Golden Distribution of Probabilities

Kazuya Hayata

Department of Socio-Informatics, Sapporo Gakuin University, Ebetsu 069-8555, Japan
E-mail address: hayata1@hotmail.co.jp

(Received May 28, 2012; Accepted August 1, 2014)

A novel concept based on the golden ratio ϕ, where the cumulative probability of the Fibonacci numbers coincides with the reciprocal of ϕ, is presented for discrete probability distributions. In addition to the binomial, Poisson, and geometrical distributions, the Benford-type as well as the inverse power distributions are considered. For the latter, in the limit of $n \to \infty$ (n being a parameter of the present distribution), the value of the power is found to approach the fractal dimension of the golden tree. Finally, examples being close surprisingly to the golden distribution are shown for the analysis of the word spectra of texts written in English.

Key words: Fibonacci Sequence, Golden Ratio, Golden Tree, Inverse Power Law, Word Spectrum

1. Introduction

Since ancient times the rectangle with the aspect ratio 1:1.618 ... has been believed to be most *beautiful*. For this reason, this special plane figure has been called the *golden rectangle*, and its proportion that can be expressed with the infinite decimal has been called the *golden ratio*. According to an established theory of the art history, it is said that since old times the idea of this ratio has been applied to a variety of art works such as, for instance, the Parthenon in Greece. In fact, this special ratio was, at the same time, a *number* in which mathematicians including Euclid and Kepler had been interested. Here we should note that the approximate value of the golden ratio can be obtained by calculating the ratio between numbers adjacent each other in the Fibonacci sequence and that in the limit of infinity it converges on $\phi = 1.6180339\cdots$. As the times proceed the golden ratio ϕ was found in unexpected fields of both mathematical and natural sciences, namely, concerning studies on the growth mechanism of nautili's shells, the arrangement of sunflower's seeds, the configuration of Penrose tiling, the quasicrystallographic structure, and the dynamics of a sequential stock market. In recent years, it appears that mysteries for the golden ratio as the world's most astonishing number have constantly been enhanced (Livio, 2002). Incidentally, one might notice later that all the frames of twelve illustrations to be seen in this paper exhibit the aspect ratio close to ϕ, supporting the validity of the Fechner's psychological experiment that was done in the 1870's, where a preference for the rectangle with the aspect ratio close to ϕ was demonstrated. In this paper a novel concept based on the golden ratio ϕ, where the cumulative probability of the Fibonacci numbers coincides with the reciprocal of ϕ, is presented for any discrete probability distribution. To date, it seems that the appearance of ϕ in the two-point distributions of categories has been reported but restricted mainly to the three cases (Schroeder, 1991; Livio, 2002), namely, 1) the mixture between long range correlated binary symbols (0 versus 1, A versus B, etc.) in strings generated by applying a prescribed procedure on the basis of a substitution algorithm; instead of the symbols, Fibonacci used two kinds of rabbits, 2) the blending between *kites* and *arrows* in the Penrose tile, for which several astonishing properties have been found (Penrose, 1979), and 3) the arrangement scheme of cells in sunflowers and in pinecones, where their cellular patterns are composed of the clockwise and the counterclockwise spirals, which were once studied by Schimper, Braun, and the Bravais brothers. In contrast to these preceding studies on the two-point distribution of qualitative data, in this paper we shall focus our attention on the golden ratio for any n-point probability distribution of quantitative data. Here, in addition to classical cases such as the binomial, Poisson, and geometrical distributions, the Benford-type as well as the inverse power functions are chosen. Finally, examples close to the golden distribution are shown for the analysis of the word spectra of texts written in several natural languages.

2. Definition

To make a definition of the golden distribution of probabilities, we shall begin with writing the cumulative probability of the Fibonacci numbers:

$$r_F = p(1) + p(2) + p(3) + p(5) + p(8) + p(13) + p(21) + p(34) + \ldots, \quad (1)$$

where $0 \le r_F \le 1$; $p(x) = P(X = x)$ stands for a discrete probability distribution. Although Eq. (1) is written in the form of an infinite series, for some cases such as the binomial distribution the series will be truncated at the finite number of terms. The golden distribution will be defined as that meeting the following condition

$$r_F = 1/\phi, \quad (2)$$

doi:10.5047/forma.2014.006

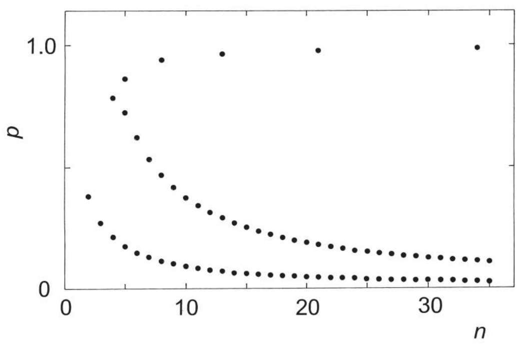

Fig. 1. Relation between two parameters of the binomial distribution $B(n, p)$ in which the golden condition, $r_F = 1/\phi$, is met.

with $1/\phi = 2/(1+\sqrt{5}) = 0.6180339\cdots$. Here Eq. (2) is a transcendental equation which will be solved for parameters included in $p(x)$.

3. Analysis

Probability distributions have been used for solving various problems in spatial statistics as well as in stochastic geometry. In recent years, one could find the stochastic approach to analyze many interesting problems in fractal geometry of nature (Schroeder, 1991; Takaki, 2003).

3.1 Binomial distribution

As the most important discrete probability distribution we first consider the binomial distribution

$$p(x) = P(X = x) = {}_nC_x p^x (1-p)^{n-x}, \qquad (3)$$

which is denoted frequently by $B(n, p)$. Here $x = 0, 1, 2, \ldots, n$; n and p $(0 < p < 1)$ are parameters with which the profile of the present distribution is determined. Substituting Eq. (3) into Eq. (2) with Eq. (1), we obtain the solution (n, p). The results are plotted in the scatter diagram of Fig. 1, where three branches are seen. For $n = 2$ and $n = 3$, there is a single solution which is located on the lowest branch; for $n \geq 4$, there are twin solutions except $n = 5, 8, 13, 21, 34, \ldots$, all of which are found to be Fibonacci numbers. Evidently, the behavior of the upper branch shows a sharp contrast to that of the lower twin branches. With increasing n, the former increases slowly and approaches $p = 1$, while the latter decreases and approaches $p = 0$. According to the Poisson's theorem, the lower branches link to the Poisson distribution that will be described in what follows. To examine the profile for the three solutions, for $n = 21$ we show in Fig. 2 the dependence of the probability $p(x)$ as a function of x. In comparison between the three cases, it is seen that with increasing p the mode (Mo) of the distribution moves considerably along the x-axis; specifically, $Mo = 1, 4$, and 21 for Figs. 2(a), (b) and (c), respectively.

3.2 Poisson distribution

Subsequently we consider the Poisson distribution

$$p(x) = P(X = x) = e^{-\lambda}\lambda^x/x!, \qquad (4)$$

where $x = 0, 1, 2, \ldots$; λ is a positive parameter. According to the Poisson's theorem, this distribution, which is denoted

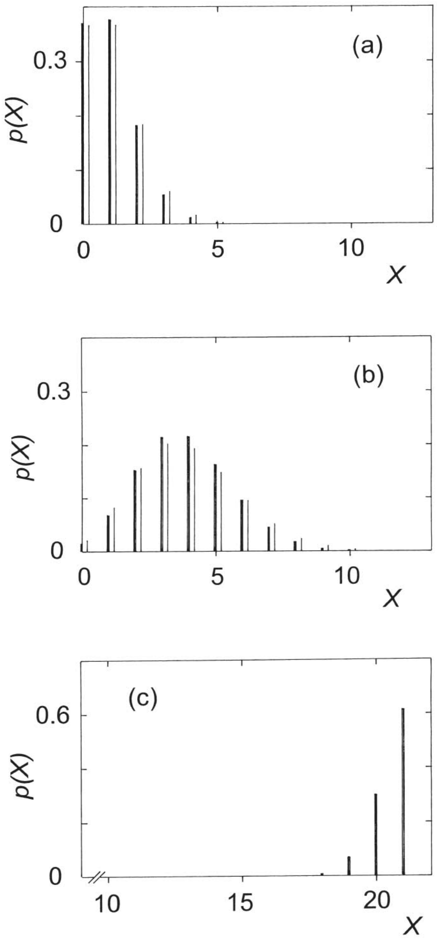

Fig. 2. The binomial distribution $B(21, p)$ for (a) $p = 0.0463$, (b) $p = 0.1822$, and (c) $p = 0.9773$, which correspond to the three dots on $n = 21$ in Fig. 1. For comparison, in (a) and (b) the Poisson distribution with $\lambda = 1.0058$ and $\lambda = 3.8464$, respectively, is plotted with fine lines.

with $Po(\lambda)$, can be obtained as a limit of $B(n, p)$. Specifically, with $np = \lambda$ remaining constant, for sufficiently large n, $B(n, p)$ can be approximated by $Po(\lambda)$. With increasing λ, the profile of the present distribution becomes less asymmetric, and eventually becomes symmetric as $\lambda \to \infty$. On substitution of Eq. (4) into Eq. (2) one obtains twin solutions:

$$\lambda = 1.00575\cdots, 3.84641\cdots.$$

For these parameters the profile of $p(x)$ against the x-axis is superposed with fine lines, respectively, in Figs. 2(a) and (b).

The Poisson distribution has been mentioned in the number density of particles randomly scattered on a plane. It has been found that the so-called Voronoi tessellation arising from the distribution contains a variety of polygons such as rectangles $(m = 4)$ to octagons $(m = 8)$, where m is the number of angles on a polygon; the average of m has been shown to be six (Takaki, 2003). To date, the Voronoi tessellation has been investigated in the boundaries of cells, the pattern of space division by territories (Hasegawa and Tanemura, 1976), and the European map (Takaki, 2003). It might be expected if the Voronoi tessellations that are gen-

Table 1. Probability G (0.193728) versus Fibonacci numbers.

x	$p(x)$
1	0.193728
2	0.156197
3	0.125938
5	0.081869
8	0.042910
13	0.014621
21	0.002611
34	0.000159
55	0.000002
89	0.000000
Sum	0.618035

erated from the distributions with the twin solutions derived above possess possibly a unique property.

3.3 Geometrical distribution

In Bernoulli's sequences of trials, the number of trials necessary for the initial occurrence of an event is known to obey the geometrical distribution $G(p)$

$$p(x) = P(X = x) = p(1 - p)^{x-1}, \tag{5}$$

with $x = 1, 2, 3, \ldots$, which includes a single parameter p ($0 < p < 1$). Substitution of Eq. (5) into Eq. (2) yields the solution for the golden distribution:

$$p = 0.19372 \cdots .$$

For the present solution, the probabilities of Eq. (5), which have been calculated for several Fibonacci numbers, are given in Table 1. Note that the cumulative probability becomes $1/\phi$.

3.4 Distribution that explains the first digit phenomenon

For certain large-scale corpora of numerical data sources, such as population, death rate, scores of baseball games, basin areas of rivers, as well as of those in the entire articles in a popular magazine, the statistical probability of a number on the first digit was found not to be uniform but to be strongly dependent upon the number. For decimal data the dependence can be expressed as (Benford, 1938)

$$p(x) = P(X = x) = \log_{10}(1 + x^{-1}), \tag{6}$$

where $x = 1, 2, \ldots, 8, 9$. It is worth mentioning that this formula, which nowadays is termed the Benford's law or the first digit phenomenon (Hill, 1998), holds also for the Fibonacci numbers themselves (Washington, 1981). The above equation was derived by integrating the inverse power-type probability density function, $f(t)$, from $t = x$ to $t = x + 1$, namely

$$p(x) = \int_x^{x+1} f(t)dt \tag{7}$$

with

$$f(t) = Ct^{-q}, \tag{8}$$

Table 2. Comparison between the Benford's law ($q = 1$) and the golden distribution ($q = 0.3742$). In the latter, $r_F = p(1) + p(2) + p(3) + p(5) + p(8) = 0.6181 \doteqdot 1/\phi$. The sum of the golden distribution slightly exceeding unity is due to the round-off error that has arisen from the finite-digit calculation.

x	$q = 1$	$q = 0.3742$
1	0.3010	0.1684
2	0.1761	0.1382
3	0.1249	0.1217
4	0.0969	0.1107
5	0.0792	0.1026
6	0.0669	0.0964
7	0.0580	0.0913
8	0.0512	0.0872
9	0.0458	0.0836
Sum	1.0000	1.0001

where $q > 0$; C is an unknown positive constant. In the derivation of Eq. (6) it has been assumed that $q = 1$, which derives, from the normalization condition of probabilities, the explicit form of the normalization factor:

$$C = \log_{10} e. \tag{9}$$

For $q \neq 1$ the probability function $p(x)$ becomes (Schroeder, 1991)

$$p(x) = P(X = x) = \{(x + 1)^{1-q} - x^{1-q}\}/(10^{1-q} - 1), \tag{10}$$

where a single parameter q is included. Substituting Eq. (10) into Eq. (2) for five Fibonacci numbers $\{1, 2, 3, 5, 8\}$ and subsequently solving it for q, one yields

$$q = 0.374196 \cdots . \tag{11}$$

The numerical results for Eq. (6) ($q = 1$) and for Eq. (10) with Eq. (11) are compared in Table 2.

3.5 Inverse power distribution for rank-ordered statistics

Here we consider the so-called inverse power distribution (Zipf, 1949; Mandelbrot, 1954; Schroeder, 1991) for rank-ordered data ($x = 1, 2, \ldots, n$):

$$p(x) = P(X = x) = Cx^{-q}, \tag{12}$$

the profile of which exhibits the form that is identical to Eq. (8), where C is a positive constant to be determined by imposing the normalization condition of the entire probability:

$$C = \left[\sum_{x=1}^{n} x^{-q}\right]^{-1}. \tag{13}$$

There are two parameters (n, q) included in Eqs. (12) and (13); in the limit of $n \to \infty$, the series in Eq. (13) leads to the Riemann's zeta function

$$\zeta(q) = \sum_{x=1}^{\infty} x^{-q}, \tag{14}$$

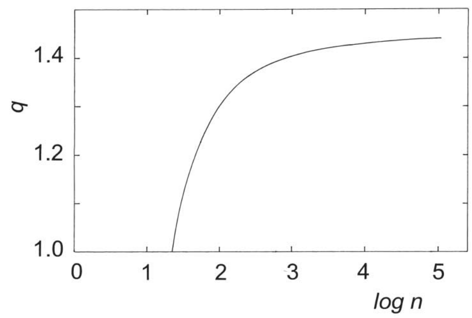

Fig. 3. Relation between two parameters of the inverse power distribution for which the golden condition, $r_F = 1/\phi$, is imposed. It is found numerically that, in the limit of $n \to \infty$, $q \to 1.4404$, i.e., the value of q approaches the fractal dimension of the golden tree. For illustrations of the tree, see Figs. 4 and 5 that follow.

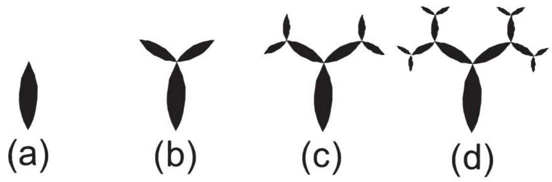

Fig. 4. Illustration of bifurcating *lunes* that will be evolving eventually to the golden tree (Walser, 1996). Here the term, lune, signifies a two-dimensional shape like a convex lens. Note that the ratio of the bifurcating lune-length over the preceding one coincides exactly with $1/\phi$. If the ratio exceeds $1/\phi$, neighboring branches on the tree overlap each other. (a) Unit lune (a trunk). (b) First generation. (c) Second generation. (d) Third generation.

which is valid for $q > 1$. An example of this distribution was initially found in linguistics. For corpora written in English, it was once demonstrated by Zipf (1949) that the relative frequencies, i.e., the statistical probabilities, of words obey Eq. (12) with $q = 1$. Later, this property, which is frequently called Zipf's law or a rank-frequency rule, has been ascertained in other diverse fields of sciences, such as demography, geography, biology, physics, and, more recently, informatics. Among them a case which might possibly be most unexpected was mentioned by Ma (1999) in the context of nuclear physics, where, for arbitrary q, Zipf's law was tested for the charge distribution of nuclear clusters in the liquid gas phase transition. In Fig. 3 the relation is shown between the two parameters of the inverse power distribution for which the golden condition, Eq. (2), is met. It is confirmed numerically that, in the limit of $n \to \infty$, $q \to 1.4404$. It seems to be much interesting and rather surprising to notice that this value of q does coincide with the fractal dimension of the golden tree (Walser, 1996):

$$D = \log 2/\log\phi = 1.4404\cdots.$$

Here the golden tree is defined as the most significant self-similar tree depicted by using the upper boundary of the coefficient of reduction, above which branches of the tree collide each other; it was verified that the boundary value of the reduction coincides exactly with $1/\phi$ (Walser, 1996). Illustrations which explain the method for generating the

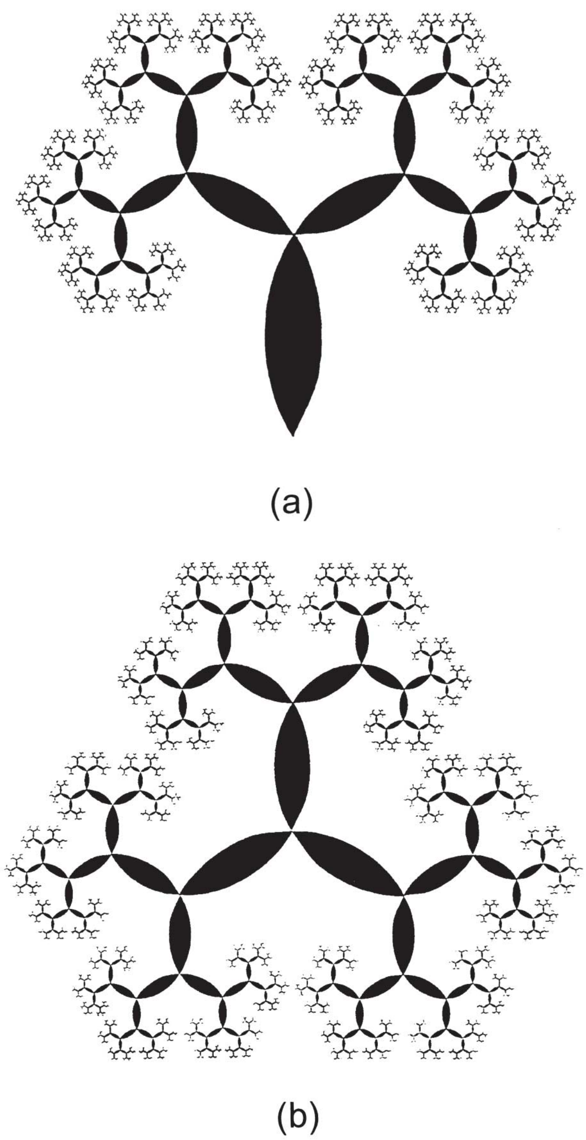

Fig. 5. The golden tree realized through the procedure of Fig. 4 (Walser, 1996). (a) Basic structure. (b) Composition of the three basic elements.

golden tree are given in Figs. 4 and 5. To conclude, the results shown in Fig. 3 suggest that, in the limit of $n \to \infty$, the concept of the golden distribution would have relevance close to that of the self-similar golden tree and, possibly, the property of the zeta function.

4. Examples in Word-Spectral Analysis

Examples of the golden distribution could be found in the word-spectrum analysis of texts in a corpus. Here the term *word spectrum*, which might be borrowed from the terminology of either physics or chemistry, can be defined by the frequency versus the length of words in a text. With these spectra being analyzed, one can obtain a stylistically important quality of texts, because their profile would depend on the writer's personality as well as the language. For all texts written by Shakespeare and by Bacon, Mendenhall (1901) analyzed their spectra and compared those of the two authors. The main conclusion was that the most

Table 3. Hellinger distance, D_{H^2} ($\times 10^{-3}$), between word spectra of English texts. To make a comparison the Poisson distribution with $\lambda = 3.8464$ is also included.

	Poissonian	Sasaki	Turney	Cohn
Poissonian	0	45.5	38.8	40.7
Sasaki	45.5	0	1.02	1.06
Turney	38.8	1.02	0	0.608
Cohn	40.7	1.06	0.608	0

frequent word length (i.e., the mode) of the former texts is four, in contrast to three being the mode of the latter. With this analysis the conjecture that Shakespeare might be none other than Bacon was rejected. Indeed the word-spectrum analysis has allowed one to make a comparative study of the statistical property of texts and has subsequently been applied to a wide range of literary texts (Brinegar, 1963; Williams, 1975). In Fig. 6 the word spectra, i.e., the statistical probabilities of words with length x, are shown of the famous Japanese novel *Botchan* (Work #N1) that was translated into English (Sasaki, 1968; Turney, 1972; Cohn, 2005). Here the length of a word is defined with the number of letters in it. First, it can be seen that the overall profile of the English spectrum bears a resemblance to the Poisson distribution already plotted in Fig. 2(b). For this reason, in Figs. 6(a)–(c) the spectrum of the Poisson distribution that meets Eq. (2) ($\lambda = 3.8464$) is juxtaposed with fine lines. The divergence between two spectra can be quantified through calculation of the Hellinger distance D_{H^2} (≥ 0; equality holds for the perfect similarity)

$$D_{H^2}(p|q) = \sum_{i=1}^{n} \left(p_i^{1/2} - q_i^{1/2} \right)^2$$

with

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} q_i = 1.$$

Here p_i and q_i ($i = 1, 2, 3, \ldots, n$) represent the relative frequencies for the length $x = i$, and n is the maximum word-length. For the results shown in Fig. 6, the distances have been calculated for all combinations of the spectra (Table 3), where the smallest divergence is seen between the spectrum of Turney (Fig. 6(b)) and that of Cohn (Fig. 6(c)). In contrast to this case, the largest divergence can be seen between the two spectra drawn with the bold and the fine lines in Fig. 6(a), namely

$$D_{H^2} = 4.55 \times 10^{-2}.$$

It has been confirmed that this value is comparable to that between the Spanish and the Filipino text of the same novel, which becomes $D_{H^2} = 4.09 \times 10^{-2}$. Here we should remember the linguistic fact that for historical reasons Filipino has a considerable part of the vocabulary in common with that of Spanish. Characteristic values of the word-length data as well as the results of r_F are summarized in Table 4, along with those of the Poisson distribution, where the relative difference between r_F and $1/\phi$ is defined by

$$\delta = \phi \left| r_F - \phi^{-1} \right|.$$

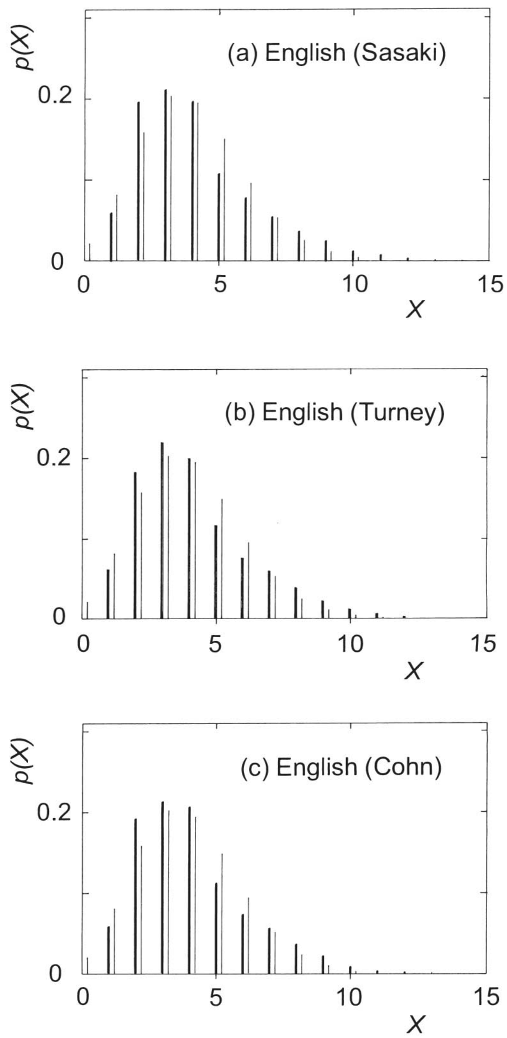

Fig. 6. Word spectrum of the famous Japanese novel *Botchan* (Work #N1) that was translated into English by (a) Sasaki (1968; first printed in 1922), (b) Turney (1972), and (c) Cohn (2005). Irrespective of the translators, a single peak with a positively distorted shape (i.e., positive skewness) is seen. There are steep walls between $x = 1$ and $x = 2$ as well as between $x = 4$ and $x = 5$. To make a comparison, the spectrum of the Poisson distribution with $r_F = 1/\phi$ ($\lambda = 3.8464$) is superimposed with fine lines.

In Table 4 one will notice the interesting fact that the magnitude of r_F for all the English texts are extremely close to $1/\phi$. In particular, it would be surprising that for the text translated by Sasaki the magnitude of δ is no more than 0.03%.

There are two reasons why Work #N1 was chosen. First, it had been translated into exceptionally many languages. Second, for several languages among them, there are different translations being available. Calculation has been made also for non-English texts currently available. In Fig. 7 the word spectra are shown of *Botchan* (#N1) that was translated into (a) Italian (Pastore, 2007), (b) Polish (Murakami, 2009), (c) Hungarian (Judit, 2003), and (d) Indonesian (Haryono, 1992); their characteristic values are listed in Table 5. Here we notice that for the Italian text (Fig. 7(a)) the magnitude of r_F would be close to $1/\phi$. In addition to the five languages listed in Tables 4 and 5, analyses

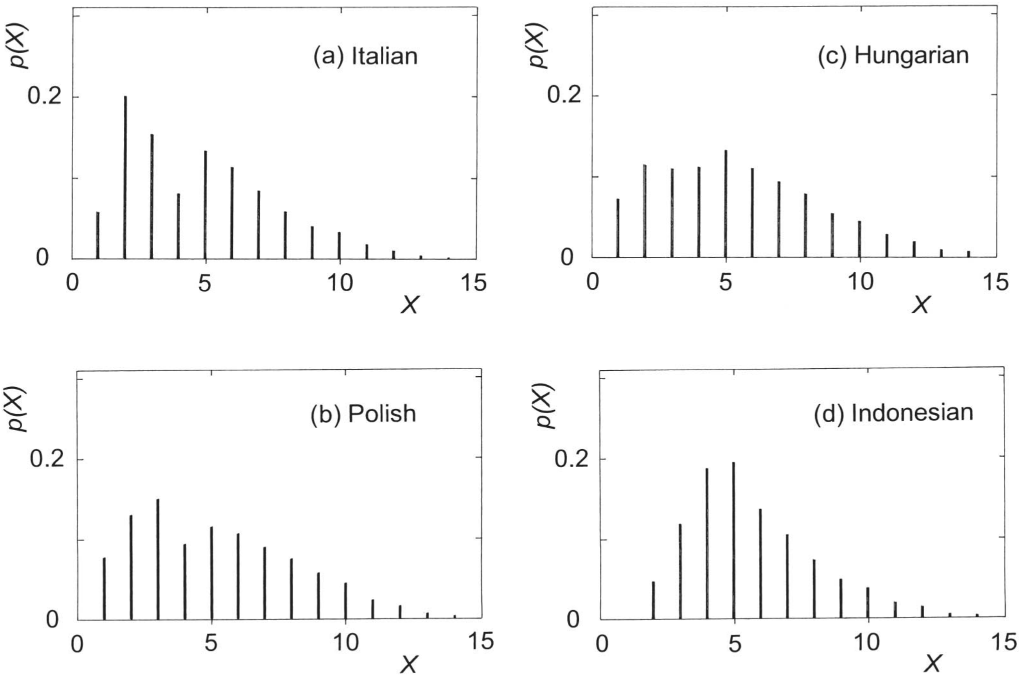

Fig. 7. Word spectrum of *Botchan* that was translated into (a) Italian, (b) Polish, (c) Hungarian, and (d) Indonesian. In sharp contrast to the uneven shape in the spectra of Italian, Polish, and Hungarian, the envelope of the Indonesian spectrum possesses a beautiful profile like a volcano, where no abrupt variation as was seen in the English spectra can be observed.

Table 4. Characteristic values for the word-length data of three translated texts of *Botchan* (#N1) by Soseki Natsume. Original texts written in Japanese are translated into English. Here Σ, L, Me, Mo, R, s, CV, α_3, and α_4 indicate, respectively, the entire number of words, mean, median, mode, range, standard deviation, coefficient of variation, skewness, and kurtosis of the data; the cumulative probability (r_F) of the Fibonacci data ($x = 1, 2, 3, 5, 8, 13, 21, 34, \ldots$) and the relative difference between r_F and $1/\phi$, respectively, are added to the second and the first column from the bottom. In order to make a comparison, those of the Poisson distribution with $\lambda = 3.8464$, for which the requirement of the golden distribution is met, are given.

	Poissonian	English (Sasaki)	English (Turney)	English (Cohn)
Σ	—	54899	53536	56868
L	3.85	4.11	4.10	4.07
Me	4	4	4	4
Mo	3	3	3	3
R	∞	18	35	16
s	1.96	2.25	2.19	2.16
CV	0.510	0.547	0.533	0.532
α_3	0.458	1.13	1.13	1.11
α_4	3.05	4.43	5.20	4.54
r_F	0.6180	0.6179	0.6206	0.6193
δ (%)	0	0.03	0.42	0.21

Table 5. Same as Table 4 but texts (#N1) translated into four non-English languages.

	Italian	Polish	Hungarian	Indonesian
Σ	45206	37619	38242	44986
L	4.83	5.26	5.56	5.74
Me	5	5	5	5
Mo	2	3	5	5
R	32	23	22	20
s	2.78	3.01	3.15	2.48
CV	0.576	0.573	0.566	0.433
α_3	0.779	0.618	0.713	1.02
α_4	3.35	2.86	3.39	4.36
r_F	0.611	0.558	0.519	0.440
δ (%)	1.1	9.6	16	29

are being made for all the texts currently available, specifically, German (55.9–57.2%), French (58.3–58.5%), Spanish (63.4–65.5%), Russian (54.6%), Turkey (45.0%), Filipino (55.9%), and Malay (45.7%) texts, where the numeral in each bracket indicates the latest estimation of r_F. Finally, preliminary results for texts including other languages but translations from other works in Japanese are given in Table 6. Evidently, among them, concerning the value of r_F, there is no case comparable to English. We would conclude that, with respect to the golden distribution of probabilities, the English texts could be regarded as cases passing along a golden mean between two extremes such as, e.g., the French and Spanish ones, for which $r_F < 1/\phi$ and $r_F > 1/\phi$, respectively.

To conclude, a method for generating artificial patterns from the word-length data will be mentioned. In this method, which was termed *spiral mapping* (Hayata, 2003), starting from the center (0, 0), one draws on the Cartesian coordinate a notched spiral with the counterclockwise rotation in accordance with the direction of a sequence. For in-

Table 6. List of preliminary results for other non-English texts. In work numbers, N and K indicate Soseki Natsume and Yasunari Kawabata, respectively.

Language	r_F	Work	Translator
Swedish	0.649	#N4	Emond (1996)
French	0.597	#N3	Cholley (1978)
French	0.597	#N4	Horiguchi (1987)
Portuguese	0.613	#N3	Teixeira (2008)
Italian	0.609	#N2	Origlia (2001)
Rumanian	0.609	#N3	Holca (2010)
Rumanian	0.607	#N4	Suzuki (1984)
Esperanto	0.609	#N5	Nishi (1960)
Serbian	0.609	#N4	Jankovic (2003)
Greek	0.594	#N4	Palantioy (2005)
Polish	0.544	#N3	Melanowicz (1977)
Latvian	0.499	#N4	Paegle (2004)
Hungarian	0.528	#N3	Erdos (1988)
Finnish	0.417	#K1	Kivimies (1968)
Mongolian	0.486	#N3	Dashdabaa (2002)

#N2: *Kusamakura*.
#N3: *Wagahai wa Neko de Aru*.
#N4: *Kokoro*.
#N5: *Londonto*.
#K1: *Yukiguni*.

stance, we consider the opening sentence by Cohn (2005), {*From the time I was a boy the reckless streak that runs in my family has brought me nothing but trouble.*}, which yields a sequence of the word-length data {4, 3, 4, 1, 3, 1, 3, 3, 8, 6, 4, 4, 2, 2, 6, 3, 7, 2, 7, 3, 7}. Applying the spiral mapping technique (see Appendix A) to this sequence, one obtains a chain of transitions

$$\begin{aligned}
&(0,0) \to (4,0) \to (4,3) \to (0,3) \to (0,4) \to (-3,4) \\
&\to (-3,3) \to (-6,3) \to (-6,0) \to (-14,0) \\
&\to (-14,-6) \to (-10,-6) \to (-10,-10) \\
&\to (-8,-10) \to (-8,-12) \to (-2,-12) \\
&\to (-2,-15) \to (5,-15) \to (5,-13) \\
&\to (12,-13) \to (12,-10) \to (19,-10). \qquad (15)
\end{aligned}$$

Here the adjacent points are joined with a segment line. The spiral pattern realized with this path is shown in Fig. 8. Note that, the more the length of a sequence increases, the more the pattern becomes complicated. Previously, it was confirmed that the shape of the notched spiral depends critically both on languages and on translators (Hayata, 2003). Namely, the two-dimensional pattern could be regarded as something like a fingerprint of a text. If necessary, in order to yield a mandala-like pattern with the four-fold rotation symmetry, the original spiral and its seven copies would be superimposed (Hayata, 2004).

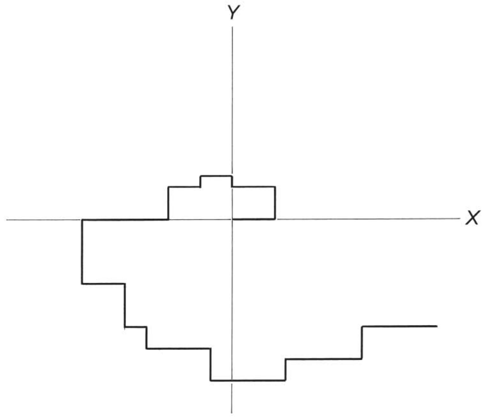

Fig. 8. Illustration that explains the spiral mapping. Sample data are selected from the opening sentence of the English text by Cohn (2005). Continued mapping of the word length data yields a complicated spiral pattern inherent in the language.

5. Conclusion

A novel concept based on the golden ratio ϕ has been presented for discrete probability distributions, where the cumulative probability of the Fibonacci numbers coincides with the reciprocal of ϕ. In addition to classical cases such as the binomial, Poisson, and geometrical distributions, the one that explains the first digit phenomenon as well as the inverse power functions have been chosen. For the latter, in the limit of $n \to \infty$, with n being one of the two parameters of the inverse power distribution, the value of the power has been found to approach the fractal dimension of the golden tree. Finally, examples being close astonishingly to the golden distribution have been shown for the analysis of the word spectra of a novel written in English.

Appendix A. Outlining the Spiral Mapping Method

1) Start from the center (0, 0) and move horizontally along the x-axis with the increment Δx. Here Δx is the length of the initial value of the sequence. For the data of Cohn (2005), $\Delta x = 4$.

2) Subsequently, move upwards with the increment Δy. Here Δy is the length of the second value from the center. For the present data, $\Delta y = 3$.

3) For the point being in the first section ($x > 0$, $y > 0$), move backward along the horizontal direction (i.e., set $\Delta x < 0$, being the decrement) and upwards along the vertical direction (set $\Delta y > 0$) until the point attains into the second section ($x < 0$, $y > 0$).

4) For the point being in the second section, move backward along the horizontal direction (set $\Delta x < 0$) and downward along the vertical direction (set $\Delta y < 0$) until the point attains into the third section ($x < 0$, $y < 0$).

5) For the point being in the third section, move forward along the horizontal direction (set $\Delta x > 0$) and downward along the vertical direction (set $\Delta y < 0$) until the point attains into the fourth section ($x > 0$, $y < 0$).

6) For the point being in the fourth section, move forward along the horizontal direction (set $\Delta x > 0$) and upwards along the vertical direction (set $\Delta y > 0$) until the point returns to the first section ($x > 0$, $y > 0$).

7) Return to Step 3 and repeat this procedure until the point attains the terminal of the data.

References

Benford, F. (1938) The law of anomalous numbers, *Proc. Amer. Phil. Soc.*, **78**, 551–572.

Brinegar, C. S. (1963) Mark Twain and the Quintus Curtius Snodgrass Letters: a statistical test of authorship, *J. Amer. Statist. Assoc.*, **58**, 85–96.

Cohn, J. (2005) *Botchan, Natsume Soseki*, Kodansha International, Tokyo.

Haryono (1992) *Natsume Soseki, Botchan*, Yayasan Karti Sarana Dalam Kerja Sama Dengan, Universitas Katolik Soegijapranata, Semarang.

Hasegawa, M. and Tanemura, M. (1976) On the pattern of space division by territories, *Annual of the Institute of Statistical Mathematics*, **28B**, 509–519.

Hayata, K. (2003) A comparative study of translated texts through the analysis of their word spectra: Application to a text in *Botchan*, *Forma*, **18**, 97–117.

Hayata, K. (2004) Generation of mandala patterns from texts that include sutras, poems and strings of words: Method and examples, *Forma*, **19**, 233–264.

Hill, T. (1998) The first digit phenomenon, *American Scientist*, **86**, 358–363.

Judit, V. (2003) *Nacume Szoszeki, A Kolyok*, Balassi Kiado, Budapest.

Livio, M. (2002) *The Golden Ratio*, Broadway Books, New York.

Ma, Y. G. (1999) Zipf's law in the liquid gas phase transition of nuclei, *Eur. Phys. J. A*, **6**, 367–371.

Mandelbrot, B. (1954) Structure formelle des textes et communication, *Word*, **10**, 1–27.

Mendenhall, T. C. (1901) A mechanical solution of a literary problem, *Popular Sci. Monthly*, **60**, 96–105.

Murakami, B. (2009) *Natsume Soseki, Panicz*, Inter Media, Sosnowiec.

Pastore, A. (2007) *Natsume Soseki, Il Signorino*, Neri Pozza Editore, Vicenza.

Penrose, R. (1979) Pentaplexity: a class of non-periodic tilings of the plane, *Math. Intelligencer*, **2**, 32–37.

Sasaki, U. (1968) *Botchan by Soseki Natsume*, Charles E. Tuttle, Tokyo.

Schroeder, M. (1991) *Fractal, Chaos, Power Laws*, W. H. Freeman, New York.

Takaki, R. (2003) *The Encyclopedia of Form*, Maruzen, Tokyo (in Japanese).

Turney, A. (1972) *Botchan, Natsume Soseki*, Kodansha International, Tokyo.

Washington, L. C. (1981) Benford's law for Fibonacci and Lucas numbers, *Fibonacci Quarterly*, **19**, 175–177.

Walser, H. (1996) *Der Goldene Schnitt*, B. G. Teubner Verlag, Leipzig. For translation into English, see Walser, H. (2001) *The Golden Section*, Mathematical Association of America, Washington, D.C.

Williams, C. B. (1975) Mendenhall's studies of word-length distribution in the works of Shakespeare and Bacon, *Biometrika*, **62**(1), 207–212.

Zipf, G. K. (1949) *Human Behavior and the Principle of Least Effort*, Addison-Wesley Press, Cambridge, MA.