Riddler April 23, 2021:
Random Voting

Mark Girard

April 26, 2021

Question 1. *Riddler Nations neighbor to the west, Enigmerica, is holding an election between two candidates, A and B. Assume every person in Enigmerica votes randomly and independently, and that the number of voters is very, very large. Moreover, due to health precautions, 20 percent of the population decides to vote early by mail.*

On election night, the results of the 80 percent who voted on Election Day are reported out. Over the next several days, the remaining 20 percent of the votes are then tallied.

What is the probability that the candidate who had fewer votes tallied on election night ultimately wins the race?

When the total number of voters is small, we can compute the answer exactly. I'll do that first before moving on to the limit of large electorates.

Exact answer for small electorates

Let N be the total number of voting residents. (For simplicity, one may assume that N is divisible by 10 to make some of the math easier, and let n be the integer for which $N = 10n$.) Define the following random variables:

- X is the number of votes for candidate A after the first $8n$ votes have been counted.
- Y is the number of votes for candidate A in the final $2n$ votes.

Assuming that each voter casts their vote independently at random, and votes for either A or B with equal probability, these variables follow a binomial distribution. That is,

$$\Pr(X = k) = \frac{1}{2^{8n}}\binom{8n}{k} \qquad \text{and} \qquad \Pr(Y = j) = \frac{1}{2^{2n}}\binom{2n}{j}$$

We want to compute the following probability

$$\Pr(X + Y > 5n \mid X < 4n)$$

(which is the probability that candidate A receives more than half of the total votes given that they received less than have of the initial 80% of the votes.) We can compute this as

$$\Pr(X + Y > 5n \mid X < 4n) = \frac{\Pr(X + Y > 5n \text{ and } X < 4n)}{\Pr(X < 4n)}.$$

Note that

$$\Pr(X < 2n/5) = \frac{1}{2^{4n/5}} \sum_{k=0}^{2n/5-1} \binom{4n/5}{k} = \Pr(X > 2n/5)$$

and that $\Pr(X < 4n) + \Pr(X > 4n) + \Pr(X = 4n) = 1$, so the probability of candidate A receiving less than half of the initlal 80% of the counted votes may be expressed as

$$\Pr(X < 4n) = \frac{1}{2}\left(1 - \frac{1}{2^{8n}}\binom{8n}{4n}\right).$$

On the other hand, the probability that candidate A receives less than half of the initial 80% but still wins the final vote count may be computed as

$$\begin{aligned}\Pr(X + Y > 5n \text{ and } X < 4n) &= \sum_{k=3n+1}^{4n-1} \Pr(X = k)\Pr(Y > 5n - k) \\ &= \frac{1}{2^{10n}} \sum_{k=3n+1}^{4n-1} \sum_{j=5n+1-k}^{2n} \binom{8n}{k}\binom{2n}{j}.\end{aligned}$$

Thus, if the total number of voters is $N = 10n$, the answer we are looking for may be expressed exactly as

$$\frac{1}{2^{10n-1} - 2^{2n-1}\binom{8n}{4n}} \sum_{k=3n+1}^{4n-1} \sum_{j=5n+1-k}^{2n} \binom{8n}{k}\binom{2n}{j}$$

For relatively small values of N, this can be computed exactly with the following code:

```
from scipy.special import comb

def sol1(N):
    b = 2**(N-1) - 2**(N//5 - 1) * comb(N*8//10, N*4//10, exact=True)

    a, c  = 0, 0
    k = N*3//10 + 1
    j = N*2//10

    while k <= (N*4//10 - 1 if N>10 else 4):
        c += comb(N*2//10, j, exact=True)
        a += comb(N*8//10, k, exact=True) * c
        k += 1
        j -= 1

    return Fraction(a, b)
```

For example, if the number of voters is equal to 10, the desired probability is computed using the above code to be equal to

$$\frac{35}{186} \approx 0.1881720430107527,$$

while, if the total number of voters is equal to 100, the desired probability is

$$\frac{241612339101337422714869594445}{2887302803308656717318208888064} \approx 0.08368098379721925.$$

Below is a table of the probability that the losing candidate (after 80% of the votes are conted) becomes the winning candidate after all the votes are counted, for different sizes of the electorate.

Size of electorate (N)	10	100	1000	10000	10^5	10^6	10^7
Probability that losing candidate ends up winning	0.1882	0.0837	0.1257	0.140514	0.14533	0.14687	0.14735

Note that using `comb(n,k)` takes a lot of time/memory for large values of n, but the binomial distribution can be well approximated using a normal distribution.

Limit of large N

In the limit of large n, a binomial distribution may be approximated by a normal distribution. If the probability of obtaining heads on a given coin is equal to p, the random variable X equal to the total number of heads after n flips is approximately normal

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$.

In our voting problem, we have $p = 1/2$ (as each voter is equally likely to vote for either A or B). As before, let X be the random variable denoting the number of votes for candidate A in the first 80% and let Y be the random variable for the number of votes for candidate A in the final 20% of the votes. These random variables have expected values

$$\mathrm{E}[X] = 0.8 \times N \times (1/2) = 0.4N \qquad \text{and} \qquad \mathrm{E}[Y] = 0.1N$$

and variances

$$\mathrm{Var}(X) = 0.8N \times (1/2)^2 = 0.2N \qquad \text{and} \qquad \mathrm{Var}(Y) = 0.05N.$$

Define now the following new random variables

$$Z_1 = \frac{Y - \mathrm{E}[Y]}{\sqrt{\mathrm{Var}(Y)}} \qquad \text{and} \qquad Z_2 = -\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}(X)}},$$

which are independent standard normal variables. Making use of the fact that $4\mathrm{Var}(Y) = \mathrm{Var}(X)$, we have that

$$\begin{aligned}\Pr(X + Y > \mathrm{E}(X) + E(Y) \text{ and } X < \mathrm{E}(X)) &= \Pr\left(\sqrt{\mathrm{Var}(Y)}Z_1 > \sqrt{\mathrm{Var}(X)}Z_2 > 0\right) \\ &= \Pr(Z_1 > 2Z_2 > 0) \\ &= \frac{\arctan(1/2)}{2\pi},\end{aligned}$$

where the last line follows from Lemma 2 below.

Lemma 2. *Let Z_1 and Z_2 be independent random variables distributed according to the standard normal distribution, and let $\alpha > 0$. It holds that*

$$\Pr(\alpha Z_1 > Z_2 > 0) = \arctan(\alpha).$$

Proof. We may evaluate the desired probability as

$$\begin{aligned}
\Pr(\alpha Z_1 > Z_2 > 0) &= \frac{1}{2\pi} \int_0^\infty \int_0^{\alpha x} \exp\left(-\frac{x^2+y^2}{2}\right) \mathrm{d}y \, \mathrm{d}x \\
&= \frac{1}{2\pi} \int_0^\infty \int_0^{\arctan(\alpha)} \exp\left(\frac{r^2}{2}\right) r \, \mathrm{d}\theta \, \mathrm{d}r \\
&= \frac{\arctan(\alpha)}{2\pi}
\end{aligned}$$

by making the change of variables $x = r\cos\theta$ and $y = r\sin\theta$. □

Finally, as it is the case that $\Pr(X < 0) = 1/2$ in the limit of large N, the desired probability is approximately

$$\frac{\Pr(X + Y > \mathrm{E}(X) + E(Y) \text{ and } X < \mathrm{E}(X))}{\Pr(X < 0)} = \frac{\arctan(1/2)}{\pi} \approx 0.147583.$$