# Using Probability to Reduce the Size of a Voting Body

M. Amorim and R.L. Robertson

**Abstract -** Suppose a large group of people, such as a university faculty, traditionally makes decisions by meeting together and voting. Because of the COVID-19 pandemic, however, large group meetings have become impractical. Having a relatively small subset of the group meet and take votes would be safer and easier to organize. The purpose of this research is to use probability theory to choose a size for this subset which guarantees that the subset has a high probability of voting the same way as the entire group. Our research involves the study of the hypergeometric probability distribution and leads to a computer algorithm for determining the subset size.

**Keywords :** probability; applied probability; hypergeometric probability

**Mathematics Subject Classification** (2020) **:** 60-04; 60C05

## 1 Introduction

At some colleges and universities, such as Drury University, the faculty vote on issues at regular meetings of the full faculty. Such meetings can be challenging to schedule. Also the COVID pandemic has made society rethink the number of people allowed in a room.

In this work, we explore how to reduce a voting body's size while maintaining a high probability that the smaller group will make the same decision as the full voting body. We consider sets of voters whose members choose between two options, usually 'yes' or 'no'. We also assume that the entire voting body always chooses the winning option by a large majority. At Drury University, for example, the full faculty often votes unanimously. Even on controversial matters, more than 60% of faculty members vote for the winning option.

If we have historical data on a voting body, it is possible to allow only a subset of voters to vote while remaining confident that the resulting decisions reflect the views of the entire voting body. First, we will determine a subset size that guarantees the same voting outcome as that of the large group. Then we will look at smaller subsets where the probabilities for the winning option to be the same are high. Throughout this paper, we assume that an individual voter will make the same decision if they are voting as part of a large group or a smaller group.

## 2 Notation

We first need to define some notation. Suppose a voting body $G$ votes 'yes' or 'no' on a proposal $V$. Here, a *proposal* could be a positive statement such as "the faculty handbook should be amended" or a negative statement such as "the faculty handbook should not be amended." Assume that everyone in $G$ votes and that the only options are to vote 'yes' or 'no.' The symbol $A$ denotes the subset containing people in $G$ who vote 'yes', and $\tilde{A}$ is the subset of those who vote 'no.' We let $a = |A|$, where $|A|$ represents the number of people in $A$. Similarly, we define $\tilde{a} = |\tilde{A}|$ and $g = |G|$. The assumptions that everyone votes and that there are only two voting options imply that $A \cup \tilde{A} = G$ and $a + \tilde{a} = g$. The probability $p$ of a randomly-selected member of $G$ voting 'yes' is then

$$p = \frac{a}{g}.$$

In this paper, we will assume that a majority of the members of $G$ vote 'yes'. Thus $a > \tilde{a}$ and $0.5 < p \leq 1$. In the numerical calculations that appear in a later section, we will use values of $p$ much larger than 0.5.

## 3 Choosing Subsets With a High Probability of Voting for $V$

Let $S$ be a subset of $G$ with size $s$. Then the members of $S$ will vote for proposal $V$ if less than half of the members of $S$ come from $\tilde{A}$, the set of members of $G$ who vote against $V$. Since $S$ can contain at most $\tilde{a}$ members who vote against $V$, $S$ is guaranteed to vote for $V$ if

$$s \geq 2\tilde{a} + 1. \tag{1}$$

Demanding that a subset $S$ vote the same as $G$ 100% of the time may produce a subset that is still quite large. For example, suppose $G$ has 100 members, and, based on historical voting patterns, we expect that at least 70% of the members of $G$ will vote for $V$. Inequality (1) says that $S$ needs at least 61 members to guarantee that $S$ will also vote for $V$.

Rather than requiring the subset to always vote for $V$, we will instead look for the minimum subset size that guarantees a high probability that the subset will vote for $V$. To find this size, we must be able to calculate the probability that a randomly-selected subset of a given size votes for $V$.

Let $n$ be a nonnegative integer, and suppose $k$ is a nonegative integer such that $k \leq n$. We use $\binom{n}{k}$ to denote the number of ways of choosing without replacement a subset of size $k$ from a set of size $n$. As is well-known from combinatorics,

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

If a subset $S$ has size $s$, what is the probability this subgroup will vote for $V$? We will first examine the question with some examples.

Suppose the entire voting body has $g = 10$ members, of which 7 will vote for $V$. If a subset has size $s = 3$, its members will vote for $V$ if at least 2 members are selected from the 7 that will vote for $V$. The number of subsets of size 3 where exactly 2 members vote for $V$ is

$$\binom{7}{2}\binom{3}{1} = 63.$$

The number of subsets of size 3 where exactly 3 members vote for $V$ is

$$\binom{7}{3}\binom{3}{0} = 35.$$

So the number of subsets of size 3 that will choose $V$ is

$$\binom{7}{2}\binom{3}{1} + \binom{7}{3}\binom{3}{0} = 98.$$

The probability of choosing a subset of size 3 that votes for $V$ equals the number of subsets that vote for $V$ divided by the number of all possible subsets of size 3:

$$\frac{\binom{7}{2}\binom{3}{1} + \binom{7}{3}\binom{3}{0}}{\binom{10}{3}} = \frac{98}{120} = .82.$$

That means 82% of subsets of size 3 will choose $V$.

One might suspect that choosing a larger subset size will always give a higher probability. This is not true, though. Consider a subset where $s = 4$. The subset will vote for $V$ if at least 3 members are selected from the 7 that will vote for $V$. The number of subsets of size 4 where exactly 3 members vote for $V$ is

$$\binom{7}{3}\binom{3}{1} = 105,$$

and the number of subsets of size 4 where exactly 4 members vote for $V$ is

$$\binom{7}{4}\binom{3}{0} = 35.$$

So the number of subsets that will choose $V$ is

$$\binom{7}{3}\binom{3}{1} + \binom{7}{4}\binom{3}{0} = 140.$$

The probability of choosing a subset of size 4 that votes for $V$ equals the number of subsets that vote for $V$ divided by all the possible subsets of size 4:

$$\frac{\binom{7}{3}\binom{3}{1} + \binom{7}{4}\binom{4}{0}}{\binom{10}{4}} = \frac{140}{210} = .67.$$

That means 67% of groups of size 4 will choose $V$. Note that this is smaller than the probability for a subgroup of size 3.

In general, suppose a voting body $G$ of size $g$ contains $a$ members that vote for $V$ and $\tilde{a}$ that vote against $V$. If a subset of size $s$ is selected from $G$, the probability that $i$ members of the subgroup vote for $V$ is given by the hypergeometric probability

$$\frac{\binom{a}{i}\binom{\tilde{a}}{s-i}}{\binom{g}{s}}. \tag{2}$$

See [1], pgs. 208-216, for a thorough discussion of hypergeometric probabilities.

A majority of the subset votes for $V$ if more than half its members vote for $V$. That is,

$$i \geq \left\lfloor \frac{s}{2} \right\rfloor + 1.$$

Here, $\left\lfloor \frac{s}{2} \right\rfloor$ is the largest integer that is less than or equal to $\frac{s}{2}$.

Suppose we randomly-select a subset of $G$. Let $Y$ (short for "yes") denote the event where a majority of the subset members vote for proposal $V$. In set-theoretic notation,

$$Y = \{S \subseteq G \,\big|\, |S \cap A| > |S \cap \tilde{A}|\}.$$

For a fixed nonnegative integer $s$, the probability of randomly-selecting a subset $S$ of size $s$ where $V$ wins is given by the conditional probability $P(Y \,\big|\, |S| = s)$. This probability equals the sum of probabilities of the form (2) where $i$ starts with $\lfloor \frac{s}{2} \rfloor + 1$ and ends with $s$:

$$P(Y \,\big|\, |S| = s) = \sum_{i=\lfloor \frac{s}{2} \rfloor + 1}^{s} \frac{\binom{a}{i}\binom{\tilde{a}}{s-i}}{\binom{g}{s}}. \tag{3}$$

For now, we use the convention that $\binom{n}{k} = 0$ if $k > n$. This convention would be needed if $s > a$. As we will see, in practice the subset sizes will be smaller than $a$.

We can now express the problem of choosing the smallest subset size that guarantees a high probability that its members vote for $V$ in terms of $P(Y \,\big|\, |S| = s)$. Let $q$ be a probability close to 1 (e.g. $q = .95$). Find the smallest $s$ such that

$$P(Y \,\big|\, |S| = s) \geq q. \tag{4}$$

If $q = .95$, for example, finding the smallest $s$ that satisfies this inequality would give a subset size that guarantees that at least 95% of subsets of that size would vote for $V$.

## 4 Odds and Evens

Solving inequality (4) algebraically is difficult. For that reason, we wrote Python code to find the smallest value of $s$ that satisfies (4). A straightforward algorithm would be to first calculate $P(Y \,\big|\, |S| = s)$ when $s = 1$, and then repeatedly increase $s$ by 1, calculating $P(Y \,\big|\, |S| = s)$ each time. The algorithm would stop when it finds the first value of $P(Y \,\big|\, |S| = s)$ that is greater than or equal to the probability $q$.

Numerical experiments suggested to us that if $s$ is odd, then $P(Y \,\big|\, |S| = s)$ is larger than $P(Y \,\big|\, |S| = s+1)$. The next theorem establishes that this is true in general, and it implies that the smallest value of $s$ that satisfies (4) will be odd.

**Theorem 4.1** *If $s$ is odd, then*

$$P(Y \,\big|\, |S| = s) > P(Y \,\big|\, |S| = s+1) \tag{5}$$

**Proof.**

Let $\tilde{Y}$ be the event where at least half of the members of a randomly-chosen subset vote against $V$. Then

$$P(Y \,\big|\, |S| = s) + P(\tilde{Y} \,\big|\, |S| = s) = 1,$$

and

$$P(Y \,\big|\, |S| = s+1) + P(\tilde{Y} \,\big|\, |S| = s+1) = 1.$$

Thus inequality (5) is equivalent to the complementary inequality

$$P(\tilde{Y} \,\big|\, |S| = s) < P(\tilde{Y} \,\big|\, |S| = s+1).$$

As in equation (3), the complementary probabilities are sums of hypergeometric probabilities:

$$P(\tilde{Y} \,\big|\, |S| = s) = \sum_{i=0}^{\lfloor \frac{s}{2} \rfloor} \frac{\binom{a}{i}\binom{\tilde{a}}{s-i}}{\binom{g}{s}}$$

and

$$P(\tilde{Y} \,\big|\, |S| = s+1) = \sum_{i=0}^{\lfloor \frac{s+1}{2} \rfloor} \frac{\binom{a}{i}\binom{\tilde{a}}{s+1-i}}{\binom{g}{s+1}}.$$

So proving the theorem is equivalent to proving

$$\sum_{i=0}^{\lfloor \frac{s}{2} \rfloor} \frac{\binom{a}{i}\binom{\tilde{a}}{s-i}}{\binom{g}{s}} < \sum_{i=0}^{\lfloor \frac{s+1}{2} \rfloor} \frac{\binom{a}{i}\binom{\tilde{a}}{s+1-i}}{\binom{g}{s+1}}. \tag{6}$$

To simplify the terms being summed, we will express $\binom{g}{s+1}$ in terms of $\binom{g}{s}$. Notice that

$$\frac{\binom{g}{s+1}}{\binom{g}{s}} = \frac{\frac{g\,!}{(s+1)\,!(g-s-1)\,!}}{\frac{g\,!}{(s)\,!(g-s)\,!}} = \frac{g-s}{s+1}.$$

So

$$\binom{g}{s+1} = \frac{g-s}{s+1}\binom{g}{s}. \tag{7}$$

Substituting (7) into inequality (6) and canceling the common factor of $\binom{g}{s}$ from both sides gives the following inequality, which is equivalent to (6):

$$\sum_{i=0}^{\lfloor \frac{s}{2} \rfloor} \binom{a}{i}\binom{\tilde{a}}{s-i} < \sum_{i=0}^{\lfloor \frac{s+1}{2} \rfloor} \frac{s+1}{g-s}\binom{a}{i}\binom{\tilde{a}}{s+1-i}. \tag{8}$$

Since $s$ is odd, $\lfloor \frac{s}{2} \rfloor = \frac{s-1}{2}$ and $\lfloor \frac{s+1}{2} \rfloor = \frac{s+1}{2}$. Thus the sum on the right side of inequality (8) has one more term than the sum on the left side (if $s$ were even, both sums would have the same number of terms). We can thus rewrite the right side of inequality (8) as

$$\sum_{i=0}^{\lfloor \frac{s+1}{2} \rfloor} \frac{s+1}{g-s}\binom{a}{i}\binom{\tilde{a}}{s+1-i} = \frac{s+1}{g-s}\binom{\tilde{a}}{s+1} + \sum_{i=0}^{\frac{s-1}{2}} \frac{s+1}{g-s}\binom{a}{i+1}\binom{\tilde{a}}{s-i}.$$

On the right side, the term outside the summation is always nonnegative. We can thus finish the proof by showing that if $0 \le i \le \frac{s-1}{2}$, then

$$\frac{s+1}{g-s}\binom{a}{i+1}\binom{\tilde{a}}{s-i} > \binom{a}{i}\binom{\tilde{a}}{s-i}.$$

A calculation similar to the one used to show (7) gives

$$\binom{a}{i+1} = \frac{a-i}{i+1}\binom{a}{i}.$$

We will finish by showing that if $0 \le i \le \frac{s-1}{2}$, then

$$\left(\frac{s+1}{g-s}\right)\left(\frac{a-i}{i+1}\right) > 1.$$

Rewrite the left side of this inequality as

$$\left(\frac{s+1}{2i+2}\right)\left(\frac{2a-2i}{g-s}\right).$$

Since $i \le \frac{s-1}{2}$, then $2i \le s-1$. So $2i+2 \le s+1$. Hence $\frac{s+1}{2i+2} \ge 1$. Also, a majority of the members of $G$ are in $A$. Thus $2a > g$. So $g - s < 2a - s \le 2a - 2i - 1 < 2a - 2i$, which implies $\frac{2a-2i}{g-s} > 1$. □

Theorem (4.1) implies that the algorithm described at the beginning of the section only needs to calculate (3) for odd values of $s$.

We implemented the algorithm using Python. The program uses the scientific calculation library Scipy to calculate the necessary hypergeometric probabilities (see [3]). For those interested in experimenting with the program, the code is available at [2].

The method to find the smallest value of $s$ has three inputs: the voting body size $g$, the percentage of votes for option $A$, and the wanted probability $q$. Taking these values,

| | |
|---|---|
| 1 | 0.698 |
| 2 | 0.485 |
| 3 | 0.784 |
| 4 | 0.649 |
| 5 | 0.839 |
| 6 | 0.745 |
| 7 | 0.879 |
| 8 | 0.810 |
| 9 | 0.908 |
| 10 | 0.857 |

| | |
|---|---|
| 11 | 0.930 |
| 12 | 0.892 |
| 13 | 0.947 |
| 14 | 0.918 |
| 15 | 0.960 |
| 16 | 0.938 |
| 17 | 0.970 |
| 18 | 0.953 |
| 19 | 0.978 |
| 20 | 0.965 |

Table 1: Probabilities for Subsets of the Drury Faculty Assuming $p = .70$

it calculates $P(Y \,\big|\, |S| = s)$ for odd values of $s$ starting at $s = 1$ until reaching the first value such that $P(Y \,\big|\, |S| = s) \geq q$. At this point, it returns $s$. The user can also choose to see lists of values of $P(Y \,\big|\, |S| = s)$.

As Theorem 4.1 implies, increasing $s$ does not necessarily increase $P(Y \,\big|\, |S| = s)$. We believe that the possibility of a tie in an even-sized subset is the cause of the up and down behavior of the values of $P(Y \,\big|\, |S| = s)$.

## 5 Numerical Examples

The algorithm can be applied to our initial example on Drury University's faculty voting problem. Drury has 116 voting faculty. Based on historical voting patterns, a reasonable minimum value for $p$ is 70%. The Python code returned the values seen in Table 1. In the table, the left column is the size of a randomly-selected subset of the faculty. The right column is the probability that the subset will vote the same as the entire faculty.

These numerical results show that even small subsets of the entire faculty have a high probability of voting the same as the full faculty. A subset of size 9 will vote the same as the full faculty 90% of the time. Increasing the subset size to 15 increases the probability to 96%. Notice that our numerical results follow Theorem 4.1, as each odd-sized subset probability is always larger than the probability for the subsequent even-sized subset.

Other numerical results show that even for large voting bodies, small subsets have a high probability of voting the same as the entire body. For example, suppose a voting body has 100000 members, and that we have data that suggests that 75% of this body will vote on option $A$. The probabilities for the smallest 20 subgroups appear in table 2.

Nearly all (99.1%) subsets of size 19 will vote for the same option as the entire body. Results are still good for subsets with as few as 9 members. These numerical calculations suggest that changing $p$ has a larger effect on the probabilities than increasing the size of the voting body. Again, we can see how the probabilities of the odds and evens follow the pattern predicted by Theorem 4.1.

| | |
|---|---|
| 1 | 0.75 |
| 2 | 0.562 |
| 3 | 0.844 |
| 4 | 0.738 |
| 5 | 0.896 |
| 6 | 0.831 |
| 7 | 0.929 |
| 8 | 0.886 |
| 9 | 0.951 |
| 10 | 0.921 |
| 11 | 0.966 |
| 12 | 0.945 |
| 13 | 0.975 |
| 14 | 0.962 |
| 15 | 0.982 |
| 16 | 0.973 |
| 17 | 0.988 |
| 18 | 0.981 |
| 19 | 0.991 |
| 20 | 0.986 |

Table 2: Probabilities for a Body of Size 10000 with $p = .75$

# 6 Possibilities for Future Work and Conclusion

The results in this paper depended on two assumptions. First, we assumed that the subset was randomly-chosen from the entire voting body. Second, we assumed that an individual in a small-group setting would always vote the same as if they were voting with the entire voting body.

Large voting bodies are typically divided into different groups, each with their own interests and agendas. University faculty, for example, are grouped into departments and colleges. Faculty in a behavioral sciences department might have stronger opinions on a specific issue than faculty in a biology department.

Randomly choosing a subset of a voting body could leave some groups with no representation. In future work, we hope to generalize the work done in this paper to a voting body that is divided into subgroups, each of which has a different probability of voting for a particular issue, and each of which must be represented in the subset that is selected.

Also, individuals in a small subset may vote differently than they would if they were voting with a larger group of people. In a small-group setting, people might be more comfortable advocating for one option over another. Lively discussions could sway some voters to change their votes. We want to explore this phenomenon further.

Theorem 4.1 says that if $s$ is odd, then $P(Y \,\big|\, |S| = s) > P(Y \,\big|\, |S| = s+1)$. Numerical results suggest that if $x = (s+1)/2$, then the following is true:

$$P(Y \,\big|\, |S| = s) = P(Y \,\big|\, |S| = s+1) + \frac{\binom{a}{x}\binom{\tilde{a}}{x}}{2\binom{g}{s+1}}. \tag{9}$$

The last term on the right is half the probability that the votes in a randomly-selected subset of size $s+1$ are tied. If (9) is true in general, then Theorem 4.1 follows. We hope to prove (9) in future work.

Given our assumptions, however, we see that tiny subsets of a voting body will often vote the same as the entire body. For a university faculty, the results suggest that a small faculty senate could represent the entire faculty.

## Acknowledgments

We thank the referee for many helpful suggestions.

## References

[1] N. Weiss, *A course in probability*, Pearson, 2006

[2] Python program code, available online at the URL: `https://github.com/mamorim00/Probability_Research`

[3] Scipy documentation, available online at the URL: `https://docs.scipy.org/doc/scipy`

*Marina Amorim*
Drury University
900 N. Benton Avenue
Springfield, Missouri 65802
E-mail: `mamorim@drury.edu`

*Robert L. Robertson*
Professor of Mathematics
Drury University
900 N. Benton Avenue
Springfield, Missourie 65802
E-mail: `rroberts@drury.edu`

**Received**: December 13, 2022 **Accepted**: February 6, 2023
**Communicated by George Jennings**