On random walks, projecting election results and statistical physics

J. Valentin Escobar

Instituto de Física, Departamento de Física Química,
Universidad Nacional Autónoma de México, México City, 04510, México.

Received 16 February 2021; accepted 12 March 2021

Several important statistical tools and concepts are covered in upper division undergraduate Statistical Physics courses, including those of random walks and the central limit theorem. However, some of their broad applicability tends to be missed by students, as well as the connection between these and other physical concepts. In this work, we apply a 1D random walk to study the evolution of the probability that a candidate will win an election given she holds some lead over her opponent and connect the result found to the concept of density of states and occupation probabilities. This paper is intended to serve as a guide to the Statistical Physics instructor who wishes to motivate students beyond the boundaries of the official syllabus.

Keywords: Random walk; density of states; statistical physics; elections; forecasting.

DOI: https://doi.org/10.31349/RevMexFisE.18.020202

1. Introduction

At the end of their undergraduate education, Physics students are familiar with random walks and the Central Limit Theorem, as well as with the concepts of Density of States (DOS) and occupation probability applied to quantum statistics. However, these two concepts - random walks and DOS- are not usually connected in regular courses. Furthermore, the usefulness of the former in both Physics and other statistical problems outside of the realm of this discipline tends to get lost.

In this paper, we present an application of random walks outside of the realm of Physics that can then be connected to the concept of DOS and the calculation of average values in Statistical Physics. Specifically, we show how a simple random walk in 1D can be used to calculate the probability a candidate will win an election given that she holds a certain lead after some percentage of the votes have been counted. The equation that yields this probability is then compared to the expression normally encountered in this course to calculate average values for a given occupation probability and DOS.

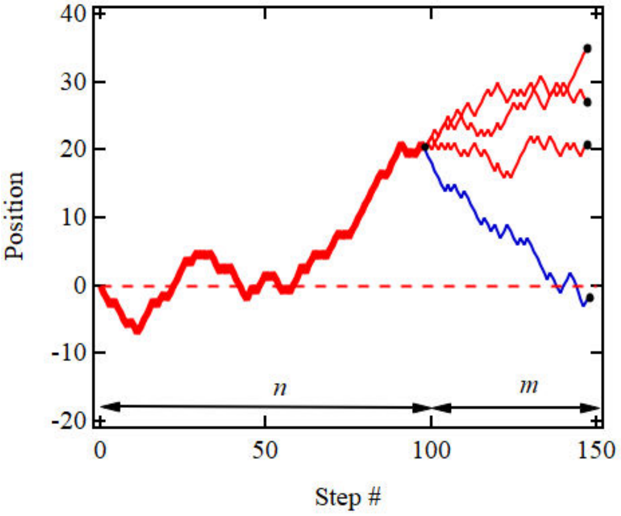

FIGURE 1. A random walk (red-thick trace) with unit steps and no hesitation is found at position $= 20$ after $n = 100$ steps. The thinner traces depict some realizations after an extra $m = 50$ steps are taken, out of which the blue one ends up at a negative position. If the odds of taking ± steps are unknown, what is the probability that such a walk will not happen?

2. An election result as a random walk

Consider a unit step random walk in 1D without hesitation. In the plot of Fig. 1, the thick red line depicts a realization of this walk after $n = 100$ steps, whose position happens to be positive (+20). If an extra $m = 50$ steps are yet to be taken, the black dots depict the final positions of different possible walks generated with the same (but in principle unknown) underlying probabilities P and $1 - P$ of taking a step forward and backward, respectively. The thin blue trace depicts a walk whose position ends up being negative, thus reverting the initial lead.

The question we would like to answer is: how likely is it that such a walk will not occur, (*i.e.* that the lead is not reverted)? The original electoral problem posed in the introduction section can be formulated as a special case of this general one. If we assign a +1 to the vote for candidate #1 and -1 to a vote for candidate #2, a partial result of an election can be given by the sign of the sum of the fraction of the votes that have been counted so far. We can regard each vote as given by a stochastic variable Xi (identical for all i) defined by the probabilities of its only two possible outcomes: $Prob(x_i = 1) = P$, and $Prob(x_i = -1) = 1 - P$, which are unknown to us. Note that we are assuming that null votes are forbidden, which translates into the condition of “no hesitation” for a random walk. Then, the partial lead after n votes have been counted is given by the position relative to the origin of a stochastic variable X defined as:

$$\begin{aligned} X &= \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} Y_i, \end{aligned} \tag{1}$$

where $Y_i \equiv X_i/n$. We point out that applications of random walks to electoral problems date back to the work by W. Whitworth in 1878 [1].

3. Solution

To find the solution to this problem, we will first calculate the probability that the observed lead after n steps (or votes) belongs to any given underlying probability P. Then, we will calculate the probability that, for a given P, a subsequent walk will not revert the lead in the remaining m steps. The sum over all P of the product of these two probabilities is the probability we are looking for. To calculate the former probability, we first obtain the statistical properties of the variable X by applying the Central Limit theorem [2], assuming that the number of steps taken so far is large. X is then well approximated by a Normal random variable, $X = N(\mu, \sigma^2)$, where the mean and variance are given, respectively, by

$$\mu \equiv \langle X \rangle = n\langle Y_i \rangle = n\langle X_i/n \rangle = \langle X_i \rangle, \tag{2}$$

and

$$\begin{aligned} \sigma^2 &\equiv VAR\{X\} = nVAR\{Y_i\} \\ &= nVAR\{X_i/n\} = \frac{1}{n}VAR\{X_i\}. \end{aligned} \tag{3}$$

If the two possible values for a vote $\langle X_i \rangle$ are $x_{i,1} = 1$ and $x_{i,2} = -1$, then, by definition,

$$\langle X_i \rangle = \sum_j x_{i,j}\text{Prob}(x_{i,j}) = 2P - 1,$$

and

$$\langle X_i^2 \rangle = \sum_j (x_{i,j})^2\text{Prob}(x_{i,j}) = 1,$$

After substituting these values in the general relation

$$VAR\{X_i\} = \langle X_i^2 \rangle - \langle X_i \rangle^2,$$

Eq. (3) leads to:

$$\sigma^2 = \frac{1}{n}(1 - [2P-1]^2) = \frac{4}{n}P(1-P), \tag{4}$$

while the mean is simply given by:

$$\mu = 2P - 1. \tag{5}$$

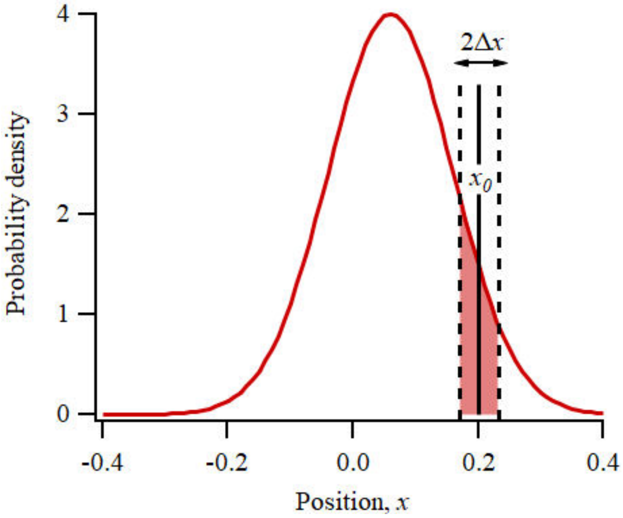

FIGURE 2. Probability density (red trace) for the variable X given by a Normal distribution with $P = 0.53$ after $n = 100$ steps, along with the relative position of the walk at that time at $x_0 = 0.2$ (black line). The integral limits of Eq. (7) are depicted by the dashed lines.

Now, the lead of candidate #1 given in terms of a percentage A (or fraction) of counted votes can be recast as some position x_0 of the walk:

$$x_0 \equiv A - (1 - A) = 2A - 1. \tag{6}$$

For example, if $A = 60\%$, then $x_0 = 0.2$, and we can visualize a certain lead related to the Normal distribution we just found for a given P (see example in Fig 2).

Now, note that by applying the Central Limit theorem, we have transformed a discrete probability distribution into a continuous one, a probability density. Consequently, the probability that a walk will end at exactly the position $x = x_0$ is strictly zero. The question we should be asking instead is: what is the probability of finding the walk in a position x in the range $(x_0 - \Delta x) < x < (x_0 + \Delta x)$? This probability is depicted by the shaded area in Fig. 2 and is given by

$$\Pi_1 \equiv \frac{1}{\sqrt{2\pi}\sigma[P]} \int_{x_0-\Delta x}^{x_0+\Delta x} e^{-(x-\mu[P])^2/2\sigma[P]^2} dx, \tag{7}$$

where the square bracket symbolizes "function of", and the mean and variance are given by Eqs. (4) and (5). But, how wide of an interval should we use? To answer this, students should first realize that, even though n is large, it is not infinite. In this respect, it does not make sense to use $\Delta x < 1/n$ since it is the resolution of data. It should also be made clear that Δx is on the order of $1/n$ but that its size is in principle arbitrary. Assuming that n is indeed large, it then follows that $\Delta x \ll 1$, which allows us to expand Eq. (7) in Taylor series to first order in Δx around zero:

$$\Pi_1[X_0] \approx \frac{2}{\sqrt{2\pi}\sigma_1[P]} e^{-(x_0-\mu[P])^2/2\sigma[P]^2} \Delta x. \tag{8}$$

Here, it is important to stress a key point: x_0 is the quantity we observe, while P is unknown. The question we should be asking instead is: what is the probability that given n and $x_0 \pm \Delta x$, the parameter with which the walk was generated is P? With this important conceptual change in mind, we reinterpret the probability π_1 as being a function of P for given n and $x_0 \pm \Delta x$ (that is, $\Pi_1 = \Pi_1[P]$). The following question immediately arises: is Π_1 a (continuous) probability density or a (discrete) probability function of P? In the former case, the normalization would necessarily be a function of Δx alone, which, as we just mentioned, is arbitrary. Therefore, Π_1 must be a discrete probability function whose normalization function is given by

$$\sum_{p} \Pi_1[P] = 1, \tag{9}$$

where the possible values of P are separated by small (but also finite!) intervals of size ΔP, a quantity we will use below to normalize Π_1. Note in Eq. (9) that we could also interpret $\Pi_1[P]$ as the average probability that the walk found at the position $x_0 \pm \Delta x$ was generated with a parameter P in the range $P + \Delta P$. To perform the summation in Eq. (9), we note first that Π_1 in Eq. (8) is itself a normal variable, but this time as a function of P. Explicitly,

$$\begin{aligned}\Pi_1[P] &= \sqrt{\frac{n}{2\pi}}\frac{e^{-(n[1-2P+x_0]^2/8[1-P]P)}}{\sqrt{(1-P)P}}\Delta x \\ &= \frac{e^{-(P-\mu_1)^2/2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}\Delta x}\end{aligned} \tag{10}$$

where $\mu_1 \equiv (1/2)(x_0 + 1) = A$ and $\sigma_1^2 \equiv (1/n)P(1-P)$. If we assume that $\Delta P \ll 1$, the normalization condition can be approximated by an integral after multiplying by $\Delta P/\Delta P$=1:

$$\begin{aligned}1 &= \sum_P \Delta x \frac{e^{-(P-\mu_1)^2/2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}}\frac{\Delta P}{\Delta P} \\ &\approx \frac{\Delta x}{\Delta P}\int_0^1 \frac{e^{-(P-\mu_1)^2/2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}}dP\end{aligned} \tag{11}$$

As long as x_0 is not too close to either 0 or 1, we can extend the integration limits to plus and minus infinity and integrate to obtain the normalization condition,

$$\begin{aligned}\sum_P \Pi_1[P] &\approx \frac{\Delta x}{\Delta P}\int_{-\infty}^{\infty} \frac{e^{-(P-\mu_1)^2/2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}}dP \\ &= \frac{\Delta x}{\Delta P} = 1.\end{aligned} \tag{12}$$

Note that we calculated the integral in this equation as if the variance σ_1^2 were constant, when, in reality, it is a function of P. However, the factor $P(1-P)$ is practically constant in the range where $\exp[-n(P-\mu 1)^2]$ varies significantly. Therefore, for large n (say, $n > 1000$), regarding σ_1^2 as a constant is an excellent approximation, and we can proceed as if the integrand was a normalized distribution. Equation (12) then imposes that $\Delta P = \Delta x = (1/n)$. After substituting $\Delta x = (1/n)$ in Eq. (10), we arrive at

$$\Pi_1[P] = \frac{1}{n}\frac{e^{-(P-\mu_1)^2/2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}}. \tag{13}$$

When this expression for Π_1 is substituted into Eq. (9) and the sum is taken overvalues of P separated by intervals of size $\Delta P = 1/n$, normalization is ensured [3]. We are now in the position to calculate the second probability necessary to answer our central question: for a given P and x_0, what is the probability that the walk will end up at a positive position, provided that m extra steps are still to be taken? In other words, what is the probability that the lead is not reverted provided a particular value for P is assumed? To this end, we construct a new variable Z analogous to the variable X we constructed in Eq. (1), but this time is given by the average of the remaining m votes:

$$Z = \frac{(X_{n+1} + X_{n+2} + \ldots + X_{m+n})}{m}.$$

Since the variables are again identical, a simple change of index leads to:

$$Z = \frac{1}{m}\sum_{k=1}^{m} X_k. \tag{14}$$

If m is large, too, we can invoke the Central Limit Theorem once again, and obtain for Z a Normal distribution with mean μ_2 and variance σ_2^2 given by Eqs. (5) and (4) respectively, but with n replaced by m. The goal is now to calculate the probability $\Pi_2[P]$ that the length of the remaining walk will be no more negative than $-x_0$ (blue trace in Fig. 1). This probability is simply given by the integral of $N(\mu_2, \sigma_2^2)$ from $z_0 \equiv x_0(n/m)$ to $+\infty$ (see Appendix A), and it can be calculated as a function of x_0, n, m, and P:

$$\begin{aligned}\Pi_2[P] &\equiv \int_{-z_0}^{\infty} N(\mu_2, \sigma_2^2[P])dz \\ &= \frac{1}{2}\left(1 + Erf\left[\frac{x_0(n/m) + 2P - 1}{\sqrt{8P(1-P)/m}}\right]\right).\end{aligned} \tag{15}$$

To illustrate how this function evolves as m decreases, a plot of Π_2 vs. P is displayed in Fig. 3. For a given P, the probability that both an observed lead was generated from a walk with that parameter and that the lead won't be reverted is $\Pi_1[P] * \Pi_2[P]$ (since these are independent events). To finally arrive at the total probability Π_T that the lead is not reverted, we must now add up the contributions from all possible values of P separated by intervals of size $\Delta P = 1/n$:

$$\Pi_T = \sum_P \Pi_1[P]\Pi[P]. \tag{16}$$

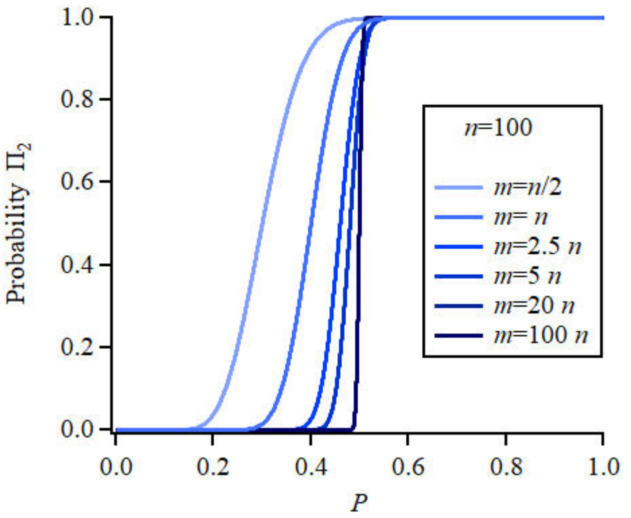

FIGURE 3. Probability π_2 vs. P of maintaining the lead given by $A = 60\%$ ($x_0 = 0.2$) for $n = 100$ and the different remaining number of votes (m) to be counted.

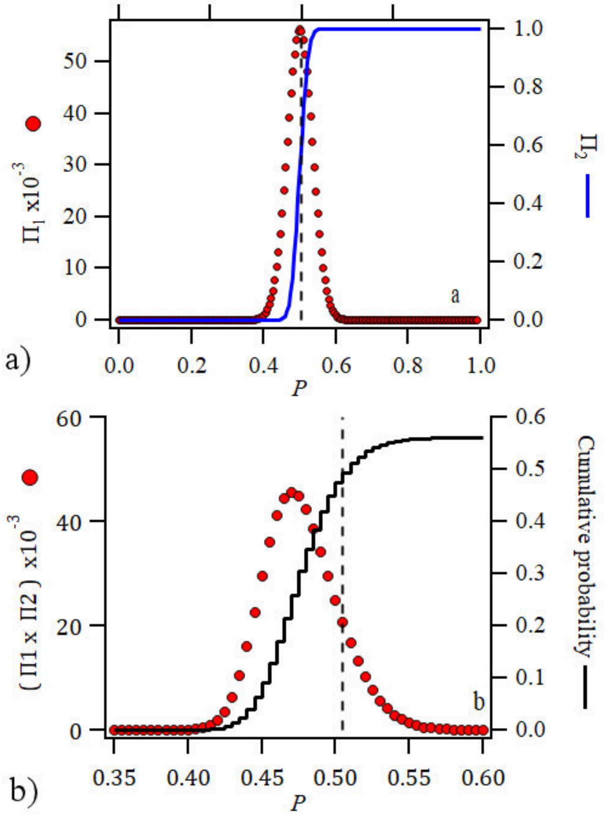

FIGURE 4. a) Functions π_1 and π_2 vs. P in Eq. (16) to obtain the probability π_T that an observed lead is not reverted ($n = 200$, $m = 800$ and $x_0 = 0.01$). b) $\pi_1 * \pi_2$ (left axis) and corresponding cumulative probability (right axis). Note that this product is not symmetrical with respect to $\mu_1 = (1/2)(x_0+1) = 0.505$, (dashed line on both figures).

It is instructive to show in the same plot the two functions that are being multiplied in Eq. (16) (Fig. 4a)), as well as their product itself (Fig. 4b)). Note in this plot that we are multiplying normal function times a sigmoidal one (except when $m/n \gg 1$). This results in an asymmetrical function for μ_1 (Fig. 4b)). However, when $m \gg n$, Π_2 becomes a step function (darkest trace in Fig. 3) that splits exactly in half the sum of the probabilities given by Π_1, so that the probability that any lead will hold is very close so 50%, as expected when there is still a relatively large number of votes to be counted.

In the opposite limit, when $m \gg n$ (lighter trace in Fig. 3), Π_2 becomes smoother, and the range in which it equals 1 extends well to the left of μ_1, the point where Π_1 is centered. This yields $\Pi_T \approx 1$, as expected when there are relatively few votes left to be counted for any given lead. Thus, in both limits, Eq. (16) gives sensible answers, and we can now explore the behavior of Π_T in-between these limits. We can ask, for example, what the probability is that a candidate will end up winning the election if her lead stays constant as the remaining votes are progressively counted. Figure 5 shows the evolution of the total probability Π_T for $N \equiv (n+m) = 10^5$ for different values of A vs. $m/N \equiv \theta$. This plot evidences that there exists a threshold value θ_T for each A below which $\Pi_T \approx 1$ (dashed lines). This is a useful limit to know since it tells us when the election can be called for the leading candidate with almost complete certainty. A good approximation for θ_T can be obtained for cases in which saturation happens when relatively few votes remain to be counted, like those shown in Fig. 5 (except for $A = 50.25\%$). This yields the following simple analytical expression for θ_T (see Appendix):

$$\theta_T \approx \frac{\alpha^2}{9} = \frac{N x_0^2}{9}. \tag{17}$$

In Fig. 6, we plot again Π_T for the same values of A presented in Fig. 5, but this time vs. (θ/θ_T), which is a rescaled relative number of remaining votes to be counted. As expected, $\Pi_T \approx 100\%$ below $(\theta/\theta_T) = 1$ for all these values for A. But note also that all curves collapse into a single one for $(\theta/\theta_T) < 5$, a range in which $\Pi_T \geq 90\%$. In other words, close to $\theta = \theta_T$, Π_T scales as $(\theta = \theta_T)$ for all systems considered here, except for the case with the largest A. We can exploit this scaling property to claim that a lead x_0 is significant (more than 90% of chances of winning) if $\theta = \theta_T = \theta/(N x_0^2/9) < 5$, which will happen as soon as $N x_0^2/\theta > 9/5$, or

$$x_0 > \sqrt{\frac{9}{5}\frac{\theta}{N}} \approx 1.34\frac{\sqrt{m}}{N}. \tag{18}$$

An implementation of Eq. (16) in Mathematica$^{©}$ is presented in Appendix C.

4. Analogy with statical physics

Note that we could have approximated the total probability Eq. (16) with the following integral:

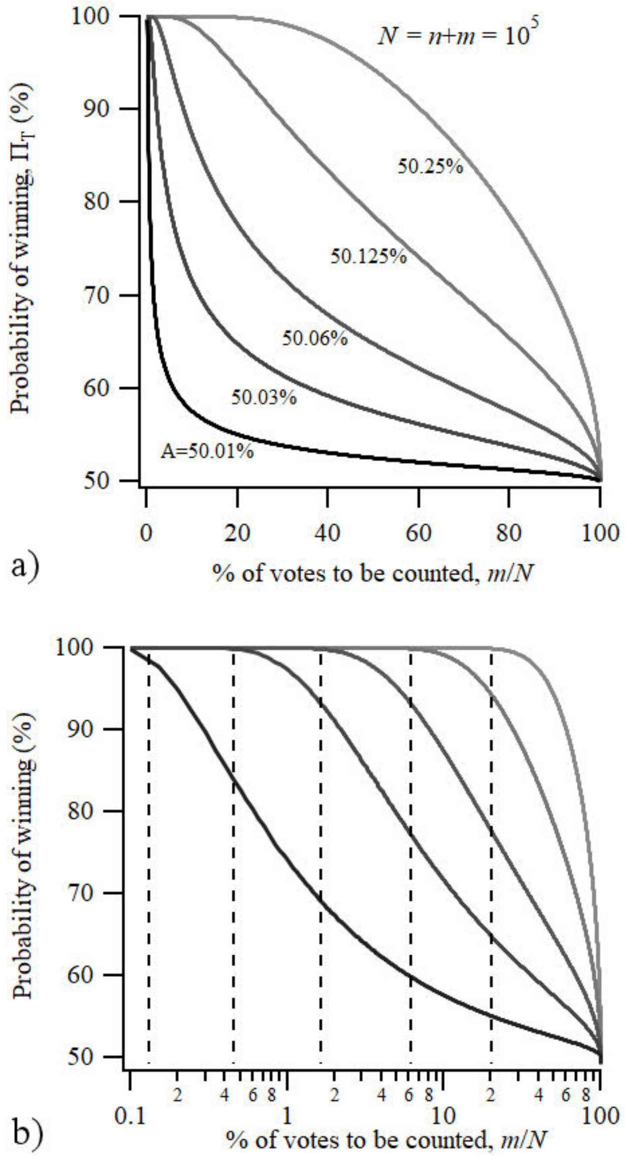

FIGURE 5. a) Probability Π_T that the leading candidate will win the election using Eq. (16), assuming the lead A does not change as the remaining votes are counted. Each curve corresponds to a different constant A, but $n + m = 10^5$ is constant. b) Same data but in lin-log scale showing the points (dashed lines) below which Π_T saturates for each A.

$$\Pi_T = \sum_{P'} \Pi_1\Pi_2 = \frac{1}{\Delta P}\sum_{P'} \Pi_1\Pi_2\Delta P$$

$$\approx \frac{1}{\Delta P}\int_0^1 \Pi_1\Pi_2 dP. \qquad (19)$$

On the other hand, consider the average number of particles obeying Maxwell-Boltzmann statistics at some temperature T. The average number of particles with energy $\varepsilon[k]$ is given by $\bar{n}[\varepsilon]$, where k is the wavenumber.

The function $\bar{n}[\varepsilon]$ can also be regarded as the probability of finding one of the N particles with energy ε. To obtain the total number of particles, we need to multiply $\bar{n}[\varepsilon[k]]$ by the number of quantum states in the range $k + dk$ (the density of states or "DOS" [4]) and sum over all possible $k's$. The total number of particles is then:

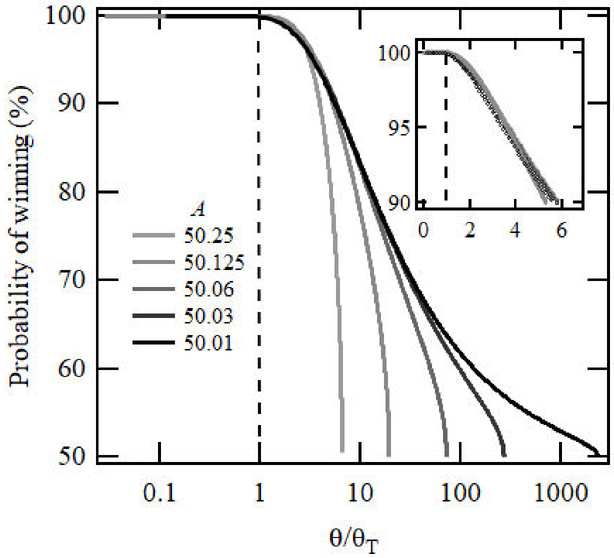

FIGURE 6. Probability Π_T vs. (θ/θ_T) that the leading candidate will win the election using Eq. (16), where the threshold value θ_T is given by Eq. (18). Inset: all curves collapse into a single one up to about $(\theta/\theta_T) \approx 5$ (the data for $A = 50.25\%$ were excluded).

$$N = \int_0^\infty D[k]\bar{n}[\varepsilon[k]]dk, \qquad (20)$$

where the discrete sum over states has been approximated by an integral. Compare now Eq. (20) with Eq. (19). Indeed, the procedure just outlined to arrive at Eq. (21) is the same one we have followed in this paper to find the answer to our electoral problem. $\Pi_1[P]$ in Eq. (19) is the probability that the system has a parameter $P + dP$. This probability is proportional to the number of walks (or "quantum-states") associated with the parameter P. Thus, in this analogy, $\Pi_1[P]$ plays the role of the DOS, while the parameter P plays that of the wavenumber k. On the other hand, $\Pi_2[P]$ is the probability that the condition we are after is fulfilled given x_0 and m, *i.e.*, that the remaining walks do not reverse the lead. In this sense, we can think of $\Pi_2[P]$ as the probability that the state defined as "not reversing the lead" is occupied for some value of P. Therefore, $\Pi_2[P]$ plays the role of $\bar{n}[\varepsilon]$. Finally, to obtain the total probability, we add over all possible values of P, while in Statistical Physics, we add over all possible values of k. In summary, the following relations hold:

$$\Pi_1[P] \to DOS, \quad \Pi_1 2[P] \to \bar{n} \quad P \to k, \qquad (21)$$

Students may find it fruitful to reflect on the similarity between the constructions of the solution of these two very different problems (one physical, one electoral) and the usefulness of the concept of random walks inside and outside of the realm of Physics.

Appendix

A. On calculating the threshold θ_c

We now calculate the threshold θ_c below which $\Pi_T \approx 1$ with help from Fig. A.1. Π_T will be close to 1 if the following two conditions are met simultaneously: 1) The peak of Π_1 (point #1) is located to the right of the point where Π_1 begins to decrease (point #2), and 2) The point where Π_1 begins to increase (point #3) is located to the right of the point where Π_2 begins to increase (point #4). Considering all positions for $\theta = 0.5$, point #1 is at $-\mu_1$, while the position of the middle point of Π_2 (point #5) is $Z_0/2$. It is important to realize that the location of points #2, #3, and #4 is somehow arbitrary because these points mark the point where an exponential function ends or begins. We circumvent this issue by using the so-called 3σ rule, which states that it is sensible to assume that the non-zero portion of a normal distribution resides in the range $3\mu - \sigma < P < 3\mu + \sigma$, (99.7% of the total area of a normal function $N(\mu, \sigma^2)$ resides in this range). To calculate this standard deviation, note now, as we mentioned earlier, that even though the variance of Π_1 is not constant, it is a very slowly varying function of P.

Thus, we can safely regard it as being independent of P and equal to the value it attains at $P = 0.5$, which is $(1/2)n^{-1/2}$. We then follow the 3σ rule, and place point #3 a distance 3σ to the right of μ_1. Explicitly, point #3 is at $-\mu_1 + (3/2)n^{-1/2} = -(1/2)(1 + x_0) + (3/2)n^{-1/2}$. Considering that Π_2 is the integral of a normal function with $\sigma_2^2 \equiv (1/n)4P(1 - P)$ and following the same reasoning as before, then point 4 is located at $(z_0/2) - (3/2)m^{-1/2} = (x_0 n/2m) - (3/2)m^{-1/2}$, while point 2 is located at $(x_0 n/2m) - (3/2)m^{-1/2}$. Then, condition 1 mentioned at the beginning of the appendix is met when:

$$\frac{x_0}{2}\left(1+\frac{n}{m}\right) - \frac{3}{2\sqrt{n}}\left(1-\sqrt{\frac{n}{m}}\right) \geq 0, \tag{A.1}$$

while, condition 2 is met when:

$$\frac{x_0}{2}\left(1+\frac{n}{m}\right) - \frac{3}{2\sqrt{m}} \geq 0. \tag{A.2}$$

Both conditions will be simultaneously met if the product of these last two expressions is positive. After simplifying this product and substituting $\theta \equiv m/N$ and $N \equiv n + m$, we arrive at:

$$\alpha^2(1-\theta) + 3\theta\Big(-3+3\Big[1+\sqrt{\theta^{-1}-1}\Big]\theta \\ - \alpha\sqrt{(1-\theta)}\Big) \geq 0. \tag{A.3}$$

where $\alpha^2 \equiv N x_0^2$. Then, θ_T can be calculated numerically for each x_0 by requiring that this expression equals zero. A good approximation can be obtained for cases in which saturation happens when relatively few votes remain to be counted, like those shown in Fig. 5. In those cases, we can estimate θ_T by expanding the above expression in the Taylor series around zero to first order in θ and α_2. This leads directly to Eq. (17) in the main text, $\theta_T \approx \alpha^2/9 = N x_0^2/9$.

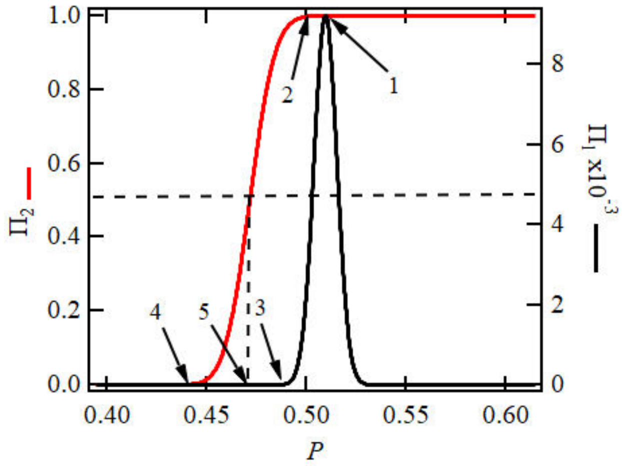

FIGURE A.1. Probabilities Π_1 (left axis) and Π_2 (right axis) with parameters such that $\Pi_T \approx 1$, showing the relevant points needed to establish this condition.

B. On why $Z_0 = x_0 n/m$

We want to calculate the probability that the position after the subsequent m steps is no more negative than minus the position after the first n steps were taken, or $-x_0$. This probability is given by the shaded are in Fig. B.1 since the new random variable Z (defined as the average of the remaining steps) is also given by a normal distribution. Recall that, by construction, both X and Z are averages instead of simple sums of steps, but also that m and n are in general different. Consequently, $-z_0$ is in not equal to $-x_0$. To clarify why this is so, suppose that after $n = 200$ votes, there is a lead given by $x_0 = 0.2$ and that there are only $m = 100$ votes left to be counted. This lead comes from a 40-vote lead since x_0 is the excess number of steps (or votes for that candidate) divided by n. But these 40 votes represent a larger percentage of the remaining m votes, $z_0 = 40/100 = 0.4$. Because this relationship is linear, in general, we will have $z_0 = x_0 n/m$.

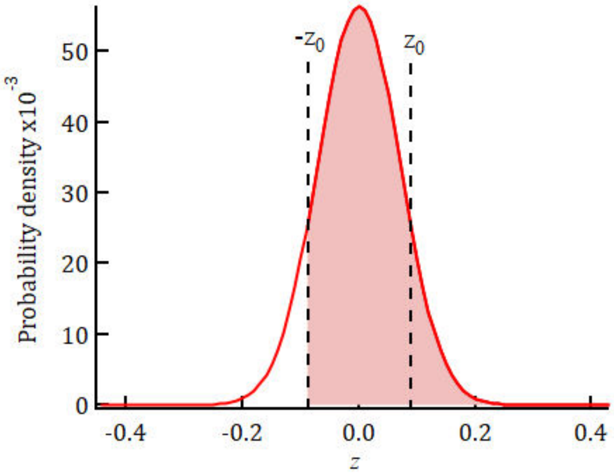

FIGURE B.1. Probability density given the random variable Z along with the position of the walk in this framework, $z_0 = x_0 n/m$. The shaded area gives the probability that the walk will not reverse the result for a given P.

C. Implementation in mathematica

Finally, we present an example of the implementation in Mathematica of Eq. (16) to calculate the probability that the leading candidate will win.

$$N_{Total} = 1000,$$

$$m = 200,$$

$$A = \frac{50.5}{100},$$

$$X_0 := 2A - 1,$$

$$Z_0 := X_0 \frac{n}{m},$$

$$\Delta x = \frac{1}{n},$$

$$\Delta P = \Delta x,$$

$$\mu_2[P_-] := 2P - 1,$$

$$\sigma 2[P_-, m_-] := \sqrt{\frac{1}{m} 4P(1-P)},$$

$$\mu_1[P_-] := 2P - 1,$$

$$\sigma 1[P_-, n_-] := \sqrt{\frac{1}{n} 4P(1-P)},$$

$$\Pi 1[P_-, n_-] := 2\Delta x \frac{e^{-(X_0 - \mu 1[P])^2 / 2\sigma 1[P,n]^2}}{\sqrt{2\pi}\sigma 1[P,n]}$$

$$\Pi 2[P_-, n_-, m_-, X0_-,] := \frac{1}{2} \times \left(\left[1 + Erf \frac{X_0 + \mu 2[P]}{\sqrt{2}\sigma 2[P,m]} \right] \right)$$

$$\text{List1} := \text{Table}[\{P, \pi 1[P,n]\}, \{p, \Delta P, 1 - \Delta P, \Delta P\}]$$

$$\text{List2} := \text{Table}[\{P, \pi 2[P,n,m,Z0]\}, \{p, \Delta P, 1 - \Delta P, \Delta P\}]$$

$$\text{ListaTotal 1} := \text{Table}[\{P, \pi 1[P,n] \pi 2[P,n,m,Z0]\} \{P, 10\Delta P, 1 - 10\Delta P, \Delta P\}]$$

Acknowledgments

The author gratefully acknowledges partial financial support from CONACyT, SNI- 52203.

1. W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd ed., Wiley, New York (1968), Chapter 3.
2. D.S. Lemons and P. Langevin, *An introduction to stochastic processes in physics* (JHU Press, USA, 2002), Chapter 5.
3. Actually, the normalization condition is not perfectly met because of the approximations made to arrive at Eq. (13). However, for $n = 1000$, the sum over all discrete values of P (Eq. (13)) yields 0.999, and the sum gets closer to 1 as n increases. Thus, we can assume that, for practical purposes, Π_1 is normalized for $n > 1000$.
4. D.V. Schroeder, *An introduction to thermal physics*, (Addison Wesley Longman, USA, 2000), Chapter 7.