



The Dual Role of Simulation

Jyotirmoy Sarkar

*Department of Mathematical Sciences, Indiana University Indianapolis, LD 270 K
402 N Blackford Street, Indianapolis, IN 46202, USA. e-mail: jsarkar@iu.edu*

ABSTRACT

Simulations provide accessible approximate solutions and verify advanced exact solutions. Oftentimes, it also generates questions worthy of theoretical developments. We illustrate these roles of simulation using a couple of examples and R codes.

Keywords: Center of mass; Weighted average; Symmetry; Doubly negative binomial distribution; Adam's law; Eve's law

AMS Subject Classifications 2020: 97K80.

Prelude

With the proliferation of accessible computer software in recent years, some problems that required advanced mathematics and were out of reach for freshman college students can now be solved using simulations. Moreover, simulations verify the exact results obtained via advanced mathematics. We discuss two problems, giving R codes (R Core Team, 2024) in the Appendix.

1. Center of Mass of a Hemisphere with a Hemispherical Hollow

Problem 1: From a hemisphere of radius 1 made of uniformly dense material, remove a hemispherical hollow of radius $\frac{1}{2}$ such that the base circles of the two hemispheres have the same center. See Figure 1. Find the center of mass (CM) of the hollow hemisphere.

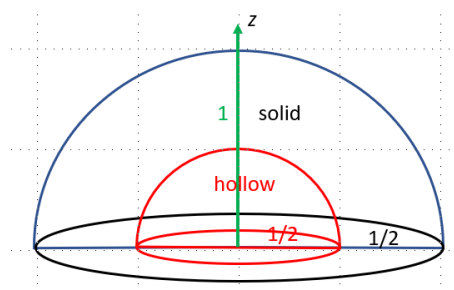


Figure 1. A hemisphere of radius 1 with a hemispherical hollow of radius $\frac{1}{2}$

CM is a familiar concept in physics and mathematics. Why should students of probability and statistics learn about the CM? See (Sarkar & Rashid, 2021).

1.1 Solution to Problem 1

Place the hollow hemisphere on a tabletop with the circular base on the xy -plane, with the center of the base at the point $(0, 0, 0)$. By symmetry, we anticipate the CM is a point on the z -axis; that is, the CM is $(0, 0, \bar{z})$. Students who have not studied calculus yet may estimate $\bar{z}$ using simulation. While the approximate solution will tentatively serve them reasonably well, it should also motivate them to learn calculus. Students who have mastered calculus can find $\bar{z}$ via integration.

1.1.1 Simulation 1. Reflect the hollow hemisphere about the xy -plane to imagine a hollow sphere. Generate a random vector (X, Y, Z) with independent and uniformly distributed coordinates in the interval $(-1, 1)$. Then the vector belongs in a $2 \times 2 \times 2$ cube. To ensure that the vector belongs inside the unit sphere, we reject the vector if $X^2 + Y^2 + Z^2 > 1$. Also, to ensure that the vector is outside the spherical hollow, we reject the vector if $X^2 + Y^2 + Z^2 < 1/4$. Otherwise, we accept the vector and store $W = |Z|$. The absolute value (or folding) ensures that the vector is in the given hollow hemisphere (before reflection). The rate of acceptance is $\frac{7}{8} \times \frac{\pi}{6}$, or about 45.8%.

Repeat the process until many (say, $N = 10^6$) values of W are generated. Then, by the law of large numbers, $\text{mean}(W)$ is a good estimate of $\bar{z}$. By the central limit theorem (CLT), see (Rohatgi & Saleh, 2015), a 95% confidence interval (CI) for $\bar{z}$ is $(\text{mean}(W) \mp 2 * \text{SD}(W)/\sqrt{N})$. Using a simulation of 10^6 values of W , our estimate of $\bar{z}$ is 0.4017, and a 95% CI for $\bar{z}$ is (.4012, .4022).

1.1.2 Calculus. Using calculus, $\bar{z}$ is calculated as follows (with an explanation given below):

$$\bar{z} = \frac{\int_0^1 r \pi(1 - r^2) dr - \int_0^{1/2} x \pi((1/2)^2 - x^2) dx}{\int_0^1 \pi(1 - r^2) dr - \int_0^{1/2} \pi((1/2)^2 - x^2) dx}.$$

Here, the denominator accounts for the volume of the hollow hemisphere — the volume of the whole hemisphere *minus* the volume of the hemispherical hollow. It is computed by integrating the cross-sectional areas when the hollow hemisphere is sliced by planes parallel to the xy -plane. The technique is called the Cavalieri's principle. See (Katz, 1998). On the other hand, the numerator equals the integration of the vertical distance multiplied by the cross-sectional area. Hence, the ratio $\bar{z}$ is a weighted average of the vertical coordinates in the hollow hemisphere, with weights proportional to the area of all points at that same vertical coordinate.

Notice that the multiplier π , which is present in every term, cancels out from the numerator and the denominator. See (Sarkar, 2023) for many other situations where π vanishes. Thereafter,

$$\begin{aligned}\bar{z} &= \frac{\int_0^1 (r - r^3) dr - \int_0^{1/2} \left(\frac{x}{4} - x^3\right) dx}{\int_0^1 (1 - r^2) dr - \int_0^{1/2} \left(\frac{1}{4} - x^2\right) dx} \\ &= \frac{\left(\frac{1}{2} - \frac{1}{4}\right) - \left(\frac{\left(\frac{1}{2}\right)^2}{4(2)} - \frac{\left(\frac{1}{2}\right)^4}{4}\right)}{\left(1 - \frac{1}{3}\right) - \left(\frac{\left(\frac{1}{2}\right)}{4} - \frac{\left(\frac{1}{2}\right)^3}{3}\right)} = \frac{\frac{15}{64}}{\frac{7}{12}} = \frac{45}{112} = 0.401785 \dots\end{aligned}$$

Since 0.401785 belongs to the 95% CI for $\bar{z}$, the simulation verifies the computed value of $\bar{z}$.

Since $\bar{z} < 1/2$, the CM $(0, 0, \bar{z})$ does not belong to the hollow hemisphere, but rather to the hollow! This is similar to the CM of a torus (doughnut shape) being in the hole. Such a feature is perfectly acceptable since the CM of any object represents the point where the entire mass of the object can be thought to be concentrated, wherever that point may be.

Remark 1.1: Suppose we fill the hollow with the same material the hemisphere is made of. From (Manthiri, no date), we know that the CM of the solid hemisphere is at $(0, 0, 3/8)$. So, the CM of the newly filled hollow is at $(0, 0, 3/8)/2$. Because the volume of the hollow is $1/8$ the volume of the hemisphere, we find the z -coordinate of CM of the hollow hemisphere by solving

$$\frac{1}{8} * \frac{3}{16} + \left(1 - \frac{1}{8}\right) * \bar{z} = \frac{3}{8}.$$

The solution is $\bar{z} = 45/112$. Hence, the CM is at $(0, 0, 45/112)$.

Remark 1.2: Suppose the hemispherical hollow has the base circle on the xy -plane, but its center has moved from $(0, 0, 0)$ to $(x_0, y_0, 0)$, where $x_0^2 + y_0^2 < 1/4$. Then the hollow lives entirely inside the hemispherical object, and the CM moves to $(-x_0/7, -y_0/7, \bar{z})$. Because the cross-sectional area at any vertical coordinate has not changed, $\bar{z}$ has not changed at all!

Remark 1.3: Suppose we cut the original hollow hemisphere into two parts by the yz -plane. Then the CM of the right (or east) half (corresponding to $\{X > 0\}$) will be at $(1, 0, 1) * 45/112$, which belongs inside the half-hollow hemisphere. This is justified by symmetry, because when you reflect the right half about the xy -plane, you effectively reconstruct the same hollow hemisphere, except its circular base is on the yz -plane. Consequently, $\bar{x} = 45/112$.

Remark 1.4: Suppose we cut the hollow hemisphere into four quarters using the yz -plane and the xz -plane. Then the CM of the northeast quarter (corresponding to $\{X > 0, Y > 0\}$) will be at $(1, 1, 1) * 45/112$, which belongs inside the quarter-hollow hemisphere (or an octant of a sphere). The justification is like that in Remark 1.3.

Remark 1.5: Suppose we cut the northeast quarter in Remark 1.4 into two halves along the plane $x = y$. We invite the reader to find the CM of each octant and then consult the postlude.

2. Toss a Fair Coin Until at Least Four Heads and Four Tails Appear

Problem 2: Keep tossing a fair coin sequentially, paying me one penny before each toss, until at least four heads and four tails have appeared. Then I will give you a dime (10 pennies) as a prize.

My undergraduate students in the *Introduction to Probability/Statistics* course had a very lively discussion on this problem, after which they summarized their findings as follows:

How many pennies do we have to pay to earn the prize of 10 pennies? It is a random variable N taking on values 8, 9, 10, After 8 tosses, we have w heads and $(8 - w)$ tails with probability (wp) $\binom{8}{w}/2^8$, for $w = 0, 1, \dots, 8$. If $w = 4$, we already won. If $w = 3$ or 5, we will toss the coin once more and win wp $\frac{1}{2}$. If we do not win, we will toss a second time and win wp $\frac{1}{2}$ of $\frac{1}{2}$. Etc. Or, if $w = 2$ or 6, we will toss the coin twice more and win wp $\frac{1}{4}$. Etc. Therefore, $N = 8, 9, 10$ wp

$$\begin{aligned} p(8) &= \frac{\binom{8}{4}}{2^8} = \frac{35}{128} \\ p(9) &= \frac{\binom{8}{3} + \binom{8}{5}}{2^8} \times \frac{1}{2} = \frac{28}{128} \\ p(10) &= \frac{\binom{8}{3} + \binom{8}{5} + \binom{8}{2} + \binom{8}{6}}{2^8} \times \frac{1}{4} = \frac{21}{128} \end{aligned}$$

Thus, we win a positive net amount wp $(35+28)/128=.4922$, break even wp $21/128=.1641$, and lose wp only $1-(63+21)/128=11/32=.34375$. This game appears to favor the player! But who will offer us such a favorable game?

Our teacher is either too generous or out of his mind!!

Dear readers, what do you think? Will you play the game 100 times?

2.1 Solution to Problem 2

To determine whether to play the game, one needs to know $\mu = E[N]$. Students who have taken only one course in *Introduction to Probability/Statistics* cannot derive μ , because N has a doubly negative binomial distribution, which they, and even some graduate students, have not even heard of. However, they, and even some high school students, can write simulation codes to find the answer approximately.

2.1.1 Simulation 2. Generate a sequence of Bernoulli($\frac{1}{2}$) random variables $X_1, X_2, X_3, \dots$, and keep the running counts of heads and tails. Let N be the *smallest* integer such that $(X_1, X_2, \dots, X_N)$ contains at least 4 heads and at least 4 tails. We call N the stopping time of the game. When the game ended, the player would have paid N pennies and won 10 pennies. By the law of large numbers, her expected payment (or the long-run average payment) per game is estimated by the mean of N , after many repetitions of the game.

Using the codes given in the Appendix, we replicated the game 10^6 times and found $\text{mean}(N)=10.1884$ and $\text{SD}(N)=2.33$. Since the player must pay slightly more *on average* than the 10 pennies she will win, her overall loss per 100 games is about \$0.1884 with an SD of \$0.233. Additionally, using the CLT, in 100 games, the player is an eventual winner wp approximately $\text{pnorm}(-0.1884/0.233, 0, 1) = 0.2094$, and an eventual loser wp 0.7906. Thus, I was neither too generous nor out of my mind in offering this game.

Where did my students go wrong in their reasoning? They correctly calculated the probability of winning, breaking even, and losing in each game. But they did not account for the amount they will lose. Whereas they can gain one or two pennies with high probabilities, they could lose an unbounded amount of money, even though that loss happens only about 34% of the time. The simulation gives the approximate probabilities of N (correct to three decimal places) as

$$p(8) \approx .2733, p(9) \approx .2185, p(10) \approx .1645, p(11) \approx .1172, p(12) \approx .0806, \\ p(13) \approx .0534, p(14) \approx .0349, p(15) \approx .0222, p(16) \approx .0141, p(17) \approx .0084, \dots$$

Had my students done the simulation, they would have known the risk of playing my game.

2.1.2 Doubly negative binomial distribution. Recall that if we count the number of tosses M needed until the first head appears, then M is said to have a geometric($\frac{1}{2}$) distribution with probability mass function (PMF) $P\{M = m\} = 2^{-m}$, for $m = 1, 2, 3, \dots$. Specifically, $\text{mean}(M)=2$ and $\text{variance}(M)=2$. See (Rohatgi & Saleh, 2015) for proofs.

However, we like to promote another proof that is missing from most textbooks, except (Ross, 2023), which uses recursion: Note that $M = 1$ wp $\frac{1}{2}$; and $M = 1 + M'$ wp $\frac{1}{2}$, where M' is another geometric($1/2$) random variable. Hence, $E[M] = \frac{1}{2}(1) + \frac{1}{2}(1 + E[M'])$, solving which we get $E[M] = 2$. Likewise,

$$E[M^2] = \frac{1}{2}(1^2) + \frac{1}{2}E[(1 + M')^2] = 1 + E[M'] + \frac{1}{2}E[M'^2] = 3 + \frac{1}{2}E[M'^2],$$

solving which, we get $E[M^2] = 6$. Hence, $\text{var}(M) = 6 - 2^2 = 2$. See (Sarkar & Tulley, 2024) for an application of a generalized geometric distribution.

The geometric($\frac{1}{2}$) distribution is also called a negative binomial($1, \frac{1}{2}$) distribution and it can be extended to a negative binomial($r, \frac{1}{2}$) distribution, which records the number N_r of tosses needed until r heads appear, by adding r independent copies of geometric($\frac{1}{2}$) distributions. As such, the $\text{mean}(N_r) = 2r = \text{variance}(N_r)$. Likewise, we call the number of tosses N needed until r heads and s tails appear a doubly negative binomial($r, s, \frac{1}{2}$) distribution. The game we offered involves a doubly negative binomial($4, 4, \frac{1}{2}$) distribution. How do we calculate its mean?

To win, the player must toss the coin *at least* 8 times. By then, by the pigeonhole principle, the player surely meets at least one of the requirements of 4 heads and 4 tails. It remains to fulfill the other requirement, unless it is already met. Therefore, we decompose N according to W , the number of heads among the first 8 tosses. Letting N_r follow a negative binomial(r ,

$\frac{1}{2}$) distribution, we have

$$N = 8 + \begin{cases} \begin{matrix} \square & \square & \square \\ 0 & \text{if } W = 4 & \text{wp } \binom{8}{4} 2^{-8} = .2734 \end{matrix} \\ \begin{matrix} N_1 & \text{if } W = 3, 5 & \text{wp } 2 \binom{8}{3} 2^{-8} = .4375 \end{matrix} \\ \begin{matrix} N_2 & \text{if } W = 2, 6 & \text{wp } 2 \binom{8}{2} 2^{-8} = .2188 \end{matrix} \\ \begin{matrix} N_3 & \text{if } W = 1, 7 & \text{wp } 2 \binom{8}{1} 2^{-8} = .0625 \end{matrix} \\ \begin{matrix} N_4 & \text{if } W = 0, 8 & \text{wp } 2 \binom{8}{0} 2^{-8} = .0078 \end{matrix} \end{cases}$$

Therefore, using the law of total expectation (or Adam's law), see (Weiss, 2005), we have

$$E[N] = E[E[N|W]] = 8 + 2^{-8} \cdot 2^2 \left[\binom{8}{3} + 2 \binom{8}{2} + 3 \binom{8}{1} + 4 \binom{8}{0} \right] = 8 + \frac{35}{16} = 10.1875.$$

Similarly,

$$E[V[N|W]] = 2^{-8} \cdot 2^2 \left[\binom{8}{3} + 2 \binom{8}{2} + 3 \binom{8}{1} + 4 \binom{8}{0} \right] = \frac{35}{16} = 2.1875$$

and

$$E[E^2[N|W]] = 2^{-8} \cdot 2^3 \left[\binom{8}{3} + 2^2 \binom{8}{2} + 3^2 \binom{8}{1} + 4^2 \binom{8}{0} \right] = 8.$$

Hence, using the law of total variance (or Eve's law), see (Weiss, 2005), we have

$$V(N) = V[E[N|W]] + E[V[N|W]]$$

or

$$V(N) = \{E[E^2[N|W]] - (E[E[N|W]])^2\} + E[V[N|W]] = 8 - \left(\frac{35}{16}\right)^2 + \frac{35}{16} = 5.4023.$$

Hence, $SD(N) = 2.3243$. The simulations justify the $\text{mean}(N)$ and the $SD(N)$ computations.

Remark 2.1: Since W follows a binomial(8, $\frac{1}{2}$) distribution, N can be thought of as a binomial mixture of several negative binomial random variables. We leave it to the reader to derive the exact probabilities: $P\{N = 8\} = \binom{8}{4}/2^8$, and thereafter, for $s = 0, 1, 2, \dots$, we have

$$P\{N = 9 + s\} = \sum_{i=0}^3 \binom{s}{i} \binom{8}{3-i} / 2^{s+8}.$$

Remark 2.2: Suppose that a player has not won after 15 tosses (this happens to about 3.5% of all players). Should he quit with no prize money? No. To earn his prize of 10 pennies, he needs *on average* 2, 4, 6, or 8 tosses, depending on how many heads (or tails) he is still missing to achieve the target of 4 heads and 4 tails. Hence, quitting before winning is an unwise choice.

Postlude

Using two examples and simple R codes, we have demonstrated that simulations provide accessible approximate solutions and verify the accuracy of advanced exact solutions. Moreover, simulations can raise and answer additional questions whose exact solutions may be tedious.

In Remark 1.5, the CMs must have the same $\bar{z} = 45/112$, and $(\bar{x}, \bar{y})$ must be on the rays in the xy -plane from $(0, 0)$ making angles $22\frac{1}{2}^\circ$ and $67\frac{1}{2}^\circ$ with the x -axis. As such, the two CMs are $(\sqrt{2}, 2 - \sqrt{2}, 1) * 45/112$ and $(2 - \sqrt{2}, \sqrt{2}, 1) * 45/112$, respectively. Verify that their average is the CM of the north-east quarter discussed in Remark 1.4.

We encourage the reader to look for other problems that require advanced mathematics but can be approximated via simulations. Happy hunting.

Acknowledgments

I thank my students for participating in several brain teasers. This paper is a tribute to my teacher (turned collaborator), Professor Bikas Kumar Sinha, who introduced me to many interesting puzzles and blessed me with the joy of solving some of them.

References

- [1] Katz, V. J. (1998). *A History of Mathematics: An Introduction* (2nd ed.). Addison-Wesley.
- [2] Manthiri, A. A. (no date). *Fundamentals of Physics — Mechanics*, 8650_Mechanics.pdf
- [3] R Core Team (2024). *R, A Language and Environment for Statistical Computing*, R Foundation, Vienna, Austria. <https://www.R-project.org/>
- [4] Rohatgi, V. K., and Saleh, A. K. M. E. (2015). *An Introduction to Probability Theory and Mathematical Statistics*, 3rd ed., Hoboken, NJ: John Wiley & Sons, Inc.
- [5] Ross, S. M. (2023). *Introduction to Probability Models*, 13 ed., San Diego, CA: Elsevier.
- [6] Sarkar, J. and Rashid, M. (2021). Visualizing the centre of gravity using the (empirical) cumulative distribution functions, *International Journal of Statistical Science*, 21(2), pp.19–43.
- [7] Sarkar, J. (2023). Watch pi appear and disappear magically, *International Journal of Mathematics, Statistics and Operations Research*, 3(2), pp. 301–326.
- [8] Sarkar, J. and Tully, C. (2024). Drawing balls from alternate baskets until success, *Resonance—Journal of Science Education*, 29(6), pp. 839–849.
- [9] Weiss, N. A. (2005). *A Course in Probability*, Addison–Wesley, pp. 380–386.

APPENDIX

The R codes needed to solve the problems are given below.

Problem 1. Center of Mass of a Hollow Hemisphere

```
## generate vertical coordinates
genW=function(){
  u=runif(3, -1, 1) # three IID uniform(-1, 1)
  ss=sum(u*u)
  if (ss >= 1/4 & ss<=1){W=abs(u[3])}
  else{W=9}
W # output W
}
## replicate generation
N=10^6
attempt=replicate(4*N, genW()) # generate a lot
refined=attempt[attempt<1] # drop when W=9
data=refined[1:N] # take only N values
m=mean(data); s=sd(data)
c(m-2*s/sqrt(N), m, m+2*s/sqrt(N)) # 95% CI
## [1] 0.4012525 0.4017425 0.4022324 # trust three decimals
## 45/112=.4017857 # true value
```

Problem 2. Toss Until at Least Four Heads and Four Tails Appear

```
# define the play: keep tossing until 4H & 4T
play=function(){
  current=rbinom(50, 1, 1/2)
  min(which( (cumsum(current)>=4)*(cumsum(1-current)>=4) == 1))
} # end play
# Repeat play many times
data=replicate(10^6, play())
table(data)
sum(table(data))
c(mean(data), sd(data))
dataT=data*(data<=20)+20*(data>20)
c(mean(dataT)+10*(1-cumsum(table(data))[20-7]/10^6), sd(dataT))
```