Randomness at large numbers: Experimental proof in coin toss and prime number

Cite as: AIP Advances 9, 125318 (2019); https://doi.org/10.1063/1.5133773
Submitted: 26 October 2019. Accepted: 04 December 2019. Published Online: 20 December 2019

Yeseul Kim, and Byung Mook Weon

© 2019 Author(s).
Randomness is a central concept to statistics and physics. Here, we conduct experimental investigations with a coin toss and prime number to show experimental evidence that tossing coins and finding last digits of prime numbers are statistically identical with respect to equally likely outcomes. The range of frequency of an outcome ($R$) is normalized by the total number of repetitions ($N$) to be the range of relative frequency ($R/N$). We find that $R/N$ has a power-law scaling $R/N \sim N^{-0.6}$, which is valid for large numbers in both cases of a coin toss and the last digit of a prime number. This analysis, indicating $R/N \to 0$ at $N \to \infty$, confirms that randomness and equally likely outcomes can be valid for large numbers.
II. ANALYSIS METHOD

There are many examples for equally likely outcomes; representatively, coin tossing is believed to occur with a probability of 50% between heads and tails. For repeated experiments with the same sample, if its frequency between expected outcomes is equal, one can say the expected outcome of the sample is random. Here, we suggest a simple way to define the randomness concerning equally likely outcomes for large numbers.

The frequency of each outcome (n_i) can vary complicatedly according to experiments and conditions. The relative frequency of an outcome (f_i) is calculated by dividing n_i by the total number of repetitions (N or equally the size of the sample). The range of frequency (R) is defined as the difference between the maximum frequency (n_{i,max}) and the minimum frequency (n_{i,min}), consequently described as R = (n_{i,max} - n_{i,min}). In statistics, it is well known that the range of frequency (R) tends to be larger for a larger size of the sample (N).\(^{10,21}\) This tendency can be described by a power-law scaling R \sim N^\alpha, where 0 < \alpha < 1. Such a power-law scaling commonly appears in statistics and physics.\(^ {22,23} \) Additionally, the range of relative frequency (R/N) between equally likely outcomes is defined as R/N = (f_{i,max} - f_{i,min}), which is equivalent to (n_{i,max} - n_{i,min})/N. From R/N \sim N^\alpha, R/N should have a simple power-law relation R/N \sim N^\beta, where \beta = \alpha - 1 \ (note that \beta < 0 because \alpha < 1). The statistical expectation of R/N \sim N^\beta (\beta < 0) implies that the frequency of each outcome should become equal (because R/N \to 0) as the total number of repetitions increases (N \to \infty). Consequently, the condition R/N \to 0 at N \to \infty explains why randomness is valid only for large numbers, which is known as the law of large numbers in probability theory. In this study, we would like to assume the \beta exponent to be approximately −0.6 in coin tosses (with two equal outcomes) and last digits of prime numbers (with four equal outcomes) with respect to equally likely outcomes.

III. RESULTS AND DISCUSSION

First, we conducted an analysis with coin tossing, as shown in Fig. 1. To rule out physical and mechanical aspects of tossed coins, we used an online virtual coin toss simulation application (http://www.virtualcointoss.com) with an ideal coin of zero thickness, where there is no bias between heads and tails, ensuring equal probabilities for heads and tails. Our experiments with perfectly thin coins enable us to consider only the statistical features of coin-tossing problems. We carried out five experiments separately. The frequency of heads (n_{H,i}) or tails (n_{T,i}) for each experiment was recorded with the number of tosses (N) (equally the size of the sample). The relative frequencies (f_{H,i} = n_{H,i}/N or f_{T,i} = n_{T,i}/N for heads or tails), the range of frequency \([R = (n_{i,max} - n_{i,min})], where i = \text{heads or tails}], and the range of relative frequency \([R/N = (n_{i,max} - n_{i,min})/N]) were summarized in Tables S1–S5 of the supplementary material. Each of experiments was illustrated with a different color.

In turn, we examined the last digits of prime numbers, as illustrated in Fig. 2. As well known, all prime numbers except 2 and 5 should end in a last digit (1, 3, 7, or 9), and the last digits are expected to be random when numbers are large enough, which suggests that the frequency of the four last digits should be equal, i.e., prob(j) = 25%. For the prime numbers in base 10 for integers up to 10^7 (where a total of 664 579 prime numbers exist), we calculated the frequency of each last digit (n_j, where j = 1, 3, 7, or 9), the range of frequency \([R = (n_{j,max} - n_{j,min})]), and the range of relative frequency \([R/N = (n_{j,max} - n_{j,min})/N]), as summarized in Table S6 of the supplementary material. Here, the number of prime numbers (N) (including 2 and 5) is equivalent to the size of the sample.

Statistical uncertainties were checked for coin tossing experiments in the plot of R/N with N [Fig. 3(a)] by measuring one standard deviation from five experiments (from five data points for R/N for a given N). However, the prime numbers and the range of relative frequency were completely deterministic for integer numbers up to 10^7, which implies no errors in the plot of R/N with N [Fig. 3(b)].

For coin tosses, the relative frequency of heads for five experiments up to 10^5 repetitions differently varies for small numbers but converges at the expected value \([\text{prob}(i) = 50\%], where i = \text{heads or tails}] for large numbers [toward the dashed lines, as shown in Fig. 1(b)], which supports the fact that coin tossing is a problem of equally likely outcomes. The well-known statistical feature that the range \((R)) tends to be larger for a larger size of the sample \((N)) suggests a power-law scaling \(R \sim N^\beta \ (0 < \alpha < 1). On this basis, we expected a simple relation for the range of relative frequency for...
heads and tails given by \( R/N = (n_{\text{max}} - n_{\text{min}})/N \) as \( R/N \sim N^\beta \), where \( \beta = \alpha - 1 < 0 \). As illustrated in Fig. 3(a), \( R/N = 3.1461N^{-0.6237} \) for the trend line; we obtained \( \beta = -0.6237 \) (the standard error = ±0.0272) for five coin tossing experiments (error bars resulting from one standard deviation). This result clearly supports the validity of \( \text{prob}(i) = 50\% \) by \( R/N \to 0 \) at \( N \to \infty \), indicating statistical evidence of randomness for coin tosses at large numbers, which is consistent with a common belief about coin tossing.

For last digits of prime numbers, the relative frequency of last digits finally approaches to the ultimately expected value \( \text{prob}(j) = 25\% \) for one of the four last digits \( j = 1, 3, 7, \) or \( 9 \) [toward the dashed lines, as shown in Fig. 2(b)]. The range of frequency among last digits increases with the total number of primes as a power-law scaling of \( R \sim N^\alpha \) with \( \alpha \approx 0.4 \), which is similar to the case of coin tossing. The range of relative frequency among last digits given by \( R/N = (n_{\text{max}} - n_{\text{min}})/N \) shows \( R/N \sim N^\beta \), where \( \beta = -0.5832 \) (the standard error = ±0.0094) for last digits [Fig. 3(b) shows \( R/N = 0.5294N^{-0.5832} \) for the trend line], which is identical to the case of coin tossing. This result supports the validity of \( \text{prob}(j) = 25\% \) for one of the four last digits by \( R/N \to 0 \) at \( N \to \infty \), indicating that the last digit of primes would occur with the same frequency for large numbers.

The above two examples of equally likely outcomes lead to the same result: as the size of the sample \( (N) \) increases, the range of relative frequency \( (R/N) \) decreases, following the power law scaling \( R/N \sim N^\beta \). Here, the \( \beta \) exponents were found to be approximately \(-0.6\) for coin tossing experiments with \( \text{prob}(i) = 50\% \) for two equal outcomes and last digits of primes with \( \text{prob}(j) = 25\% \) for four equal outcomes. Interestingly, there is a slight difference in the prefactor of the power-law scaling. This result \( R/N \sim N^{-0.6} \) indicates \( R/N \to 0 \) at \( N \to \infty \) and confirms that randomness can be valid for large numbers for both cases, supporting that tossing coins and finding last digits of prime numbers are statistically identical with respect to equally likely outcomes.
IV. CONCLUSION

In conclusion, we introduced a simple expression of randomness for large numbers. From statistical analyses of coin tosses and last digits of primes, we showed that the range of relative frequency between equally likely outcomes ($R/N$) decreases as the total repetition number ($N$) increases. A power-law scaling for $R/N$ vs $N$ in both cases was found as $R/N \sim N^{-\beta}$ ($\beta \approx -0.6$), implying that the frequency of each outcome becomes equal ($R/N \to 0$) as the total number of repetitions increases ($N \to \infty$). The condition $R/N \to 0$ at $N \to \infty$ explains why randomness is valid only for large numbers. This result experimentally confirms that finding last digits of primes is intrinsically identical to tossing coins in statistics: the problems of equally likely outcomes are the same in both cases. Finally, our finding of the power-law relation between the range of relative frequency among equally likely outcomes and the total number of repetitions would be significant to understand the validity of randomness for large numbers (known as the law of large numbers), which would be important in statistics, physics, and mathematics.

SUPPLEMENTARY MATERIAL

See the supplementary material for Tables S1–S6 of coin tossing experiments and last digits of prime numbers.

ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant Nos. NRF-2016R1D1A1B01007133 and 2019R1A6A1A0303215).

REFERENCES

1. J. Ford, Phys. Today 36(4), 40–47 (1983).
2. P. Diaconis, S. Holmes, and R. Montgomery, SIAM Rev. 49, 211–235 (2007).
3. J. Strzalko et al., Phys. Rep. 469, 59–92 (2008).
4. L. Mahadevan and E. H. Yong, Phys. Today 64(7), 66–67 (2011).
5. M. Falcioni, L. Palatella, S. Pigolotti, and A. Vulpiani, Phys. Rev. E 72, 016220 (2005).
6. T. E. Murphy and R. Roy, Nat. Photonics 2, 714–715 (2008).
7. C. Ferrie and J. Combes, Phys. Rev. Lett. 113, 120404 (2014).
8. T. O’Hagan, Significance 1, 132–133 (2004).
9. B. Hayes, Am. Sci. 99, 282–287 (2011).
10. V. Z. Vulovic and R. E. Prange, Phys. Rev. A 33, 576–582 (2008).
11. J. Nagler and P. Richter, Phys. Rev. E 78, 036207 (2008).
12. J. Strzalko, J. Grabski, A. Stefanski, and T. Kapitaniak, Int. J. Bifurcation Chaos 20, 1175 (2010).
13. M. Kapitaniak, J. Strzalko, J. Grabski, and T. Kapitaniak, Chaos 22, 047504 (2012).
14. M. Le Bellac, Prog. Biophys. Mol. Biol. 110, 97–105 (2012).
15. A. Granville and G. Martin, Am. Math. Mon. 113, 1–33 (2006).
16. J. C. Pain, Phys. Rev. E 77, 021102 (2008).
17. B. Luque and L. Lacasa, Proc. R. Soc., Ser. A 465, 2197–2216 (2009).
18. L. Shao and B.-Q. Ma, Physica A 389, 3109–3116 (2010).
19. T. Tao, in An Invitation to Mathematics, edited by D. Schleicher and M. Lackmann (Springer-Verlag, Berlin, Heidelberg, 2011).
20. S. P. Hozo, B. Djulbegovic, and I. Hozo, BMC Med. Res. Methodol. 5, 13 (2005).
21. X. Wan, W. Wang, J. Liu, and T. Tong, BMC Med. Res. Methodol. 14, 135 (2014).
22. M. E. J. Newman, Contemp. Phys. 46, 323–351 (2005).
23. S. A. Frank, J. Evol. Biol. 22, 1563–1585 (2009).