ISSN 1066-5307, Mathematical Methods of Statistics, 2018, Vol. 27, No. 1, pp. 60–70. © Allerton Press, Inc., 2018.

Entropic Moments and Domains of Attraction on Countable Alphabets

S. Molchanov[1*], Z. Zhang[1], and L. Zheng[2***]**

[1]*Univ. of North Carolina at Charlotte, Charlotte, USA*

[2]*Tennessee Technological University, Cookeville, USA*

Received August 30, 2017; in final form, November 1, 2017

Abstract—Modern information theory is largely developed in connection with random elements residing in large, complex, and discrete data spaces, or alphabets. Lacking natural metrization and hence moments, the associated probability and statistics theory must rely on information measures in the form of various entropies, for example, Shannon's entropy, mutual information and Kullback–Leibler divergence, which are functions of an entropic basis in the form of a sequence of entropic moments of varying order. The entropic moments collectively characterize the underlying probability distribution on the alphabet, and hence provide an opportunity to develop statistical procedures for their estimation. As such statistical development becomes an increasingly important line of research in modern data science, the relationship between the underlying distribution and the asymptotic behavior of the entropic moments, as the order increases, becomes a technical issue of fundamental importance. This paper offers a general methodology to capture the relationship between the rates of divergence of the entropic moments and the types of underlying distributions, for a special class of distributions. As an application of the established results, it is demonstrated that the asymptotic normality of the remarkable Turing's formula for missing probabilities holds under distributions with much thinner tails than those previously known.

Keywords: entropic moments, domains of attraction, countable alphabets, Turing's formula.

2000 Mathematics Subject Classification: 60B10, 62G32.

DOI: 10.3103/S1066530718010040

Dedicated to Dmitry M. Chibisov

1. INTRODUCTION

The asymptotic methods in statistics, as well as in general mathematics, require some regularity conditions on the objects of the study. However the powerful utility of asymptotic methods can also bring fundamental understanding to statistics theories even under highly discrete models. A good illustration of such a case is the "missing probability" estimator, known as Turing's formula, attributed mainly to Alan Turing. A partial goal of this article is to demonstrate that the central limit theorem for Turing's formula holds under tail conditions that were previously unknown.

Toward that end, consider a countably infinite alphabet $\mathscr{X} = \{\ell_k; k \geq 1\}$ along with an associated random element X and its probability distribution $\mathbf{p} = \{p_k; k \geq 1\}$. Let, for every v, $v = 1, 2, \cdots$,

$$\zeta_v = \sum_{k\geq 1} p_k (1-p_k)^v \tag{1}$$

be referred to as the v^{th} entropic moment. The sequence $\boldsymbol{\zeta} = \{\zeta_v; \geq 1\}$, known as the entropic basis, plays a fundamental role in the theory of probability and statistics on alphabets.

*E-mail: smolchan@uncc.edu
**E-mail: zzhang@uncc.edu
***E-mail: lzheng@tntech.edu

Zhang and Grabchak [10] established a characterization theorem with regard to $\boldsymbol{\zeta}$, which states that, assuming $\mathbf{p}$ is non-increasing, $\mathbf{p}$ and $\boldsymbol{\zeta}$ uniquely determine each other. This characterization theorem, bypassing the fact that there exit no natural moments for the random element X on alphabet $\mathscr{X}$, allows statistical inference on the underlying distribution $\mathbf{p}$ by means of estimating the entropic moments. The entropic moments as defined in (1) offer a fundamental shift in the probabilistic and statistical paradigm on alphabets, from the one based on likelihood (the p_k's) to the one based on entropic moments (the ζ_v's). The shifted paradigm brings about many advantages in statistical inference on alphabets, particularly with regard to modern information theory where data space is of very high dimension often without a universally suitable metric. Let $\{X_i; i = 1, \cdots, n\}$ be an identically and independently distributed (iid) sample of size n drawn from $\mathscr{X}$ under $\mathbf{p}$. Let the sample data be summarized into letter frequencies $\{Y_k; k \geq 1\}$ and relative letter frequencies $\{\hat{p}_k = Y_k/n; k \geq 1\}$. Zhang and Zhou [9] established unbiased estimators Z_v of ζ_v for all v from $v = 1$ up to $v = n - 1$. The availability of these unbiased estimators provide an opportunity to further explore the practical and the theoretical potential of the new paradigm.

As in all probability and statistical development, an issue of essential importance is to understand and describe the tail tightness of underlying probability distributions. Within the new paradigm, it is of primary interest to understand the relationship between the tail decay rate of the underlying distribution $\mathbf{p}$ as $k \to \infty$ and the decay rate of ζ_v as $v \to \infty$.

To serve that interest, Zhang [8] made an attempt to define domains of attraction for probability distributions on the alphabet using

$$\tau_v = v\zeta_v \tag{2}$$

as a tail index. To see ζ_v, and hence equivalently τ_v, is tail-relevant, one needs only to consider a two-stage sampling scheme, in which an iid sample of size n, say $\{X_1, \cdots, X_n\}$, is taken first and then an additional observed X_{n+1} is taken. The probability that X_{n+1} is a letter not observed in the iid sample of size n is

$$\begin{aligned} \mathrm{P}\left(\cap_{i=1}^{n}\{X_{n+1} \neq X_i\}\right) &= \sum_{k\geq 1} \mathrm{P}\left(X_i \neq \ell_k \mid X_{n+1} = \ell_k\right) \mathrm{P}(X_{n+1} = \ell_k) \\ &= \sum_{k\geq 1} p_k(1-p_k)^n = \zeta_n. \end{aligned}$$

In this sampling scheme, one may think of ζ_n as the probability of a new discovery.

Another way to view ζ_v as a tail-relevant index is to associate it with the remarkable Turing's formula. Letting

$$\pi_0 = \sum_{k\geq 1} p_k 1[Y_k = 0], \tag{3}$$

the total probability associated with the letters of the alphabet not observed in an iid sample of size n, and

$$N_1 = \sum_{k\geq 1} 1[Y_k = 1],$$

the number of letters of $\mathscr{X}$ observed exactly once in the sample, Turing's formula, introduced by Good (1953) but largely credited to Alan Turing, is defined to be

$$T_n = \frac{N_1}{n}. \tag{4}$$

It is well known that Turing's formula is a good estimator of π_0. Statistical properties of Turing's formula have been studied for quite a long time, more notable ones include [4, 2, 1, 5]. For a comprehensive introduction to the topic, interested readers may refer to [6]. π_0 is also known as the "missing probability" and as such is a characteristic of the tail of the underlying distribution. Yet,

$$\mathrm{E}(\pi_0) = \sum_{k\geq 1} p_k(1-p_k)^n = \zeta_n.$$

Surrogating the tail properties of $\{p_k; k \geq 1\}$ in $\{\zeta_v; v \geq 1\}$, Zhang [8] established that a distribution has finitely many positive $p_k > 0$, that is, the alphabet is finite, if and only if τ_n converges to zero exponentially fast in n, that for distributions with exponentially decaying tails, e.g., $p_k \propto e^{-\lambda k}$ or thinner tails, τ_n perpetually oscillates between two distinct positive constants without a limit, and that for distributions with thicker tails, $\limsup \tau_n \to \infty$. These results allow a definition of domains of attraction on the alphabet. The domain with $\limsup \tau_n \to \infty$ is named the Turing-Good family which includes many thicker tailed distributions. As it turned out, the Turing-Good family can be fruitfully further partitioned into sub-domains with various divergence rates of $\tau_v = v\zeta_v$ in v for different classes of $p_k$characterized by the decay rates in k.

To put the above mentioned problem in a broader perspective, the *iid* sample $\{X_i; i = 1, \cdots, n\}$ may be considered as a result of a generalized Maxwell—Boltzmann scheme of allocating n particles one by one to a sequence of boxes labelled ℓ_k, $k = 1, 2, \cdots$, according to $\mathbf{p}$. In this case, Y_k is the number of particles landed in the k^{th} box. Y_k, $k \geq 1$, are also known as the occupation numbers and satisfy $\sum_{k=1}^{\infty} Y_k = n$. Consider the variable

$$N_r = \sum_{k=1}^{\infty} 1[Y_k(n) = r] \tag{5}$$

for $r \geq 1$. N_r represents the number of the boxes with occupation number being exactly r. Many statistical and probability problems are related to the vector of the occupation numbers,

$$(N_1, \cdots, N_m)' \tag{6}$$

for a fixed integer $m > 0$, specifically noting that for any fixed r, $r = 1, \cdots, m$,

$$\mathrm{E}\, N_r = \binom{n}{r} \sum_{k \geq 1} p_k^r (1 - p_k)^{n-r} \asymp (n - r)^r \sum_{k \geq 1} p_k^r (1 - p_k)^{n-r} =: \tau_{r,n-r},$$

where "$\asymp$" indicates equality in rate of convergence or divergence, i.e., if $a_n \asymp b_n$ then there exist constants c_1 and c_2 satisfying $c_1 c_2 > 0$ such that $c_1 \leq a_n/b_n \leq c_2$ to be distinguished from "$\sim$", which indicates equality in relative limit, i.e., if $a_n \sim b_n$ then $a_n/b_n \to 1$.

The main objective of this paper is to introduce a general analytical methodology to capture and describe the divergence rates of $\tau_{r,v}$ in general and $\tau_v = \tau_{1,v}$ in specific for well-behaved underlying distributions in the Turing-Good family. The main results are presented in Section 2. The paper is concluded with an application to the condition of asymptotic normality for Turing's formula, which is only known to be supported by thick tailed distributions with, e.g., power tails, such as $p_k \propto k^{-\lambda}$ where $\lambda > 1$, in the existing literature. In the application given in Section 3, it is demonstrated that the said normality holds for distributions with much thinner tails than previously known, e.g., those with near-exponential tails such as $p_k \propto \exp(-k/(\log k)^{\beta})$ where $\beta > 2$. It is also demonstrated that the said normality does not hold for distributions with exponentially decaying tails such as, e.g., $p_k \propto e^{-k}$.

2. ASYMPTOTIC ANALYSIS OF ENTROPIC MOMENTS

2.1. Conditions

The following conditions are assumed in order to establish the asymptotic properties of the entropic moments.

(A1) Assume that the sequence $\{p_k\}$ is decreasing no faster than exponential decay, i.e., there exists an appropriate constant α in $(0, 1)$ such that

$$\alpha \leq \frac{p_{k+1}}{p_k} \leq 1. \tag{7}$$

(A2) Assume that there exists a smooth $C^2(R_+)$ interpolation $p(x)$, $x \geq 0$, for the probabilities p_k, $k = 1, 2, \ldots$, such that $p_k = p(k)$ for all k, $p(0) < \infty$, $p'(x) < 0$ for $x \geq x_0$ where x_0 is a sufficiently large number, and $p'(x)$ is monotone increasing.

(A3) Assume the underlying interpolation $p(x)$ satisfies

$$(\log p(x))' = \frac{p'(x)}{p(x)} \uparrow 0, \quad x \to \infty, \quad \text{and} \tag{8}$$

$$\lim_{x\to\infty} \frac{p^2(x)}{p'(x)} = 0. \tag{9}$$

(A4) Assume

$$\left(\frac{p'(x)}{p(x)}\right)' = \frac{p''(x)p(x) - (p'(x))^2}{p^2(x)} \geq 0.$$

(A5) Assume that there exists a constant γ in the interval $[0, 1]$, such that

$$\limsup_{x\to\infty} \frac{p''(x)p(x) - [p'(x)]^2}{(p'(x))^2} = 1 - \gamma.$$

It will be shown that "typical" distributions $\mathbf{p}$ satisfy these conditions in Section 3.

2.2. Main Results

Let

$$S_{r,n} = n^r \sum_{k\geq 1} p_k^r e^{-np_k} \qquad \text{and} \qquad \tau_{r,n} = n^r \sum_{k\geq 1} p_k^r (1-p_k)^n. \tag{10}$$

Lemma 1. *For any $r \geq 1$, there exists a positive constant $C_r(\alpha)$ (with α involved in (7)) such that*

$$S_{r,n} \geq C_r(\alpha). \tag{11}$$

Proof. The function $f_r(x) = x^r e^{-x}$, $x \geq 0$, takes the maximum value $r^r e^{-r}$ at the point $x = r$. Let $x_r = x_r(n)$ be the point at which $np(x_r) = r$. Let $k_r(n)$ be the greatest integer less or equal to x_r. It follows from condition (A1) that

$$p_{k_r(n)} \geq \frac{r}{n} > p_{k_r(n)+1} \geq \alpha p_{k_r(n)} \qquad \text{and} \qquad r \leq np_{k_r(n)} \leq \frac{r}{\alpha},$$

and hence

$$S_{r,n} = \sum_k (np_k)^r e^{-np_k} \geq \left(\frac{r}{\alpha}\right)^r e^{-\frac{r}{\alpha}}.$$

□

Proposition 1. *For every $r > 0$, as $n \to \infty$, $\tau_{r,n} \sim S_{r,n}$.*

A proof of Proposition 1 is given in the Appendix.

Proposition 2. *If $S_{r,n} \to \infty$, then, for any fixed $r \geq 1$,*

$$S_{r,n} \sim \int_0^\infty (np(x))^r e^{-np(x)}\, dx + \mathcal{O}(1) = I_{r,n} + \mathcal{O}(1), \tag{12}$$

where

$$I_{r,n} = \int_0^\infty (np(x))^r e^{-np(x)}\, dx. \tag{13}$$

This is the simplest case of Euler–MacLaurin formula, based on the monotonicity of $f_r(x) = x^r e^{-x}$ for $0 \leq x \leq r$ and $x > r$.

Proposition 3. *Under conditions* (A1)–(A5),

$$I_{r,n} \asymp \frac{p(x_r)}{|p'(x_r)|}, \tag{14}$$

where $x_r = x_r(n)$ *is the root of* $np(x_r) = r$.

A proof of Proposition 3 is given in the Appendix.

Propositions 1, 2, and 3 lead to the following theorem.

Theorem 1. *Under conditions* (A1)–(A5),

$$\tau_{r,n} \sim S_{r,n} \asymp \frac{p(x_r)}{|p'(x_r)|}, \tag{15}$$

where x_r *is the value of* x *such that* $np(x) = r$.

3. CLASSES OF FUNCTIONS

In this section, it is demonstrated that conditions (A1)–(A5) are satisfied by several classes of well-behaved distributions $\mathbf{p}$.

3.1. Power-Decaying Functions

Consider a power-decaying function

$$p(x) = \frac{A}{(x+T)^\beta}, \quad \beta > 1,$$

where A and T are appropriate positive constants. Then

$$p'(x) = -A\beta(x+T)^{-\beta-1}, \qquad p''(x) = A^2\beta(\beta+1)(x+T)^{-\beta-2},$$
$$p''(x)p(x) - [p'(x)]^2 = A^2\beta(x+T)^{-2\beta-2}.$$

Therefore

$$(\log p(x))' = \frac{p'(x)}{p(x)} = -\beta(x+T)^{-1} \uparrow 0,$$
$$\lim_{x\to\infty} \frac{p^2(x)}{p'(x)} = \lim_{x\to\infty} -\frac{A}{\beta}(x+T)^{1-\beta} = 0,$$
$$\left(\frac{p'(x)}{p(x)}\right)' = \frac{p''(x)p(x) - (p'(x))^2}{p^2(x)} \geq 0,$$
$$\limsup_{x\to\infty} \frac{p''(x)p(x) - [p'(x)]^2}{(p'(x))^2} = \frac{1}{\beta}.$$

Therefore all the conditions (A1)–(A5) are satisfied.

3.2. Super Power-Decaying Functions

Let

$$p(x) = A\exp\big[-\log^{\beta}(x+T)\big], \quad \beta > 1,$$

where A and T are appropriate positive constants. Then $p(x)$ may be re-expressed as

$$p(x) = A(x+T)^{-[\log(x+T)]^{\beta-1}},$$

and therefore may be viewed as a power-decaying function with an increasing power, albeit very slowly. Then

$$p'(x) = -\frac{A\beta\big[\log(x+T)\big]^{\beta-1}}{x+T}\exp\big[-\log^{\beta}(x+T)\big],$$

$$p''(x) = A\left[\frac{\beta^2\big[\log(x+T)\big]^{2\beta-2}}{(x+T)^2} + \frac{\beta\big[\log(x+T)\big]^{\beta-2}\big[\log(x+T)-(\beta-1)\big]}{(x+T)^2}\right]\exp\big[-\log^{\beta}(x+T)\big],$$

and

$$p''(x)p(x) - [p'(x)]^2 = \frac{A^2\beta\big\{[\log(x+T)]^{\beta-2}[\log(x+T)-(\beta-1)]\big\}}{(x+T)^2}\exp\big[-2\log^{\beta}(x+T)\big].$$

Therefore

$$(\log p(x))' = \frac{p'(x)}{p(x)} = -\frac{\beta\big[\log(x+T)\big]^{\beta-1}}{x+T} \uparrow 0,$$

$$\lim_{x\to\infty}\frac{p^2(x)}{p'(x)} = \lim_{x\to\infty} -\frac{(x+T)\exp\big[-\log^{\beta}(x+T)\big]}{\beta\big[\log(x+T)\big]^{\beta-1}} = 0,$$

$$\left(\frac{p'(x)}{p(x)}\right)' = \frac{p''(x)p(x)-(p'(x))^2}{p^2(x)} \geq 0,$$

$$\limsup_{x\to\infty}\frac{p''(x)p(x)-(p'(x))^2}{(p'(x))^2} = \limsup_{x\to\infty}\frac{\log(x+T)-(\beta-1)}{\beta\log^{\beta}(x+T)} = 0,$$

that is, all the conditions (A1)–(A5) are satisfied.

3.3. Sub-Exponential Functions

Let

$$p(x) = Ae^{-x^{\rho}},\ 0<\rho<1,$$

where A is an appropriate positive constant. Then

$$p'(x) = -A\rho x^{\rho-1}e^{-x^{\rho}},$$

$$p''(x) = A\rho^2 x^{2(\rho-1)}e^{-x^{\rho}} - A\rho(\rho-1)x^{\rho-2}e^{-x^{\rho}},$$

$$p''(x)p(x) - [p'(x)]^2 = -A^2\rho(\rho-1)x^{\rho-2}e^{-2x^{\rho}}.$$

Therefore

$$(\log p(x))' = \frac{p'(x)}{p(x)} = -\rho x^{\rho-1} \uparrow 0,$$

$$\lim_{x\to\infty}\frac{p^2(x)}{p'(x)} = \lim_{x\to\infty} -\frac{A}{\rho}x^{1-\rho}e^{-x^{\rho}} = 0,$$

$$\left(\frac{p'(x)}{p(x)}\right)' = \frac{p''(x)p(x)-(p'(x))^2}{p^2(x)} \geq 0,$$

$$\limsup_{x\to\infty}\frac{p''(x)p(x)-(p'(x))^2}{(p'(x))^2} = \limsup_{x\to\infty}\frac{\rho-1}{\rho x^{\rho}} = 0,$$

that is, all the conditions (A1)–(A5) are satisfied.

3.4. Near-Exponential Functions

Let $p(x) = Ae^{-\frac{x+T}{\log^{\beta}(x+T)}},\ \beta > 0$, where A and T are appropriate positive constants. Then

$$p'(x) = -Ae^{-\frac{x+T}{[\log(x+T)]^{\beta}}}\left[\frac{1}{[\log(x+T)]^{\beta}} - \frac{\beta}{[\log(x+T)]^{\beta+1}}\right] \quad \text{and}$$

$$p''(x) = Ae^{-\frac{x+T}{[\log(x+T)]^{\beta}}}\left[\frac{1}{[\log(x+T)^{\beta}} - \frac{\beta}{[\log(x+T)]^{\beta+1}}\right]^2 + Ae^{-\frac{x+T}{[\log(x+T)]^{\beta}}}\frac{\beta}{x+T}\left[\frac{1}{[\log(x+T)]^{\beta+1}} - \frac{\beta+1}{[\log(x+T)^{\beta+2}}\right].$$

Therefore

$$(\log p(x))' = \frac{p'(x)}{p(x)} = \Big[\frac{\beta}{[\log(x+T)]^{\beta+1}} - \frac{1}{[\log(x+T)]^{\beta}}\Big] \uparrow 0,$$

$$\lim_{x\to\infty}\frac{p^2(x)}{p'(x)} = \lim_{x\to\infty}\frac{-A[\log(x+T)]^{\beta}}{e^{\frac{x+T}{[\log(x+T)]^{\beta}}}\left(\frac{\beta}{\log(x+T)} - 1\right)} = 0,$$

$$\Big(\frac{p'(x)}{p(x)}\Big)' = \frac{\beta}{x+T}\left[\frac{1}{[\log(x+T)]^{\beta+1}} - \frac{\beta+1}{[\log(x+T)^{\beta+2}]}\right] \geq 0,$$

$$\limsup_{x\to\infty}\frac{p''(x)p(x) - (p'(x))^2}{(p'(x))^2} = \limsup_{x\to\infty}\left(\frac{\beta[\log(x+T)]^{\beta-1}}{x+T}\frac{(1-\frac{\beta+1}{\log(x+T)})}{(1-\frac{\beta}{\log(x+T)})^2}\right) = 0,$$

that is, all the conditions (A1)–(A5) are satisfied.

3.5. Distributions on the Margin

Consider several distributions on the "margins", very slow-decaying distributions or the distributions with the exponential decaying tails. Two examples are given below.

Example 1. Let $p_k = c(k\log^2 k)^{-1},\ k > e$, be a probability distribution for some constant c. Then

$$S_{1,n} = \sum_k np_k e^{-np_k} \sim \int_e^{\infty}\frac{nc}{t\log^2 t}\exp\left(-\frac{nc}{t\log^2 t}\right)\,dt.$$

Put $t\log^2 t = y$. Then $t = y(\log^2 t)^{-1}$ and successive approximations yield the following:

$$t_0(y) \sim \frac{y}{\log^2 y}, \quad t_1(y) = \frac{y}{\log^2(y\log^2 y)} = \frac{y}{(\log y + 2\log\log y)^2} = \frac{y}{\log^2 y}\left(1 + \frac{2\log\log y}{\log y}\right)^{-1} = \frac{y}{\log^2 y}\left(1 - \frac{2\log\log y}{\log y} - \cdots\right).$$

Therefore

$$S_{1,n} \sim nc\int_e^{\infty}\frac{e^{-\frac{nc}{y}}}{y\log^2 y}\,dy \quad \left(dx = \frac{dy}{\log^2 y} - \frac{2dy}{\log^3 y}\right)$$
$$= nc\int_{e/n}^{\infty}\frac{e^{-\frac{c}{z}}}{z(\log n + \log z)^2}\,dz \quad (y = nz)$$
$$= nc\int_{e/n}^{\delta}\frac{e^{-\frac{c}{z}}}{z(\log n + \log z)^2}\,dz + nc\int_{\delta}^{1/\delta}\frac{e^{-\frac{c}{z}}}{z(\log n + \log z)^2}\,dz + nc\int_{1/\delta}^{\infty}\frac{e^{-\frac{c}{z}}}{z(\log n + \log z)^2}\,dz$$

for some small $\delta > 0$.

As it is easy to verify, the first term in above expression is $\mathcal{O}\big(n(\log^2 n)^{-1}\big)$, the middle term is also $\mathcal{O}\big(n(\log^2 n)^{-1}\big)$. For the last term,

$$nc\int_{1/\delta}^{\infty}\frac{e^{-\frac{c}{z}}}{z(\log n+\log z)^2}\,dz \sim nc\int_{1/\delta}^{\infty}\frac{1}{z(\log n+\log z)^2}\,dz = nc\int_{1/\delta}^{\infty}\frac{d\log z}{(\log n+\log z)^2} \sim \frac{cn}{\log n}.$$

Also note that

$$\frac{p''(x)p(x)}{(p'(x))^2} \to 2,\ x\to\infty$$

(i.e., condition (A5) is not satisfied) and $x_0 \sim \frac{n}{\log^2 n}$.

Example 2. Consider the geometric distribution with $p_k = pq^{k-1}$, $p+q=1$, $p>0$, $q>0$, and $k=1,2,\ldots$. For fixed n, one can find $k_0 = k_0(n)$ such that

$$npq^{k_0-1} > 1 \ \text{ and } \ npq^{k_0} \le 1$$

from which one gets

$$k_0 = \left\lfloor \log_{1/q} n + \log_{1/q}\left(\frac{p}{q}\right)\right\rfloor, \tag{16}$$

where $\lfloor x\rfloor$ denotes the largest integer no greater than x.

Putting $npq^{k_0-1} = e^{\alpha(n)}$, it follows that

$$\alpha(n) = \log\left(\frac{1}{q}\right)\left\{\left[\log_{1/q} n + \log_{1/q}\left(\frac{p}{q}\right)\right] - \left\lfloor \log_{1/q} n + \log_{1/q}\left(\frac{p}{q}\right)\right\rfloor\right\},$$

noting that $x - \lfloor x\rfloor$ is the fractional part of x.

The function $\alpha(n)$, $0 \le \alpha(n) < \log(1/q)$, is a periodic function of the argument $\log_{1/q} n$ with periodicity 1. It follows that

$$S_{1,n} = \sum_k np_k e^{-np_k} = e^{\alpha}e^{-e^{\alpha}} + qe^{\alpha}e^{-qe^{\alpha}} + q^2e^{\alpha}e^{-q^2e^{\alpha}} + \cdots$$
$$+\frac{1}{q}e^{\alpha}e^{-e^{\frac{e^{\alpha}}{q}}} + \frac{1}{q^2}e^{\alpha}e^{-e^{\frac{e^{\alpha}}{q^2}}} + \cdots.$$

This function is bounded, strictly positive and periodic with periodicity 1. The same is true for $S_{2,n} = \sum_{k\ge1}(np_k)^2 e^{-np_k}$ and for its higher order counterpart $S_{r,n}$, $r \ge 3$.

4. CENTRAL LIMIT THEOREM OF TURING'S FORMULA

Let T_n be Turing's formula given in (4) and π_0 be the missing probability as defined in (3). The following sufficient normality condition is due to [7].

Proposition 4. *For any $\varepsilon > 0$, let $k_{\varepsilon,n} = \max\{k : p_k > \varepsilon\sqrt{\tau_n}/n\}$. Then*

$$\frac{n(T_n-\pi_0)}{\sqrt{\mathrm{E}(N_1)+2\,\mathrm{E}(N_2)}} \xrightarrow{L} N(0,1) \tag{17}$$

if $\tau_n \to \infty$ and

$$k_{\varepsilon,n}e^{-\varepsilon\sqrt{\tau_n}} \to 0. \tag{18}$$

Suppose $p_k = ce^{-k/(\log k)^\beta}$ for $k > k_0$ where $k_0 \geq 1$ is a positive integer, and $c > 0$ and $\beta > 2$ are constants. It is easily verified, by Theorem 1, that

$$\tau_n = n\zeta_{1,n} \asymp (\log\log n)^\beta \to \infty. \tag{19}$$

Therefore to establish the asymptotic normality of (17), it suffices to show that (18) holds. For any given $\varepsilon > 0$, let $x_{\varepsilon,n}$ be the root of

$$p(x) = ce^{-x/(\log x)^\beta} = \varepsilon \frac{\sqrt{\tau_n}}{n}$$

and let $k_{\varepsilon,n} = \max\{k : k \leq x_{\varepsilon,n}\}$. Let x_n be the root of

$$p(x) = ce^{-x/(\log x)^\beta} = \frac{1}{n}$$

and let $k_n = \max\{k : k \leq x_n\}$. It may be verified that

$$x_n \asymp (\log\log n)(\log n).$$

Since, for sufficiently large n, $x_n \geq x_{\varepsilon,n}$, it follows that $k_n \geq k_{\varepsilon,n}$. By Proposition 4, the desired result follows from the fact that, for sufficiently large n,

$$k_{\varepsilon,n}e^{-\varepsilon\sqrt{\tau_n}} \leq k_n e^{-\varepsilon\sqrt{\tau_n}} \asymp (\log\log n)(\log n)e^{-\varepsilon\sqrt{\tau_n}} \asymp (\log\log n)(\log n)e^{-\varepsilon(\log\log n)^{\beta/2}} \to 0. \tag{20}$$

To see (20) is true, it suffices to note that

$$\log\left[(\log\log n)(\log n)e^{-\varepsilon(\log\log n)^{\beta/2}}\right] = \log\log\log n + \log\log n - \varepsilon(\log\log n)^{\beta/2} \to -\infty$$

for every given $\varepsilon > 0$ when $\beta/2 > 1$ or $\beta > 2$.

This example puts the distributions with near-exponential tails in the sub-domain supporting asymptotic normality for Turing's formula.

5. APPENDICES

5.1. Proof of Proposition 1

Since $p \leq -\log(1-p)$ for all $p \in (0,1)$, it follows that $-np_k \geq n\log(1-p_k)$, that $e^{-np_k} \geq (1-p_k)^n$, and therefore

$$\tau_{r,n} = n^r \sum_{k\geq 1} p_k^r (1-p_k)^n \leq \sum_{k\geq 1}(np_k)^r e^{-np_k} = S_{r,n}. \tag{21}$$

Note that for any $\delta \in (0,1)$,

$$\begin{aligned}
\sum_{k\geq 1}(np_k)^r\left[e^{-np_k} - (1-p_k)^n\right] &= \sum_{k\geq 1}(np_k)^r\left[e^{-np_k} - e^{n\log(1-p_k)}\right] \\
&= \sum_{k\geq 1}(np_k)^r\left(e^{-np_k} - e^{-np_k - \frac{np_k^2}{2} - \cdots}\right) = \sum_{k\geq 1}(np_k)^r e^{-np_k}\left(1 - e^{-\frac{np_k^2}{2} - \cdots}\right) \\
&= \sum_{k:p_k > n^{-(1-\delta)}}(np_k)^r e^{-np_k}\left(1 - e^{-\frac{np_k^2}{2} - \cdots}\right) + \sum_{k:p_k \leq n^{-(1-\delta)}}(np_k)^r e^{-np_k}\left(1 - e^{-\frac{np_k^2}{2} - \cdots}\right) \\
&\leq ne^{-n^\delta} + \sum_{k:p_k \leq n^{-(1-\delta)}} np_k e^{-np_k}\frac{1}{n^{1-2\delta}}.
\end{aligned}$$

Let $\delta \in (0, 1/10)$ be a constant. Let $K_\delta(n)$ be the largest integer such that $p_k \geq 1/n^{1-\delta}$ for every $k \leq K_\delta(n)$. Noting that $p_k = o(1/k)$, it follows that $K_\delta = o(n^{1-\delta})$.

Therefore

$$S_{r,n} - \tau_{r,n} = \sum_{k\geq 1}(np_k)^r\left[e^{-np_k} - (1-p_k)^n\right] = \sum_{k\geq 1}(np_k)^r\left[e^{-np_k} - e^{n\log(1-p_k)}\right]$$

$$= \sum_{k\ge 1}(np_k)^r\Big(e^{-np_k} - e^{-np_k-\frac{np_k^2}{2}-\cdots}\Big) = \sum_{k\ge 1}(np_k)^r e^{-np_k}\Big(1-e^{-\frac{np_k^2}{2}-\cdots}\Big)$$
$$= \sum_{k\le K_\delta}(np_k)^r e^{-np_k}\Big(1-e^{-\frac{np_k^2}{2}-\cdots}\Big) + \sum_{k> K_\delta}(np_k)^r e^{-np_k}\Big(1-e^{-\frac{np_k^2}{2}-\cdots}\Big).$$

However,

$$\sum_{k\le K_\delta}(np_k)^r e^{-np_k}\Big(1-e^{-\frac{np_k^2}{2}-\cdots}\Big) \le K_\delta (n^\delta)^r e^{-n^\delta} \le n^{1-\delta} n^{r\delta} e^{-n^\delta} = o(1), \quad n\to\infty,$$

which is negligible since $S_{r,n} \ge C_r(\alpha)$ by Lemma 1.

For the second summation, it follows that

$$\sum_{k> K_\delta}(np_k)^r e^{-np_k}\Big(1-e^{-\frac{np_k^2}{2}-\cdots}\Big) \le \sum_{k> K_\delta}(np_k)^r e^{-np_k}\big(1-e^{-np_k^2}\big)$$
$$\le \sum_{k\ge K_\delta}(np_k)^r e^{-np_k}\frac{1}{n^{1-2\delta}} = \big(S_{r,n}-o(1)\big)\frac{1}{n^{1-2\delta}}.$$

That is to say,

$$S_{r,n} - \tau_{r,n} \le (S_{r,n}-o(1))\frac{1}{n^{1-2\delta}},$$

which leads to

$$\Big(1-\frac{1}{n^{1-2\delta}}\Big)S_{r,n} \le \tau_{r,n}. \tag{22}$$

The proposition is a direct result from equations (21) and (22). □

5.2. Proof of Proposition 3

Consider the point $x_r = x_r(n)$ such that

$$np(x_r) = r, \ \text{ or } \ p(x_r) = \frac{r}{n}, \quad r = 1, 2, \ldots . \tag{23}$$

Write

$$I_{r,n} = \int_{x_r}^{\infty}[np(x)]^r e^{-np(x)}\,dx + \int_0^{x_r}[np(x)]^r e^{-np(x)}\,dx = I'_{r,n} + I''_{r,n}. \tag{24}$$

Consider $I'_{r,n}$ first. It is clear that

$$n^r e^{-r}\int_{x_r}^{\infty} p^r(x)\,dx \le \int_{x_r}^{\infty}[np(x)]^r e^{-np(x)}\,dx \le n^r\int_{x_r}^{\infty}p^r(x)\,dx. \tag{25}$$

Hence

$$I'_{r,n} = \int_{x_r}^{\infty}(np(x))^r e^{-np(x)}dx \asymp n^r\int_{x_r}^{\infty}p^r(x)\,dx = n^r\int_{x_r}^{\infty}p(x)\frac{dp^r(x)}{rp'(x)}$$
$$= \frac{n^r(p(x))^{r+1}}{rp'(x)}\Big|_{x_r}^{\infty} - \frac{n^r}{r}\int_{x_r}^{\infty}\Big(\frac{p(x)}{p'(x)}\Big)'(p(x))^r\,dx$$
$$= -\frac{p(x_r)}{p'(x_r)} + \frac{1}{r}\int_{x_r}^{\infty}\Big(\frac{p''(x)p(x)-[p'(x)]^2}{(p'(x))^2}\Big)(np(x))^r\,dx.$$

Due to inequality (25), conditions (A2) and (A5), it follows that

$$I'_{r,n} \asymp \frac{p(x_r)}{|p'(x_r)|}. \tag{26}$$

Now consider $I''_{r,n}$:

$$\begin{aligned} I''_{r,n} &= \int_0^{x_r} (np(x))^r e^{-np(x)} dx = -\int_{x_r}^0 (np(x))^r e^{-np(x)} \frac{d(np(x))}{np'(x)} \\ &= \int_{x_r}^0 \frac{1}{-np'(x)} (np(x))^r e^{-np(x)} d(np(x)) \\ &\le \frac{1}{n|p'(x_r)|} \int_{x_r}^0 (np(x))^r e^{-np(x)} d(np(x)) = \frac{1}{n|p'(x_r)|} \int_{np(x_r)}^{np(0)} u^r e^{-u} du \\ &\le \frac{np(x_r)}{n|p'(x_r)|} \frac{\Gamma(r+1)}{r} = \frac{\Gamma(r+1)}{r} \frac{p(x_r)}{|p'(x_r)|}, \end{aligned}$$

where $\Gamma(r+1) = \int_0^\infty u^r e^{-u} du$ is the Gamma function. The first inequality above is due to the fact that $p'(x) < 0$ and $p'(x)$ is monotone increasing. The second inequality above is due to the fact that $np(x_r) = r$ and $\int_{np(x_r)}^{np(0)} u^r e^{-u} du \le \Gamma(r+1)$.

Hence, it follows that

$$I''_{l,n} \asymp \frac{p(x_l)}{p'(x_l)}. \tag{27}$$

Equations (26) and (27) lead to the proposition. □

ACKNOWLEDGMENTS

This study was partially supported by Russian Science Foundation (Project No. 17-11-01098) and National Science Foundation (Grant No. 530552).

REFERENCES

1. A. Chao, S.-M. Lee, and T.-C. Chen, "A generalized Good's nonparametric coverage estimator", Chinese J. Math. **16** (3), 189–199 (1988).
2. W. W. Esty, "A Normal Limit Law for a Nonparametric Estimator of the Coverage of a Random Sample", Ann. Statist. **11**, 905–912 (1983).
3. I. J. Good, "The Population Frequencies of Species and the Estimation of Population Parameters", Biometrika **40** (3, 4), 237–264 (1953).
4. H. E. Robbins, "Estimating the Total Probability of the Unobserved Outcomes of an Experiment", Ann. Math. Statist. **39** (1), 256–257 (1968).
5. C.-H. Zhang and Z. Zhang, "Asymptotic Normality of a Nonparametric Estimator of Sample Coverage", Ann. Statist. **37** (5A), 2582–2595 (2009).
6. Z. Zhang, *Statistical Implications of Turing's Formula*. (Wiley, New York, 2017).
7. Z. Zhang, *On Normal Law Conditions for Turing's Formula*. in *Wiley StatsRef: Statistics Reference Online* (2018).
8. Z. Zhang, "Domains of Attraction on Countable Alphabets", Bernoulli, **24** (2), 873–894 (2018).
9. Z. Zhang and J. Zhou, "Re-Parametrization of Multinomial Distribution and Diversity Indices", J. Statist. Planning and Inference **140** (7), 1731–1738 (2010).
10. Z. Zhang and M. Grabchak, "Entropic Representation and Estimation of Diversity Indices", J. Nonparam. Statist. **28** (3), 563–575 (2016).