Received: 2 May 2022 | Revised: 11 September 2022 | Accepted: 11 September 2022

DOI: 10.1111/test.12321

ORIGINAL ARTICLE

WILEY

Lessons from the famous 17th-century paradox of the Chevalier de Méré

José Daniel López-Barrientos[1] | Eliud Silva[1] | Enrique Lemus-Rodríguez[2]

[1]Facultad de Ciencias Actuariales, Universidad Anáhuac México, Huixquilucan, Mexico

[2]Facultad de Ciencias Físico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico

Correspondence
José Daniel López-Barrientos, Facultad de Ciencias Actuariales, Universidad Anáhuac México, Edo. de México, Huixquilucan 52786, México.
Email: daniel.lopez@anahuac.mx

Abstract

We take advantage of a combinatorial misconception and the famous paradox of the Chevalier de Méré to present the multiplication rule for independent events; the principle of inclusion and exclusion in the presence of disjoint events; the median of a discrete-type random variable, and a confidence interval for a large sample. Moreover, we pay tribute to our original bibliographic sources by providing two computational tools to facilitate the students' insights on these topics.

KEYWORDS

confidence intervals, median, multiplication rule, principle of inclusion and exclusion

1 | INTRODUCTION

Independence between events and disjoint events are a couple of difficult concepts for students to understand and distinguish. Indeed, even when their respective definitions and early illustrations are straightforward enough, it is not difficult to find examples where pupils take one for the other.

This article is about faulty reasoning in probability, and how to use it to learn about three important results in this discipline. We survey two classic examples to exhibit the use of the multiplication rule (Reference [1], section 1.4), and when the Principle of inclusion and exclusion (Reference [2], theorem 3.8) can be simplified to a simple sum without substractions. One of the

paradox, posed by Joseph Bertrand in Reference [3], p. 2 (see also Reference [4]). In this puzzle, there are three boxes: a box containing two gold coins, a box with two silver coins; and a box with one of each. After randomly choosing a box and withdrawing one coin at random that turns out to be gold, the question is what is the probability that the other coin is gold as well. The most celebrated puzzle related to this example is, of course, Monty Hall's problem (see, References [5], section 1.5, [6], p. 104; episode 177 of 2011's season of *Mythbusters* and Robert Luketic's film *21*); and a never-ending list of derivations that includes (but is not restricted to) Ferguson's *N*-door generalization of the original problem [7], Martin Gardner's *Three Prisoners problem* [8,9] and a quantum version of the paradox that illustrates the relation between

The second one is a small part of a renowned collaboration between two of the greatest contributors to probability theory, which also serves the purpose of motivating the use of the median as a measure of central tendency. We are aware that teaching the median of a discrete random variable to students of nine to 19 years old is not customary. However, in our involvement as instructors in a University with an Actuarial Program, the archetype we provide is of great help when it comes to counting filed claims for the purpose of insurance (see Reference [12], section 17.5.3), or to find measures of central tendency for non-scalar random variables. Moreover, to aid our students' understanding of the amount of spare time needed to obtain some insight into results as (for instance) the Law of Large Numbers and the Central Limit Theorem, we have coded a couple of computational tools. We confess how hard it was *not* to imagine being able to time-travel all the way back to the 17th-century to bring our tools to Pascal, Fermat, and the Chevalier de Méré with the sincere expectation that the result differed from the one obtained by the protagonist referred to in 1979's tale *Newton's gift* (this work originally appeared in Reference [13] and was republished in Reference [14] more recently).

The following section is devoted to present the principle of inclusion and exclusion by means of a modification of an illustration presented in Reference [15], p. 69. We wrap that section by introducing the multiplication rule. Section 3 presents the well-known paradox of the Chevalier de Méré in the spirit of References [16], pp. 248 to 250, [17], pp. 28,29 and 44,45 and [5], chapters 6 and 7.3, and pays a homage to our original sources by updating Pascal and Fermat's approach by means of two computational tools specially coded for the publication of this paper. We present computerized versions of the Chevalier's gambles in Section 4 and take advantage of them to include a discussion on confidence intervals for the difference in proportions. Section 5 presents our conclusions.

2 | AS CERTAIN AS 459.48%!

We start this section with a simplification of the example

before. Finally, we divide the number of ways of choosing all six members of the committee among the 27 persons. This results in the following:

$$\frac{\binom{7}{1}\binom{8}{1}\binom{12}{1}\binom{24}{3}}{\binom{27}{6}} = 4.5948.$$

The problem is that this approach does not return a number in the interval [0, 1], so, it clearly cannot represent a probability. As a matter of fact, our reasoning is defective because the product in the numerator assumes a population of size $7 + 8 + 12 + 24 = 51$, which is not the case, for there are only 27 staff members.

The following result will help us to give a proper answer to the question posed in the former example (A proof can be found, for instance, in Reference [2], theorem 3.8).

Theorem 1. *Principle of inclusion and exclusion. Let* $E_1, E_2, \ldots, E_n$ *be events of the sample space. Then*

$$\mathrm{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathrm{P}(E_i) - \sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{P}(E_i \cap E_j) + \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n} \mathrm{P}(E_i \cap E_j \cap E_k) - \cdots\cdots$$

We use Theorem 1 to find the answer to the question posed by the example in this section. Define the events

- E_1 = {There is no principal researcher among the committee's ranks},
- E_2 = {There are no associate professors among the committee's ranks},
- E_3 = {There are no instructors among the committee's ranks}.

The event $E_1 \cup E_2 \cup E_3$ stands for the case where at

$$\mathrm{P}\left[(E_1 \cup E_2 \cup E_3)^C\right] = 1 - \mathrm{P}(E_1 \cup E_2 \cup E_3). \tag{1}$$

Now, Theorem 1 yields

$$\begin{aligned}\mathrm{P}(E_1 \cup E_2 \cup E_3) &= \mathrm{P}(E_1) + \mathrm{P}(E_2) + \mathrm{P}(E_3) \\ &\quad - \mathrm{P}(E_1 \cap E_2) - \mathrm{P}(E_1 \cap E_3) - \mathrm{P}(E_2 \cap E_3) \\ &\quad + \mathrm{P}(E_1 \cap E_2 \cap E_3).\end{aligned}$$

Each of these probabilities can be easily computed. For instance, since there are $\binom{20}{6}$ ways to make a committee without any principal researcher, then $\mathrm{P}(E_1) = \frac{\binom{20}{6}}{\binom{27}{6}}$. It is also straightforward that, since there are $\binom{12}{6}$ ways to form a committee without principal researcher, nor associate professors, $\mathrm{P}(E_1 \cap E_2) = \frac{\binom{12}{6}}{\binom{27}{6}}$. Finally, $\mathrm{P}(E_1 \cap E_2 \cap E_3) = 0$, due to the fact that the event $E_1 \cap E_2 \cap E_3$ is void. Then

$$\begin{aligned}\mathrm{P}(E_1 \cup E_2 \cup E_3) &= \frac{\binom{20}{6} + \binom{19}{6} + \binom{15}{6}}{\binom{27}{6}} \\ &\quad - \frac{\binom{12}{6} + \binom{8}{6} + \binom{7}{6}}{\binom{27}{6}} \\ &= \frac{69{,}938}{296{,}010} = 0.23627.\end{aligned}$$

Plugging this into (1), gives us that $\mathrm{P}\left[(E_1 \cup E_2 \cup E_3)^C\right] = 0.76373$.

The problem studied in this section is tricky and leads to an absurd result because it deals with chances of the unions of non-disjoint events. Theorem 1 gives us an

Definition 1. Let E and F be events in a sample space, and $\mathrm{P}(F) > 0$. We denote the conditional probability of E given F by $\mathrm{P}(E|F)$ and use it to define the simultaneous occurrence of E and F by means of the *multiplication rule*:

$$\mathrm{P}(E \cap F) = \mathrm{P}(F)\mathrm{P}(E|F). \tag{2}$$

The multiplication rule yields a way to compute the probability of the simultaneous occurrence of two events. Actually, as it is well-known, we say that the two events E and F are *independent to each other* (or simply *independent*) when $\mathrm{P}(E|F) = \mathrm{P}(E)$, in which case, (2) reduces to

$$\mathrm{P}(E \cap F) = \mathrm{P}(F)\mathrm{P}(E). \tag{3}$$

We point out at the fact that, if instead of being independent, E and F were *disjoint* (as in the example in this section), we would end up with

$$\mathrm{P}(E \cap F) = \mathrm{P}(\emptyset) = 0,$$

which would match (2) only if $\mathrm{P}(E|F) = 0$. Note that this holds when $E = \emptyset$. This means, in particular, that $\emptyset$ is both disjoint and independent of every set.

3 | DES PENSÉES RATÉES

Once upon a time of Swords and Musketeers, in France, a well-educated fellow, of scarce means and acute mind, fond of discussion, and welcome in the circles of the French nobility, had a troubling matter weighing on his heart. A matter of gambling and earnings.

Antoine Gombaud, the Chevalier de Méré, as he was known to his friends [19], claimed that after *much practice*, experience had shown him that gambling on at least one six in at most four rolls of a fair die was a winning bet.

Figure 1 delves into the "much practice" business.

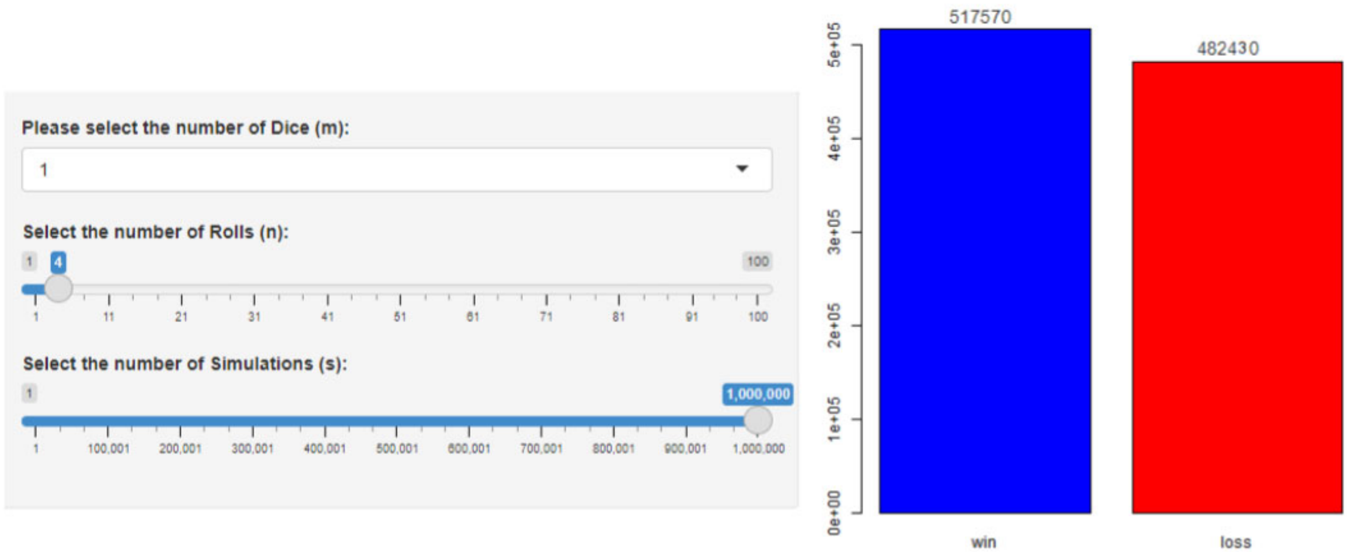


FIGURE 1 A million experiments strongly suggest that it is favorable to bet on an outcome of at least one six in four rolls of a single fair die

The Chevalier de Méré also argued that if getting a six was replaced by getting a double six, he would need only $6 \times 4 = 24$ rolls of two fair dice to restore the advantage. He believed that, since the new event of our interest is six times more difficult to come by than the one six in one die roll, six times more opportunities would re-establish his odds. Indeed, *the theory said that*, as the chance of a double six in a toss of a pair of dice is $\frac{1}{36}$, then the chance of winning in this new game should be $24 \times \frac{1}{36} = \frac{2}{3}$. So, these two games should be won with equal chance. Figure 2 shows what happens for ten thousand simulations of 24 rolls of two dice.

Is this a proof that mathematics is inconsistent? How is it possible that such clean and straightforward reasoning by proportionality fails?

Fortunately this was a well-connected chap, somehow in talking terms with the famous and wise M. Blaise Pascal. But Pascal, being moderately sociable, was a practitioner of the Art of Epistles and wrote a letter about this (and other questions posed by the Chevalier) to a singular acquaintance: a lawyer, partial of numerical conundrums, Pierre Fermat.

In the epistolary exchange that followed, numbers

That would entail four consecutive failures in the first game discussed. Each failure is a 5/6 probability event. Failures are independent. With these ideas in mind, we use a generalized version of the multiplication rule (3) to see that the probability of losing, P(*Loss*), equals the probability of failing four times in a row, that is:

$$\begin{aligned} \mathrm{P}(Loss) &= \mathrm{P}(F_1 \cap F_2 \cap F_3 \cap F_4) \\ &= \mathrm{P}(F_1) \cdot \mathrm{P}(F_2) \cdot \mathrm{P}(F_3) \cdot \mathrm{P}(F_4) \\ &= \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \\ &= \left(\frac{5}{6}\right)^4. \end{aligned} \tag{4}$$

Consequently,

$$\mathrm{P}(Win) = 1 - \mathrm{P}(Loss) = 1 - \left(\frac{5}{6}\right)^4 \approx 51.77\%. \tag{5}$$

The reader may easily conclude that this bet is indeed favorable to the Chevalier, for this probability is

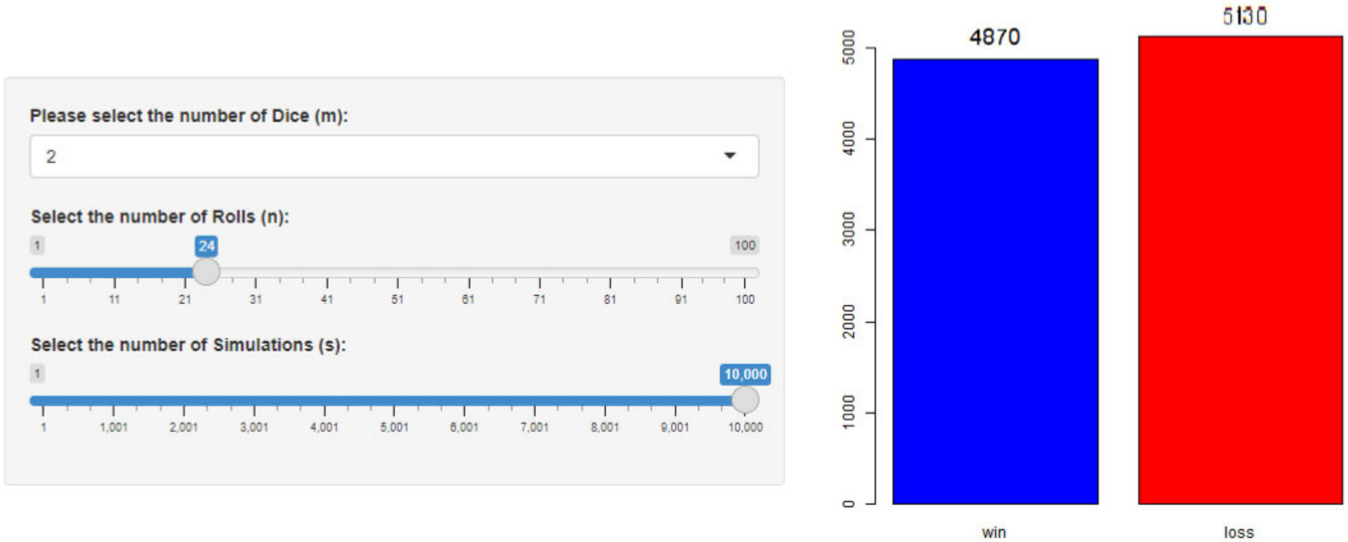


FIGURE 2 We simulated 24 rolls of two fair dice ten thousand times. One might have thought that betting on obtaining a double six would have a probability greater than a half. But applying this strategy yields 5130 losses and only 4870 wins

FIGURE 3 The probability of winning is greater than 50%

It is clear from (6) that winning in (at most) 24 attempts has a probability of less than 50%, and therefore the idea of asking only for 24 rolls of two dice should be improved. As a matter of fact, if we concede one more attempt to the Chevalier, the new bet is favorable (by a minute margin) see Figure 3:

rolls to six, it is certain that you will obtain the one six, for $6\times\frac{1}{6}=1$. Moreover, according to his logic, if you rolled seven times, it turns out that the probability of winning if of $\frac{7}{6}>1$ (as in the example in Section 2)!

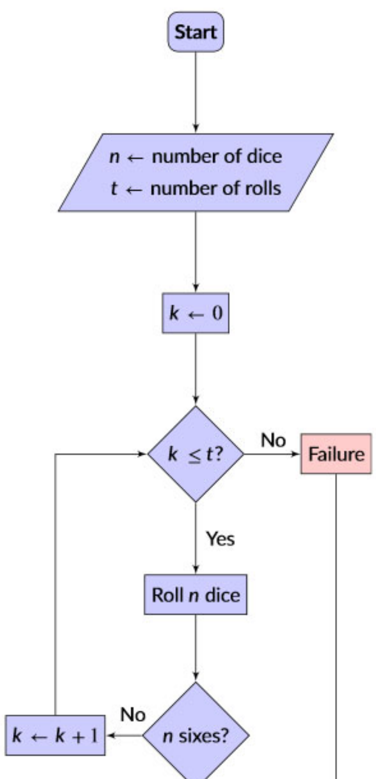
4 | DIGITAL DICE

be confident that you actually win the first gamble with greater chance than the second would have taken a very large number of plays of these two games. There were some people here with serious leisure on their hands!

We devote this section to showing computational procedures to find the least number of rolls that the Chevalier needs to demand so that his probability of winning is barely superior to 50%. We hope our tools will enable people of leisure to spend less time crunching data, and more doing math.

4.1 | Let us roll the dice!

The following figure represents the experience the Chevalier had when he designed his gambling plan. With this in mind, all he had to do was to repeat these steps *a few (thousand) times* to see if his choices for the number of dice to roll and the number of rolls that his fellows allowed him were favorable or not.



This is the schema we followed to produce Figures 1 and 2. Indeed, in Figure 1, we chose $n = 1$, $t = 4$, and repeated the steps a million times; while, for Figure 2, we used $n = 2$ and $t = 24$, but we repeated the simulation ten thousand times.

We invite the interested reader to see an implementation of the steps we just displayed in R v.4.2.1 in https://github.com/DonDisparates/chevalier [21].

The approach used by Pascal and Fermat can be used to generalize the two games discussed to t rolls of n six-sided dice. This is what we did to produce Figure 3 by means of the code we just quoted. Indeed, let p be the probability that you obtain at least one roll with all sixes. Then, the probability that none of the n dice gives the desired outcome in a single roll is $1-\frac{1}{6^n}$. Therefore, an analogous argument to (4) yields

$$1-p=\left(1-\frac{1}{6^n}\right)^t.$$

In other words,

$$p=1-\left(1-\frac{1}{6^n}\right)^t. \tag{7}$$

With this in mind, we generate the following table to help our friend design more games, which are slightly favorable to the gambler.

n	t	p
1	3	0.4213
1	4	0.5177
2	24	0.4914
2	25	0.5055
3	149	0.4991
3	150	0.5015
4	897	0.49962
4	898	0.50001

n were appropriate, he had to partake in *a large number* of gambles. (After all, he was not precisely keen in matters of disjoint, nor independent events.)

In general, when we make an *estimation* of a number (in this case, the probability of winning the wager), we do not go as far as to just give our estimate. What we do, is that we talk about a *confidence interval* where that number should be. That is, a range of numbers we are pretty sure the number must be contained in. These intervals depend on two things: the variation of the outcomes of the experiment, and the size of the sample. A measure of the former is the variance, and the latter is just the number of times the Chevalier gambled. As it is to be expected, the more he played the games, the more he could be certain of the outcome.

There are several ways to give confidence intervals. The most common is to use estimations that use the so-called *Normal distribution* and the Central Limit Theorem (Reference [2], pp. 330-334). If we compute the probability of winning the wager as the proportion $\overline{p}$ of the times we win to the number of times we gamble, N, it turns out that we can be 95.4% confident that the actual probability of winning lies in the interval

$$\left(\overline{p}-2\sqrt{\frac{\overline{p}(1-\overline{p})}{N}},\overline{p}+2\sqrt{\frac{\overline{p}(1-\overline{p})}{N}}\right). \qquad (8)$$

That is $P\left(\overline{p}-2\sqrt{\frac{\overline{p}(1-\overline{p})}{N}}\leq p\leq\overline{p}+2\sqrt{\frac{\overline{p}(1-\overline{p})}{N}}\right)=95.4\%$.
Example:

a. From Figure 1, we know that, after $N = 1,000,000$ trials, the Chevalier de Méré found that the probability he won the one-die bet with four attempts was *close to* $\overline{p}=\frac{517,570}{1,000,000}=51.75\%$ (according to (5), the actual value is 51.77%). The 95.4%-confidence interval, in this case, is (51.66%, 51.86%). This is an interval of size 0.199%.
b. Figure 2 states that, after betting $N = 10,000$ times to at least a double six on 24 rolls of two dice, the Chevalier de Méré obtained an estimate of the probability of winning of $\overline{p}=\frac{4,870}{10,000}=48.7\%$ (by (6), we know that

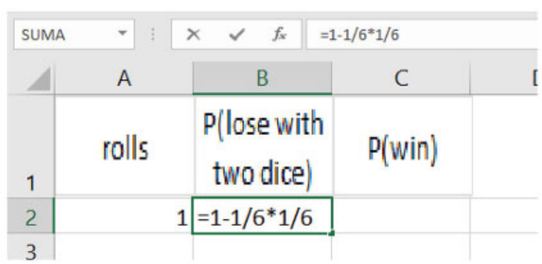
$N = 1,000,000$ times, to gamble *only* $\frac{N}{100}=10,000$ times, you should expect to obtain a confidence interval $\sqrt{100}=10$ times larger. In any case, the sample size needed by the Chevalier de Méré to be confident that he was more likely to win with the one-die game in four rolls than the two-dice game with 24 rolls is *gargantuan*. The bottom line here is as follows: he was most definitely a true person of leisure. However, we must emphasize the fact that the Chevalier and his friends did not have access to the tools we do (eg, confidence intervals or computer simulation). Indeed, there is a distinction between what might constitute evidence for different game probabilities for the Gombaud, Pascal, and Fermat, and what might constitute evidence for different game probabilities for us.

4.3 | Spread *cheating*

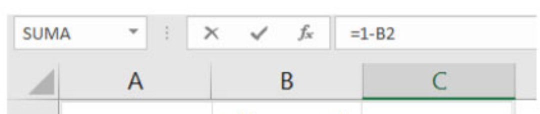
We will use a spreadsheet program to find the number of rolls in a two-dice bet that turns the game barely favorable to the gambler. That is, setting $n = 2$, we will solve (7) for t so that p is just a bit larger than 50%.

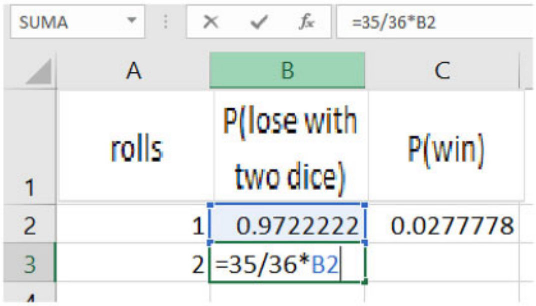
We start by adding labels to the headers. The leftmost column represents the number of rolls; the second, the probability of losing when the Chevalier rolls two dice; and the rightmost column stands for the complement to one, *that is*, the probability of winning.

Probability of losing in a roll of two dice.

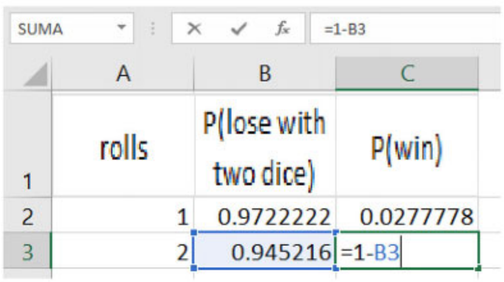


Probability of winning when rolling two dice once.

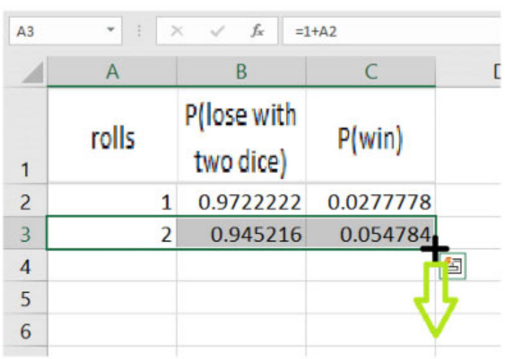




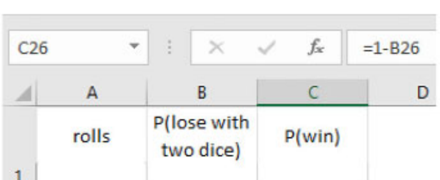
Probability of winning in two rolls of two dice.



Drag the formula downwards until P(Win) > 50.



The number of rolls that make the game barely favorable to the gambler is 25, not



As the reader surely noticed, *t*, the number of rolls referred to in this section is no other than the median of the random variable of rolls of *n* six-sided dice.

5 | FINAL REMARKS

This work presents two failed threads of thought to introduce the idea of independence; the principle of inclusion and exclusion; and the notion of a confidence interval. Our first example is combinatorial, and the second one is a classic paradox, which has been well-documented along the past three centuries. Each deals with the chance of a union of non-disjoint events.

We also take advantage of the paradox of the Chevalier de Méré to present two computational tools. One of them allows the student to play with the number of six-sided dice they will use, the number of rolls, and the number of times the Chevalier will bet so as to compute both the theoretical and empirical probability of winning the wager. The second tool is an implementation on a spreadsheet of the computations performed by Pascal and Fermat to help the Chevalier to find the median of the number of rolls needed to win his bets.

We hope you enjoy the material just as much as we enjoyed preparing it.

ACKNOWLEDGEMENTS

We sincerely thank Mrs. Élida Estrada-Barragán, for providing the authors with a copy of the book [16]; the anonymous referee appointed by Dr. Helen MacGillivray, Editor in Chief of *Teaching Statistics*, and Dr. MacGillivray herself, for their valuable suggestions to improve the original draft of our manuscript.

ORCID

José Daniel López-Barrientos https://orcid.org/0000-0001-9315-1227

REFERENCES

1. G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*, Oxford Science Publications, New York, NY 1994.
2. C. M. Grinstead and J. L. Snell, *Introduction to Probability*, American Mathematical Society, Rhode Island 1997.

7. S. Selvin, *On the Monty Hall problem (letter to the editor)*, Am. Stat. **29** (1975), no. 3, 134.
8. M. Gardner Mathematical Games. Scientific American; October:180–182, 1959.
9. M. Gardner Mathematical Games. Scientific American; November:188, 1959.
10. A. P. Flitney and D. Abbott, *Quantum version of the Monty Hall problem*, Phys. Rev. A **656** (2002), 62318.
11. G. M. D'Ariano, R. D. Gill, M. Keyl, B. Kuemmerer, H. Maassen, and R. F. Werner, *The Quantum Monty Hall Problem*, Quant. Inf. Comput. **2** (2002), no. 5, 355–366.
12. D. S. Promislow, *Fundamentals of Actuarial Mathematics*, Wiley, Chichester 2011.
13. P. J. Nahin, *Newton's gift*, OMNI, New York, NY 1979, 50–53.
14. P. J. Nahin, *Number-Crunching: Taming Unruly Computational Problems from Mathematical Physics to Science Fiction*, Princeton University Press, New Jersey 2011.
15. S. Ghahramani, *Fundamentals of Probability*, Prentice Hall, N.J, 1996.
16. D. Freedman, R. Pisani, and R. Purves, *Statistics*, W.W. Norton and Company, New York, NY 2007.
17. L. Gonick and W. Smith, *The cartoon guide to statistics*, Harpercollins Publishers, New York, NY 1993.
18. P. J. Nahin, *Digital Dice*, Princeton University Press, New Jersey 2008.
19. *Chamaillard E. Le Chevalier de Méré: Rival de Voiture, Ami de Pascal, Precepteur de Mme. de Maintenon; Étude Biographique et Litteraire, Suivi D'Un Choix de Lettres Et de Pensées Du Chevalier*, Wentworth Press, California 2016.
20. B. Pascal, *Les Lettres Provinciales de Blaise Pascal*, Nabu Press, South Carolina 2010.
21. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2022, https://www.R-project.org/

How to cite this article: J. D. López-Barrientos, E. Silva, and E. Lemus-Rodríguez, *Lessons from the famous 17th-century paradox of the Chevalier de Méré*, Teach. Stat. **45** (2023), 36–44. https://doi.org/10.1111/test.12321

ANNOUNCEMENT

Announcement of Special Issue 2023 in *Teaching Statistics*

1 | RE-THINKING LEARNERS' REASONING WITH NON-TRADITIONAL DATA

This Special Issue will showcase work that was presented at SRTL-12. Many ubiquitous forms of data do not clearly fit the sample-population assumptions that underpin the statistical reasoning that has been the focus of much in statistical education. For example, data collected in real time (GPS, live traffic, tweets), image-based (photographs, drawings, facial recognition), semi-structured (scraped from social media posts), repurposed (school testing data to estimate housing prices) and big data data as well as how they model, analyze and make predictions from these forms of data. This special issue focuses on empirical studies that investigate or nurture learners' understanding and reasoning with non-traditional, messy and/or complex data and models. Papers will focus on practical advice and implications for good practice in teaching statistics using non-traditional data.

This special issue will appear in mid-2023.

Guest Editors:
Jennifer Noll, TERC (USA).
Sibel Kazak, Pamukkale University (Turkey).

(open access internet data, civic databases) are all examples of non-traditional data.

While non-traditional forms of data have been with us for some time, the digital age has led to a pervasive culture of data in all aspects of life, including those of our students. Widespread availability and access to myriad of non-conventional, repurposed, massive or messy data sets necessitate broadening educational knowledge to better understand how learners make sense of and interrogate

Lucia Zapata, University of Antioquia (Colombia).
Katie Makar, University of Queensland (Australia).

Helen MacGillivray, Editor-in-Chief

Correspondence
Email: h.macgillivray@qut.edu.au
DOI 10.1111/test.12326