopted above for the analysis of sum-of-digits function differs from the original approach by Stein [129]. First, he used $Y_{\lambda+1}$ in lieu of Y_λ, as a good approximation to X_n, and derived a uniform bound for the point probability. Second, instead of exploiting the fact that X_n is a mixture of binomial distributions his analysis is more subtle and is based on the construction of an exchangeable pair. By doing so he managed to simplify the right-hand side of (37) to

$$\begin{aligned}\mathbb{E}(h(X_n)) - \mathbb{E}(h(Y_{\lambda+1})) &= \mathbb{E}\big((X_n - (\lambda+1))g_h(X_n) + X_n g_h(X_n - 1)\big) \\ &= \mathbb{E}(Q g_h(X_n)),\end{aligned} \tag{40}$$

where Q is a random variable such that $0 \leqslant \mathbb{E}(Q) \leqslant 2$ and g_h is the solution to the recurrence equation

$$h(x) - \mathbb{E}(h(Y_{\lambda+1})) = \big(x - (\lambda+1)\big)g(x) + xg(x-1),$$

for $x \in \{0, 1, \ldots, \lambda\}$ whose precise expression is given by Lemma 3.5 with $k = \lambda + 1$. The estimate of Lemma 3.7 together with the property $0 \leqslant \mathbb{E}(Q) \leqslant 2$ now immediately give the estimate of the total variation distance $d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(Y_{\lambda+1})) = O(\lambda^{-1/2})$. Further applications of similar ideas will be explored elsewhere.

3.4. The Krawtchouk-Parseval approach: χ^2-distance

We examine in this section yet another approach based on properties of the Krawtchouk polynomials and the Parseval identity (or more generally Plancherel's formula). The approach is the binomial analogue of the Charlier-Parseval approach we developed and explored earlier in [143]. We consider only the simplest case of deriving the χ^2-distance, leaving the extension to other distances to the interested reader, which follows readily from the framework developed in [143].

3.4.1. Krawtchouk polynomials

Krawtchouk (or Kravchuk) polynomials, introduced in the late 1920s, are polynomials orthogonal with respect to the binomial distribution. Over the years, they were frequently encountered in a variety of areas, including combinatorics, number theory, asymptotic analysis, image analysis, coding theory, etc. In probability theory, their appearance is perhaps even more anticipated than in other areas due to the prevalence of binomial distribution, and sometimes without noticing the explicit connection; see Diaconis's monograph [32] for more information and applications. See also [70, 90, 122, 136] for more recent update. Despite the large literature on diverse properties of Krawtchouk polynomials and the high usefulness of the Parseval identity, we did not find application of the corresponding Parseval identity similar to ours; see however [33, 103] for a direct manipulation of Fourier integrals.

We start with reviewing the definition of Krawtchouk polynomials and some of their well-known properties (see [133, pp. 35–37]).

Assume that p and q are nonnegative integers such that $p+q=1$.
Introduce the notation

$$B(N,t)=\binom{N}{t}p^t q^{N-t}.$$

The Krawtchouk polynomials $K_n(t)=K_n(N,t)$ are defined by

$$\sum_{0\leqslant j\leqslant N} K_j(t)w^j=\left(\frac{1+qw}{1-pw}\right)^t(1-pw)^N. \tag{41}$$

Multiplying both sides by $B(N,t)z^t$ and summing over all t from 0 to N, we obtain

$$\begin{aligned}\sum_{0\leqslant t\leqslant N} B(N,t)z^t \sum_{0\leqslant j\leqslant N} K_j(x)w^j &= \big(pz(1+qw)+q(1-pw)\big)^N\\ &= \big(pz+q+wpq(z-1)\big)^N\\ &= \sum_{0\leqslant j\leqslant N}\binom{N}{j}w^j(pq)^j(z-1)^j(pz+q)^{N-j}.\end{aligned}$$

Taking the coefficients of w^n on both sides, we get

$$\sum_{0\leqslant t\leqslant N} B(N,t)K_n(t)z^t=\binom{N}{n}(pq)^n(z-1)^n(pz+q)^{N-n}.$$

On the other hand, by (41), we have

$$\begin{aligned}&\sum_{0\leqslant n,m\leqslant N}\left(\sum_{0\leqslant t\leqslant N} B(N,t)K_n(t)K_m(t)\right)w^n z^n\\ &\quad=\sum_{0\leqslant t\leqslant N} B(N,t)\big((1+qw)(1+qz)\big)^t\big((1-pw)(1-pw)\big)^{N-x}\\ &\quad=\big(p(1+qw)(1+qz)+q(1-pw)(1-pw)\big)^N\\ &\quad=(1+pqzw)^N.\end{aligned}$$

Accordingly, we obtain the orthogonality relation

$$\sum_{0\leqslant t\leqslant N} B(N,t)K_n(t)K_m(t)=\delta_{m,n}\binom{N}{n}(pq)^n.$$

3.4.2. The Parseval identity for Krawtchouk polynomials

Let $F(z)$ be a polynomial of degree not greater than N. Thus

$$f(z)=\sum_{0\leqslant t\leqslant N} f_t z^t,$$

and we have the expansion

$$\frac{f_t}{B(N,t)} = \sum_{0\leqslant j\leqslant N} b_j K_j(t). \tag{42}$$

Taking square of the above identity, multiplying it by $B(N,t)$ and summing the resulting identity with respect to t, we obtain

$$\sum_{0\leqslant t\leqslant N} \left|\frac{f_t}{B(N,t)}\right|^2 B(N,t) = \sum_{0\leqslant j\leqslant N} |b_j|^2 \binom{N}{j} (pq)^j. \tag{43}$$

By the definition (41), we deduce that

$$(1-pw)^N f\left(\frac{1+qw}{1-pw}\right) = \sum_{j=0}^{N} b_j \binom{N}{j} (pq)^j w^j.$$

Comparing this identity with (43), we conclude that

$$\sum_{0\leqslant t\leqslant N} \left|\frac{f_t}{B(N,t)}\right|^2 B(N,t) = \sum_{0\leqslant j\leqslant N} \frac{|c_j|^2}{\binom{N}{j}(pq)^j},$$

where c_j is defined by

$$(1-pw)^N f\left(\frac{1+qw}{1-pw}\right) = \sum_{0\leqslant j\leqslant N} c_j w^j.$$

Now by the Parseval identity

$$\begin{aligned} J(f,N;r) &:= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|(1-pre^{it})^N f\left(\frac{1+qre^{it}}{1-pre^{it}}\right)\right|^2 \mathrm{d}t \\ &= \sum_{0\leqslant j\leqslant N} |c_j|^2 r^{2j}. \end{aligned} \tag{44}$$

Comparing (44) with (43), and using the relation

$$(N+1)\int_0^\infty \frac{u^j}{(1+u)^{N+2}} = \binom{N}{j}^{-1},$$

we obtain the *Krawtchouk-Parseval identity*

$$\sum_{0\leqslant t\leqslant N} \left|\frac{f_t}{B(N,t)}\right|^2 B(N,t) = (N+1)\int_0^\infty \frac{J\left(f,N;\sqrt{\frac{u}{pq}}\right)}{(1+u)^{N+2}}\,\mathrm{d}u, \tag{45}$$

which is crucial for deriving the asymptotics of the χ^2-distance.

3.4.3. The χ^2-distance

For any non-negative integer-valued random variables Z and W, the χ^2-distance is defined by

$$\chi^2(\mathscr{L}(Z), \mathscr{L}(W)) = \sum_{j\geqslant 0} \left(\frac{\mathbb{P}(Z=j)}{\mathbb{P}(W=j)} - 1\right)^2 \mathbb{P}(W=j),$$

provided that the series on the right-hand side has a meaning. It possesses two important properties. First, its square root upper-bounds the total variation distance

$$d_{\mathrm{TV}}(\mathscr{L}(Z), \mathscr{L}(W)) \leqslant \frac{1}{2}\sqrt{\chi^2(\mathscr{L}(Z), \mathscr{L}(W))}.$$

Second, it also provides an effective upper bound for the Kullback-Leibler divergence (or information divergence)

$$d_{\mathrm{KL}}(\mathscr{L}(Z), \mathscr{L}(W)) := \sum_{j\geqslant 0} \mathbb{P}(Z=j) \log \frac{\mathbb{P}(Z=j)}{\mathbb{P}(W=j)} \leqslant \chi^2(\mathscr{L}(Z), \mathscr{L}(W)),$$

a very useful measure in information theory and related applications.

Theorem 3.12. *The χ^2-distance between the distribution of X_n and the binomial distribution Y_λ satisfies*

$$\chi^2(\mathscr{L}(X_n), \mathscr{L}(Y_\lambda)) = O\left(\lambda^{-1}\right).$$

Proof. Let

$$f(z) := P_n(z) - \left(\frac{1+z}{2}\right)^\lambda,$$

where P_n is the probability generating function (23) of X_n. Then, by (13),

$$\left(1-\frac{w}{2}\right)^\lambda f\left(\frac{1+\frac{w}{2}}{1-\frac{w}{2}}\right) = \frac{1}{n}\sum_{2\leqslant j\leqslant s} 2^{\lambda_j}\left(\left(1+\frac{w}{2}\right)^{j-1}\left(1-\frac{w}{2}\right)^{\lambda-\lambda_j-j+1} - 1\right).$$

We need the elementary inequality

$$\left|(1+z)^a(1-z)^b - 1\right| \leqslant (1+|z|)^{a+b} - 1 \leqslant (a+b)|z|(1+|z|)^{a+b-1},$$

for nonnegative integers a, b with $a+b\geqslant 1$. Applying this inequality, we get, with $p=q=1/2$ and $N=\lambda$,

$$J(f,\lambda;r) \leqslant \frac{r^2}{4}\left|\sum_{2\leqslant j\leqslant s} \frac{\lambda-\lambda_j}{2^{\lambda-\lambda_j}}\left(1+\frac{r}{2}\right)^{\lambda-\lambda_j-1}\right|^2.$$

Substituting this estimate into the Krawtchouk-Parseval identity (45), we have

$$
\begin{aligned}
&\left(\sum_{0\leqslant m\leqslant\lambda}\left|\frac{\mathbb{P}\big(X_n=m\big)}{\frac{1}{2^\lambda}\binom{\lambda}{m}}-1\right|^2\frac{1}{2^\lambda}\binom{\lambda}{m}\right)^{1/2}\\
&\quad\leqslant\left((\lambda+1)\int_0^\infty\frac{J\left(f,\lambda;2\sqrt{u}\right)}{(1+u)^{\lambda+2}}\,\mathrm{d}u\right)^{1/2}\\
&\quad\leqslant\sum_{2\leqslant j\leqslant s}\frac{\lambda-\lambda_j}{2^{\lambda-\lambda_j}}\left((\lambda+1)\int_0^\infty\frac{u^2}{(1+u)^{\lambda_j+3}}\cdot\frac{(1+\sqrt{u})^{2(\lambda-\lambda_j-1)}}{(1+u)^{\lambda-\lambda_j-1}}\,\mathrm{d}u\right)^{1/2}\\
&\quad=\sum_{2\leqslant j\leqslant s}\frac{\lambda-\lambda_j}{2^{\lambda-\lambda_j}}\left((\lambda+1)\int_0^\infty\frac{u^2}{(1+u)^{\lambda_j+3}}\left(1+\frac{2\sqrt{u}}{1+u}\right)^{\lambda-\lambda_j-1}\mathrm{d}u\right)^{1/2}.
\end{aligned}
$$

Since the function $\sqrt{u}/(1+u)$ reaches the maximum at the point $u=1$, we see that

$$1+2\sqrt{u}/(1+u)\leqslant 2.$$

It follows that

$$
\begin{aligned}
&\big(\chi^2(\mathscr{L}(X_n),\mathscr{L}(Y_\lambda))\big)^{1/2}\\
&\quad\leqslant\sum_{2\leqslant j\leqslant s}\frac{\lambda-\lambda_j}{2^{(\lambda-\lambda_j+1)/2}}\left((\lambda+1)\int_0^\infty\frac{u^2}{(1+u)^{\lambda_j+3}}\,\mathrm{d}u\right)^{1/2}\\
&\quad\leqslant\sum_{2\leqslant j\leqslant s}\frac{\lambda-\lambda_j}{2^{(\lambda-\lambda_j+1)/2}}\left(\frac{\lambda+1}{(\lambda_j+2)(\lambda_j+1)}\right)^{1/2}\\
&\quad\leqslant\sqrt{\lambda+1}\sum_{2\leqslant j\leqslant s}\frac{\lambda-\lambda_j}{2^{(\lambda-\lambda_j+1)/2}(1+\lambda_j)}\\
&\quad\leqslant\frac{1}{\sqrt{\lambda+1}}\sum_{\lambda-\lambda_2\leqslant k\leqslant\lambda-\lambda_s}\frac{k(\lambda+1)}{2^{k/2}(1+\lambda-k)}.
\end{aligned}
$$

It is clear that

$$\sum_{\lambda-\lambda_2\leqslant k<\lambda}\frac{k(\lambda+1)}{2^{k/2}(1+\lambda-k)}=O(1).$$

This proves the theorem. □

Finer results can be derived by developing similar techniques as those used in [143] for Poisson approximation.

4. A general numeration system and applications

The properties we studied above can be readily extended to a more general framework of numeration system in which we encode each integer by a different binary string and impose the sole condition that

$$Z_{2n} \stackrel{d}{=} Z_n + I \qquad (n \geqslant 1), \tag{46}$$

where Z_n denotes the number of 1s in the resulting coding string for a random integer, assuming that each of the first n nonnegative integers is equally likely, and $I \sim \text{Bernoulli}(1/2)$. For definiteness, let $Z_0 = Z_1 = 0$. This simple scheme covers in particular the binary coding of X_n above (as can be easily checked) and binary reflected Gray code, which will be discussed in more detail later. Let $\mu(n)$ denote the number of 1s in the coding of n in such a numeration system. All our results below roughly say that this numeration system does not differ much from the binary coding although the codings inside each 2^k block can be rather flexible.

Theorem 4.1 (Local limit theorem)**.** *Assume that Z_n satisfies (46). Then Z_n is asymptotically normally distributed:*

$$\mathbb{P}\left(Z_n = \left\lfloor \frac{\lambda}{2} + x\frac{\sqrt{\lambda}}{2} \right\rfloor\right) = \frac{\sqrt{2}\, e^{-x^2/2}}{\sqrt{\pi\lambda}} \left(1 + O\left(\frac{1+|x|^3}{\sqrt{\lambda}}\right)\right), \tag{47}$$

uniformly for $x = o(\lambda^{1/6})$, with mean and variance satisfying

$$\begin{aligned} \mathbb{E}(Z_n) &= \frac{\log_2 n}{2} + G_1(\log_2 n), \\ \mathbb{V}(Z_n) &= \frac{\log_2 n}{4} + G_2(\log_2 n). \end{aligned} \tag{48}$$

Here G_1, G_2 are bounded periodic functions.

For results related to moderate deviations, see [18]. We can derive more precise Fourier expansions for the periodic functions G_1, G_2 when more information is available.

Theorem 4.2. *Assume that Z_n satisfies (46). Then*

$$\begin{aligned} &\sum_{0\leqslant k\leqslant\lambda} \left| \mathbb{P}\big(Z_n = k\big) - \sum_{0\leqslant r<m} (-1)^r b_r(n) 2^{-\lambda} \Delta^r \binom{\lambda}{k} \right| \\ &\qquad = \frac{h_m |b_m(n)|}{\lambda^{m/2}} + O\left(\lambda^{-(m+1)/2}\right), \end{aligned}$$

for $m = 1, 2, \ldots$, where the sequence $b_r(n) = b_r(2n)$ is defined by (see (13))

$$\mathbb{E}(y^{Z_n}) \left(\frac{1+y}{2}\right)^{-\lambda} = \sum_{r\geqslant 0} b_r(n)(y-1)^r. \tag{49}$$

In particular, $b_0 = 1$, $b_1(n) = \mathbb{E}(Z_n) - \lambda/2$, and

$$b_2(n) = \frac{\mathbb{E}(Z_n^2)}{2} - \frac{\lambda+1}{2}\,\mathbb{E}(Z_n) + \frac{\lambda(\lambda+1)}{8}.$$

Corollary 4.3.

$$d_{\mathrm{TV}}\big(\mathscr{L}(Z_n), \mathscr{L}(\mathrm{Y}_\lambda)\big) = \frac{\sqrt{2}|\bar{G}_1(\log_2 n)|}{\sqrt{\pi \log_2 n}} + O\left(\frac{1}{\log n}\right), \tag{50}$$

where $\bar{G}_1(\log_2 n) = \mathbb{E}(Z_n) - \lambda/2$ *is periodic* $\bar{G}_1(x+1) = \bar{G}_1(x)$ *and continuous on the set* $\mathbb{R} \setminus \mathbb{N}$.

The periodic function $\bar{G}_1(x)$ can be defined as follows. Write $2^x = \sum_{j\geqslant 0} \xi_j 2^{-j} \in [1,2)$.

$$\bar{G}_1(x) = 2^{-x} \sum_{j\geqslant 0} \frac{1-(-1)^{\lfloor 2^{j+x} \rfloor}}{2} 2^{-j} \left(\mu\left(\lfloor 2^{j-1+x} \rfloor\right) - \frac{j}{2}\right).$$

Note that $\bar{G}_1(x) = G_1(x) - \{x\}/2$.

Theorem 4.4. *Assume, as above, that* $n = 2^\lambda + 2^{\lambda_2} + \cdots + 2^{\lambda_s}$ *with* $\lambda > \lambda_2 > \cdots > \lambda_s \geqslant 0$. *Then*

$$d_{\mathrm{TV}}(\mathscr{L}(Z_n), \mathscr{L}(\mathrm{Y}_\lambda)) \asymp \frac{1}{2^{\lambda-\lambda_2}} \min\left\{1, \frac{\lambda-\lambda_2}{\sqrt{\lambda}}\right\},$$

whenever $\lambda - \lambda_2 \geqslant c$, *where* c *is sufficiently large.*

4.1. Sketches of proofs

Most of our analysis is based on the following explicit expression; cf. (13).

Lemma 4.5. *If* Z_n *satisfies the condition (46), then the probability generating function of* Z_n *satisfies*

$$\mathbb{E}\left(y^{Z_n}\right) = \frac{1}{n} \sum_{1\leqslant j\leqslant s} y^{\mu\left(\lfloor n/2^{\lambda_k} \rfloor -1\right)} (1+y)^{\lambda_k}, \tag{51}$$

where $n = 2^\lambda + 2^{\lambda_2} + \cdots + 2^{\lambda_s}$ *with* $\lambda > \lambda_2 > \cdots > \lambda_s \geqslant 0$.

Proof. Observe that the crucial condition (46) implies the recurrence

$$\mathbb{E}\left(y^{Z_{2n}}\right) = \frac{1+y}{2}\, \mathbb{E}\left(y^{Z_n}\right),$$

the same as (24) for X_n. Consequently, we also have, following the same analysis there,

$$\mathbb{E}\left(y^{Z_n}\right) = \frac{1}{n} \sum_{0\leqslant j\leqslant \lambda} \frac{1-(-1)^{\lfloor n/2^j \rfloor}}{2} y^{\mu\left(\lfloor n/2^j \rfloor -1\right)} (1+y)^j,$$

for $n \geqslant 1$; compare (25). Since 2 divides $\lfloor n/2^j \rfloor$ if and only if $j \notin \{\lambda_1, \lambda_2, \ldots, \lambda_s\}$, we obtain (51). □

From the expression (51), we easily obtain

$$\mathbb{E}(Z_n) = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j} \left(\bar{\mu}_j + \frac{\lambda_j}{2} \right),$$

$$\mathbb{E}(Z_n^2) = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j} \left(\bar{\mu}_j^2 + \lambda_j \bar{\mu}_j + \frac{\lambda_j^2 + \lambda_j}{4} \right),$$

where $\bar{\mu}_j := \mu(\lfloor n/2^{\lambda_j} \rfloor - 1)$. The identity for the mean in (48) then follows with

$$G_1(\log_2 n) = \frac{2^\lambda}{n} \sum_{1 \leqslant j \leqslant s} 2^{-(\lambda - \lambda_j)} \left(\bar{\mu}_j - \frac{\lambda - \lambda_j}{2} \right) - \frac{\{\log_2 n\}}{2},$$

which is periodic and bounded since $\bar{\mu}_j \leqslant j + 1$. Also the definition of G_1 here for $\log_2 n$ can be readily extended to all reals.

We now prove the identity in (48) for the variance as the proof is very simple. Let

$$\bar{Z}_n := Z_n - \frac{\lambda}{2} = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j} \left(\bar{\mu}_j - \frac{\lambda - \lambda_j}{2} \right).$$

Then

$$\begin{aligned}
&\mathbb{V}(Z_n) - \frac{\lambda}{4} \\
&\quad = \mathbb{E}(Z_n^2) - \left(\mathbb{E}(\bar{Z}_n) + \frac{\lambda}{2} \right)^2 - \frac{\lambda}{4} \\
&\quad = \frac{1}{n} \sum_{1 \leqslant j \leqslant s} 2^{\lambda_j} \left(\bar{\mu}_j (1 + \lambda_j - \lambda) + \frac{\lambda_j(\lambda_j + 1) - \lambda(\lambda + 1) + 2\lambda(\lambda - \lambda_j)}{4} \right) \\
&\qquad - \left(\mathbb{E}(\bar{Z}_n) \right)^2,
\end{aligned}$$

and we obtain the identity for the variance in (48) with

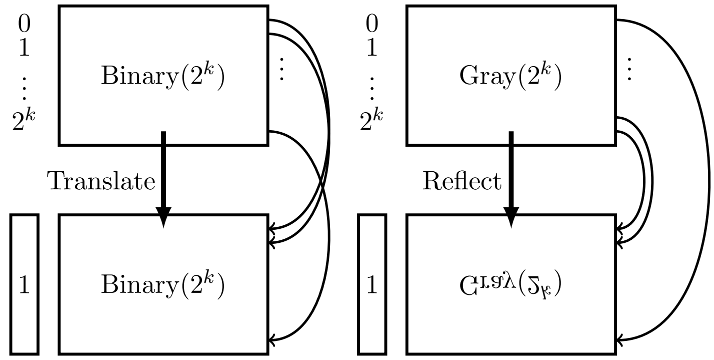
$$\begin{aligned}
G_2(\log_2 n) = \frac{2^\lambda}{n} \sum_{1 \leqslant j \leqslant s} 2^{-(\lambda - \lambda_j)} &\left(\bar{\mu}_j^2 - \bar{\mu}_j \, (\lambda - \lambda_j) + \frac{(\lambda - \lambda_j)(\lambda - \lambda_j - 1)}{4} \right) \\
&- G_1(\log_2 n)^2 - G_1(\log_2 n)\{\log_2 n\} - \frac{\{\log_2 n\}^2 + \{\log_2 n\}}{4},
\end{aligned}$$

which is also bounded and periodic, and extendible to all $x \in \mathbb{R}$.

The local limit theorem (47) is proved in a way similar to the proof of Proposition 3.2.

In terms of probabilities, the identity (51) means that the random variable Z_n can be expressed as a mixture of shifted binomial random variables. Its distribution can be described in the following way. Let ζ_n be a random variable defined by

$$\mathbb{P}(\zeta_n = j) = \frac{2^{\lambda_j}}{n} \qquad (1 \leqslant j \leqslant s).$$



0			0
1			1
2		1	1
3		1	0
4	1	1	0
5	1	1	1
6	1	0	1
7	1	0	0

Fig 12. *Constructions of binary code (left) and Gray code (middle), and the Gray code of the first few integers (right).*

Then

$$\mathbb{P}(Z_n \in A | \zeta_n = j) = \mathbb{P}(Y_{\lambda_j} + r_j \in A),$$

for any $A \subset \mathbb{R}$, where $r_j := \mu(\lfloor n/2^{\lambda_j} \rfloor - 1) \leqslant \lambda - \lambda_j + 1$ and $r_0 := 0$. By the same arguments used above, we see that the identity

$$\mathbb{E}\left(h\big(\mu(X_n)\big)\right) = \sum_{2 \leqslant j \leqslant s} \frac{2^{\lambda_j}}{n} \mathbb{E}\big((\lambda_j - \lambda + r_j) g(Y_{\lambda_j} + r_j) + r_j g(Y_{\lambda_j} + r_j - 1)\big)$$

holds for any function $h : \mathbb{R} \to \mathbb{R}$, where g is the solution to Stein's equation (37).

We skip all details of the proofs as they are almost identical to those for X_n.

4.2. Gray code

The Gray code is characterized by the property that the codings of any two successive integers differ by exactly one bit. It is named after Frank Gray's 1947 patent, although the same construction had been introduced in telegraphy in the late nineteenth century by the French engineer Émile Baudot; see Wikipedia's page on Gray code for more information. The coding notion with two neighboring objects differing at one location has turned out to be extremely useful in many scientific disciplines beyond the original communication motivations such as experimental designs, job scheduling in computer systems, and combinatorial generation; see the survey paper [119] and the references therein.

The binary reflected Gray code is constructed by reflecting (or mirroring) the first 2^k codings of the first 2^k nonnegative integers and then adding 1 at the beginning for each coding, resulting in the Gray code for the first 2^{k+1} nonnegative integers; see Figure 12 for an illustration.

By construction, the Gray code, say $\mathcal{G}(2^k+j)$ of 2^k+j with $0 \leqslant j < 2^k$ is equal to $10^\ell \mathcal{G}(2^k - 1 - j)$ (string concatenation), where $\ell := k - 1 - \lfloor \log_2(2^k - 1 - j) \rfloor$ and 0^ℓ means 0 written ℓ times. For example,

$$\mathcal{G}(19) = 1\mathcal{G}(12) = 110\mathcal{G}(3) = 11010.$$

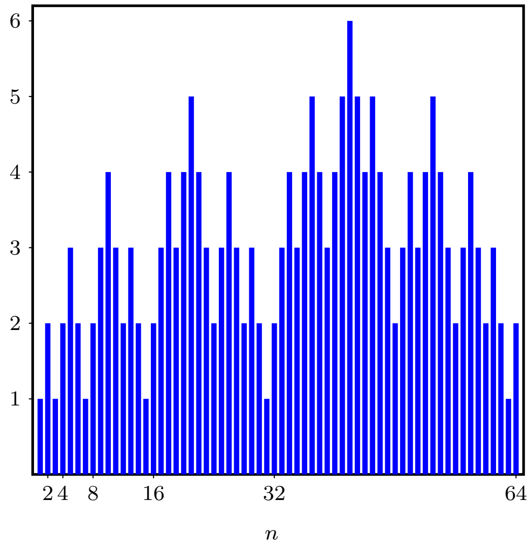


FIG 13. $\gamma(n)$.

Thus the number of 1s, denoted by $\gamma(n)$, of n under such a coding system satisfies the recurrence

$$\gamma(2^k + j) = 1 + \gamma(2^k - 1 - j), \tag{52}$$

for $0 \leqslant j < 2^k$ and $k \geqslant 1$. Another interesting type of recurrence is (by induction)

$$\gamma(n) = \gamma\left(\lfloor n/2 \rfloor\right) + \frac{1 - (-1)^{\lceil n/2 \rceil}}{2},$$

for $k \geqslant 1$, in contrast to

$$\nu(n) = \nu\left(\lfloor n/2 \rfloor\right) + \frac{1 - (-1)^{n}}{2},$$

for binary coding.

Let now

$$R_n(z) = \sum_{0 \leqslant j < n} z^{\gamma(j)}.$$

Then, obviously,

$$R_{2^k}(z) = (1+z)^k,$$

and, by (52),

$$\begin{aligned} R_{2^k+j}(z) &= R_{2^k}(z) + z(R_{2^k}(z) - R_{2^k-j}(z)) \\ &= R_{2^k}(z)(1+z) - zR_{2^k-j}(z) \\ &= (1+z)^{k+1} - zR_{2^k-j}(z). \end{aligned}$$

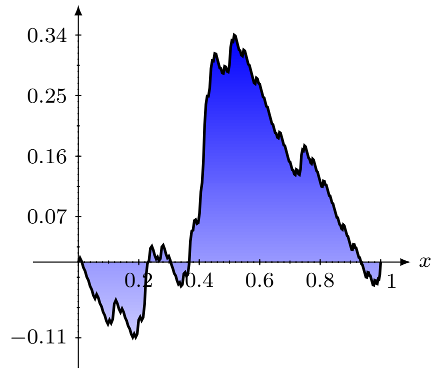


FIG 14. $G_1(x)$.

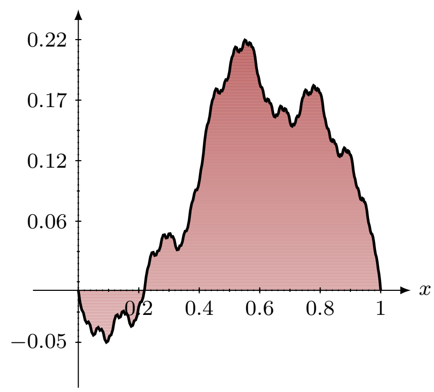


FIG 15. $G_2(x)$.

From this recurrence relation, we deduce by induction that

$$R_{2n}(z) = (1 + z)R_n(z),$$

for all $n \geqslant 1$. Thus the the sum-of-digits function Z_n of random integers under the Gray coding satisfies (46), and thus Theorems 47, 4.2 and 4.4 and Corollary 4.3 all hold. In addition to the mean and the variance, all results are new. The mean of Z_n was first studied by Flajolet and Ramshaw [48] where more precise characterizations of G_1 (including a Fourier series expansion) are given. A closed-form expression for $\mathbb{E}(y^{Z_n})$ was derived by Kobayashi et al. [79] by singular measures, together with exact expressions for all moments (non-centered).

For other properties related to $\gamma(n)$ and Z_n, see [36,66,76,77,79,80,113,115].

So far, we considered only the goodness of approximations to $\mathscr{L}(X_n)$ and $\mathscr{L}(Z_n)$ by the binomial distribution Y_λ. It is also natural to consider approxi-

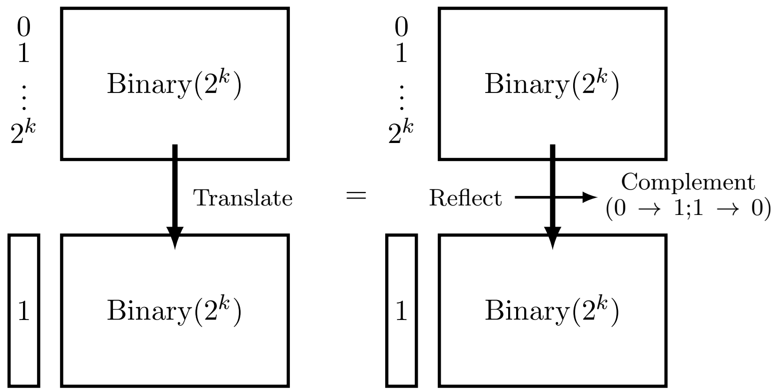


FIG 16. *Two different ways of constructing the same binary code.*

mations of $\mathscr{L}(Z_n)$ by $\mathscr{L}(X_n)$, and the result is as follows.

$$d_{\mathrm{TV}}\left(\mathscr{L}(X_n), \mathscr{L}(Z_n)\right) = \frac{\sqrt{2}|F(\log_2 n) - G_1(\log_2 n)|}{\sqrt{\pi \log_2 n}} + O\left(\frac{1}{\log_2 n}\right), \qquad (53)$$

where the difference $F(x) - G_1(x)$ is a *continuous function for all x*.

4.3. Beyond binary and Gray codings

We give here another simple binary coding system for integers satisfying the condition (46). We start with the observation that binary coding can be constructed not only in the usual translation way, but also using reflection and complement (first reflect the whole block of 2^k numbers as in Gray code, and then change every 1 to 0 and every 0 to 1); see Figure 16.

We now consider a coding system using translation and complement. Let $\mu(n)$ denote the number of 1s in the coding of n. Then by construction

$$\mu(2^k + j) = k + 1 - \mu(j),$$

for $0 \leqslant j < 2^k$ and $k \geqslant 1$. From this recurrence, we see that

$$\sum_{0 \leqslant \ell < 2^k + j} y^{\mu(\ell)} = (1+y)^k + y^{k+1} \sum_{0 \leqslant \ell < j} y^{-\mu(\ell)},$$

and it is straightforward to see that (46) holds in such a coding system. Thus Z_n satisfies all properties stated in the beginning of this section.

To obtain other examples for which (46) holds, one may combine more block operations (such as translation, horizontal or vertical reflection, reversal, flip, etc.) and string operations (complement, reversal, cyclic rotation, rewriting, etc.). A simple example is the block translation or reflection followed by any cyclic rotation of each coding (which does not change the number of 1s). Such a coding scheme also satisfies (46).

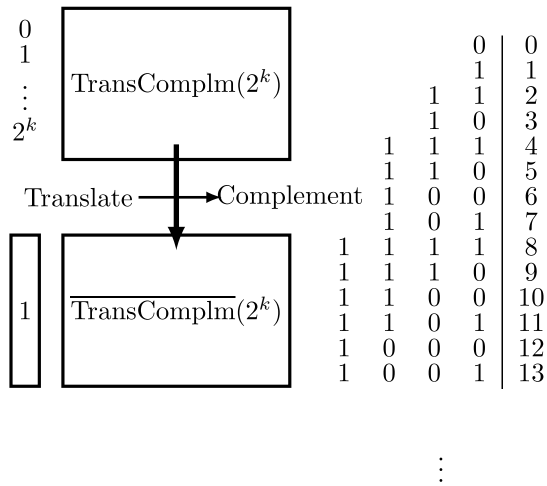


FIG 17. *Yet another code.*

Acknowledgement

We thank the referee for their helpful comments.

References

[1] AGNARSSON, G., On the number of hypercubic bipartitions of an integer. *Discrete Math.*, 313(24):2857–2864, 2013. MR3115296

[2] ALKAUSKAS, G., Dirichlet series associated with strongly q-multiplicative functions. *Ramanujan J.*, 8(1):13–21, 2004. MR2068427

[3] ALLOUCHE, J.-P. and SHALLIT, J., *Automatic Sequences.* Cambridge University Press, Cambridge, 2003. MR1997038

[4] BARAT, G., BERTHÉ, V., LIARDET, P., and THUSWALDNER, J., Dynamical directions in numeration. *Ann. Inst. Fourier (Grenoble)*, 56(7):1987–2092, 2006. MR2290774

[5] BARBOUR, A. D., Stein's method and poisson process convergence. *J. Appl. Probab.*, 25:175–184, 1988. MR0974580

[6] BARBOUR, A. D., Stein's method for diffusion approximations. *Probab. Th. Related Fields*, 84(3):297–322, 1990. MR1035659

[7] BARBOUR, A. D. and CHEN, L. H. Y., On the binary expansion of a random integer. *Statist. Probab. Lett.*, 14(3):235–241, 1992. MR1173624

[8] BARBOUR, A. D., HOLST, L., and JANSON, S., *Poisson Approximation.* The Clarendon Press, Oxford University Press, New York, 1992. MR1163825

[9] BASSILY, N. L. and KÁTAI, I., Distribution of the values of q-additive functions on polynomial sequences. *Acta Math. Hungar.*, 68(4):353–361, 1995. MR1333478

[10] Bassily, N. L. and Kátai, I., Distribution of consecutive digits in the q-ary expansions of some subsequences of integers. In *Proceedings of the XVI Seminar on Stability Problems for Stochastic Models, Part II (Eger, 1994)*, volume 78, pages 11–17, 1996. MR1381030

[11] Bellman, R. and Shapiro, H. N., On a problem in additive number theory. *Ann. of Math. (2)*, 49:333–340, 1948. MR0023864

[12] Berthé, V. and Rigo, M., editors, *Combinatorics, Automata and Number Theory*. Cambridge University Press, 2010. MR2742574

[13] Bowden, J., *Special Topics in Theoretical Arithmetic*. J. Bowden, Garden City, New York, 1936.

[14] Brown, T. C., Powers of digital sums. *Fibonacci Quart.*, 32(3):207–210, 1994. MR1285747

[15] Bush, L. E., An asymptotic formula for the average sum of the digits of integers. *Amer. Math. Monthly*, 47:154–156, 1940. MR0001225

[16] Chen, F.-J., A problem in the r-adic representation of positive integers (Chinese). *J. Nanjing University (Natural Sciences)*, 40(1):89–93, 2004. MR2370534

[17] Chen, L. H. Y., Poisson approximation for dependent trials. *Ann. Probability*, 3(3):534–545, 1975. MR0428387

[18] Chen, L. H. Y., Fang, X., and Shao, Q.-M., From Stein identities to moderate deviations. *Ann. Probab.*, 41(1):262–293, 2013. MR3059199

[19] Chen, L. H. Y. and Shao, Q.-M., Stein's method for normal approximation. In *An Introduction to Stein's Method*, volume 4 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–59. Singapore Univ. Press, Singapore, 2005. MR2235448

[20] Chen, L. H. Y. and Soon, S. Y. T., On the number of ones in the binary expansion of a random integer. Unpublished manuscript, 1994.

[21] Chen, W.-M., Hwang, H.-K., and Chen, G.-H., The cost distribution of queue-mergesort, optimal mergesorts, and power-of-2 rules. *J. Algorithms*, 30(2):423–448, 1999. MR1671856

[22] Cheo, P.-H. and Yien, S.-C., A problem on the k-adic representation of positive integers. *Acta Math. Sinica*, 5:433–438, 1955. MR0075979

[23] Clements, G. F. and Lindström, B., A sequence of $(\pm 1)-$determinants with large values. *Proc. Amer. Math. Soc.*, 16:548–550, 1965. MR0178001

[24] Cooper, C. and Kennedy, R. E., Digital sum sums. *J. Inst. Math. Comput. Sci. Math. Ser.*, 5(1):45–49, 1992. MR1182467

[25] Cooper, C. N. and Kennedy, R. E., A generalization of a theorem by Cheo and Yien concerning digital sums. *Internat. J. Math. Math. Sci.*, 9(4):817–820, 1986. MR0870542

[26] Coquet, J., Power sums of digital sums. *J. Number Theory*, 22(2):161–176, 1986. MR0826949

[27] Dartyge, C., Luca, F., and Stănică, P., On digit sums of multiples of an integer. *J. Number Theory*, 129(11):2820–2830, 2009. MR2549536

[28] Deheuvels, P. and Pfeifer, D., A semigroup approach to Poisson approximation. *Ann. Probab.*, 14(2):663–676, 1986. MR0832029

[29] Delange, H., Sur les fonctions q-additives ou q-multiplicatives. *Acta Arith.*, 21:285–298 (errata insert), 1972. MR0309891

[30] Delange, H., Sur la fonction sommatoire de la fonction "somme des chiffres". *Enseignement Math. (2)*, 21(1):31–47, 1975. MR0379414

[31] Diaconis, P., The distribution of leading digits and uniform distribution mod 1. *Ann. Probability*, 5(1):72–81, 1977. MR0422186

[32] Diaconis, P., *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988. MR0964069

[33] Diaconis, P., Graham, R. L., and Morrison, J. A., Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms*, 1(1):51–72, 1990. MR1068491

[34] Dickson, L. E., *History of the Theory of Numbers. Vol. I: Divisibility and Primality.* Chelsea Publishing Co. (unaltered reprintings of the 1919 original), New York, 1966.

[35] d'Ocagne, M., Sur certaines sommations arithmétiques. *Jornal de Sciencias Mathematicas e Astronomicas (de M. Gomes Teixeira. Coimbre)*, 7:117–128, 1886.

[36] Doran, R., The Gray code. *J. Universal Comput. Sci.*, 13(11):1573–1597, 2007. MR2390238

[37] Drazin, M. P. and Griffith, J. S., On the decimal representation of integers. *Proc. Cambridge Philos. Soc.*, 48:555–565, 1952. MR0049959

[38] Drmota, M., The joint distribution of q-additive functions. *Acta Arith.*, 100(1):17–39, 2001. MR1864623

[39] Drmota, M., Fuchs, M., and Manstavičius, E., Functional limit theorems for digital expansions. *Acta Math. Hungar.*, 98(3):175–201, 2003. MR1956755

[40] Drmota, M. and Gajdosik, J., The distribution of the sum-of-digits function. *J. Théor. Nombres Bordeaux*, 10(1):17–32, 1998. MR1827283

[41] Dumont, J.-M. and Thomas, A., Systèmes de numération et fonctions fractales relatifs aux substitutions. *Theoret. Comput. Sci.*, 65(2):153–169, 1989. MR1020484

[42] Dumont, J.-M. and Thomas, A., Digital sum moments and substitutions. *Acta Arith.*, 64(3):205–225, 1993. MR1225425

[43] Dumont, J. M. and Thomas, A., Gaussian asymptotic properties of the sum-of-digits function. *J. Number Theory*, 62(1):19–38, 1997. MR1430000

[44] Ettestad, D. J. and Carbonara, J. O., Formulas for the number of states of an interesting finite cellular automaton and a connection to Pascal's triangle. *J. Cell. Autom.*, 5(1–2):157–166, 2010. MR2583067

[45] Fang, Y., A theorem on the k-adic representation of positive integers. *Proc. Amer. Math. Soc.*, 130(6):1619–1622 (electronic), 2002. MR1887007

[46] Flajolet, P. and Golin, M., Mellin transforms and asymptotics. The mergesort recurrence. *Acta Inform.*, 31(7):673–696, 1994. MR1300060

[47] Flajolet, P., Grabner, P., Kirschenhofer, P., Prodinger, H., and Tichy, R. F., Mellin transforms and asymptotics: Digital sums. *Theoret. Comput. Sci.*, 123(2):291–314, 1994. MR1256203

[48] Flajolet, P. and Ramshaw, L., A note on Gray code and odd-even merge. *SIAM J. Comput.*, 9(1):142–158, 1980. MR0557835

[49] Foster, D. M. E., Estimates for a remainder term associated with the sum of digits function. *Glasgow Math. J.*, 29(1):109–129, 1987. MR0876156

[50] Foster, D. M. E., A lower bound for a remainder term associated with the sum of digits function. *Proc. Edinburgh Math. Soc. (2)*, 34(1):121–142, 1991. MR1093181

[51] Foster, D. M. E., Averaging the sum of digits function to an even base. *Proc. Edinburgh Math. Soc. (2)*, 35(3):449–455, 1992. MR1187007

[52] Gel'fond, A. O., Sur les nombres qui ont des propriétés additives et multiplicatives données. *Acta Arith.*, 13:259–265, 1967/1968. MR0220693

[53] Gilbert, E. N., Games of identification or convergence. *SIAM Review*, 4(1):16–24, 1962.

[54] Gittenberger, B. and Thuswaldner, J. M., Asymptotic normality of b-additive functions on polynomial sequences in the gaussian number field. *Journal of Number Theory*, 84(2):317–341, 2000. MR1796518

[55] Glaisher, J. W. L., On the residue of a binomial-theorem coefficient with respect to a prime modulus. *Quart. J. Pure and Appl. Math.*, 30:150–156, 1899.

[56] Glaser, A., *History of Binary and Other Nondecimal Numeration*. Tomash Publishers, Los Angeles, Calif., second edition, 1981. MR0666393

[57] Grabner, P. J., Completely q-multiplicative functions: the Mellin transform approach. *Acta Arith.*, 65(1):85–96, 1993. MR1239244

[58] Grabner, P. J. and Hwang, H.-K., Digital sums and divide-and-conquer recurrences: Fourier expansions and absolute convergence. *Constr. Approx.*, 21(2):149–179, 2005. MR2107936

[59] Grabner, P. J., Kirschenhofer, P., Prodinger, H., and Tichy, R. F., On the moments of the sum-of-digits function. In *Applications of Fibonacci Numbers, Vol. 5 (St. Andrews, 1992)*, pages 263–271. Kluwer Acad. Publ., Dordrecht, 1993. MR1271366

[60] Graham, R. L., On primitive graphs and optimal vertex assignments. *Ann. New York Acad. Sci.*, 175:170–186, 1970. MR0269533

[61] Greene, D. H. and Knuth, D. E., *Mathematics for the Analysis of Algorithms*. Modern Birkhäuser Classics. Birkhäuser Boston Inc., Boston, MA, 2008. MR2381155

[62] Hadjicostas, P. and Lakshmanan, K. B., Recursive merge sort with erroneous comparisons. *Discrete Appl. Math.*, 159(14):1398–1417, 2011. MR2823899

[63] Hart, S., A note on the edges of the n-cube. *Discrete Math.*, 14(2):157–163, 1976. MR0396293

[64] Hata, M. and Yamaguti, M., The Takagi function and its generalization. *Japan J. Appl. Math.*, 1(1):183–199, 1984. MR0839313

[65] Heppner, E., Über die Summe der Ziffern natürlicher Zahlen. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*, 19:41–43 (1977), 1976. MR0506024

[66] Hofer, R., Larcher, G., and Pillichshammer, F., Average growth-behavior and distribution properties of generalized weighted digit-block-counting functions. *Monatsh. Math.*, 154(3):199–230, 2008. MR2413302

[67] Holmes, S., Stein's method for birth and death chains. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 45–67. Inst. Math. Statist., Beachwood, OH, 2004. MR2118602

[68] Hong, Z. and Sedgewick, R., Notes on merging networks (preliminary version). In *Proc. ACM Symposium on Theory of Computing*, pages 296–302, 1982.

[69] Ifrah, G., *The Universal History of Numbers*. John Wiley & Sons Inc., New York, 2000. From prehistory to the invention of the computer, Translated from the 1994 French original by David Bellos, E. F. Harding, Sophie Wood and Ian Monk. MR1725387

[70] Ismail, M. E. H., *Classical and Quantum Orthogonal Polynomials in One Variable*, volume 98 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, Cambridge, 2009. MR2542683

[71] Kano, H., On the sums of digits in integers. *Proc. Japan Acad. Ser. A Math. Sci.*, 67(5):148–150, 1991. MR1114959

[72] Kátai, I., On the sum of digits of primes. *Acta Math. Acad. Sci. Hungar.*, 30(1–2):169–173, 1977. MR0472747

[73] Kátai, I. and Mogyoródi, J., On the distribution of digits. *Publ. Math. Debrecen*, 15:57–68, 1968. MR0236139

[74] Kennedy, R. E. and Cooper, C. N., An extension of a theorem by Cheo and Yien concerning digital sums. *Fibonacci Quart.*, 29(2):145–149, 1991. MR1119401

[75] Kirschenhofer, P., On the variance of the sum of digits function. In *Number-Theoretic Analysis (Vienna, 1988–89)*, volume 1452 of *Lecture Notes in Math.*, pages 112–116. Springer, Berlin, 1990. MR1084640

[76] Kirschenhofer, P. and Prodinger, H., Subblock occurrences in positional number systems and Gray code representation. *J. Inform. Optim. Sci.*, 5(1):29–42, 1984. MR0737164

[77] Klavžar, S., Milutinović, U., and Petr, C., Stern polynomials. *Adv. in Appl. Math.*, 39(1):86–95, 2007. MR2319565

[78] Knuth, D. E., *Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley, third edition, November 1997.

[79] Kobayashi, Z., Digital sum problems for the Gray code representation of natural numbers. *Interdiscip. Inform. Sci.*, 8(2):167–175, 2002. MR1972038

[80] Kobayashi, Z. and Sekiguchi, T., On a characterization of the standard Gray code by using it edge type on a hypercube. *Inform. Process. Lett.*, 81(5):231–237, 2002. MR1879645

[81] Krüppel, M., De Rham's singular function, its partial derivatives with respect to the parameter and binary digital sums. *Rostock. Math. Kolloq.*, 64:57–74, 2009. MR2605000

[82] Kummer, E. E., Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen. *J. Reine Angew. Math.*, 44:93–146, 1852.

[83] Laczay, B. and Ruszinkó, M., Collision channel with multiplicity feedback. In E. Biglieri and L. Györfi, editors, *Proceedings of the NATO Advanced Study Institute on Coding and Analysis of Multiple Access Channels. Theory and Practice*, volume D. 10, pages 250–270. IOS Press, 2007.

[84] Lagarias, J. C., The Takagi function and its properties. In *Functions in Number Theory and Their Probabilistic Aspects*, RIMS Kôkyûroku Bessatsu, B34, pages 153–189. Res. Inst. Math. Sci. (RIMS), Kyoto, 2012. MR3014845

[85] Legendre, A., *Théorie des Nombres.* Firmin Didot Frères, fourth edition, 1900.

[86] Li, S.-Y. R., Binary trees and uniform distribution of traffic cutback. *J. Comput. System Sci.*, 32(1):1–14, 1986. MR0844201

[87] Lindström, B., On a combinatorial problem in number theory. *Canad. Math. Bull.*, 8:477–490, 1965. MR0181604

[88] Loh, W.-L., Stein's method and multinomial approximation. *Ann. Appl. Probab.*, 2(3):536–554, 1992. MR1177898

[89] Lucas, É., Sur les congruences des nombres eulériens et des coefficients différentiels des fonctions trigonométriques, suivant un module premier. *Bull. Soc. Math. France*, 6:49–54, 1878. MR1503769

[90] MacWilliams, F. J. and Sloane, N. J. A., *The Theory of Error-Correcting Codes.* North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

[91] Madritsch, M. and Pethő, A., Asymptotic normality of additive functions on polynomial sequences in canonical number systems. *J. Number Theory*, 131(9):1553–1574, 2011. MR2802135

[92] Madritsch, M. G., Asymptotic normality of b-additive functions on polynomial sequences in number systems. *Ramanujan J.*, 21(2):181–210, 2010. MR2593247

[93] Manstavičius, E., Probabilistic theory of additive functions related to systems of numeration. In *New Trends in Probability and Statistics, Vol. 4 (Palanga, 1996)*, pages 413–429. VSP, Utrecht, 1997. MR1653626

[94] Mauclaire, J.-L., Sur la répartition des fonctions q-additives. *J. Théor. Nombres Bordeaux*, 5(1):79–91, 1993. MR1251228

[95] Mauclaire, J.-L. and Murata, L., On q-additive functions. I. *Proc. Japan Acad. Ser. A Math. Sci.*, 59(6):274–276, 1983. MR0718620

[96] Mauclaire, J.-L. and Murata, L., On q-additive functions. II. *Proc. Japan Acad. Ser. A Math. Sci.*, 59(9):441–444, 1983. MR0732606

[97] Mauduit, C. and Rivat, J., Propriétés q-multiplicatives de la suite $\lfloor n^c \rfloor$, $c > 1$. *Acta Arith.*, 118(2):187–203, 2005. MR2141049

[98] Mauduit, C. and Rivat, J., La somme des chiffres des carrés. *Acta Math.*, 203(1):107–148, 2009. MR2545827

[99] Mauduit, C. and Rivat, J., Sur un problème de Gelfond: la somme des chiffres des nombres premiers. *Ann. of Math.*, 171(3):1591–1646, 2010. MR2680394

[100] McIlroy, M. D., The number of 1's in binary integers: Bounds and extremal properties. *SIAM J. Comput.*, 3:255–261, 1974. MR0436687

[101] Mehrabian, A., Mitsche, D., and Prałat, P., On the maximum density of graphs with unique-path labelings. *SIAM J. Discrete Math.*, 27(3):1228–1233, 2013. MR3072758

[102] Mirsky, L., A theorem on representations of integers in the scale of r. *Scripta Math.*, 15:11–12, 1949. MR0030991

[103] Morrison, J. A., Weighted averages of Radon transforms on Z_2^k. *SIAM J. Algebraic Discrete Methods*, 7(3):404–413, 1986. MR0844043

[104] Muramoto, K., Okada, T., Sekiguchi, T., and Shiota, Y., Digital sum problems for the p-adic expansion of natural numbers. *Interdiscip. Inform. Sci.*, 6(2):105–109, 2000. MR1839805

[105] Muramoto, K., Okada, T., Sekiguchi, T., and Shiota, Y., Power and exponential sums of digital sums with information per digits. *Math. J. Toyama Univ.*, 26:35–44, 2003. MR2048391

[106] Murata, L. and Mauclaire, J.-L., An explicit formula for the average of some q-additive functions. In *Prospects of Mathematical Science (Tokyo, 1986)*, pages 141–156. World Sci. Publishing, Singapore, 1988. MR0948466

[107] Okada, T., Sekiguchi, T., and Shiota, Y., Applications of binomial measures to power sums of digital sums. *J. Number Theory*, 52(2):256–266, 1995. MR1336748

[108] Okada, T., Sekiguchi, T., and Shiota, Y., An explicit formula of the exponential sums of digital sums. *Japan J. Indust. Appl. Math.*, 12(3):425–438, 1995. MR1356664

[109] Okada, T., Sekiguchi, T., and Shiota, Y., A generalization of Hata-Yamaguti's results on the Takagi function. II. Multinomial case. *Japan J. Indust. Appl. Math.*, 13(3):435–463, 1996. MR1415064

[110] Osbaldestin, A. H., Digital sum problems. In *Fractals in the Fundamental and Applied Sciences*, pages 307–328. Elsevier Science, B. V., North-Holland, Amsterdam, 1991.

[111] Panny, W. and Prodinger, H., Bottom-up mergesort—A detailed analysis. *Algorithmica*, 14(4):340–354, 1995. MR1343320

[112] Prodinger, H., Generalizing the sum of digits function. *SIAM J. Algebraic Discrete Methods*, 3(1):35–42, 1982. MR0644955

[113] Prodinger, H., Nonrepetitive sequences and Gray code. *Discrete Math.*, 43(1):113–116, 1983. MR0680311

[114] Prodinger, H., A subword version of d'Ocagne's formula. *Utilitas Math.*, 24:125–129, 1983. MR0724766

[115] Prodinger, H., Digits and beyond. In *Mathematics and Computer Science, II (Versailles, 2002)*, Trends Math., pages 355–377. Birkhäuser, Basel, 2002. MR1940147

[116] Roberts, J. B., On binomial coefficient residues. *Canad. J. Math.*, 9:363–370, 1957. MR0086828

[117] Roos, B., Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion. *Theory Probab. Appl.*, 45(2):258–272, 2001. MR1967760

[118] Sándor, J. and Crstici, B., *Handbook of Number Theory. II.* Kluwer Academic Publishers, Dordrecht, 2004. MR2119686

[119] Savage, C., A survey of combinatorial Gray codes. *SIAM Rev.*, 39(4):605–629, 1997. MR1491049

[120] Schmid, J., The joint distribution of the binary digits of integer multiples. *Acta Arith.*, 43(4):391–415, 1984. MR0756290

[121] Schmidt, W. M., The joint distribution of the digits of certain integer s-tuples. In *Studies in Pure Mathematics*, pages 605–622. Birkhäuser, Basel, 1983. MR0820255

[122] Schoutens, W., *Stochastic Processes and Orthogonal Polynomials*, volume 146 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. MR1761401

[123] Shiokawa, I., On a problem in additive number theory. *Math. J. Okayama Univ.*, 16:167–176, 1973/1974. MR0357352

[124] Shiokawa, I., g-adical analogues of some arithmetical functions. *Math. J. Okayama Univ.*, 17:75–94, 1974. MR0364069

[125] Soon, Y.-T., *Some Problems in Binomial and Compound Poisson Approximations*. Ph.D. Thesis, National University of Singapore, 1993.

[126] Stein, A. H., Exponential sums related to binomial coefficient parity. *Proc. Amer. Math. Soc.*, 80(3):526–530, 1980. MR0581019

[127] Stein, A. H., Exponential sums of sum-of-digit functions. *Illinois J. Math.*, 30(4):660–675, 1986. MR0857218

[128] Stein, C., A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory*, pages 583–602. Univ. California Press, Berkeley, Calif., 1972. MR0402873

[129] Stein, C., *Approximate Computation of Expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986. MR0882007

[130] Steiner, W., *The Distribution of Digital Expansions on Polynomial Sequences*. Dissertation, TU-Wien, 2002.

[131] Stolarsky, K. B., Power and exponential sums of digital sums related to binomial coefficient parity. *SIAM J. Appl. Math.*, 32(4):717–730, 1977. MR0439735

[132] Stolarsky, K. B., Integers whose multiples have anomalous digital frequencies. *Acta Arith.*, 38(2):117–128, 1980/81. MR0604228

[133] Szegő, G., *Orthogonal Polynomials*. AMS, Providence, R.I., fourth edition, 1975. MR0372517

[134] Tang, S. C., An improvement and generalization of Bellman-Shapiro's theorem on a problem in additive number theory. *Proc. Amer. Math. Soc.*, 14:199–204, 1963. MR0150082

[135] Tenenbaum, G., Sur la non-dérivabilité de fonctions périodiques associées à certaines formules sommatoires. In *The Mathematics of Paul Erdős, I*, volume 13 of *Algorithms Combin.*, pages 117–128. Springer, Berlin, 1997. MR1425180

[136] Terras, A., *Fourier Analysis on Finite Groups and Applications*, volume 43 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999. MR1695775

[137] Thim, J., *Continuous Nowhere Differentiable Functions*. Master Thesis, Luleå Tekniska Universitet, 2003.

[138] Trollope, J. R., Generalized bases and digital sums. *Amer. Math. Monthly*, 74:690–694, 1967. MR0211950

[139] Trollope, J. R., An explicit expression for binary digital sums. *Math. Mag.*, 41:21–25, 1968. MR0233763

[140] Wolfram, S., Statistical mechanics of cellular automata. *Rev. Modern Phys.*, 55(3):601–644, 1983. MR0709077

[141] Wolfram, S., Geometry of binomial coefficients. *Amer. Math. Monthly*, 91(9):566–571, 1984. MR0764797

[142] Yu, X. Y., On the mean-value of the powers of digital sums. *Kexue Tongbao (Chinese)*, 41(7):581–585, 1996. MR1418096

[143] Zacharovas, V. and Hwang, H.-K., A Charlier-Parseval approach to Poisson approximation and its applications. *Lithuanian Math. J.*, 50(1):88–119, 2010. MR2607681