The Optimal Size of Congress Under the Epistemic Approach and Consequences for Modern Models of Governance

Class project for CS238 by Tao Lin, Daniel Halpern, and Manon Revel

April 28, 2021

However small the Republic may be, the Representatives must be raised to a certain number, in order to guard against the cabals of a few; and however large it may be, they must be divided to certain number, in order to guard against the confusion of a multitude. (Federalist Paper No. 10)

– James Madison

Abstract

We analyze the optimal size of a congress in a representative democracy. Prior work has evaluated several models for society and voting behavior reaching conclusions that the best size of a congress should be proportional to the cube-root or the square-root of the size of the population. We take an epistemic view where voters decide on a binary issue with one ground truth outcome, and each voter votes correctly according to their competence levels in $[0, 1]$. Under mild assumptions, we find that the optimal congress size should be linear in the population size. We then analyze real world data, finding that the actual sizes of congresses are much smaller than the optimal size our theoretical results suggest. We conclude by analyzing under what conditions congresses of sub-optimal sizes would still outperform direct democracy.

1 Introduction

Modern governments often take the form of a representative democracy, that is, a college of chosen representatives form a congress to make decisions on behalf of the citizenry. Clearly, the performance of the congress depends the number of representatives, and the optimal number of representatives has been subject to great debates. A wealth of research dating back to the 1960s concluded that the number of representatives should follow a "cube-root law" as a function of the population size [Taagepera, 1972]. Yet, recent work using machinery from physics and economics revisited these claims and showed that, under different assumptions, the optimal number should be larger, proportional to the square-root of the population size [Auriol and Gary-Bobo, 2012, Margaritondo, 2021]. We contribute to this literature by offering a different perspective: finding the optimal number of representatives through an epistemic approach. In our setting, society decides on a binary issue and aims at differentiating between the ground truth correct choice and its alternative. Each voter has a competence level in $[0, 1]$ representing the probability that the voter votes correctly. Further, the competence levels of the population are drawn according to some fixed distribution. We take the idealized view that given a target congress size k, we can identify the k most competent voters in society to form the congress and vote on the issue, following the majority opinion. We conclude that, should voters' competence levels follow uniform distribution $\mathcal{U}(0, 1)$, the optimal size of congress is lower bounded by $\Omega(\frac{n}{\log n})$, with experimental evidence suggesting that the true optimal size is $\Theta(n)$. For arbitrary distributions where the maximum competence level is bounded away from 1 and the inverse cumulative distribution function is Lipschitz continuous, the optimal size is $\Theta(n)$. We note that these results suggest an optimal size much larger than that in previous work. Even relaxing the assumption that the most competent voters can be reliably identified, this should only increase the optimal size. We then turn to studying real-world data on the sizes of countries' representative bodies. Here, we notice that congresses in real world are much smaller than the optimal our theoretical results suggest. We then ask under what conditions on the distribution of competence level could a smaller congress still outperform the majority. We find that, for population whose average level of competence is biased above 0.5, a relatively small congress can be better than the majority as long as the bias is small enough.

The rest paper is structured as follow: After discussing related works, we present our model and the problems of interest in Section 2. Then, Section 3 states that the optimal size of congress is near-linear and linear for different competence level distributions. Studying real-world data in Section 4, we find that the sizes of congresses around the world are of order $n^{0.36}$. Hence, we identify in Section 5 conditions under which that a congress of sub-optimal size outperforms majority voting. All proofs can be found in the Appendix.

1.1 Related Work

There has been much research on the optimal size of congresses (or parliaments). We present some of the most important works here.

The first work [Taagepera, 1972] about the optimal size of parliaments focused on maximizing parliament's efficiency, which was said to be related to the time congress-members spent communicating with their constituents. Maximizing efficiency was equivalent in the paper to minimizing the communication time spent on discussions with constituents — the authors ultimately stated that the average time spent talking to the constituents per congress-members should be equal to the time spent talking to the other congress-members. Hence, Taagepera [1972] argued that the optimal congress size should follow a "cube-root law". However, recently, Margaritondo [2021] revisited this work and found a flaw in the original proof, and argued that the optimal size under this model should in fact be $\Theta(\sqrt{n})$.

These findings are regarded as seminal [Jacobs and Otjes, 2015] and have influenced political decisions and referendums [Margaritondo, 2021, De Sio and Angelucci, 2018]. Interestingly, empirical papers [Taagepera, 1972, Auriol and Gary-Bobo, 2007] that focused on finding the optimal number of representatives used country data to back up the "square-root law" – however, such conclusions are by no means causal. Jacobs and Otjes [2015] aims at investigating potential causal effects.

The work of Auriol and Gary-Bobo [2012] also aims to derive the optimal number of representatives for a society. However, their criteria for optimality lies in stark contrast to ours. At the heart of their criteria is a preference-based utilitarian rather than epistemic approach to decision making. They assume a very uninformative prior, while we assume a fully-informative and restricted distribution over competencies. Further, they assume that the representatives are chosen uniformly at random from society, while we only take the best. Finally, their set of feasible mechanisms is incomparable to ours, including some beyond simple majority, but requiring additional strategic restrictions. They reach the conclusion that the optimal size of congress is proportional the square-root of the population size.

The most recent work on the topic [Zhao and Peng, 2020] looks at the optimal number of representatives as the minimum size of a node set such that all nodes in that set can reach other nodes in at most m steps (where m is an exogenous threshold). In this set up, the authors argue that the optimal number of representatives is n^γ where $\frac{1}{3} \leq \gamma \leq \frac{5}{9}$ assuming $m = \Theta(\log n)$.

Finally, the use of an *epistemic* approach, using voting to aggregate objective opinions, is well studied in computational social choice [Brandt et al., 2016]. One particularly important result is known as the Condorcet Jury Theorem [Grofman et al., 1983], which our work loosely builds on.

2 Problem Statement

Let n be the number of voters in our society. Following the *epistemic* approach, voters choose between two options, 0 and 1, where 1 is assumed to be the ground truth. Each voter i is endowed with a level of expertise (or competence) $p_i \in [0, 1]$ that amounts to the probability that person i votes correctly, that is, votes for the option 1. We further assume that the p_i's are sampled from some distribution $\mathcal{D}$ whose support is contained in $[0, 1]$. We sort voters by decreasing order of competence level, denoted by $p_{(1)} \geq \cdots \geq p_{(n)}$, so $p_{(i)}$ is the competence of the i^{th} best voter. Let $X_{(1)}, \ldots, X_{(n)}$ be Bernoulli random variables denoting the votes of the voters, with $X_{(i)} = 1$ meaning a correct vote and 0 otherwise; $X_{(i)}$'s are conditionally independent given $p_{(i)}$'s, and $\Pr[X_{(i)} = 1] = p_{(i)}$. Our goal is to choose a size k for our *congress*, which is assumed to be the k

most competent voters of the society. A congress of size k makes a correct decision when $\sum_{i=1}^{k} X_{(i)} > k/2$.[1] We define $K^\star$ to be the optimal size of congress, which is the size k that maximizes the probability that the congress makes a correct decision. Formally,

$$K^\star = \underset{1\le k\le n}{\arg\max} \left\{ \Pr\left[\sum_{i=1}^{k} X_{(i)} > \frac{k}{2} \;\middle|\; X_{(i)} \sim \text{Bernoulli}(p_{(i)}) \right] \right\}, \tag{1}$$

which is itself a random variable over the random draw of $p_{(i)}$'s (or p_i's) from $\mathcal{D}$.

For tractability reasons, we will sometimes assume that each $p_{(i)}$ is deterministically equal to its expectation since these values are close to their expectation with high probability. We conjecture that this assumption will not affect the optimal k asymptotically. This makes $K^\star$ a deterministic value.

Finally, in Section 5, we will also be interested in when congresses of certain sizes k outperform majority. We let $\Gamma_n^{\mathcal{D}}(k)$ be the gain in probability of correctness by using a congress of size k instead of the entire population. More formally,

$$\Gamma_n^{\mathcal{D}}(k) = \Pr\left[\sum_{i=1}^{k} X_{(i)} > \frac{k}{2} \right] - \Pr\left[\sum_{i=1}^{n} X_{(i)} > \frac{n}{2} \right]. \tag{2}$$

Note that unlike in the definition of $K^\star$, we are not treating Γ as a random variable, rather, the randomness of the p_i samples is captured in the probability. We will be interested in for certain values of k, what distributions $\mathcal{D}$ (which may depend on n) $\Gamma_n^{\mathcal{D}}(k)$ is positive as n grows large.

3 Optimal Congress Size

In this section, we prove theoretical bounds on the optimal size of congress for several natural distributions.

3.1 Uniform $[0,1]$ Distribution

Here, we will focus on the case where competence levels are drawn from uniform distribution $\mathcal{U}(0,1)$. As mentioned in Section 2, we will assume that the competence levels are exactly equal to their expectation; in other words, $p_{(i)} = \frac{n+1-i}{n+1}$ [Ma, 2010]. In this case, the top experts have competencies very close to 1. We find that even with top experts becoming arbitrarily accurate, the optimal congress size still remains very large. In particular, it is much larger than the cube-root (or even square-root) results of previous work.

Theorem 3.1. *Let $\mathcal{D} = \mathcal{U}(0,1)$, then $K^\star = \Omega(\frac{n}{\log n})$.*

We provide a proof sketch here; the full proof is relegated to Appendix A.1. The proof relies on the observation that for an odd $k = O(\frac{n}{\log n})$, one can always add two experts and increase accuracy of the congress. The two added experts are only relevant to the outcome when (i) exactly $\frac{k+1}{2}$ out of the k initial experts are correct, so adding two incorrect expert reverses the majority's decision or when (ii) exactly $\frac{k+1}{2}$ out of the k initial experts are incorrect, so adding two correct expert reverses the majority's decision. The rest of the proof follows from comparing the probability of the first scenario (exactly $\frac{k+1}{2}$ out of k experts are correct and two added experts are incorrect) and the second scenario (exactly $\frac{k+1}{2}$ out of k experts are incorrect and two added experts are correct); if the probability of the second scenario is larger, adding two experts increases the accuracy of the congress.

We in fact conjecture something even stronger.

Conjecture 1. *For $\mathcal{D} = \mathcal{U}(0,1)$, $K^\star = \Omega(n)$.*

[1] Here we are asking for a strict majority for the correct outcome, which forces the optimal k to be odd. Weak majority would force the optimal k to be even and would likely not change the asymptotic results.

Such a result would suggest that even with arbitrarily accurate experts, it is still beneficial to include those that have a constant probability of getting the wrong answer. In support of Conjecture 1, we have computed for several values of n (up to $n = 102$) the optimal congress sizes using a brute force method. The results, along with a line of best fit, can be found in Figure 1.

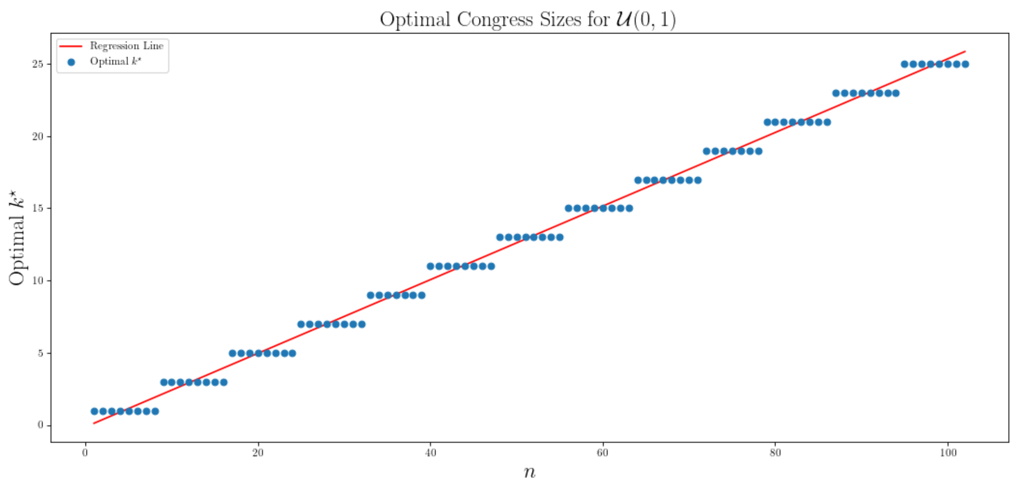


Figure 1: Optimal value of k for $\mathcal{U}(0,1)$ competence levels following their expectation.

3.2 Distributions Bounded Away From 1

In this section, we will consider a broad class of distributions which do not allow for arbitrarily accurate experts. We do not fix $p_{(i)}$ to be their expectation; instead, they are random draws from $\mathcal{D}$. Under relatively mild conditions, we show that the optimal size $K^\star$ grows linearly in the population size with high probability.

Theorem 3.2. *Let $\mathcal{D}$ be any distribution supported by $[L, H]$ with constants $L \geq 0$ and $H < 1$, with cumulative distribution function $F(\cdot)$. If $\frac{2}{3} < H < 1$ and $F^{-1}(\cdot)$ is M-Lipschitz continuous for some constant $0 < M < \infty$, then $K^\star = \Omega(n)$ with probability at least $1 - \frac{1}{n}$.*

The proof technique is similar to Theorem 3.1's, and is relegated to Appendix A.2. To see this in action, we consider a specific distribution as follows.

Example 3.3. *Let $\mathcal{D} = \mathcal{U}(L, H)$ with $L = 0.1$ and $H = 0.9 > \frac{2}{3}$, then $F(x) = \frac{x-L}{H-L}$, $M = H - L$, and asymptotically,*

$$\frac{K^\star}{n} > 1 - \frac{\left(1 - \frac{1}{1+\sqrt{\frac{2H}{1-H}}}\right) - L}{H - L} = 0.113$$

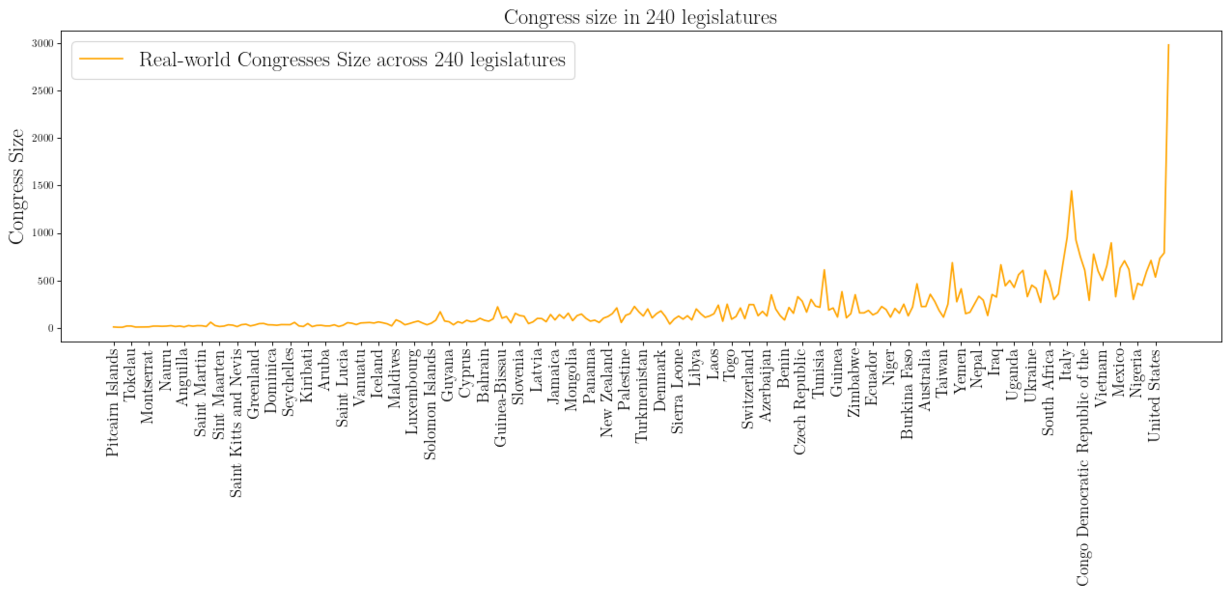
with probability at least $1 - \frac{1}{n}$.

4 Congresses Around The World

As discussed in the introduction, prior work has suggested that the size of congress should be near the cube-root of the population size. Exploring real-world data[2], we notice that this rule seems to be followed

[2]These data come from Wikipedia: `https://en.wikipedia.org/wiki/List_of_legislatures_by_number_of_members`.

quite often by real legislatures. To support this claim, we ran a linear regression on the log of the congress sizes of many countries compared to the log of the population size, which yields a slope of 0.36 (with an intercept of -0.65 and a coefficient of determination $R^2 = 0.85$), suggesting $k = cn^{0.36}$. The results can be seen from Figure 2. (This is an update of the regression analysis done by Auriol and Gary-Bobo [2012] using more recent data.)



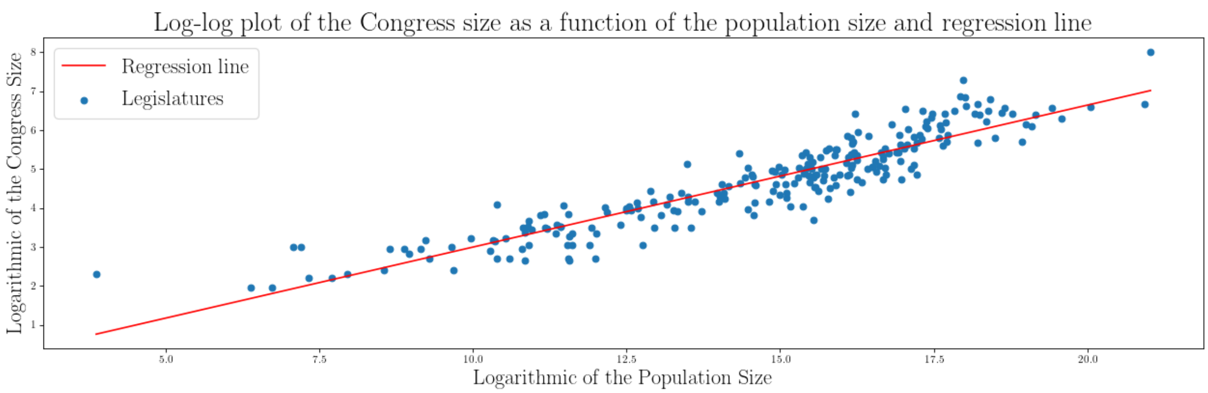


Figure 2: Congress sizes in 240 legislatures (top) and log-log plot of the Congress size as a function of the Population size. The regression line yields $\log k = 0.36 \log n - 0.65$, or $k = cn^{0.36}$, with a coefficient of determination $R^2 = 0.85$.

Our theoretical results suggest that under the epistemic framework, these real world congress sizes are not optimal. Some natural followup questions arise: How much worse are these sizes than the optimal? And when can these smaller congresses still outperform direct democracy?

5 Can a Small Congress Outperform Direct Voting?

Let us again focus on the case where the expertise levels are drawn from a (slightly more general) uniform distribution $\mathcal{U}(L, H)$. We will now ask under what conditions a smaller congress can at least outperform majority. However, there is one caveat. If the expected competence level of the distribution $\frac{L+H}{2} \leq 0.5$, letting the entire population vote will lead to the correct answer at most 1/2 of the time (and approaching 0 as n grows large if the inequality is strict). Hence, we will instead focus on scenarios where the expected competence level is strictly larger than 0.5, that is, the society as a whole is biased towards the correct answer.

5.1 Dictatorship

First, we will consider an extreme case: when can a single voter outperform the entire society? In particular, when there are n voters, we assume that their competencies are drawn from distribution $\mathcal{U}(2\varepsilon_n, 1)$ for some sequence $(\varepsilon_n)_{n=1}^{\infty} > 0$. This ensures that society is (slightly) biased towards the correct outcome. Further, since the upper end of the distribution's support is 1, the top expert becomes arbitrarily accurate in expectation as the society grows. Under these assumptions, we find the following boundary condition on ε_n for the congressional benefit to occur.

Theorem 5.1.

- *If* $\varepsilon_n = o\left(\sqrt{\frac{\log n}{n}}\right)$*, then for sufficiently large* n*,* $\Gamma_n^{\mathcal{U}(2\varepsilon_n,1)}(1) > 0$.
- *If* $\varepsilon_n = \omega\left(\sqrt{\frac{\log n}{n}}\right)$*, then for sufficiently large* n*,* $\Gamma_n^{\mathcal{U}(2\varepsilon_n,1)}(1) < 0$.

The proof of this result can be found in Appendix A.3.

5.2 Real-world Congress

Now, let us consider the case suggested by the regression results in Section 4, $k = n^{0.36}$. Let $p_i \sim \mathcal{U}(L + \varepsilon_n, 1 - L)$ for some $L \in [0, 1/2)$ and for all $i \in [n]$. This captures scenarios where $\mathbb{E}[p_i] = \frac{1+\varepsilon_n}{2}$, so the entire society is slightly biased toward the correct answer. Herein, we identify numerically sequences $(\varepsilon_n)_{n=1}^{\infty}$ such that the congress of size k outperforms direct democracy, i.e., $\Gamma_n^{\mathcal{U}(L+\varepsilon_n,1-L)}(n^{0.36}) > 0$, for sufficiently large n. As this choice of k does not offer clean analytical solutions, we turn to experimental simulation.

Our simulations were run as follows. First, for a given distribution, we sample n competencies and votes associated with these competencies. Next, we check whether a majority of all voters, and a majority of the top k voters, were correct. By repeating this $1,000$ times, we can estimate the difference between the probabilities that the congress was correct and all voters were correct (which is Γ). We test this over distributions with different values of L and orders of ε_n. The results of these simulations are in Figure 3. Unsurprisingly, we see that the larger the bias, the smaller the gain. For $L \leq 0.1$ and a bias of order $\sqrt{\frac{\log n}{n}}$, there is a no gain from relying on the congress, while on the other hand, if the bias is of order $\sqrt{\frac{\log \log n}{n}}$, there is positive gain. Yet, for $L = 0.4$, a bias of order $\sqrt{\frac{\log n}{n}}$ systematically yields a strictly negative gain for $n \leq 10^6$ (see dotted line at $y = 0$). In this regime, the plot suggests that there exists a critical population size such that above it, the gain is positive, and below, the gain is negative.

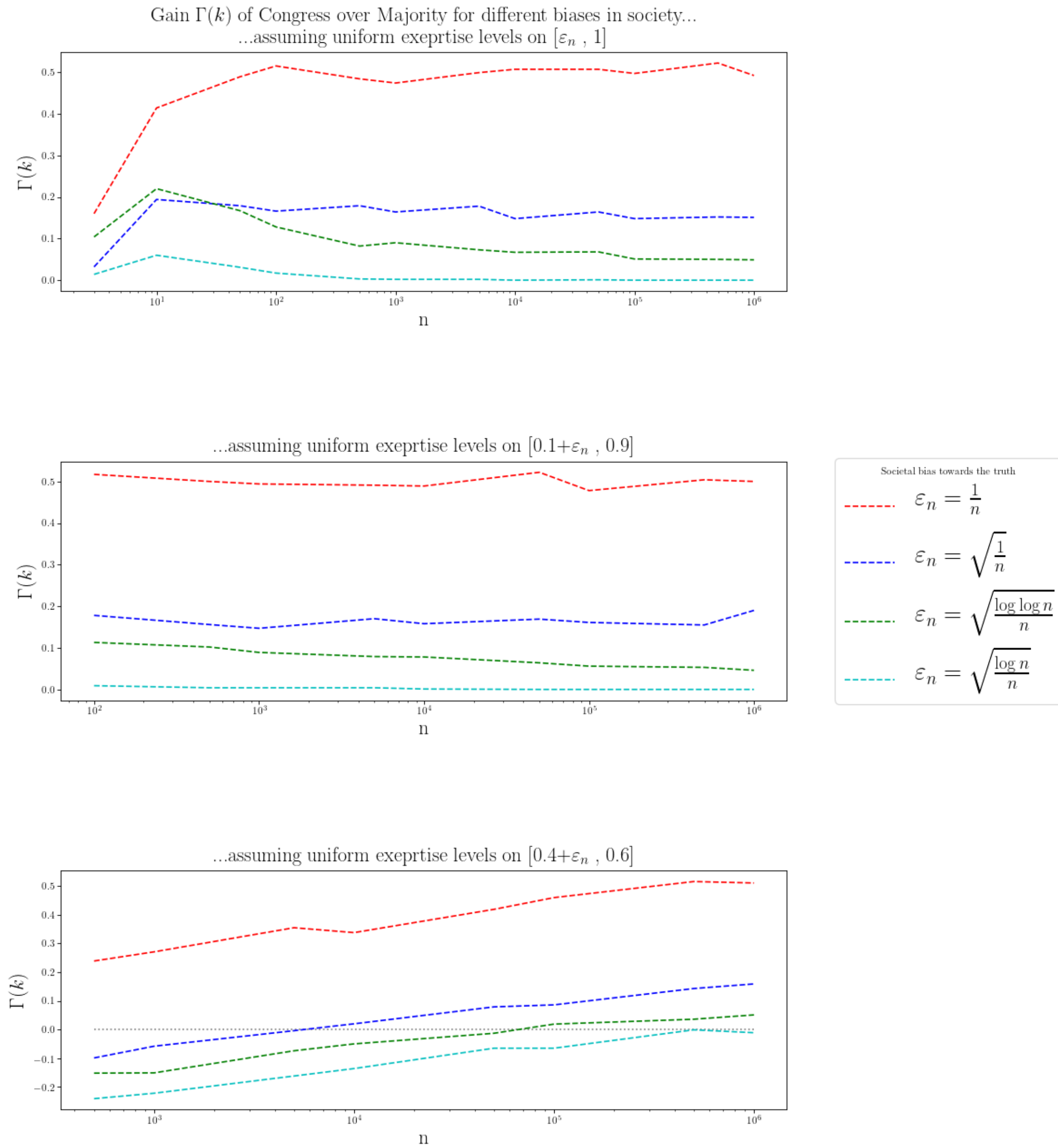


Figure 3: $\Gamma_n^{\mathcal{U}(L+\varepsilon_n,1-L)}(n^{0.36})$ as a function of the population size for different values of ε_n. Recall that $\Gamma_n^{\mathcal{D}}(k) = \mathbb{P}(\sum_{i=1}^k X_{(i)} > \frac{k}{2}) - \mathbb{P}(\sum_{i=1}^n X_{(i)} > \frac{n}{2})$ is the gain from using a congress of size k over the majority. Note that $\mathbb{E}[p_i] = \frac{1+\varepsilon_n}{2}$ so ε_n can be thought of as the bias of society towards the correct answer. The top image is for $L = 0$, the middle one is for $L = 0.1$ and the bottom one for $L = 0.4$.

From these simulations, we conjecture the following.

Conjecture 2. *There is some critical sequence $\eta_1, \eta_2, \dots$ with $\sqrt{\frac{\log\log n}{4n}} < \eta_n < \sqrt{\frac{\log n}{4n}}$ such that the following holds. For all $l_0 \in (0, 0.5)$, there is some $n_0 \in \mathbb{N}$ such that, when $L \le l_0$ and $\varepsilon_n \le \eta_n$ for all n, for $n \ge n_0$, $\Gamma_n^{\mathcal{U}(L+\varepsilon_n,1-L)}(n^{0.36}) > 0$.*

6 Discussion

We proved that under mild conditions, through the lens of an epistemic approach, current congresses are run with a sub-optimal size. However, despite this, it seems that these smaller congresses can still be cogent by at least beating majority under appropriate conditions.

What is somewhat striking is that the current debates about the number of representatives in democracies tend to be about reducing their size, not increasing.[3] Indeed, even under the assumption that a larger congress would lead to a "correct" answer more often, this is clearly not the only factor to consider. Even under the strong assumption that the congress-members' votes reflect those of the top experts in society, congress-members are costly for the taxpayers. Beyond this, the legitimacy and representativeness of the institution are constantly under scrutiny. While these factors are related to the congress-members accuracy, they are not equivalent.[4] Hence, regulating the size of congress should couple mathematical, cognitive, and economical insights to reach a reasonable trade-off, rather than simply optimizing a single factor.

We conclude by discussing how this work relates to development of modern models of governance. Propositions to constitute assemblies of experts under liquid democracy [Miller, 1969, Blum and Zuber, 2016, Kahng et al., 2018] is supported, to some extent, by our result. Indeed, our findings suggest that a committee size of $\Theta(n)$ is more accurate than an informed large population of size n, under the assumption that the votes are delegated uniformly at random to the experts.

[3] A 2020 referendum in Italy yield a plebiscite for reducing the Congress size from 945 to 600 [De Sio and Angelucci, 2018].

[4] The history of apportionment in the U.S. taught us legitimacy can arise from considerations about the absolute number of seats of the States — and not necessarily about correctness [Szpiro, 2010].

References

Emmanuelle Auriol and Robert J Gary-Bobo. On the optimal number of representatives. *Public Choice*, 153 (3-4):419–445, 2012.

Emmanuelle Auriol and Robrt J Gary-Bobo. The more the merrier? choosing the optimal number of representatives in modern democracies. *Retrieved*, 16(5):2008, 2007.

Christian Blum and Christina Isabel Zuber. Liquid democracy: Potentials, problems, and perspectives. *Journal of Political Philosophy*, 24(2):162–182, 2016.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

Harald Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. *Actual. Sci. Ind.*, 736:5–23, 1938.

Lorenzo De Sio and Davide Angelucci. *945 sono troppi? 600 sono pochi? Qual è il numero "ottimale" di parlamentari?* Cise, 2018. URL `https://cise.luiss.it/cise/2019/10/09/945-sono-troppi-600-sono-pochi-qual-e-il-numero-ottimale-di-parlamentari/`.

Bernard Grofman, Guillermo Owen, and Scott L Feld. Thirteen theorems in search of the truth. *Theory and decision*, 15(3):261–278, 1983.

Kristof Jacobs and Simon Otjes. Explaining the size of assemblies. a longitudinal analysis of the design and reform of assembly sizes in democracies around the world. *Electoral Studies*, 40:280–292, 2015.

Anson Kahng, Simon Mackenzie, and Ariel D Procaccia. Liquid democracy: An algorithmic perspective. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Dan Ma. *The order statistics and the uniform distribution.* A Blog on Probability and Statistics, 2010.

Giorgio Margaritondo. Size of national assemblies: The classic derivation of the cube-root law is conceptually flawed. *Frontiers in Physics*, 8:606, 2021.

James C Miller. A program for direct and proxy voting in the legislative process. *Public choice*, 7(1):107–113, 1969.

George Szpiro. *Numbers rule: the vexing mathematics of democracy, from Plato to the present.* Princeton University Press, 2010.

Rein Taagepera. The size of national assemblies. *Social science research*, 1(4):385–401, 1972.

Liang Zhao and Tianyi Peng. An allometric scaling for the number of representative nodes in social networks. In *International Conference on Network Science*, pages 49–59. Springer, 2020.

A Proofs

A.1 Proof Theorem 3.1

By our choice of tiebreaking (a strict majority is needed for correctness), we have that the optimal k must be odd, as for any even k increasing it be one can only improve the probability of correctness. Fix arbitrary odd k and n with $k \leq n$ where $k = 2\ell + 1$ for some nonnegative integer ℓ. We will give sufficient conditions as a function of n and k for which $k + 2$ experts is strictly preferable to k.

Let p_n^k be the probability that the top k experts have a majority correct where experts have uniform order statistic competencies. More formally, for each $S \subseteq [k]$, let

$$w(S) = \prod_{i \in [k]} \left(\frac{n+1-i}{n+1}\right)^{\mathbb{I}[i \in S]} \left(\frac{i}{n+1}\right)^{\mathbb{I}[i \notin S]}$$

be the probability that exactly the agents in S (which are sampled from the k experts) are correct.

Note that here, we assume that the ith order of statistics has probability deterministically equals to its expectation to vote correctly, that is, $p_{(i)} = \frac{n+1-i}{n+1}$ [Ma, 2010].

Then,

$$p_n^k = \sum_{\substack{S \subseteq [k] \\ |S| \geq \ell+1}} w(S).$$

Fix some arbitrary odd value $k = 2\ell + 1$. We are interested in sufficient conditions for $p_n^{k+2} > p_n^k$. Let us consider $p_n^{k+2} - p_n^k$. Note that when there are k experts, $\ell + 1$ correct opinions are needed for a majority, while when there are $k + 2$, $\ell + 2$ are needed. Hence, the only way the two new experts can change the outcome from incorrect to correct is when exactly ℓ of the top k experts were correct (so the majority of k were incorrect), and the two new experts are correct. Conversely, the only scenario in which a correct outcome becomes incorrect is when exactly $\ell + 1$ of the top k experts are correct while the two new experts are incorrect. To state this more formally, we will introduce some new notation. let $\mathcal{E}_k^j$ be the event that exactly j of the top k experts out of n are correct. Note that while $\mathcal{E}_k^j$ does depend on n, we suppress it for ease of notation (it will be clear from context). We can now restate the preceding argument more formally as

$$p_n^{k+2} - p_n^k = \Pr[\mathcal{E}_k^\ell] \cdot \Pr[\text{Two extra experts correct}] - \Pr[\mathcal{E}_k^{\ell+1}] \cdot \Pr[\text{Two extra experts incorrect}].$$

Hence, $p_n^{k+2} - p_n^k > 0$ is equivalent to

$$\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^\ell]} < \frac{\Pr[\text{Two extra experts correct}]}{\Pr[\text{Two extra experts incorrect}]}. \tag{3}$$

Since the two extra experts have competencies $\frac{n+1-(k+1)}{n+1}$ and $\frac{n+1-(k+2)}{n+1}$, we have that

$$\frac{\Pr[\text{Two extra experts correct}]}{\Pr[\text{Two extra experts incorrect}]} = \frac{\frac{n+1-(k+1)}{n+1} \cdot \frac{n+1-(k+2)}{n+1}}{\frac{k+1}{n+1} \cdot \frac{k+2}{n+1}} = \frac{(n-k)(n-k-1)}{(k+1)(k+2)}$$

where the second equality is because the $n + 1$ terms cancel.

For each j,

$$\Pr[\mathcal{E}_k^j] = \sum_{\substack{S \subseteq [k] \\ |S| = j}} w(S)$$

Additionally,

$$
\begin{aligned}
(j+1)\cdot \Pr[\mathcal{E}_k^{j+1}] &= (j+1)\cdot \sum_{\substack{S\subseteq[k]\\|S|=j+1}} w(S)\\
&= \sum_{\substack{S\subseteq[k]\\|S|=j}} \sum_{i\in[k]\setminus S} w(S\cup\{i\})\\
&= \sum_{\substack{S\subseteq[k]\\|S|=j}} \sum_{i\in[k]\setminus S} w(S)\cdot \frac{n+1-i}{i}\\
&\le \sum_{\substack{S\subseteq[k]\\|S|=j}} \sum_{i\in[k]\setminus S} w(S)\cdot \frac{n}{i}\\
&= \sum_{\substack{S\subseteq[k]\\|S|=j}} w(S)\cdot n\cdot \sum_{i\in[k]\setminus S} \frac{1}{i}\\
&\le \sum_{\substack{S\subseteq[k]\\|S|=j}} w(S)\cdot n\cdot \sum_{i\in[k]} \frac{1}{i}\\
&= n\cdot H_k \cdot \sum_{\substack{S\subseteq[k]\\|S|=j}} w(S)\\
&= n\cdot H_k\cdot \Pr[\mathcal{E}_k^{j}]
\end{aligned}
$$

where the second transition holds because each set S of size $j+1$ will be counted exactly $j+1$ times in the subsequent sum once for each $i\in S$, when the smaller set equals $S\setminus\{i\}$. Plugging in ℓ for j, we have that

$$\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^{\ell}]} \le \frac{n\cdot H_k}{\ell+1} < \frac{2n\cdot H_k}{k}$$

Suppose $k \le \frac{1}{4}\cdot\frac{n}{\log(n)}$. Then, note that for sufficiently large n, $H_k \le \log(n)$, so $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^{\ell}]} < 2\cdot\frac{n\log(n)}{k}$. Additionally, for sufficiently large n, $\frac{(n-k)(n-k-1)}{(k+1)(k+2)} \ge \frac{1}{2}\cdot\frac{n^2}{k^2} \ge 2\cdot\frac{n\log(n)}{k}$. Hence, for these values of k, Equation (3) will be satisfied, implying that it cannot be optimal. This implies that for sufficiently large n, $k > \frac{1}{4}\cdot\frac{n}{\log(n)}$, or in other words, the optimal $K^\star = \Omega(\frac{n}{\log(n)})$. $\square$

A.2 Proof of Theorem 3.2

The proof reuses a lot of notation introduced in the proof of Theorem 3.1. We begin with the following claim.

Claim A.1. *Given a sequence of levels of expertise $p_{(1)} \ge p_{(2)} \ge \cdots \ge p_{(n)}$, the difference between the probabilities that $k+2$ experts are correct and k experts are correct satisfies*

$$p_n^{k+2} - p_n^k \ge \Pr[\mathcal{E}_k^{\ell}]\left[p_{(k+2)}^2 - \frac{1}{\ell+1}\sum_{i=1}^{k}\frac{p_{(i)}}{1-p_{(i)}}(1-p_{(k+2)})^2\right] \tag{4}$$

Proof.

$$
\begin{aligned}
p_n^{k+2} - p_n^k &= \Pr[\mathcal{E}_k^{\ell}]\cdot \Pr[\text{Two extra experts correct}] - \Pr[\mathcal{E}_k^{\ell+1}]\cdot \Pr[\text{Two extra experts incorrect}]\\
&= \Pr[\mathcal{E}_k^{\ell}]\cdot p_{(k+1)}p_{(k+2)} - \Pr[\mathcal{E}_k^{\ell+1}]\cdot (1-p_{(k+1)})(1-p_{(k+2)})
\end{aligned}
$$

For each j,

$$\Pr[\mathcal{E}_k^j] = \sum_{\substack{S \subseteq [k] \\ |S|=j}} w(S)$$

Additionally,

$$\begin{aligned}
(j+1) \cdot \Pr[\mathcal{E}_k^{j+1}] &= (j+1) \cdot \sum_{\substack{S \subseteq [k] \\ |S|=j+1}} w(S) \\
&= \sum_{\substack{S \subseteq [k] \\ |S|=j}} \sum_{i \in [k] \setminus S} w(S \cup \{i\}) \\
&= \sum_{\substack{S \subseteq [k] \\ |S|=j}} \sum_{i \in [k] \setminus S} w(S) \cdot \frac{p_{(i)}}{1-p_{(i)}} \\
&= \sum_{\substack{S \subseteq [k] \\ |S|=j}} w(S) \sum_{i \in [k] \setminus S} \frac{p_{(i)}}{1-p_{(i)}} \\
&\leq \sum_{\substack{S \subseteq [k] \\ |S|=j}} w(S) \sum_{i \in [k]} \frac{p_{(i)}}{1-p_{(i)}} \\
&= \Pr[\mathcal{E}_k^j] \sum_{i \in [k]} \frac{p_{(i)}}{1-p_{(i)}},
\end{aligned}$$

where the second transition holds because each set S of size $j+1$ will be counted exactly $j+1$ times in the subsequent sum once for each $i \in S$, when the smaller set equals $S \setminus \{i\}$. Letting $j = \ell$, we get

$$\Pr[\mathcal{E}_k^{\ell+1}] \leq \frac{1}{\ell+1} \sum_{i \in [k]} \frac{p_{(i)}}{1-p_{(i)}} \Pr[\mathcal{E}_k^\ell].$$

So,

$$\begin{aligned}
p_n^{k+2} - p_n^k &\geq \Pr[\mathcal{E}_k^\ell] \cdot p_{(k+1)} p_{(k+2)} - \frac{1}{\ell+1} \sum_{i \in [k]} \frac{p_{(i)}}{1-p_{(i)}} \Pr[\mathcal{E}_k^\ell] \cdot (1-p_{(k+1)})(1-p_{(k+2)}) \\
&= \Pr[\mathcal{E}_k^\ell] \left[p_{(k+1)} p_{(k+2)} - \frac{1}{\ell+1} \sum_{i \in [k]} \frac{p_{(i)}}{1-p_{(i)}} (1-p_{(k+1)})(1-p_{(k+2)}) \right] \\
&\geq \Pr[\mathcal{E}_k^\ell] \left[p_{(k+2)}^2 - \frac{1}{\ell+1} \sum_{i \in [k]} \frac{p_{(i)}}{1-p_{(i)}} (1-p_{(k+2)})^2 \right],
\end{aligned}$$

as desired. $\square$

Since $H \geq p_{(1)} \geq \cdots \geq p_{(n)}$, we have

$$\frac{1}{\ell+1} \sum_{i=1}^{k} \frac{p_{(i)}}{1-p_{(i)}} \leq \frac{k}{\ell+1} \frac{H}{1-H} \leq \frac{2H}{1-H}. \tag{5}$$

Combining with (4), we get

$$
\begin{aligned}
p_n^{k+2} - p_n^k &\geq \Pr[\mathcal{E}_k^\ell]\left[p_{(k+2)}^2 - \frac{2H}{1-H}(1-p_{(k+2)})^2\right] \\
&= \Pr[\mathcal{E}_k^\ell]\left(p_{(k+2)} + \sqrt{\frac{2H}{1-H}}(1-p_{(k+2)})\right)\left(p_{(k+2)} - \sqrt{\frac{2H}{1-H}}(1-p_{(k+2)})\right) \\
&= \Pr[\mathcal{E}_k^\ell]\left(p_{(k+2)} + \sqrt{\frac{2H}{1-H}}(1-p_{(k+2)})\right)\left(1+\sqrt{\frac{2H}{1-H}}\right)\left(p_{(k+2)} - \frac{\sqrt{\frac{2H}{1-H}}}{1+\sqrt{\frac{2H}{1-H}}}\right)
\end{aligned}
$$

Now, if we want $p_n^{k+2} - p_n^k > 0$, it suffices to require that

$$
p_{(k+2)} > \frac{\sqrt{\frac{2H}{1-H}}}{1+\sqrt{\frac{2H}{1-H}}}. \tag{6}
$$

For the next part, we will need the following concentration inequality, the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality.

Lemma A.2 (DKW inequality)**.** *Let $p_1, \ldots, p_n$ be n i.i.d. draws from a distribution $\mathcal{D}$ (we use $F(\cdot)$ to denote the cumulative distribution function of $\mathcal{D}$), and denote by $p_{(1)} \geq \cdots \geq p_{(n)}$ the order statistics, then*

$$
\Pr\left[\forall i \in [n], \left|F(p_{(i)}) - \frac{n-i}{n}\right| > \varepsilon\right] < 2e^{-2n\varepsilon^2}, \tag{7}
$$

for every $\varepsilon > 0$.

Lemma A.3. *Let $\mathcal{D}$ be a distribution supported on interval $[L, H]$ with $0 \leq L < H < 1$. Assume that its inverse CDF $F^{-1}(\cdot)$ is M-Lipschitz continuous. Let H and $\varepsilon \in (0,1)$ be such that $1 - \frac{1}{1+\sqrt{\frac{2H}{1-H}}} + M\varepsilon < H$. Then, with probability at least $1 - 2\exp(-2n\varepsilon^2)$ over the random draw of p_i's, the optimal size $K^\star$ satisfies*

$$
\frac{K^\star}{n} \geq 1 - F\left(1 - \frac{1}{1+\sqrt{\frac{2H}{1-H}}} + M\varepsilon\right) - \frac{2}{n}. \tag{8}
$$

Proof. By DKW inequality (Lemma A.2), with probability at least $1 - 2\exp(-2n\varepsilon^2)$, for every $i \in [n]$,

$$
\left|F(p_{(i)}) - \frac{n-i}{n}\right| \leq \varepsilon.
$$

Since F^{-1} is M-Lipschitz continuous,

$$
\left|p_{(i)} - F^{-1}(\frac{n-i}{n})\right| \leq M\varepsilon.
$$

For a congress of size k, we let $i = k+2$ and get

$$
p_{(k+2)} \geq F^{-1}(\frac{n-k-2}{n}) - M\varepsilon.
$$

If we require

$$
F^{-1}(\frac{n-k-2}{n}) - M\varepsilon > \frac{\sqrt{\frac{2H}{1-2H}}}{1+\sqrt{\frac{2H}{1-2H}}}, \tag{9}
$$

then (6) is satisfied and hence $p_n^{k+2} - p_n^k > 0$, which implies that such k cannot be optimal. Solving (9) gives $\frac{k}{n} < 1 - F\left(1 - \frac{1}{1+\sqrt{\frac{2H}{1-H}}} + M\varepsilon\right) - \frac{2}{n}$. □

Now we prove the theorem. Letting $2\exp(-2n\varepsilon^2) = \frac{1}{n}$, we get $\varepsilon = \sqrt{\frac{\ln 2n}{2n}} = o(1)$. Moreover, we note that $1 - \frac{1}{1+\sqrt{\frac{2H}{1-H}}} < H$ is satisfied when $H > \frac{2}{3}$. Then from the above lemma, with probability at least $1 - \frac{1}{n}$ over the random draw of p_i's, we have

$$\frac{K^\star}{n} \geq 1 - F\left(1 - \frac{1}{1+\sqrt{\frac{2H}{1-H}}} + M\sqrt{\frac{\ln 2n}{2n}}\right) - \frac{2}{n} = \Omega(1), \tag{10}$$

as needed. □

A.3 Proof of Theorem 5.1

Recall, for $k = 1$,

$$\Gamma_n^{\mathcal{D}}(1) = \Pr[X_{(1)} = 1] - \Pr\left[\sum_{i=1}^n X_{(i)} > \frac{n}{2}\right].$$

Note that, by Cramér [1938],

$$\Pr\left[\sum_{i=1}^n X_i > \frac{n}{2}\right] \approx 1 - \exp(-nI(\frac{1}{2}))$$

where $I(\frac{1}{2})$ is the entropy function. Plugging in our distribution $\mathcal{D} = \mathcal{U}(2\varepsilon_n, 1)$, $I(\frac{1}{2})$ is actually the Kullback–Leibler divergence between Bernoulli(0.5) and Bernoulli(p) with $p = \mathbb{E}[\mathcal{D}] = \frac{1}{2} + \varepsilon_n$. So,

$$\begin{aligned} I\left(\frac{1}{2}\right) &= KL(\frac{1}{2}||p)_{\text{Ber}} \\ &= \frac{1}{2}\log\left(\frac{1}{2p}\right) + \frac{1}{2}\log\left(\frac{1}{2(1-p)}\right) \\ &= \frac{1}{2}\log\left(\frac{1}{4p(1-p)}\right) \end{aligned}$$

and

$$\begin{aligned} \Pr\left[\sum_{i=1}^n X_i > \frac{n}{2}\right] &\approx 1 - \exp\left(-\frac{n}{2}\log\left(\frac{1}{4p(1-p)}\right)\right) \\ &= 1 - (4p(1-p))^{n/2} \\ &= 1 - ((1-\varepsilon_n)(1+\varepsilon_n))^{n/2} \\ &= 1 - (1-\varepsilon_n^2)^{n/2}. \end{aligned}$$

On the other hand, we can compute the closed-form probability for the top agent $X_{(1)}$ to vote correctly, using the following lemma.

Lemma A.4. *Let $p_{(1)}$ be the largest order statistics of a sequence drawn from $\mathcal{U}[a, 1]$, then $\mathbb{E}[p_{(1)}] = 1 - \frac{1-a}{n+1}$.*

Proof. Let $X \sim \mathcal{U}[a,1]$, f be the probability distribution function of X and F be the cumulative distribution function of X. It follows that $f(x) = \frac{1}{1-a}1_{\{x\in[a,1]\}}$ and $F(x) = \frac{x-a}{1-a}1_{\{x\in[a,1]\}} + 1_{\{x>1\}}$. It is know that the probability distribution function of $p_{(1)}$ is such that $f_{X_{(1)}}(x) = nf(x)F(x)^{n-1}$. As a result,

$$\begin{aligned}
\mathbb{E}[p_{(1)}] &= \int_{\mathbb{R}} nf(x)F(x)^{n-1}dx \\
&= n\int_a^1 x\frac{(x-a)^{n-1}}{(1-a)^n} \\
&= \frac{n}{(1-a)^n}\int_0^{1-a}(x+a)x^{n-1} \\
&= \frac{n}{(1-a)^n}\left(\frac{(1-a)^{n+1}}{n+1} + a\frac{(1-a)^n}{n}\right) \\
&= n\left(\frac{(1-a)}{n+1} + a\frac{(1}{n}\right) \\
&= \frac{n+a}{n+1} \\
&= 1 - \frac{1-a}{n+1}
\end{aligned}$$

where the third line is obtained from the change of variable $x - a \to x$. □

As a result of Lemma A.4, with $a = 2\varepsilon_n$,

$$\Pr[X_{(1)} = 1] = \mathbb{E}[p_{(1)}] = 1 - \frac{1-2\varepsilon_n}{n+1}.$$

Hence,

$$\Gamma_n^{\mathcal{U}(2\varepsilon_n,1)}(1) = \Pr[X_{(1)} = 1] - \Pr\left[\sum_{i=1}^n X_i > \frac{n}{2}\right] = (1-\varepsilon_n^2)^{n/2} - \frac{1-2\varepsilon_n}{n+1}. \tag{11}$$

Fix some constant c. We will substitute $c \cdot \sqrt{\frac{\log n}{n}}$ for ε_n in Equation (11) to compute some bounds. Note that when $\varepsilon = o\left(\sqrt{\frac{\log n}{n}}\right)$, $\varepsilon < c \cdot \sqrt{\frac{\log n}{n}}$ for sufficiently large n for any constant c we choose, and symmetrically, when $\varepsilon = \omega\left(\sqrt{\frac{\log n}{n}}\right)$, $\varepsilon > c \cdot \sqrt{\frac{\log n}{n}}$ for sufficiently large n for any constant c we choose.

We have that

$$\begin{aligned}
\left(1 - \left(c\cdot\sqrt{\frac{\log n}{n}}\right)^2\right)^{n/2} - \frac{1 - c\cdot\sqrt{\frac{\log n}{n}}}{n+1} &= \left(1 - c^2\cdot\frac{\log n}{n}\right)^{n/2} - \frac{1 - c\cdot\sqrt{\frac{\log n}{n}}}{n+1} \\
&\underset{n\to\infty}{\approx} \exp\left(-\frac{c^2\log(n)}{2}\right) - \frac{1}{n} \\
&= \frac{1}{n^{c^2/2}} - \frac{1}{n}.
\end{aligned}$$

This implies that when $\varepsilon = o\left(\sqrt{\frac{\log n}{n}}\right)$, by choosing c sufficiently small, for sufficiently large n, $\frac{1}{n^{c^2/2}} - \frac{1}{n} > 0$ and hence $\Gamma_n^{\mathcal{U}(2\varepsilon_n,1)}(1) > 0$. Symmetrically, when $\varepsilon = \omega\left(\sqrt{\frac{\log n}{n}}\right)$, by choosing c sufficiently large, for sufficiently large n, $\Gamma_n^{\mathcal{U}(2\varepsilon_n,1)}(1) < 0$, as needed. □