Nearest-neighbor-spacing Distribution of People Sitting along the Bank of Kyoto's Kamogawa River

Ryo Yasufuku[1,2], Yamato Ueno[1,3] and Syuji Miyazaki[1,4,*]

[1]*ELCAS, Kyoto University, Kyoto 606-8501, Japan*
[2]*Hyogo Prefectural Nagata Upper Secondary School, 2-5 Ikeda-tanimachi, Nagata-ku, Kobe-shi, Hyogo 653-0821, Japan*
[3]*Kouyou Junior & Senior High School, 127 Ohta, Wakayama 640-8323, Japan*
[4]*Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan*
*E-mail address: miyazaki.syuji.8m@kyoto-u.ac.jp

Nearest-neighbor-spacing distributions of people sitting along the west bank of the Kamogawa River flowing through the central part of Kyoto City are derived, based on a dual partitioning process with a fixed ratio. The theoretical asymmetric bell-shaped distribution is compared with previous measurements. This was performed as an advanced study of high school students that does not follow the curriculum guidelines of Japanese high schools.
Key words: Binomial coefficient, Stirling's formula, Gamma function, Level repulsion

1. Introduction

In the framework of the Experience-based Learning Course for Active Students, abbreviated as ELCAS, at Kyoto University, two high-school students, Yamato Ueno and Ryo Yasufuku, under the supervision of Syuji Miyazaki, performed advanced studies that do not follow the curriculum guidelines of high schools set by the Ministry of Education, Culture, Sports, Science and Technology. Two of the other advanced studies have already been reported (Maeda-Miyazaki, 2019; Ueno *et al*., 2022). A related topic was also considered (Miyazaki, 2019).

A person or groups of people are said to keep equal distance to one another regardless of the number of people and sit along the bank of the Kamogawa River in the central part of Kyoto City, as shown in Fig. 1. This phenomenon is sometimes called *the principle of equal distance at the Kamogawa River*. Some measurements of the distance were reported (Morita, 1987; Saitoh *et al*., 1994). In this article, we derived a theoretical distribution function and compared it with the above measurements.

The principle of equal distance at the Kamogawa River is a kind of space-filling problem, which is an important problem in the research field of *Science on Form*. It is recommended for lay experts and high school students to weigh the equally spaced principle with other similar phenomena, for example, birds perching on an electric wire or level repulsion in the research field of quantum chaos (Haake, 2010).

In Sec. 2, we derive the distribution function of the distance based on the binomial expansion. In Sec. 3, we compared our distribution function with previous measurements. The final section is devoted to concluding remarks.

Fig. 1. The west bank of the Kamogawa River seen from the Sanjo Ohashi Bridge.

2. Derivation of the Distribution Function Based on a Dual Partitioning Process with a Fixed Ratio

A person who wants to sit on the bank decides his/her place not right beside an entirely unrelated person but nearly at the midpoint between already seated people with a longer distance L, where "a person" can be replaced by a couple or a group. In such a case, the person is regarded as a barycenter of the couple or group. Since it is difficult to find the exact midpoint, the sitting place divides L at a ratio of $r : 1 - r$ or $1 - r : r$, where r is nearly equal to 1/2. We consider without loss of generality that r satisfies $1/2 < r < 1$. The ratio r is different from person to person. For the sake of simplicity, r is regarded as its average and fixed in the following. We consider the simple dual partitioning process with a fixed ratio. The initial interval with a length of L is divided into two intervals: rL and $(1 - r)L$, which are divided into four intervals: $r^2 L$, $r(1 - r)L$, $r(1 - r)L$, and $(1 - r)^2 L$. The partitioning is repeated N times and

2^N intervals are created. This partitioning is expressed by a binomial expansion

$$
1 = 1^N = (r + 1 - r)^N
$$
$$
= \sum_{n=0}^{N} \binom{N}{n} r^n (1 - r)^{N-n}
$$
$$
= \sum_{n=0}^{N} 2^{-N} \binom{N}{n} \cdot 2^N r^n (1 - r)^{N-n}, \qquad (1)
$$

which indicates that the length $r^n (1 - r)^{N-n}$ normalized by the average 2^{-N}, namely, $2^N r^n (1 - r)^{N-n}$, appears at a frequency of $2^{-N} \binom{N}{n}$ and the normalization condition of the frequency is satisfied.

The variance σ^2 of the normalized distance is given by

$$
\sigma^2 = \sum_{n=0}^{N} 2^{-N} \binom{N}{n} \cdot (2^N r^n (1 - r)^{N-n} - 1)^2
$$
$$
= \sum_{n=0}^{N} 2^{-N} \binom{N}{n} \left(2^{2N} (r^2)^n [(1 - r)^2]^{N-n} \right.
$$
$$
\left. -2 \cdot 2^N r^n (1 - r)^{N-n} + 1 \right)
$$
$$
= \sum_{n=0}^{N} \binom{N}{n} \left(2^N (r^2)^n [(1 - r)^2]^{N-n} \right.
$$
$$
\left. -2 \cdot r^n (1 - r)^{N-n} + 2^{-N} \right)
$$
$$
= \sum_{n=0}^{N} \binom{N}{n} 2^N (r^2)^n [(1 - r)^2]^{N-n} - 2 + 1
$$
$$
= 2^N [r^2 + (1 - r)^2]^N - 1, \qquad (2)
$$

where Eq. (1) and the relationship $\sum_{n=0}^{N} \binom{N}{n} = 2^N$ are used. Let us define the positive deviation Δ from the equal partition $1/2$ by $r = \dfrac{1 + \Delta}{2}$. Then, we have

$$
\sigma^2 = (1 + \Delta^2)^N - 1. \qquad (3)
$$

and its inverse function

$$
\Delta = \sqrt{(1 + \sigma^2)^{1/N} - 1}. \qquad (4)
$$

Since the number of the intervals M is not always in the form of 2^N in a real measurement, we define the real number X as $2^X = M$ and replace N by X. Equation (4) tells us the value of $r = \dfrac{1 + \Delta}{2}$ from σ^2 and $X = \dfrac{\log M}{\log 2}$ obtained from the measured data, and enables us to compare the measured distribution with the theoretical one. For small Δ, we can expand Eq. (2) as $\sigma^2 \sim X\Delta^2$.

For the sake of comparison with histograms of previous measurements, the binomial coefficient $\binom{N}{n}$ is replaced by the continuous function $\dfrac{\Gamma(X + 1)}{\Gamma(x + 1)\Gamma(X - x + 1)}$, where the real number x corresponds to n satisfying $0 \leq x \leq X$ and Γ denotes the gamma function. The normalized length of the interval $l(x)$ as a function of x is given by

$$
l(x) = 2^X r^x (1 - r)^{X-x}. \qquad (5)
$$

Hence, the distribution function $P(l)$ of the normalized length of the interval $l(x)$ and the approximate expression $P_s(l)$ using the Stirling formula are expressed by

$$
P(l) = \frac{C\Gamma(X + 1)}{2^X \Gamma(x(l) + 1)\Gamma(X - x(l) + 1)}, \qquad (6)
$$
$$
P_s(l) = C_s e^{-[x(l)\log x(l) + (X - x(l))\log(X - x(l))]}, \qquad (7)
$$

where $x(l)$ is an inverse function of Eq. (5) and given by

$$
x(l) = \frac{\log l - X(\log 2 + \log(1 - r))}{\log r - \log(1 - r)}. \qquad (8)
$$

x and l satisfy $0 \leq x \leq X$ and $[2(1 - r)]^X \leq l \leq (2r)^X$, respectively. The constants C and C_s are determined by the normalization condition. Although X in a real measurement might be not large enough to apply the Stirling formula, we will compare $P(l)$ and $P_s(l)$ in the next section.

If we could find the exact midpoint, the nearest neighbor spacing distribution should be a delta function, which corresponds to the case of $r = 1/2$. Continuous distributions originate from a human error in finding the midpoint. The distribution reflects and characterizes its statistics.

3. Comparison with Previous Measurements

Visual reading of Morita's data of the distances between people sitting on the west bank of the Kamogawa River between Sanjo and Shijo Ohashi Bridges at nine PM on June 16, 1986 (Morita, 1987) yields the histogram in the upper panel of Fig. 2, where the distance normalized by the average 1.29 m on the abscissa and its probability density on the ordinate are plotted. The distance between the two bridges is nearly 556 m, the number of intervals of people sitting is 89, leading to $X = \log 89 / \log 2 \simeq 6.48$. The variance σ^2 of the normalized distance is calculated as $\sigma^2 \simeq 0.22$, leading to $\Delta \simeq 0.18$ and $r \simeq 0.59$. The theoretical distribution functions of Eq. (6) and Eq. (7) with these parameters are plotted with solid and dashed lines in the upper panel of Fig. 2, respectively.

A histogram of the most congested time at seven PM on the same date (Morita, 1987) is shown in the lower panel of Fig. 2, where the distance normalized by the average 0.94 m on the abscissa and its probability density on the ordinate are plotted. The number of intervals of people sitting is equal to 92, leading to $X = \log 92 / \log 2 \simeq 6.52$. The variance σ^2 of the normalized distance is calculated as $\sigma^2 \simeq 0.021$, leading to $\Delta \simeq 0.056$ and $r \simeq 0.53$. The theoretical distribution functions of Eq. (6) and Eq. (7) with these parameters are plotted with solid and dashed lines in the lower panel of Fig. 2, respectively.

Saitoh *et al.* measured the distances to the right neighbor l_R and the left neighbor l_L separately between Sanjo and Shijo Ohashi Bridges (Saitoh *et al.*, 1994). Using a pointing and coordinate display of the software Acrobat applied to the scanned graph, we obtained data on the distances l_R and l_L. Then, we directly calculated the ratio of the maximum of the right and left distances $\max(l_R, l_L)$ to its sum $l_R + l_L$. The average $r_{direct} = \langle \max(l_R, l_L)/(l_R + l_L) \rangle$ is an estimation of the value of r. The value of r_{direct} is 0.58. We use data showing that $\min(l_R, l_L)$ is less than 5 m only.
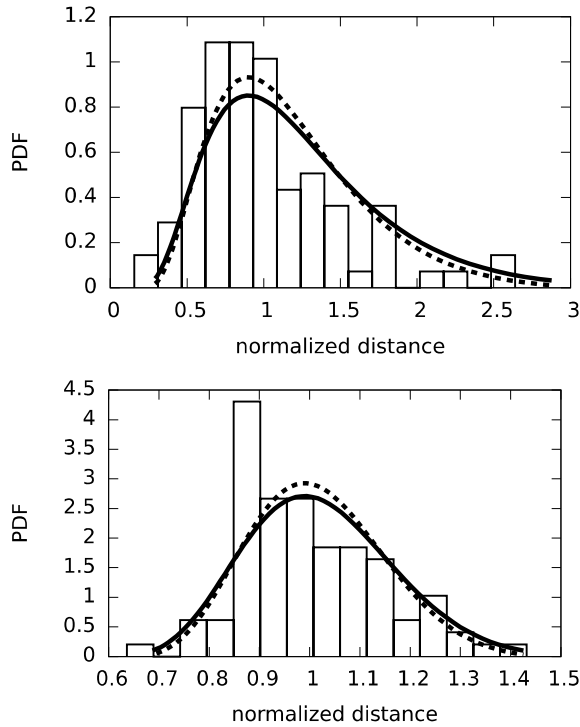
Fig. 2. Histograms of Morita's data of the distances between people sitting on the west bank of the Kamogawa River between Sanjo and Shijo Ohashi Bridges with theoretical distributions Eq. (6) (solid line) and Eq. (7) (dashed line) at nine PM on June 16, 1986 (upper) and seven PM on the same date (lower).
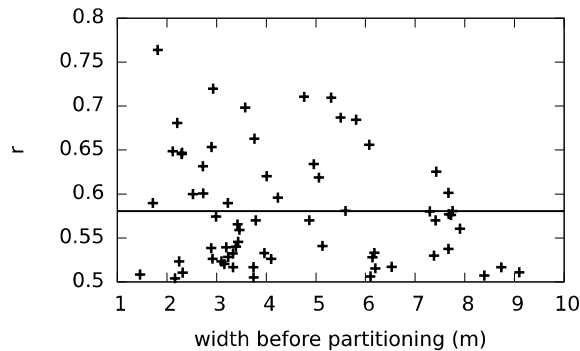

Fig. 3. The ratio $r = \max(l_R, l_L)/(l_R + l_L)$ dependence on the width of the interval between which a person is about to sit (symbol +) with the average $r_{direct} = \langle r \rangle$ (solid line) based on Saitoh et al.'s data of the distances between people sitting on the west bank of the Kamogawa River between Sanjo and Shijo Ohashi Bridges at five PM on October 10, 1993.

Figure 3 shows how the ratio $r = \max(l_R, l_L)/(l_R + l_L)$ depends on the width of the interval between which a person is about to sit (symbol +) in units of meters with the average r_{direct} (solid line). Since the dependence largely fluctuates around the average, the assumption that r is constant cannot be fully justified from these data in this stage. The number of intervals of people sitting is equal to 72, leading to $X = \log 72/\log 2 \simeq 6.17$. The variance σ^2 of the normalized distance is calculated as $\sigma^2 \simeq 0.25$, leading to $\Delta \simeq 0.19$ and $r \simeq 0.60$. Since this estimation of r from the distribution of the distance is nearly equal to $r_{direct} = 0.58$, the introduction of the fixed r turns out to be a good approx-
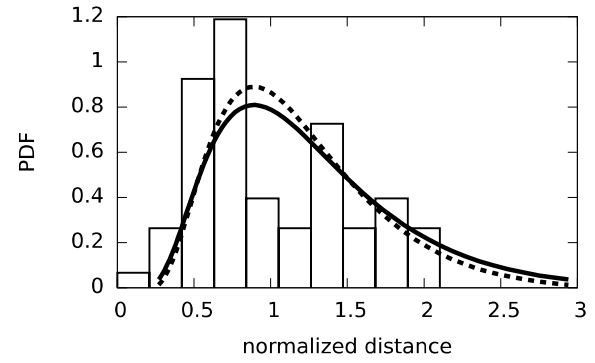

Fig. 4. Histograms of Saitoh et al.'s data of the distances between people sitting on the west bank of the Kamogawa River between Sanjo and Shijo Ohashi Bridges with theoretical distributions Eq. (6) (solid line) and Eq. (7) (dashed line) at five PM on October 10, 1993.

imation in the sense that $r_{direct} = 0.58$ is close to $r \simeq 0.60$ obtained from the values of X and σ^2 of the raw data consisting of individual values of l_R and l_L. The theoretical distribution functions of Eq. (6) and Eq. (7) with these parameters are plotted with solid and dashed lines in the upper panel of Fig. 4, respectively.

4. Concluding Remarks

Starting from a binormal coefficient included in high school mathematics, we derived the nearest-neighbor spacing distribution of the people sitting on the bank of the Kamogawa River with the help of the Gamma function and Stirling's formula, beyond the curriculum guidelines of the high school.

As shown in Figs. 2 and 4, we observed a systematic discrepancy whereby the maximum of the real distribution is located at a smaller value of the normalized distance than our theoretical estimation, and that the right side of the real distribution decayed faster than our theoretical estimation. Our assumption of a fixed r might have caused this discrepancy. In other words, the question is whether one can determine the midpoint of a five-meter section and that of a twenty-meter section with the same precision. Shortage of data give us less conclusive results. If the constant-r assumption causes the discrepancy, Eq. (1) could be replaced with an expression in the form of

$$1 = \prod_{n=1}^{N} (r_n + 1 - r_n),$$

where r_n depends on the depth n of the dual partitioning process. The expression of r_n is still unclear. Let us consider two simple cases, where the common parameters are given by $N = 2$, $r_1 = r$, and $r_2 = ar$ and $r = 3/5$. The first and second cases are specified by $a = 1$ and $a = 6/5$, respectively. The lengths of the four intervals are $r_1 r_2, r_1(1 - r_2), r_2(1 - r_1)$ and $(1 - r_1)(1 - r_2)$, whose values are given by 0.36, 0.24, 0.24, and 0.16 for the first case; and 0.432, 0.168, 0.288, and 0.112 for the second case, respectively. We can construct a histogram (distribution of the length), where the width of the bin is equal to 0.1. The probability density (height of the bar) in the range $[0, 0.1)$, $[0.1, 0.2)$, $[0.2, 0.3)$, $[0.3, 0.4)$, and $[0.4, 0.5)$ is equal to
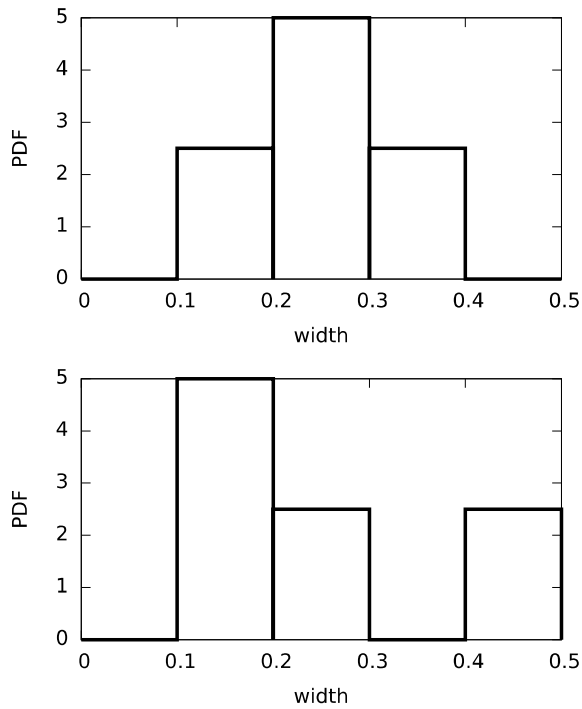
Fig. 5. The probability density of the width between people sitting for two simple cases with $N = 2$, $r_1 = r$, and $r_2 = ar$ and $r = 3/5$. (Upper) $a = 1$. (Lower) $a = 6/5$.

0.0, 2.5, 5.0, 2.5, and 0.0 for the first case; and 0.0, 5.0, 2.5, 0.0, and 2.5 for the second case, respectively, as shown in Fig. 5. The normalization condition that the area of the histogram is equal to unity is satisfied. In this way, the n dependence of r may change the function form of the nearest neighbor spacing distribution. We can also treat r as a stochastic parameter, as shown in Fig. 3, which is another idea for improvement.

Detailed measurements of the nearest neighbor spacing of people sitting on the bank are difficult. Promising objects that can be used for measurement are long benches provided by railroad stations, for example, at Tofukuji Station of the

Fig. 6. The long bench at Tofukuji Railroad Station.

Keihan Electric Railway, which operates a route partially along the Kamogawa River, as shown in Fig. 6.

References

Haake, F. (2010) Quantum Signatures of Chaos (Springer Series in Synergetics) 3rd ed., Springer, ISBN-13:978-3642054273.

Maeda, Y. and Miyazaki, S. (2019) Spectroscopy of Hydrogen, Helium, Neon and Mercury Low-Pressure Discharge and a Barcode-like Periodic Table Based on Energy Levels, *Forma* **34**, 1–3.

Miyazaki, S. (2020) A Pale View of Shizuo Ishiguro's Research –Abiki, the 1953 North Sea Flood, and Beyond– *Forma* **35**, 37–39.

Morita, T. (1987) 5373 On Territorial Behavior of the People Enjoying the Evening on the Bank of the Kamogawa River, *Summaries of technical papers of Annual Meeting Architectural Institute of Japan. Architectural planning and design rural planning* (written in Japanese) 745–746.

Saitoh, H., Hashizume, S. and Fuzoku Kenkyu Network (1994) Kyoto Ren'ai Kuukan Gendai Kappuru Kou, (written in Japanese) Gakugei Shuppansha, ISBN 4-7615-1145-1, Fig. 5, 71.

Ueno, Y., Yasufuku, R. and Miyazaki, S. (2022) Spectroscopy of plasma produced in a kitchen microwave oven, *Forma* **37**, 1–3.