sciendo

uniform distribution theory

DOI: 10.2478/udt-2019–0011
Unif. Distrib. Theory **14** (2019), no.2, 27–32

CHAINS OF TRUNCATED BETA DISTRIBUTIONS AND BENFORD'S LAW

Pongpol Ruankong[1] — Songkiat Sumetkijakan[2]

[1]Kasetsart University, Bangkok, THAILAND

[2]Chulalongkorn University, Bangkok, THAILAND

ABSTRACT. It was proved by Jang et al. that various chains of one-parameter distributions converge to Benford's law. We study chains of truncated distributions and propose another approach, using a recent convergence result of the Lerch transcendent function, to proving that they converge to Benford's law for initial Beta distributions with parameters α and 1.

Communicated by Friedrich Pillichshammer

1. Introduction

Since its discovery many decades ago, it has been widely observed that the first digit of a large naturally occurring data set, e.g. time between earthquakes, is distributed according to the Benford's law [2, 4], i.e. the probability of the first digit in base-10 representation being d is $\log_{10}\left(\frac{d+1}{d}\right)$. Theoretically, the first digit of a random variable, as a model for a data set, can be far from Benford's law. However, supported by broad empirical evidences, it is generally believed that "more randomness" tends to bring the first digit distribution closer to Benford's law. A simple way to increase the degree of randomness of a parametrically distributed random variable is to randomly pick a value of the parameter according

2010 Mathematics Subject Classification: 11K06, 60A10.
Keywords: Benford's law, beta distribution.
The second author is grateful for the financial support from the Thailand Research Fund through Grant No. RSA6080022.

to another parametric distribution. Alex E. Kossovsky [4] conjectured that if the process is repeated iteratively, the first digit distributions of the resulted *chain of distributions* should converge to the Benford distribution.

If the randomly (and iteratively) chosen parameter shrinks the support of the distribution, the resulted chain is called a *chain of truncated distributions*. An evident real-world example is the remaining credit of a credit card. Chains of truncated distributions can also be viewed as the first step in the further study of chains of distributions where each iteration might be truncating or appending. A potential application is to explain first digit behavior of bank account balances resulting from a series of withdrawals and deposits.

In 2009, Jang, Kang, Kruckman, Kudo, and Miller [3] confirmed the conjecture for chains of one-parameter distributions satisfying two conditions. The main condition is a natural compatibility condition among distributions and the second condition is a technical assumption. For examples, chains of exponential (respectively, uniform $(0, b)$) distributions, whose parameters are randomly drawn from a previous exponential (respectively, uniform $(0, b)$) distribution, were shown to have first digit distributions converging to Benford's law. Though the first compatibility condition in [3] is natural and straightforward to verify, the second technical condition is quite difficult to check and it is unclear whether it holds for certain chains of truncated distributions. The second condition is needed in proving the convergence to Benford's law. But for chains of uniform and exponential distributions, the condition can be replaced by convergence of some related special functions.

In this manuscript, we give a precise proof that chains of truncated Beta distributions as defined in Section 2 converge to Benford's law. For the most part, the proof is quite straightforward while the convergence part is proved via a recent convergence result on Lerch transcendent [5]. This stresses the relationship between convergence to Benford's law and convergence of certain special functions. An advantage of this approach is that, having been studied for centuries, a myriad of special functions with known asymptotic behaviors might lead to more results on convergence of chains of (truncated) distributions.

2. Main result

Loosely speaking, an infinite sequence of distributions $\{F_n\}_{n=0}^{\infty}$ is called a *chain of truncated distributions* if, for each $n = 0, 1, 2, \dots$, F_{n+1} is the distribution of an F_n-truncated random variable. For brevity, we also call the corresponding sequence of random variables $X_n \sim F_n$ a chain of truncated distributions.

In this manuscript, we limit our scope to chains of truncated distributions $\{X_n\}$ with parameter $\alpha > 0$ defined as follows. Put

$$X_0 \sim U_\alpha(0,1) \quad \text{and} \quad X_{n+1} \sim U_\alpha(0, X_n) \quad \text{for} \;\; n = 0,1,2,\dots,$$

where $X \sim U_\alpha(0,b)$, for $b > 0$, if its probability density function is

$$f_X(x) = \begin{cases} \dfrac{\alpha x^{\alpha-1}}{b^\alpha} & \text{if } 0 \le x \le b, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Observe that $X \sim U_\alpha(0,b)$ means exactly that X/b follows the Beta distribution [7] with parameters α and 1.

The following lemma is used in deriving the explicit formulas of the probability density functions of X_n's stated in Theorem 2.

Lemma 1. *Let X be a continuous random variable with density f_X supported on $[0,1]$ and $X' \sim U_\alpha(0,X)$. Then*

$$f_{X'}(x) = \alpha x^{\alpha-1} \int_x^1 \frac{1}{t^\alpha} f_X(t)\, \mathrm{d}t \quad \textit{for } x \in [0,1], \quad \textit{and} \quad f_{X'}(x) = 0, \;\; \textit{elsewhere.}$$

Proof. Observe that for $x \in [0,1]$ if $t < x$, then $P(X' \le x \,|\, X = t) = 1$ and if $x \le t \le 1$, then

$$P(X' \le x \,|\, X = t) = \int_0^x f_{X'}(s)\, \mathrm{d}s = \int_0^x \frac{\alpha s^{\alpha-1}}{t^\alpha}\, \mathrm{d}s = \frac{x^\alpha}{t^\alpha}.$$

By the observation, the distribution function of X' is

$$\begin{aligned} F_{X'}(x) &= \int_0^1 P(X' \le x \,|\, X = t)\, \mathrm{d}F_X(t) \\ &= \int_0^x 1\, \mathrm{d}F_X(t) + \int_x^1 P(X' \le x \,|\, X = t)\, \mathrm{d}F_X(t) \\ &= F_X(x) + x^\alpha \int_x^1 \frac{1}{t^\alpha}\, \mathrm{d}F_X(t). \end{aligned}$$

Differentiating both sides using the Fundamental Theorem of Calculus yields the desired result. □

As a consequence, we obtain the following result. Here, ln denotes the natural logarithm and log denotes $\log_{10}$.

Theorem 2. *Put $X_0 \sim U_\alpha(0,1)$ and $X_n \sim U_\alpha(0, X_{n-1})$ for $n \geq 1$. Then the probability density function f_n of X_n, $n \geq 0$, is given by*

$$f_n(x) = \begin{cases} \dfrac{\alpha^{n+1}}{n!}\, x^{\alpha-1} \left(\ln \dfrac{1}{x}\right)^n & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

P r o o f. This will be proved by induction on n. By assumption, the pdf of X_0 is f_0. We assume that $f_{X_n} = f_n$. By Lemma 1, for $0 \leq x \leq 1$,

$$\begin{aligned} f_{X_{n+1}}(x) &= \frac{\alpha^{n+2}\, x^{\alpha-1}}{n!} \int_x^1 \frac{1}{t} \left(\ln \frac{1}{t}\right)^n \mathrm{d}t \\ &= \frac{\alpha^{n+2}\, x^{\alpha-1}}{(n+1)!} \left(\ln \frac{1}{x}\right)^{n+1} = f_{n+1}(x). \end{aligned}$$ □

Next, let Y_n be the first digit of X_n. Recall that the first digit of a positive number a is defined as $\left\lfloor \dfrac{a}{10^{\lfloor \log a \rfloor}} \right\rfloor$. Then the distribution of Y_n is given by

$$\begin{aligned} G_n(d) = P(Y_n = d) &= \sum_{\ell=1}^{\infty} P\left(\frac{d}{10^\ell} \leq X_n < \frac{d+1}{10^\ell}\right) \\ &= \sum_{\ell=1}^{\infty} \int_{d/10^\ell}^{(d+1)/10^\ell} \frac{\alpha^{n+1}}{n!}\, t^{\alpha-1} \left(\ln \frac{1}{t}\right)^n \mathrm{d}t, \end{aligned}$$

where

$$d = 1, 2, \dots, 9\,.$$

We will show that $\lim_{n\to\infty} G_n(d) = \log\left(\frac{d+1}{d}\right)$. Substituting t by $s/10^\ell$ and applying Tonelli's theorem [1], we have

$$G_n(d) = \int_d^{d+1} h_n(s)\, \mathrm{d}s \quad \text{where} \quad h_n(s) = \sum_{\ell=1}^{\infty} \frac{\alpha^{n+1}\, s^{\alpha-1}\, (\ell \ln 10 - \ln s)^n}{10^{\alpha\ell}\, n!}. \tag{2}$$

In order to compute the limit of G_n, the following Lemmas will be helpful.

Lemma 3. *For a positive real number α and a positive integer n,*

$$\sum_{\ell=1}^{\infty} \frac{(\alpha\ell \ln 10)^n}{10^{\alpha\ell}\, n!} \leq \frac{1}{\alpha \ln 10} + 1.$$

P r o o f. Put $K = \alpha \ln 10$ and write $g(x) = \frac{(Kx)^n}{e^{Kx}\, n!}$ for $x \geq 0$. Its derivative, $g'(x) = \frac{K^n (n-Kx) x^{n-1}}{e^{Kx}\, n!}$ for $x > 0$, vanishes if and only if $x = \frac{n}{K}$. Moreover, g is increasing if $0 \leq x \leq \frac{n}{K}$ and decreasing if $x \geq \frac{n}{K}$.

With $I_n = \left[\left\lfloor \frac{n}{K} \right\rfloor, \left\lfloor \frac{n}{K} \right\rfloor + 1\right]$, the integral of g on $[0, \infty) \setminus I_n$ is bounded below by the integral of step functions g_n defined on $[0, \infty) \setminus I_n$ by $g_n(x) = g(\ell)$ if $\ell \le x < \ell + 1 \le \left\lfloor \frac{n}{K} \right\rfloor$ and $g_n(x) = g(\ell+1)$ if $\left\lfloor \frac{n}{K} \right\rfloor + 1 \le \ell < x \le \ell + 1$. Hence,

$$\sum_{\ell=1}^{\infty} \frac{(K\ell)^n}{e^{K\ell}\, n!} \le \int_0^{\infty} \frac{(Kx)^n}{e^{Kx}\, n!}\, dx + g\left(\frac{n}{K}\right) \le \frac{\Gamma(n+1)}{n! \cdot K} + \frac{n^n}{e^n\, n!}.$$

Finally, by Stirling's formula (see [6, p.17]),

$$\sum_{\ell=1}^{\infty} \frac{(\alpha\ell \ln 10)^n}{10^{\alpha\ell}\, n!} \le \frac{1}{K} + \frac{1}{\sqrt{2\pi n}} \le \frac{1}{\alpha \ln 10} + 1. \qquad \square$$

Lemma 4 ([5, Corollary 4])**.** *The Lerch transcendent* $\Phi(\lambda, m, z) = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k+z)^m}$ *is considered as a function of the complex variable* z*, with* λ *and* m *as real-valued parameters. Then, for* $0 < \lambda < 1$*,*

$$\lim_{m \to -\infty} \frac{(-\log \lambda)^{1-m} \Phi(\lambda, m, z)}{\Gamma(1-m)} = \lambda^{-z} \tag{3}$$

uniformly for z *on compact subsets of* $\mathbb{C}$*.*

Theorem 5. *Put* $\alpha > 0$*,* $X_0 \sim U_\alpha(0,1)$ *and* $X_n \sim U_\alpha(0, X_{n-1})$ *for* $n \ge 1$*. Let* G_n *be the distribution of the first digit of* X_n *and* h_n *be defined as in* (2)*. Then, for* $s \in [1, 10)$*,*

$$\lim_{n \to \infty} h_n(s) = \frac{1}{s \ln(10)}$$

and, consequently,

$$\lim_{n \to \infty} G_n(d) = \log\left(\frac{d+1}{d}\right).$$

That is, Y_n *converges in distribution to Benford's law.*

P r o o f. By Lemma 3, h_n is bounded on $[1, 10)$:

$$h_n(s) \le \sum_{\ell=1}^{\infty} \frac{\alpha^{n+1}\, 10^{\alpha-1}\, (\ell \ln 10)^n}{10^{\alpha\ell}\, n!} \le 10^{\alpha-1} \left(\frac{1}{\ln 10} + \alpha\right). \tag{4}$$

Thus,

$$\lim_{n \to \infty} \int_d^{d+1} h_n(s)\, \mathrm{d}s = \int_d^{d+1} \lim_{n \to \infty} h_n(s)\, \mathrm{d}s$$

by the dominated convergence theorem [1]. It is straightforward to verify that

$$h_n(s) = \frac{(\alpha \ln 10)^{n+1} s^{\alpha-1}}{n! \ln 10} \left[\Phi(10^{-\alpha}, -n, -\log s) - (-\log s)^n\right].$$

Applying an asymptotic approximation of Φ (Lemma 4)

$$\lim_{n\to\infty} \frac{(\alpha \ln 10)^{n+1}}{n!} \Phi(10^{-\alpha}, -n, -\log s) = s^{-\alpha}.$$

The remaining term goes to zero as $\lim_{n\to\infty} \frac{x^n}{n!} = 0$ for all x. Therefore, $h_n(s)$ converges to $\frac{1}{s \ln 10}$. □

Acknowledgement. The authors would like to thank Alex Kossovsky for eye- -opening lectures at our universities and several subsequent talks.

REFERENCES

[1] BARTLE, R. G.—SHERBERT, D. R.: *Introduction to Real Analysis* (4th eddition). John Wiley & Sons, Inc., New York, 2011.

[2] BENFORD, F.: *The law of anomalous numbers*, Proc. Am. Philos. Soc., **78(4)** (1938), 551–572.

[3] JANG, D.—KANG, J. K.—KRUCKMAN, A.—KUDO, J.—MILLER, S. J.: *Chains of distributions, hierarchical Bayesian models and Benford's law*, J. Algebra, Number Theory, Adv. Appl. **1** (2009), no. 1, 37–60.

[4] KOSSOVSKY, A. E.: *Benford's Law. Theory, The General Law of Relative Quantities, And Forensic Fraud Detection Applications*, World Scientific Publishing Company, Hackensack, NJ, 2015.

[5] NAVAS, L. M.—RUIZ, F. J.—VARONA, J. L.: *Asymptotic behavior of the Lerch transcendent function*, J. Approx. Theory, **170** (2013), 21–31.

[6] RIVLIN, T. J.: *An Introduction to the Approximation of Functions*, Blaisdell Publishing Co. Ginn and Co., Waltham, Mass.-Toronto, Ont.-London, 1969.

[7] ROHATGI, V. K.—SALEH, A. K. MD. EHSANES.: *An Introduction to Probability and Statistics* (3rd edition). In: *Wiley Series in Probability and Statistics.* John Wiley & Sons, Inc., Hoboken, NJ, 2015.

Received April 2, 2018
Accepted February 28, 2019

Pongpol Ruankong
Department of Mathematics
Faculty of Science
Kasetsart University
50 Ngam Wong Wan Rd.
Ladyaow, Chatuchak, Bangkok 10900
THAILAND
E-mail: fscippru@ku.ac.th

Songkiat Sumetkijakan
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
254 Phayathai Rd.
Pathumwan, Bangkok 10330
THAILAND
E-mail: songkiat.s@chula.ac.th