How Many Representatives Do We Need? The Optimal Size of an Epistemic Congress

Manon Revel,[*1] **Tao Lin,**[*2] **Daniel Halpern**[*2]

[1] Massachusetts Institute of Technology
[2] Harvard University
mrevel@mit.edu, tlin@g.harvard.edu, dhalpern@g.harvard.edu

Abstract

Aggregating opinions of a collection of agents is a question of interest to a broad array of researchers, ranging from ensemble-learning theorists to political scientists designing democratic institutions. This work investigates the optimal number of agents needed to decide on a binary issue under majority rule. We take an *epistemic* view where the issue at hand has a ground truth "correct" outcome and each one of n voters votes correctly with a fixed probability, known as their competence level or competence. These competencies come from a fixed distribution $\mathcal{D}$. Observing the competencies, we must choose a specific group that will represent the population. Finally, voters sample a decision (either correct or not), and the group is correct as long as more than half the chosen representatives voted correctly. Assuming that we can identify the best experts, i.e., those with the highest competence, to form an *epistemic congress* we find that the optimal congress size should be linear in the population size. This result is striking because it holds even when allowing the top representatives to become arbitrarily accurate, choosing the correct outcome with probabilities approaching 1. We then analyze real-world data, observing that the actual sizes of representative bodies are much smaller than the optimal ones our theoretical results suggest. We conclude by examining under what conditions congresses of sub-optimal sizes would still outperform direct democracy, in which all voters vote. We find that a small congress would beat direct democracy if the rate at which the societal bias towards the ground truth decreases with the population size fast enough, and we quantify the speed needed for constant and polynomial congress sizes.

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 on the number of representatives, and this optimal number of representatives has been subject to great debate.[1] In the Federalist Paper No. 56, Madison argued that there should be *a representative for every thirty thousand inhabitants*. In response, every ten years between 1785 and 1913, the American congress was enlarged–in aggregate from 65 to 435–adapting to evolving state populations (Szpiro 2010). However, since 1913, this number of representatives has remained constant, bringing the current number of inhabitants per representative well below Madison's objective.

Quantitative research rationalizing the optimal congress size dates back to at least the 1970s. Taagepera (1972) concluded that the number of representatives should be the cube root of the population size. These findings are regarded as seminal (Jacobs and Otjes 2015) and have influenced political decisions and referendums, such as the 2020 Italian referendum to reduce the size of both chambers from 945 to 600 parliamentary (Margaritondo 2021; De Sio and Angelucci 2019).

Yet, recent work using machinery from physics and economics revisited these claims and showed that the optimal number should be larger, at least proportional to the square root of the population size (e.g., Auriol and Gary-Bobo 2012; Margaritondo 2021). In particular, Magdon-Ismail and Xia (2018) explored an *epistemic* setup that groups voters into pods of size L, and each pod selects one representative. The authors find that the congress size ought to be linear under this model when voting is cost-less.

Note that our setup for decision-making in a congress on binary issues, applied here in a democratic context, resembles an ensemble of classifiers in machine learning: classifiers are "voters" who predict a binary outcome, and they collectively decide, through a majority rule, on the decision's outcome (Magdon-Ismail and Xia 2018). To obtain a good ensemble of classifiers, one can measure the accuracy of all classifiers and keep only the most accurate ones. Designers have used these ideas to reduce uncertainty in decisions and increase the classifiers' performance by combining their predictions (Yang 2011; Polikar 2012; Sagi and Rokach 2018).

We can now reformulate our research question in these terms: how many agents should we include to maximize the accuracy of the decision?

Our Contribution

Through novel proofs techniques, we strengthen the pessimistic results of Magdon-Ismail and Xia (2018) for congress under the epistemic approach, finding that even

[*] These authors contributed equally.

[1] See activists at https://thirty-thousand.org who advocate for enlarging the congress.

with the ability to identify the most accurate agents to form a congress, the optimal number of representatives remains linear in the size of the population. However, we find that all is not lost for congresses of a more practical nature: we follow this up with comparisons of different sizes and identify conditions for smaller congresses to be more accurate than when the entire society votes.

In the epistemic setting, voters decide on a binary issue and aim at differentiating between the ground truth correct choice, the value 1, and its alternative, 0. Each voter has a competence level in $[0, 1]$ representing the probability they vote correctly. Further, the competence levels of the population are drawn according to some distribution $\mathcal{D}$. We take the idealized view that, given a target size k, we can identify the k most competent voters in society (that is, the first k order statistics from the competence levels $p_1, \ldots, p_n$ where each $p_i \sim \mathcal{D}$) to form the congress. These members then vote on the issue, and the outcome is that of the majority. We first observe that, if voters' competence levels are the expected values of the order statistics from uniform distribution $\mathcal{U}(0, 1)$, the optimal size of congress is between $(3 - 2\sqrt{2})n$ and $\frac{n}{2}$. 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 $\Theta(n)$ with more refined bounds based on the distribution.

The assumption that we can identify the k best agents is purposefully idealistic; we give the congress its best shot at favoring smaller sizes by granting it an unrealistically powerful selection procedure. This assumption strengthens our claim: the optimal congress size is linear with the population size *even* under the generous assumption that only the most competent agents represent. Second, our first analysis of the uniform distribution in $[0, 1]$ allows the best agents' competence to converge towards 1 as the population size increases, again providing an unrealistic assumption favoring small congress sizes. Under this extra generous assumption, we prove that the optimal congress size remains, strikingly, linear with the population size — later, we generalize this to more realistic distributions.

We then turn to study real-world data on the sizes of countries' representative bodies. Here, we notice that congresses in the real world are on the order of the cube root of the population size, much smaller than the optimal (linear) size our theoretical results suggest. We then seek to understand when real-world congress sizes can be deemed effective: we identify conditions on the distribution of competence level under which a smaller congress outperforms the majority. If the population is unbiased or biased towards 0, a congress composed of experts with expertise levels above 0.5 trivially outperforms the majority. We further find that, for a population whose average level of competence is biased above 0.5, a relatively small congress can still be better than the majority as long as the bias is small enough. We characterize this threshold for both single-agent and n^r-sized congresses.

Related Work

The use of an epistemic approach, relying on voting to aggregate objective opinions, is well studied in computational social choice (Brandt et al. 2016). One particularly significant result is the Condorcet Jury Theorem (De Condorcet 1785; Grofman, Owen, and Feld 1983), which shows that in the limit, a majority vote by an increasing number of independent voters biased towards the correct outcome will be correct with probability approaching 1. Note that this epistemic setup models legislative decisions or referendums in which one choice is inherently more desirable for society — yet, this *correct* outcome is not known a priori, and agents are trying to uncover it. Subsequent work has studied extensions of the Condorcet Jury Theorem in instances where the voters are inhomogeneous, dependent, or strategic, as summarized in a survey paper by Nitzan and Paroush (2017).

The first work about the optimal size of parliaments focused on maximizing parliament's efficiency (Taagepera 1972). For them, maximizing efficiency was equivalent 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". Margaritondo (2021) revisited this work and found a flaw in the original proof, arguing that the optimal size under this model should instead be $\Theta(\sqrt{n})$. Empirical papers (Taagepera 1972; Auriol and Gary-Bobo 2007) focused on finding the optimal number of representatives have used country data to back up the "square root law" result. Jacobs and Otjes (2015), on the other hand, investigate potential causal effects of different congress sizes.

The work of Auriol and Gary-Bobo (2012) also aims to derive the optimal number of representatives for a society. However, their model lies in stark contrast to the epistemic one: they assume that voters have preference-based utilities, with an uninformative prior, and the representatives are chosen uniformly at random from society. They too reach the conclusion that the optimal size of congress is proportional to the square root of the population size. Zhao and Peng (2020) look at the optimal number of representatives in a social network. They consider a set of nodes representative if together they can reach all other nodes in at most m steps (where $m = \Theta(\log n)$ is an exogenous threshold). The goal is to find the minimum size of such a set. Under a certain class of realistic social networks, they find that the minimum should be proportional to n^γ for some $\frac{1}{3} \leq \gamma \leq \frac{5}{9}$.

Finally, we build upon the work of Magdon-Ismail and Xia (2018). There, the authors consider a model for representative democracy where agents are grouped into K groups of sizes L and each chooses one representative per group. Importantly, the competencies are drawn from a distribution $\mathcal{D}$ only after the agents are grouped. The authors then derive the group size that maximizes the probability that the representatives make the correct decision. They show the optimal group size is constant, so the optimal number of representatives (which is, in the simplest setup, the population size divided by the number of groups) should then be linear in the population size. The fact that the level of competence is drawn after grouping people imposes a tradeoff between how accurate the representatives will be and how many representatives (n/L) there are. Indeed, the best agent in each group has a competence level that is the top order statistic of the distribution with L draws. For instance, the top level of competence from a uniform distribution is in expectation $1 - \frac{1}{L+1}$, which gets large only if the number of groups n/L is small. Importantly, the trade-off implied by the model favors large congresses. Yet, one could wonder whether the optimal congress size remains linear if one breaks with this implicit trade-off allowing the highest competencies to become arbitrarily large. This is precisely the gap we fill.

2 Model

Let n be the number of voters in the society. Following the epistemic approach, voters need to 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) $p_i \in [0, 1]$, which is the probability that she votes "correctly" (i.e., votes for option 1). Depending on the instance, we will sometimes assume that the p_is are sampled from some distribution $\mathcal{D}$ whose support is contained in $[0, 1]$ and other times assume the p_is are deterministic (perhaps also depending on n which will always be clear from context).

Given $p_1, \ldots, p_n$, we sort voters by decreasing competence level, denoted by $p_{(1)} \geq \cdots \geq p_{(n)}$, where $p_{(i)}$ is the competence level of the i^{th} best voter. (Note that, for notational convenience, this is the reverse of normal order statistics.) Let $X_{(1)}, \ldots, X_{(n)}$ be Bernoulli random variables denoting their votes, with $X_{(i)} = 1$ meaning a correct vote for the i^{th} best voter and 0 otherwise; the $X_{(i)}$s are conditionally independent given $p_{(i)}$s, and $\Pr[X_{(i)} = 1 \mid p_{(i)}] = p_{(i)}$.

A congress of size k is composed of the k best voters in society and makes a correct decision when a strict majority is correct, $\sum_{i=1}^{k} X_{(i)} > k/2$.[2] One may envision other rules to select the congress members, for example, the group representatives analyzed by Magdon-Ismail and Xia (2018). Here we take the best k voters, and this can be seen as a best-case scenario for accuracy. Strikingly, as we will show, even under this strong assumption, the optimal number of representatives is already very large, which suggests that the optimal number would even be larger in more realistic scenarios.

Although the assumption that we can sample the k best experts is unrealistic if one thinks at the democratic context, we could argue that finding the k most accurate classifiers in an ensemble is indeed realistic. In any case, this sampling method favors small congresses while allowing the sampled congress to reach the maximal probability of making the correct decision. Our conclusions hence read that *despite* the generous assumptions, a large number of voters are needed to maximize a collective's chance to make a correct decision.

3 Optimal Congress Size

In this section, we prove theoretical bounds on the optimal size of congress for several natural distributions. We begin by formally stating our problem.

For fixed voter competencies $p_{(1)} \geq \cdots \geq p_{(n)}$, we define $K^\star$ to be the optimal size of congress, the size k that maximizes the probability that the representatives make a correct decision (for convenience breaking ties in favor of an arbitrary odd k[3]). Formally,

$$K^\star \in \operatorname*{arg\,max}_{1 \leq k \leq n} \left\{ \Pr\left[\sum_{i=1}^{k} X_{(i)} > \frac{k}{2} \;\middle|\; X_{(i)} \sim \text{Bern}(p_{(i)}) \right] \right\}.$$

We note that since $K^\star$ is a function of the voter competencies, if these competencies are random samples, then $K^\star$ is a random variable. However, we sometimes assume for tractability that the competencies match their expectation, that is, $p_{(i)}$ is exactly equal to the expectation of the $(n+1-i)$'th order statistic of n draws from $\mathcal{D}$. In this case, $K^\star$ is a deterministic value for each n.

For fixed voter competencies $p_{(1)} \geq \cdots \geq p_{(n)}$, let $\mathcal{E}_k^j$ be the event that exactly j of the top experts out of k are correct. Our characterization of the optimal size $K^\star$ relies on the following key lemma.

Lemma 1. *For fixed competencies $p_{(1)} \geq \cdots \geq p_{(n)}$, for all odd $k \leq n$ with $k = 2\ell + 1$,*

- *If* $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^{\ell}]} < \frac{p_{(k+1)} p_{(k+2)}}{(1-p_{(k+1)})(1-p_{(k+2)})}$, *then* $K^\star \neq k$.
- *If* $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^{\ell}]} > \frac{p_{(k+1)} p_{(k+2)}}{(1-p_{(k+1)})(1-p_{(k+2)})}$, *then* $K^\star \neq k+2$.

The proof of the lemma involves comparing a congress of some specific size k to one of size $k+2$ (recall that we chose $K^\star$ to be odd, so we may as well restrict ourselves to odd k). Clearly, if the top $k+2$ experts have a higher chance of being correct than k, then k cannot be optimal (and vice-versa). Importantly, this gives us a sufficient condition to rule out certain values of k. For example, if we know that for all $k < c$ the first condition of the lemma holds, then that implies $K^\star \geq c$.

Proof of Lemma 1. For any $k \leq n$, let $q_k = \sum_{j=\lfloor k/2 \rfloor + 1}^{k} \Pr[\mathcal{E}_k^j]$ be the probability that a congress of size k will be correct. We have that $K^\star \in \arg\max_{k \leq n} q_k$. Fix $p_{(1)} \geq \cdots \geq p_{(n)}$ and a specific $k = 2\ell + 1$. We will show that $q_{k+2} > q_k$ (resp. $<$) is equivalent to $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^{\ell}]} < \frac{p_{(k+1)} p_{(k+2)}}{(1-p_{(k+1)})(1-p_{(k+2)})}$ (resp. $>$). If $q_{k+2} > q_k$ (resp. $<$), then $K^\star \neq k$ (resp. $k+2$) as that would imply $K^\star$ is not optimal.

Let us now consider $q_{k+2} - q_k$. 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

[2] A strict rather than weak majority here corresponds to tie-breaking in favor of the incorrect outcome. Tie-breaking in the other direction would not asymptotically change our results.

[3] Note that there must always be an optimal k that is odd, as for any even k, due to our strict majority constraint, $k-1$ must have overall accuracy at least as high.

experts are correct while the two new experts are incorrect. Since $\mathcal{E}_k^j$ is the event that exactly j of the top k experts out of n are correct, we can formally write the above as

$$q_{k+2} - q_k = \Pr[\mathcal{E}_k^\ell] \cdot p_{(k+1)} p_{(k+2)} - \Pr[\mathcal{E}_k^{\ell+1}] \cdot (1 - p_{(k+1)})(1 - p_{(k+2)}).$$

Rearranging this yields the two equivalent inequalities previously stated. $\square$

For a set of representatives $S \subseteq [k]$, let $w(S) = \prod_{i \in S} p_{(i)} \cdot \prod_{i \in [k] \setminus S} (1 - p_{(i)})$ be the probability that exactly those in S are correct (and those in $[k] \setminus S$ are incorrect). We then have the following.

Lemma 2. *For each $\mathcal{E}_k^j$,*

$$\Pr[\mathcal{E}_k^j] = \frac{1}{k-j} \sum_{\substack{S \subseteq [k] \\ |S| = j+1}} w(S) \sum_{i \in S} \frac{1 - p_{(i)}}{p_{(i)}}.$$

Proof. By the definition of $\mathcal{E}_k^j$, $\Pr[\mathcal{E}_k^j] = \sum_{\substack{S \subseteq [k] \\ |S| = j}} w(S)$. We then note that

$$\sum_{\substack{S \subseteq [k] \\ |S| = j}} w(S) = \frac{1}{k-j} \sum_{\substack{S \subseteq [k] \\ |S| = j+1}} \sum_{i \in S} w(S \setminus \{i\})$$

because when we count the sets S of size j by first selecting a set of size $j+1$ and then removing one of its $j+1$ elements, each set of size j is counted exactly $k - j$ times. Therefore,

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

as needed. $\square$

Armed with these lemmas, we can now move to proving bounds on the optimal congress size.

Standard Uniform Distribution

First, we focus on the case where competence levels are drawn from uniform distribution $\mathcal{U}(0, 1)$. For tractability, as discussed in the problem statement, we assume that the competence levels are exactly equal to their expectation, i.e., $p_{(i)} = \frac{n+1-i}{n+1}$ (see e.g., Ma (2010)). In this case, the competence levels of the top experts approach 1 asymptotically. Note that this is an unrealistic assumption that, again, acts in favor of small congresses. Including it emphasizes the striking nature of the result: Even with top experts becoming arbitrarily accurate and the ability to identify the most accurate members of society, the optimal size of congress remains a constant fraction of the population.

Theorem 1. *Suppose $p_{(i)} = \frac{n+1-i}{n+1}$. Then, $(3 - 2\sqrt{2}) \cdot n - O(1) \leq K^\star \leq \frac{1}{2} \cdot n + O(1)$.*

Proof. Recall that we can focus only on odd k. Fix some odd $k \leq n$ where $k = 2\ell + 1$ for some non-negative integer ℓ. Our goal will be to compare $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^\ell]}$ and $\frac{p_{(k+1)} p_{(k+2)}}{(1 - p_{(k+1)})(1 - p_{(k+2)})} = \frac{(n-k)(n-k-1)}{(k+1)(k+2)}$ in order to apply Lemma 1.

By Lemma 2 with $j = \ell$ and using the fact that $k - \ell = \ell + 1$,

$$\Pr[\mathcal{E}_k^\ell] = \frac{1}{\ell+1} \sum_{\substack{S \subseteq [k] \\ |S| = \ell+1}} w(S) \sum_{i \in S} \frac{i}{n+1-i}. \tag{1}$$

We begin with the lower bound. Let us consider the inner sum of Equation (1). We have that for all S,

$$\sum_{i \in S} \frac{i}{n+1-i} \geq \frac{1}{n} \sum_{i \in S} i \geq \frac{1}{n} \sum_{i=1}^{\ell+1} i = \frac{(\ell+1)(\ell+2)}{2n}$$

where the first inequality holds because $i \geq 1$ for all i and the second inequality holds because $|S| = \ell + 1$ and $S \subseteq [k]$ hence the minimum it could sum to is that of the smallest $\ell + 1$ positive integers. As this bound is independent of S, we can pull it out of the outer sum to yield

$$\Pr[\mathcal{E}_k^\ell] \geq \frac{\ell+2}{2n} \sum_{\substack{S \subseteq [k] \\ |S| = \ell+1}} w(S) = \frac{k+3}{4n} \cdot \Pr[\mathcal{E}_k^{\ell+1}]$$

where the last inequality holds because $\ell + 2 = \frac{k-1}{2} + 2 = \frac{k+3}{2}$. This allows us to write $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^\ell]} \leq \frac{4n}{k+3}$, so in order to invoke the first item of Lemma 1 to show a specific value of k is not optimal, we need a sufficient condition for k to guarantee

$$\frac{4n}{k+3} < \frac{(n-k)(n-k-1)}{(k+1)(k+2)}. \tag{2}$$

Note that Equation (2) is implied by $4n < \frac{(n-k-1)^2}{k+1}$ which we can rearrange to $(k+1)^2 - 6n(k+1) + n^2 > 0$. The left hand side of the inequality is a quadratic in $(k+1)$ with roots at $(3 \pm 2\sqrt{2}) \cdot n$. Since the squared term is positive and hence the quadratic is only non-positive between the two roots, as long as $(k+1) < (3 - 2\sqrt{2}) \cdot n$, the inequality holds. Along with the first item of Lemma 1, this implies the desired $(3 - 2\sqrt{2}) \cdot n - O(1)$ lower bound.

Next, we will show the upper bound. In the inner summand of Equation (1), $i \in [k]$ so $i \leq k$, and hence $\frac{i}{n+1-i} \leq \frac{k}{n+1-k}$. This yields

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

Here, we get that $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^\ell]} \geq \frac{k}{n+1-k}$. As with the lower bound, to invoke the second item of Lemma 1, we need a sufficient condition for

$$\frac{k}{n+1-k} > \frac{(n-k)(n-k-1)}{(k+1)(k+2)}. \tag{3}$$

Equation (3) is equivalent to

$$k(k+1)(k+2) > (n-k-1)(n-k)(n-k+1).$$

As both sides are the product of three consecutive integers, this will be true as long as $n-k-1 < k$, or equivalently $k+2 > \frac{n}{2}+\frac{3}{2}$. Applying Lemma 1 yields the desired upper bound. □

Hence, we have proved that for competencies equal to the expectation of $\mathcal{U}[0,1]$ order statistics, a constant fraction of the total population is necessary to maximize the probability the representatives make the correct decision. We conjecture that $K^\star$ is in fact close to $n/4$ in this set up (see simulations in Appendix B).

Distributions Bounded Away From 1

Next, we consider a broad class of distributions that do not allow for arbitrarily accurate experts. Unlike in the previous section, 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 2. *Let $\mathcal{D}$ be any continuous distribution supported by $[L,H]$ with cumulative distribution function $F(\cdot)$. If $0 < L < \frac{1}{2} < H < 1$, and $F^{-1}(\cdot)$ is M-Lipschitz continuous with $0 < M < \infty$,[4] then, with probability at least $1-4e^{-2n\varepsilon^2}$ the competency draws will yield an optimal $K^\star$ such that*

$$c_H n - O(1) \leq K^\star \leq c_L n + O(1)$$

for all n and $\varepsilon > 0$, where $c_H = 1 - F\left(\frac{1}{1+\sqrt{\frac{1-H}{H}}} + M\varepsilon\right)$ and $c_L = 1 - F\left(\frac{1}{1+\sqrt{\frac{1-L}{L}}} - M\varepsilon\right)$.

We remark that $L \geq 0$ is sufficient for the lower bound $c_H n - O(1) \leq K^\star$ to hold and vice-versa, $H \leq 1$ is sufficient for the upper bound to hold. Both of these bounds individually hold with probability at least $1-2e^{-2n\varepsilon^2}$.

To prove Theorem 2, we will make use of the following well-known concentration inequality.

Lemma 3 (Dvoretzky–Kiefer–Wolfowitz inequality, see e.g., Massart 1990)**.** *Let $p_{(1)} \geq \cdots \geq p_{(n)}$ be n sorted i.i.d. draws from $\mathcal{D}$. For every $\varepsilon > 0$, $\Pr\left[\forall i \in [n], \left|F(p_{(i)}) - \frac{n-i}{n}\right| \leq \varepsilon\right] \geq 1-2e^{-2n\varepsilon^2}$.*

[4] This condition is satisfied when the PDF of $\mathcal{D}$ is lower bounded by $1/M$, which is satisfied by, e.g., uniform, normal, and beta distributions truncated to $[L,H]$.

Lemma 3 implies that, with probability at least $1 - 2e^{-2n\varepsilon^2}$, for every $i \in [n]$, $\left|F(p_{(i)}) - \frac{n-i}{n}\right| \leq \varepsilon$. Since F^{-1} is assumed to be M-Lipschitz continuous,

$$\left|p_{(i)} - F^{-1}\left(\frac{n-i}{n}\right)\right| \leq M\varepsilon. \tag{4}$$

We are now ready to prove Theorem 2. We show the lower bound here; the proof for the upper bound uses similar techniques and is relegated to Appendix A.

Proof of Theorem 2. We will show that both the lower bound $c_H n - O(1) \leq K^\star$ and the upper bound $K^\star \leq c_L n + O(1)$ each occur with probability at least $1-2e^{-2n\varepsilon^2}$ which, by a union bound, proves the desired probability. As previously mentioned, we will only prove the lower bound here. Fix arbitrary odd k and n with $k \leq n$ where $k = 2\ell+1$ for some non-negative integer ℓ. We will give sufficient conditions as a function of n and k for which we can apply Lemma 1.

First, by Lemma 2 with $j = \ell$, $\Pr[\mathcal{E}_k^\ell] = \frac{1}{k-\ell}\sum_{\substack{S\subseteq[k]\\|S|=\ell+1}} w(S)\sum_{i\in S}\frac{1-p_{(i)}}{p_{(i)}}$. Because the support of $\mathcal{D}$ is upper-bounded by H, $p_{(i)} \leq H$ for all i with probability one. So, $\sum_{i\in S}\frac{1-p_{(i)}}{p_{(i)}} \geq (\ell+1)\frac{1-H}{H}$. Noting that $\ell+1 = \frac{k+1}{2} = k-\ell$ and $\Pr[\mathcal{E}_k^{\ell+1}] = \sum_{\substack{S\subseteq[k]\\|S|=\ell+1}} w(S)$, after rearranging we have $\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^\ell]} \leq \frac{H}{1-H}$. Further, we note that $\frac{p_{(k+1)}p_{(k+2)}}{(1-p_{(k+1)})(1-p_{(k+2)})} \geq \frac{p_{(k+2)}^2}{(1-p_{(k+2)})^2}$.

Now, if we want to apply the first item of Lemma 1 to show some k is not optimal, it suffices to require that

$$\frac{p_{(k+2)}^2}{(1-p_{(k+2)})^2} > \frac{H}{1-H} \iff p_{(k+2)} > \frac{1}{1+\sqrt{\frac{1-H}{H}}}. \tag{5}$$

Relying on Equation (4), it holds that $p_{(k+2)} \geq F^{-1}\left(\frac{n-k-2}{n}\right) - M\varepsilon$. If we require

$$F^{-1}\left(\frac{n-k-2}{n}\right) - M\varepsilon > \frac{1}{1+\sqrt{\frac{1-H}{H}}}, \tag{6}$$

then Equation (5) is satisfied and hence so will the condition of Lemma 1, which implies that such k cannot be optimal. Equation (6) gives $\frac{k}{n} \leq 1 - F\left(\frac{1}{1+\sqrt{\frac{1-H}{H}}} + M\varepsilon\right) - \frac{2}{n}$, so

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

Multiplying by n yields the desired lower bound. □

This proves that for competencies drawn from an arbitrary distribution whose support is bounded away from 1, a constant fraction of the total population is needed to maximize the probability that the representatives make the correct decision on behalf of the entire population.

We illustrate Theorem 2 by distribution $\mathcal{D} = \mathcal{U}(0.1, 0.9)$. Letting $\varepsilon = \sqrt{\frac{\log n}{2n}}$, one can check that $0.186n \le K^\star \le 0.813n$ with probability at least $1 - \frac{4}{n}$ for sufficiently large ns.

4 Can a Small Congress Outperform Direct Voting?

Our theoretical results from the previous section suggest that the optimal size of a congress should be linear in the size of the population. However, this may not be feasible for many scenarios, and there are other desiderata one must consider in choosing an "optimal" size. Hence, we now turn to *comparing* how well different sizes of congresses perform in the epistemic model.

As a baseline, we will compare the accuracy of a congress to the accuracy of *direct democracy* in which all n members of society vote. This comparison is well-motivated by classic results such as the *Condorcet Jury Theorem* and extensions thereof, which show that the entire society will converge to the correct answer if and only if the competency distribution is *biased* toward the correct answer, that is, $\mathbb{E}_{p\sim\mathcal{D}}[p] > 1/2$. We aim to find bounds on *how* biased this distribution must be for congresses of different sizes to outperform the entire society.

We now state our problem formally. We will be interested in how the cutoff of the bias of the competency distribution varies with n; hence, we will allow the distribution $\mathcal{D}$ to depend on n by having a distribution $\mathcal{D}_n$ for each n. We use F_n and f_n to denote the CDF and PDF of $\mathcal{D}_n$ respectively. Let $\Gamma_n^{\boldsymbol{p}}(k)$ be the gain in probability of correctness by using a congress of size k instead of the entire population, given competence levels $\boldsymbol{p} = (p_{(1)}, \dots, p_{(n)})$:

$$\Gamma_n^{\boldsymbol{p}}(k) = \Pr\left[\sum_{i=1}^{k} X_{(i)} > \frac{k}{2} \;\middle|\; X_{(i)} \sim \mathrm{Bern}(p_{(i)})\right] - \Pr\left[\sum_{i=1}^{n} X_{(i)} > \frac{n}{2} \;\middle|\; X_{(i)} \sim \mathrm{Bern}(p_{(i)})\right].$$

Similar to the definition of $K^\star$, $\Gamma_n^{\boldsymbol{p}}(k)$ is a random variable whose randomness comes from the random draws of $p_i \sim \mathcal{D}_n$. We aim at identifying, for certain values of k, for what kinds of distributions $\mathcal{D}_n$ we have $\Gamma_n^{\boldsymbol{p}}(k) > 0$ with high probability as n grows large.

Dictatorship

First, we consider an extreme case: when can a single voter outperform the entire society? In particular, we identify conditions under which $\Gamma_n^{\boldsymbol{p}}(1) > 0$ or $\Gamma_n^{\boldsymbol{p}}(1) < 0$. We show that if the distributions $\mathcal{D}_n$ put high enough probability mass on competence levels near 1 and its mean $\mathbb{E}_{\mathcal{D}_n}[p]$ is not much larger than $1/2$, then $\Gamma_n^{\boldsymbol{p}}(1) > 0$ with high probability as n grows large, and $\Gamma_n^{\boldsymbol{p}}(1) < 0$ on the contrary. The probability mass conditions are satisfied by many natural classes of distributions; we give several examples (e.g., uniform and beta distributions) in Appendix C.

Theorem 3. *Let* $k = 1$.

- *Suppose* $\mathbb{E}_{\mathcal{D}_n}[p] \le \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$ *and* $f_n(x) \ge \underline{C}(1-x)^{\underline{\beta}-1}$ *for* $x \in [1-\underline{\delta}, 1]$ *for some constants* $a, \underline{C}, \underline{\beta}, \underline{\delta} > 0$. *If* $a < \sqrt{\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \cdot \min\{1, 2/\underline{\beta}\}}$, *then, with probability at least* $1 - n^{-\Omega(1)}$, $\Gamma_n^{\boldsymbol{p}}(1) > 0$.
- *Suppose* $\mathbb{E}_{\mathcal{D}_n}[p] \ge \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$ *and* $f_n(x) \le \overline{C}$ *for* $x \in [1-\overline{\delta}, 1]$ *for some constants* $a, \overline{C}, \overline{\delta} > 0$. *If* $a > \frac{1}{\sqrt{2}}$, *then with probability at least* $1 - n^{-\Omega(1)}$, $\Gamma_n^{\boldsymbol{p}}(1) < 0$.

We sketch a proof of the theorem; the full proof is in Appendix A. When $\mathbb{E}_{\mathcal{D}_n}[p] = \frac{1}{2} + O(\sqrt{\frac{\log n}{n}})$, by Hoeffding's inequality, the entire population makes a correct decision with probability $1 - O(n^{-c_1})$ for some constant c_1, while by our assumption on $\mathcal{D}_n$ the top expert is correct with probability $p_{(1)} = 1 - O(n^{-c_2})$. We identify conditions on $\mathcal{D}_n$ for which $c_1 < c_2$ or $c_1 > c_2$.

Real-world and Polynomial-sized Congress

We now turn our attention to more practical congress sizes. 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 for 240 legislatures (the data comes from https://en.wikipedia.org/wiki/List_of_legislatures_by_number_of_members; we considered the number of representatives to be the total number of representatives in both chambers), we re-ran regression analysis of Auriol and Gary-Bobo (2012) 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 intercept -0.65 and coefficient of determination $R^2 = 0.85$)), suggesting $k = \Theta(n^{0.36})$. See results in Appendix B.

Next, we numerically investigate how congresses of this size perform compared to direct democracy with different levels of bias. We consider $k = n^{0.36}$ and $\mathcal{D}_n = \mathcal{U}(L + \varepsilon_n, 1 - L)$ such that $\mathbb{E}_{\mathcal{D}_n}[p] = \frac{1+\varepsilon_n}{2}$. So the society is slightly biased toward the correct answer. We identify sequences $(\varepsilon_n)_{n=1}^{\infty}$ such that a congress of size k outperforms direct democracy for sufficiently large n.

The simulations were run on a MacBook Pro as follows: for a given distribution, we sample n competencies and votes associated with these competencies. We perform two majority votes — with all the voters and with the top k voters. Repeating this operation $1,000$ times, we estimate the probabilities that the majority of all voters (Direct Democracy) and k voters (Representative Democracy) are correct. Figure 1 displays the probabilities and 95% confidence intervals for different population sizes, with $L = 0.4$. Additional simulations are located in Appendix B.

Let us now formalize and prove this result for general distributions. If the average competence level of the population, $\mathbb{E}_{\mathcal{D}_n}[p]$, is larger than $\frac{1}{2}$ by a constant margin, then both the entire population and a congress of size n^r will be correct with probabilities that are exponentially close to 1. Hence, again, to make things more interesting, we are concerned with the case where $\mathbb{E}_{\mathcal{D}_n}[p] = \frac{1}{2} + \varepsilon_n$ with $0 < \varepsilon_n < o(1)$. We identify conditions on ε_n, n and $\mathcal{D}_n$

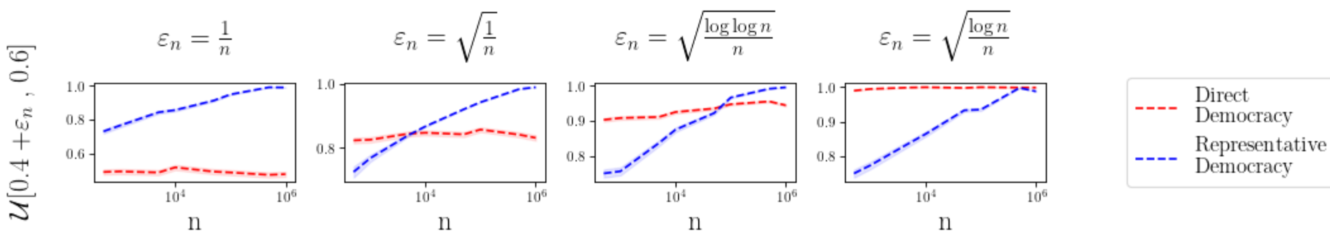

Figure 1: Estimates of $\Pr[\sum_{i=1}^{k} X_{(i)} > \frac{k}{2} \mid \boldsymbol{p}]$ (Representative Democracy) and $\Pr[\sum_{i=1}^{n} X_{(i)} > \frac{n}{2} \mid \boldsymbol{p}]$ (Direct Democracy) as a function of the population size for different values of ε_n, with $k = n^{0.36}$ and $\mathcal{D}_n = \mathcal{U}[0.4 + \varepsilon_n, 0.6]$. For large ε_n, the population size needs to reach a critical mass for the congress to outperform direct democracy.

under which $\Gamma_n^{\boldsymbol{p}}(k) > 0$ or $\Gamma_n^{\boldsymbol{p}}(k) < 0$. The following result is proved in Appendix A.

Theorem 4. *Let $k = n^r$ for some constant $0 < r < 1$.*

- *Suppose $\mathbb{E}_{\mathcal{D}_n}[p] \leq \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$, and $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \geq \frac{k}{n} + \Omega(\sqrt{\frac{\log n}{n}})$ for some constants $a, \alpha > 0$. If $a < \sqrt{\mathbb{E}_{\mathcal{D}_n}[p(1-p)]}$ and $\alpha > \frac{a}{2\sqrt{r \cdot \mathbb{E}_{\mathcal{D}_n}[p(1-p)]}}$, then, with probability at least $1 - n^{-\Omega(1)}$, $\Gamma_n^{\boldsymbol{p}}(k) > 0$.*
- *Suppose $\mathbb{E}_{\mathcal{D}_n}[p] \geq \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$ and $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq \frac{1}{n^{1+\Omega(1)}}$ for some constants $a, \alpha > 0$. If $\alpha < \frac{1}{2}$ and $a > \sqrt{r}\alpha$, then, with probability at least $1 - n^{-\Omega(1)}$, $\Gamma_n^{\boldsymbol{p}}(k) < 0$.*

Intuitively, in the first item above, the condition on the CDF,

$$1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \geq \frac{k}{n} + \Omega(\sqrt{\frac{\log n}{n}}),$$

and the condition on α imply that $\mathcal{D}_n$ assigns large enough probability to high competence levels $p > \frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}$, so a congress of size n^r will be composed of competent enough experts and hence will beat the entire population. The conditions in the second item are in the opposite direction.

We remark that the above conditions on the relation between a and α are sharp: for distributions $\mathcal{D}_n$ that are concentrated around $1/2$, we have $\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \approx 1/4$, so the first condition becomes $\alpha > \frac{a}{2\sqrt{r \cdot 1/4}} = \frac{a}{\sqrt{r}}$, or equivalently $a < \sqrt{r}\alpha$, while the second condition is the opposite: $a > \sqrt{r}\alpha$.

Finally, we note that the conditions in Theorem 4 on the distributions $\mathcal{D}_n$ are satisfied by many natural classes of distributions, e.g., beta distributions and normal distributions truncated to $[0, 1]$. We identify more examples in Appendix C.

5 Discussion

We have 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 the majority under appropriate conditions.

Current debates about the number of representatives in democracies tend to discuss reductions in size, not increases, as embodied by a 2020 Italian referendum approved reducing congress' size from 945 to 600 (De Sio and Angelucci 2019). Indeed, even under the assumption that a larger congress would lead to a "correct" answer more often, this is clearly not the only desiderata 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 (Michener, Amorim Neto, and Civitarese 2021) of the institution are constantly under scrutiny. Designing political institutions relying solely on mathematical insights could yield unforeseen negative externalities (did Madison not warn against the *confusion of the multitude*?). Cognitive, sociological, and economic knowledge should be coupled with mathematical analyses to reach a reasonable trade-off rather than optimizing a single factor.

Incorporating a cost analysis, similar to Magdon-Ismail and Xia (2018) also seems particularly relevant to quantify the trade-off between the congress accuracy and its costs for the constituents. They find that adding a cost polynomial in the number of representatives and a benefit of choosing the correct outcome polynomial in the number of voters decreases the optimal congress size to $O(\log n)$. Finally, this work supports, to some extent, propositions to constitute assemblies of citizens under fluid democracy (Miller 1969; Blum and Zuber 2016; Green-Armytage 2015; Christoff and Grossi 2017; Kahng, Mackenzie, and Procaccia 2021; Gölz et al. 2018; Halpern et al. 2021) that would vote on behalf of the entire population. Indeed, fluid democracy could yield very large citizen assemblies deemed desirable by our findings. Further research on the accuracy of such citizen assemblies could discuss the influence of the voters' weight in the weighted majority's performance.

Acknowledgments

We thank Ariel Procaccia for his helpful feedback. This paper evolved from class projects in Elchanan Mossel's Fall 2019 course, Special Subject in Mathematics, and Procaccia's Spring 2021 course, Optimized Democracy.

References

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

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

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

Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D. 2016. *Handbook of computational social choice*. Cambridge University Press.

Christoff, Z.; and Grossi, D. 2017. Binary voting with delegable proxy: An analysis of liquid democracy. In *Proceedings of the 16th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, 134–150.

De Condorcet, N. 1785. Essai sur l'application de l'analyse la probabilite des decisions rendues a la pluralite des voix. *Paris: L'Imprimerie Royale.*

De Sio, L.; and Angelucci, D. 2019. 945 sono troppi? 600 sono pochi? Qual è il numero "ottimale" di parlamentari? *Cise*.

Gölz, P.; Kahng, A.; Mackenzie, S.; and Procaccia, A. D. 2018. The fluid mechanics of liquid democracy. In *Proceedings of the 14th Conference on Web and Internet Economics (WINE)*, 188–202.

Green-Armytage, J. 2015. Direct voting and proxy voting. *Constitutional Political Economy*, 26(2): 190–220.

Grofman, B.; Owen, G.; and Feld, S. L. 1983. Thirteen theorems in search of the truth. *Theory and decision*, 15(3): 261–278.

Halpern, D.; Halpern, J. Y.; Jadbabaie, A.; Mossel, E.; Procaccia, A. D.; and Revel, M. 2021. In defense of fluid democracy. arXiv:2107.11868.

Jacobs, K.; and Otjes, S. 2015. 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.

Kahng, A.; Mackenzie, S.; and Procaccia, A. 2021. Liquid democracy: An algorithmic perspective. *Journal of Artificial Intelligence Research*, 70: 1223–1252.

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

Magdon-Ismail, M.; and Xia, L. 2018. A mathematical model For optimal decisions in a representative democracy. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS)*, 4707–4716.

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

Massart, P. 1990. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The annals of Probability*, 1269–1283.

Michener, G.; Amorim Neto, O.; and Civitarese, J. 2021. The remoteness of democratic representation. *Party Politics*, 13540688211049545.

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

Nitzan, S.; and Paroush, J. 2017. Collective decision making and jury theorems. *The Oxford Handbook of Law and Economics*, 1.

Polikar, R. 2012. Ensemble learning. In *Ensemble machine learning*, 1–34. Springer.

Sagi, O.; and Rokach, L. 2018. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4): e1249.

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

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

Yang, L. 2011. Classifiers selection for ensemble learning based on accuracy and diversity. *Procedia Engineering*, 15: 4266–4270.

Zhao, L.; and Peng, T. 2020. An allometric scaling for the number of representative nodes in social networks. In *Proceedings of the 6th International Winter School and Conference on Network Science (NetSci-X)*, 49–59.

Appendix

A Missing Proofs

Missing Portion of Proof of Theorem 2

Symmetric to the lower bound, we have that

$$\frac{\Pr[\mathcal{E}_k^{\ell+1}]}{\Pr[\mathcal{E}_k^{\ell}]} \geq \frac{L}{1-L}.$$

Further,

$$\frac{p_{(k+1)}p_{(k+2)}}{(1-p_{(k+1)})(1-p_{(k+2)})} \leq \frac{p_{(k+1)}^2}{(1-p_{(k+1)})^2}.$$

Hence, to prove a certain value $k+2$ is not optimal using Lemma 1, it suffices that

$$\frac{p_{(k+1)}^2}{(1-p_{(k+1)})^2} < \frac{1-L}{L},$$

which is equivalent to

$$p_{(k+1)} < \frac{1}{1+\sqrt{\frac{L}{1-L}}} \tag{7}$$

Now, relying on Equation (4), it holds that

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

If we require

$$F^{-1}(\frac{n-k-1}{n}) + M\varepsilon < \frac{1}{1+\sqrt{\frac{1-L}{L}}}, \tag{8}$$

then Equation (7) is satisfied and hence $p_n^{k+2} - p_n^k < 0$, which implies that such k cannot be optimal. Solving Equation (8) gives $\frac{k}{n} > 1 - F\left(\frac{1}{1+\sqrt{\frac{1-L}{L}}} - M\varepsilon\right) - \frac{1}{n}$.

Hence, as long as $\frac{k}{n} > 1 - F\left(\frac{1}{1+\sqrt{\frac{1-L}{L}}} - M\varepsilon\right) - \frac{1}{n}$, the condition of Lemma 1 will be satisfied. Multiplying through by n yields the desired upper bound. □

Proof of Theorem 3

For the proof, we will need the following lemmas, the first and third are well-known concentration inequalities, and the second is a standard bound on the standard normal CDF, which we prove here for completeness.

Lemma 4 (Berry-Esseen Theorem). *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = \sigma_i^2 > 0$, and $\mathbb{E}[|X_i|^3] = \rho_i < \infty$. Let F_{S_n} be the CDF of $S_n = \frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$ and Φ be the CDF of the standard normal distribution. Then, there exists an absolute constant C_1 such that*

$$|F_{S_n}(x) - \Phi(x)| \leq \frac{C_1}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \max_{1\leq i\leq n} \frac{\rho_i}{\sigma_i^2}, \quad \forall x \in \mathbb{R}$$

Lemma 5 (Bounds on standard normal CDF). *Let $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t$ be the CDF of the standard normal distribution. Then we have for any $x > 0$,*

$$\frac{1}{\sqrt{2\pi}} \frac{x}{x^2+1} e^{-x^2/2} \leq \Phi(-x) = 1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.$$

Proof. The right inequality is because

$$\begin{aligned} 1 - \Phi(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\ &\leq \int_x^\infty \frac{1}{\sqrt{2\pi}} \frac{t}{x} e^{-\frac{t^2}{2}} \mathrm{d}t = \frac{1}{\sqrt{2\pi}} \frac{1}{x} \left(-e^{-\frac{t^2}{2}}\right)\Big|_{t=x}^\infty \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}. \end{aligned}$$

The left inequality is because

$$\begin{aligned}1-\Phi(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,\mathrm{d}t \\ &\geq \int_x^\infty \frac{1}{\sqrt{2\pi}} \frac{(t^2+1)^2-2}{(t^2+1)^2} e^{-\frac{t^2}{2}}\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}}\left(-\frac{t}{t^2+1}e^{-\frac{t^2}{2}}\right)\Big|_{t=x}^{\infty} \\ &= \frac{1}{\sqrt{2\pi}} \frac{x}{x^2+1} e^{-\frac{x^2}{2}}.\end{aligned}$$

□

Lemma 6 (Hoeffding's Inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables bounded by $0 \leq X_i \leq 1$. Then*

$$\Pr\left[\sum_{i=1}^n X_i \geq \mathbb{E}[\sum_{i=1}^n X_i] + t\right] \leq \exp(-\frac{2t^2}{n}),$$

for any $t > 0$. The other direction also holds:

$$\Pr\left[\sum_{i=1}^n X_i \leq \mathbb{E}[\sum_{i=1}^n X_i] - t\right] \leq \exp(-\frac{2t^2}{n}).$$

Now we prove Theorem 3.

Proof of Theorem 3. To simplify notations we write $\Pr\left[\sum_{i=1}^k X_{(i)} > \frac{k}{2} \mid X_{(i)} \sim \mathrm{Bern}(p_{(i)})\right]$ as $\Pr\left[\sum_{i=1}^k X_{(i)} > \frac{k}{2} \mid \boldsymbol{p}\right]$. Recalling the definition of $\Gamma_n^{\boldsymbol{p}}(k)$, since $\Pr\left[\sum_{i=1}^k X_{(i)} > \frac{k}{2} \mid \boldsymbol{p}\right] = 1 - \Pr\left[\sum_{i=1}^k X_{(i)} \leq \frac{k}{2} \mid \boldsymbol{p}\right]$, $\Gamma_n^{\boldsymbol{p}}(k)$ can be equivalently written as

$$\Gamma_n^{\boldsymbol{p}}(k) = \Pr\left[\sum_{i=1}^n X_{(i)} \leq \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] - \Pr\left[\sum_{i=1}^k X_{(i)} \leq \frac{k}{2} \,\middle|\, \boldsymbol{p}\right].$$

To show either $\Gamma_n^{\boldsymbol{p}}(k) > 0$ or $\Gamma_n^{\boldsymbol{p}}(k) < 0$, we will compare $\Pr\left[\sum_{i=1}^n X_{(i)} \leq \frac{n}{2} \mid \boldsymbol{p}\right]$ with $\Pr\left[\sum_{i=1}^k X_{(i)} \leq \frac{k}{2} \mid \boldsymbol{p}\right]$. To do this, we prove the following lemmas:

Lemma 7. *Suppose $\mathbb{E}_{p\sim\mathcal{D}_n}[p] \leq \frac{1}{2} + \varepsilon_n$ where $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$ for some constant $a > 0$. Let $\varepsilon = b\sqrt{\frac{\log n}{n}}$ for some constant $b > 0$. Suppose $\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] > \varepsilon$. Let $c = \frac{a+b}{\sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)]-\varepsilon}}$. Then we have: with probability at least $1 - 2n^{-2b^2}$ (over the random draw of $\boldsymbol{p} \sim \mathcal{D}_n$),*

$$\Pr\left[\sum_{i=1}^n X_{(i)} \leq \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] \geq \frac{1}{\sqrt{2\pi}} \cdot \frac{c\sqrt{\log n}}{(c^2\log n + 1)} \cdot \frac{1}{n^{c^2/2}} - \frac{C_1}{\sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)]-\varepsilon}} \frac{1}{\sqrt{n}},$$

where C_1 is the constant in Berry-Esseen theorem (Lemma 4).

Proof. Given $\boldsymbol{p} = (p_{(1)}, \dots, p_{(n)})$, each $X_{(i)}$ independently follows $\mathrm{Bern}(p_{(i)})$. We use Berry-Esseen theorem (Lemma 4) for $Y_i = X_{(i)} - p_{(i)}$, $i = 1, \dots, n$. Noticing that $\mathbb{E}[Y_i] = 0$, $\sigma_i^2 = \mathbb{E}[Y_i^2] = p_{(i)}(1-p_{(i)})$, and $\rho_i = \mathbb{E}[|Y_i|^3] = p_{(i)}(1-p_{(i)})[(1-p_{(i)})^2 + p_{(i)}^2] \leq \sigma_i^2$, the theorem implies

$$\left|\Pr\left[\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n \sigma_i^2} \leq x\right] - \Phi(x)\right| \leq \frac{C_1}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \max_{1\leq i\leq n} \frac{\rho_i}{\sigma_i^2} \leq \frac{C_1}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \stackrel{\text{def}}{=} \Delta_1$$

for any $x \in \mathbb{R}$, where $\Phi(x)$ is CDF of the standard normal distribution. Therefore,

$$\begin{aligned}\Pr\left[\sum_{i=1}^n X_{(i)} \leq \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] &= \Pr\left[\sum_{i=1}^n X_{(i)} - \sum_{i=1}^n p_{(i)} \leq \frac{n}{2} - \sum_{i=1}^n p_{(i)} \,\middle|\, \boldsymbol{p}\right] \\ &= \Pr\left[\frac{\sum_{i=1}^n Y_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq \frac{\frac{n}{2} - \sum_{i=1}^n p_{(i)}}{\sqrt{\sum_{i=1}^n \sigma_i^2}}\right] \\ &\geq \Phi\left(\frac{\frac{n}{2} - \sum_{i=1}^n p_{(i)}}{\sqrt{\sum_{i=1}^n \sigma_i^2}}\right) - \Delta_1 \qquad (9)\end{aligned}$$

We note that $\sum_{i=1}^n p_{(i)} = \sum_{i=1}^n p_i$ is the sum of n i.i.d. draws from distribution $\mathcal{D}_n$, with mean $\mathbb{E}[\sum_{i=1}^n p_i] = n\mathbb{E}_{p\sim\mathcal{D}_n}[p]$. By Hoeffding's inequality (Lemma 6), letting $t = n\varepsilon$, we have

$$\sum_{i=1}^n p_i \le n\mathbb{E}_{p\sim\mathcal{D}_n}[p] + n\varepsilon \tag{10}$$

with probability at least $1 - \exp(-\frac{2(n\varepsilon)^2}{n}) = 1 - n^{-2b^2}$. Also, $\sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n p_i(1-p_i)$ is the sum of n i.i.d. draws from a distribution, with mean $n\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)]$, so

$$\sum_{i=1}^n \sigma_i^2 \ge n\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - n\varepsilon \tag{11}$$

also with probability at least $1 - \exp(-\frac{2(n\varepsilon)^2}{n}) = 1 - n^{2b^2}$. By a union bound, we have with probability at least $1 - 2n^{-2b^2}$, both Equation (10) and Equation (11) hold, which imply

$$\begin{aligned}
\Pr\left[\sum_{i=1}^n X_{(i)} \le \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] &\ge \Phi\left(\frac{\frac{n}{2} - \sum_{i=1}^n p_{(i)}}{\sqrt{\sum_{i=1}^n \sigma_i^2}}\right) - \Delta_1 \\
&\ge \Phi\left(\frac{\frac{n}{2} - n\mathbb{E}_{p\sim\mathcal{D}_n}[p] - n\varepsilon}{\sqrt{\sum_{i=1}^n \sigma_i^2}}\right) - \Delta_1 \\
&\ge \Phi\left(\frac{-n\varepsilon_n - n\varepsilon}{\sqrt{\sum_{i=1}^n \sigma_i^2}}\right) - \Delta_1 \\
&\ge \Phi\left(\frac{-n\varepsilon_n - n\varepsilon}{\sqrt{n\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - n\varepsilon}}\right) - \Delta_1 \\
&= \Phi\left(-\sqrt{n}\frac{\varepsilon_n + \varepsilon}{\sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - \varepsilon}}\right) - \Delta_1 \\
&= \Phi\left(-\sqrt{n}\frac{(a+b)\sqrt{\frac{\log n}{n}}}{\sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - \varepsilon}}\right) - \Delta_1 \\
&= \Phi\left(-c\sqrt{\log n}\right) - \Delta_1.
\end{aligned}$$

Using Lemma 5 with $x = c\sqrt{\log n}$, we get

$$\Pr\left[\sum_{i=1}^n X_{(i)} \le \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] \ge \Phi\left(-c\sqrt{\log n}\right) - \Delta_1 \ge \frac{1}{\sqrt{2\pi}}\frac{c\sqrt{\log n}}{c^2\log n + 1}\frac{1}{n^{c^2/2}} - \Delta_1,$$

concluding the proof. $\square$

Lemma 8. *Suppose $\mathbb{E}_{p\sim\mathcal{D}_n}[p] \ge \frac{1}{2} + \varepsilon_n$ where $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$ for some constant $a > 0$. Let b be a constant with $0 < b < a$. Then we have: with probability at least $1 - n^{-2b^2}$ (over the random draw of $\boldsymbol{p} \sim \mathcal{D}_n$),*

$$\Pr\left[\sum_{i=1}^n X_{(i)} \le \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] \le \frac{1}{n^{2(a-b)^2}}.$$

Proof. We note that $\sum_{i=1}^n p_{(i)} = \sum_{i=1}^n p_i$ is the sum of n i.i.d. draws from distribution $\mathcal{D}_n$, with mean $\mathbb{E}[\sum_{i=1}^n p_i] = n\mathbb{E}_{p\sim\mathcal{D}_n}[p]$. Let $\varepsilon = b\sqrt{\frac{\log n}{n}} < \varepsilon_n$. By Hoeffding's inequality (Lemma 6), with probability at least $1 - \exp(-\frac{2(n\varepsilon)^2}{n}) = 1 - n^{-2b^2}$, it holds that

$$\sum_{i=1}^n p_i \ge n\mathbb{E}_{p\sim\mathcal{D}_n}[p] - n\varepsilon \ge \frac{n}{2} + n\varepsilon_n - n\varepsilon > \frac{n}{2}.$$

Assuming $\sum_{i=1}^n p_i \geq n\mathbb{E}_{p\sim\mathcal{D}_n}[p] - n\varepsilon$ holds, we consider the conditional probability $\Pr[\sum_{i=1}^n X_{(i)} \leq \frac{n}{2} \mid \boldsymbol{p}]$. Given $\boldsymbol{p}$, $X_{(i)}$'s are independent Bernoulli random variables with means $\mathbb{E}[X_{(i)}] = p_{(i)}$. Hence, by Hoeffding's inequality (Lemma 6),

$$\Pr\left[\sum_{i=1}^n X_{(i)} \leq \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] \leq \exp\left(-\frac{2(\sum_{i=1}^n p_{(i)} - \frac{n}{2})^2}{n}\right)$$
$$\leq \exp\left(-\frac{2(n\varepsilon_n - n\varepsilon)^2}{n}\right) = \exp\left(-2n(\varepsilon_n - \varepsilon)^2\right) = \frac{1}{n^{2(a-b)^2}}.$$

□

Lemma 9. *Suppose the PDF of $\mathcal{D}_n$ satisfies $f_n(x) \geq \underline{C}(1-x)^{\underline{\beta}-1}$ for $x \in [1-\underline{\delta}, 1]$ for some constants $\underline{C}, \underline{\beta}, \underline{\delta} > 0$. Then, for sufficiently large n, with probability at least $1 - n^{-d}$ over the random draw of $\boldsymbol{p} \sim \mathcal{D}_n$,*

$$\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] \leq \left(\frac{\underline{\beta} d \log n}{\underline{C} n}\right)^{1/\underline{\beta}}.$$

Proof. We note that $\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] = 1 - p_{(1)}$, so for any $x \in [0,1]$,

$$\begin{aligned}\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] \leq x] = \Pr[1 - p_{(1)} \leq x] = \Pr[p_{(1)} \geq 1 - x] &= 1 - \Pr[p_{(1)} < 1 - x] \\ &= 1 - \Pr[\max_{1\leq i\leq n} p_i < 1 - x] \\ &= 1 - F_n(1-x)^n.\end{aligned}$$

We let x be such that $F_n(1-x) = 1 - \frac{d\log n}{n}$, i.e., $x = 1 - F_n^{-1}(1 - \frac{d\log n}{n})$, then $F_n(1-x)^n = (1 - \frac{d\log n}{n})^n \leq e^{-d\log n} = n^{-d}$. So, with probability at least $1 - F_n(1-x)^n \geq 1 - n^{-d}$, we have

$$\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] \leq x = 1 - F_n^{-1}\left(1 - \frac{d\log n}{n}\right).$$

We then show that $1 - F_n^{-1}\left(1 - \frac{d\log n}{n}\right) \leq \left(\frac{\underline{\beta} d\log n}{\underline{C} n}\right)^{1/\underline{\beta}}$. Define $G(t) = 1 - F_n(1-t)$ for $t \in [0,1]$. This implies

$$1 - F_n^{-1}(1-y) = G^{-1}(y)$$

for any $y \in [0,1]$. We note that for t sufficiently close to 1, $f_n(x) \geq \underline{C}(1-x)^{\underline{\beta}-1}$ for any $x \in [1-t, 1]$, implying

$$G(t) = 1 - F_n(1-t) = \int_{1-t}^1 f_n(x)\mathrm{d}x \geq \int_{1-t}^1 \underline{C}(1-x)^{\underline{\beta}-1}\mathrm{d}x = \int_0^t \underline{C} u^{\underline{\beta}-1}\mathrm{d}u = \frac{\underline{C}}{\underline{\beta}} t^{\underline{\beta}}.$$

Let $\underline{G}(t) = \frac{\underline{C}}{\underline{\beta}} t^{\underline{\beta}}$. We have $G(t) \geq \underline{G}(t)$ and $\underline{G}^{-1}(y) = (\frac{\underline{\beta}}{\underline{C}} y)^{1/\underline{\beta}}$. Since $G(t) \geq \underline{G}(t)$ and $\underline{G}^{-1}(y)$ is increasing in y, we have

$$G(t) \geq \underline{G}(t) \implies \underline{G}^{-1}(G(t)) \geq t \implies \underline{G}^{-1}(y) \geq G^{-1}(y).$$

Therefore,

$$1 - F_n^{-1}(1-y) = G^{-1}(y) \leq \underline{G}^{-1}(y) = (\frac{\underline{\beta}}{\underline{C}} y)^{1/\underline{\beta}}.$$

Letting $y = \frac{d\log n}{n}$, we conclude that

$$\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] \leq 1 - F^{-1}\left(1 - \frac{d\log n}{n}\right) \leq \left(\frac{\underline{\beta} d\log n}{\underline{C} n}\right)^{1/\underline{\beta}}.$$

□

Lemma 10. *Suppose the PDF of $\mathcal{D}_n$ satisfies $f_n(x) \leq \overline{C}$ for $x \in [1-\overline{\delta}, 1]$ for some constants $\overline{C}, \overline{\delta} > 0$. Then, for sufficiently large n, with probability at least $1 - n^{-d}$ over the random draw of $\boldsymbol{p} \sim \mathcal{D}_n$,*

$$\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] \geq \frac{1}{\overline{C} n^{d+1}}.$$

Proof. We note that $\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] = 1 - p_{(1)}$, so for any $x \in [0,1]$,

$$\begin{aligned}\Pr[\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] \geq x] = \Pr[1 - p_{(1)} \geq x] = \Pr[p_{(1)} \leq 1 - x] &= \Pr[\max_{1 \leq i \leq n} p_i < 1 - x] \\ &= F_n(1-x)^n.\end{aligned}$$

We let $x = \frac{1}{\overline{C} n^{d+1}}$. Then for sufficiently large n, $x \geq 1 - \overline{\delta}$, and hence $f_n(t) \leq \overline{C}$ for $t \in [1-x, 1]$, which implies

$$1 - F_n(1-x) = \int_{1-x}^{1} f_n(t)\mathrm{d}t \leq \int_{1-x}^{1} \overline{C}\mathrm{d}t = x\overline{C} = \frac{1}{n^{d+1}},$$

or equivalently

$$F_n(1-x) \geq 1 - \frac{1}{n^{d+1}}.$$

Using inequality $(1 - \frac{x}{n})^n \geq 1 - x$ (for $n \geq 1, 0 \leq x \leq n$), we get

$$F_n(1-x)^n \geq \left(1 - \frac{1}{n^{d+1}}\right)^n \geq 1 - \frac{1}{n^d}.$$

Therefore, with probability at least $1 - \frac{1}{n^d}$, $\Pr[X_{(1)} = 0 \mid \boldsymbol{p}] \geq x = \frac{1}{\overline{C} n^{d+1}}$ holds. ◻

To prove $\Gamma_n^{\boldsymbol{p}}(1) > 0$, we use Lemma 7 and Lemma 9 to get

$$\begin{aligned}\Gamma_n^{\boldsymbol{p}}(1) &= \Pr\left[\sum_{i=1}^{n} X_{(i)} \leq \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] - \Pr\left[X_{(1)} = 0 \,\middle|\, \boldsymbol{p}\right] \\ &\geq \frac{1}{\sqrt{2\pi}} \frac{c\sqrt{\log n}}{(c^2 \log n + 1)} \frac{1}{n^{c^2/2}} - \frac{C_1}{\sqrt{\mathbb{E}_{p \sim \mathcal{D}_n}[p(1-p)] - \varepsilon}} \frac{1}{\sqrt{n}} - \left(\frac{\underline{\beta} d \log n}{\underline{C} n}\right)^{1/\underline{\beta}}\end{aligned}$$

with probability at least $1 - 2n^{-2b^2} - n^{-d}$, where $c = \frac{a+b}{\sqrt{\mathbb{E}_{p \sim \mathcal{D}_n}[p(1-p)] - \varepsilon}}$, $\mathbb{E}_{p \sim \mathcal{D}_n}[p] \leq \frac{1}{2} + \varepsilon_n$ with $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$ for some $a > 0$, and $\varepsilon = b\sqrt{\frac{\log n}{n}}$ for some $b > 0$, and $\underline{C}$ and $\underline{\beta}$ are constants. If $c^2/2$ is a constant such that

$$c^2/2 < \min\left\{1/2, 1/\underline{\beta}\right\},$$

then $\Gamma_n^{\boldsymbol{p}}(1) = O(\frac{1}{n^{c^2/2}}) > 0$ for sufficiently large n. Requiring $c^2/2 < \min\left\{1/2, 1/\underline{\beta}\right\}$ is equivalent to requiring

$$a + b < \sqrt{(\mathbb{E}_{p \sim \mathcal{D}_n}[p(1-p)] - \varepsilon) \cdot \min\{1, 2/\underline{\beta}\}},$$

which can be satisfied when a and b are constants such that $a < \sqrt{\mathbb{E}_{p \sim \mathcal{D}_n}[p(1-p)] \cdot \min\{1, 2/\underline{\beta}\}}$, $0 < b < \sqrt{\mathbb{E}_{p \sim \mathcal{D}_n}[p(1-p)] \cdot \min\{1, 2/\underline{\beta}\}} - a$, and n is sufficiently large (so $\varepsilon = b\sqrt{\frac{\log n}{n}}$ is sufficiently small).

To prove $\Gamma_n^{\boldsymbol{p}}(1) < 0$, we use Lemma 8 and Lemma 10 to get

$$\begin{aligned}\Gamma_n^{\boldsymbol{p}}(1) &= \Pr\left[\sum_{i=1}^{n} X_{(i)} \leq \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] - \Pr\left[X_{(1)} = 0 \,\middle|\, \boldsymbol{p}\right] \\ &\leq \frac{1}{n^{2(a-b)^2}} - \frac{1}{\overline{C} n^{d+1}}\end{aligned}$$

with probability at least $1 - n^{-2b^2} - n^{-d}$, where $\mathbb{E}_{p \sim \mathcal{D}_n}[p] \geq \frac{1}{2} + \varepsilon_n$ with $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$ for some constant $a > 0$, with any $b < a$, and $\overline{C}$ is a constant. When

$$2(a-b)^2 > d + 1,$$

we have $\Gamma_n^{\boldsymbol{p}}(1) = -O(\frac{1}{n^{d+1}}) < 0$ for sufficiently large n. The inequality $2(a-b)^2 > d+1$ is satisfied when $a > \frac{1}{\sqrt{2}}$ and b, d are sufficiently close to 0. ◻

Proof of Theorem 4

Similar to the proof of Theorem 3 (in Appendix A), we write $\Gamma_n^{\boldsymbol{p}}(k)$ as

$$\Gamma_n^{\boldsymbol{p}}(k) = \Pr\left[\sum_{i=1}^{n} X_{(i)} \le \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] - \Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right].$$

To show either $\Gamma_n^{\boldsymbol{p}}(k) > 0$ or $\Gamma_n^{\boldsymbol{p}}(k) < 0$, we will compare $\Pr\left[\sum_{i=1}^{n} X_{(i)} \le \frac{n}{2} \mid \boldsymbol{p}\right]$ with $\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \mid \boldsymbol{p}\right]$.

Lemma 11. *Suppose* $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \ge \frac{k}{n} + \varepsilon$ *where* $\varepsilon = b\sqrt{\frac{\log n}{n}}$ *for some constants* $\alpha, b > 0$. *Then, with probability at least* $1 - 2n^{-2b^2}$ *(over the random draw of* $\boldsymbol{p} \sim \mathcal{D}_n$*),*

$$\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] \le \frac{1}{k^{2\alpha^2}}.$$

Proof. By DKW inequality (Lemma 2.5), with probability at least $1 - 2e^{-2n\varepsilon^2} = 1 - 2n^{-2b^2}$ over the random draw of $\boldsymbol{p} \sim \mathcal{D}_n$, it holds that $|F_n(p_{(i)}) - \frac{n-i}{n}| \le \varepsilon$ for every $i \in [n]$. In particular, for $i = 1, \dots, k$, we have

$$F_n(p_{(i)}) \ge \frac{n-i}{n} - \varepsilon \ge \frac{n-k}{n} - \varepsilon = 1 - \frac{k}{n} - \varepsilon,$$

This implies

$$1 - F_n(p_{(i)}) \le \frac{k}{n} + \varepsilon \le 1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}})$$

and hence

$$p_{(i)} \ge \frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}.$$

Assuming the above inequalities hold, we consider the conditional probability $\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \mid \boldsymbol{p}\right]$. Given $\boldsymbol{p}$, the $X_{(i)}$'s are independent draws from $\mathrm{Bern}(p_{(i)})$ distributions, with means $\mathbb{E}[X_{(i)}] = p_{(i)}$, hence, by Hoeffding's inequality (Lemma 6),

$$\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] \le \exp\left(-\frac{2(\sum_{i=1}^{k} p_{(i)} - \frac{k}{2})^2}{k}\right).$$

Plugging in $p_{(i)} \ge \frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}$, we get

$$\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] \le \exp\left(-\frac{2(\frac{k}{2} + \alpha\sqrt{k\log k} - \frac{k}{2})^2}{k}\right) = \frac{1}{k^{2\alpha^2}}.$$

$\square$

Proof of the first item of Theorem 4. By Lemma 7 and Lemma 11, we have

$$\begin{aligned}\Gamma_n^{\boldsymbol{p}}(k) &= \Pr\left[\sum_{i=1}^{n} X_{(i)} \le \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] - \Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] \\ &\ge \frac{1}{\sqrt{2\pi}} \frac{c\sqrt{\log n}}{(c^2\log n + 1)} \frac{1}{n^{c^2/2}} - \frac{C_1}{\sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - \varepsilon}} \frac{1}{\sqrt{n}} - \frac{1}{k^{2\alpha^2}}\end{aligned}$$

with probability at least $1 - 4n^{-2b^2}$, where $c = \frac{a+b}{\sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - \varepsilon}}$, $\mathbb{E}_{p\sim\mathcal{D}_n}[p] \le \frac{1}{2} + \varepsilon_n$ with $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$ for some $a > 0$, and $\varepsilon = b\sqrt{\frac{\log n}{n}}$ for some $b > 0$, and α is a constant. Since $k = n^r$,

$$\Gamma_n^{\boldsymbol{p}}(k) \ge \frac{1}{\sqrt{2\pi}} \frac{c\sqrt{\log n}}{(c^2\log n + 1)} \frac{1}{n^{c^2/2}} - \frac{C_1}{\sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - \varepsilon}} \frac{1}{\sqrt{n}} - \frac{1}{n^{2r\alpha^2}}$$

When $c^2/2 < 1/2$ and $c^2/2 < 2r\alpha^2$, we have $\Gamma_n^{\boldsymbol{p}}(k) = O(\frac{1}{n^{c^2/2}}) > 0$ for sufficiently large n. The latter requirement $c^2/2 < 2r\alpha^2$ is satisfied when $\alpha > \frac{c}{2\sqrt{r}}$. The former requirement $c^2/2 < 1/2$ is equivalent to $a+b < \sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)] - \varepsilon\}}$, which is satisfied when constants $a < \sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)]\}}$, $0 < b < \sqrt{\mathbb{E}_{p\sim\mathcal{D}_n}[p(1-p)]\}} - a$, and n is sufficiently large. $\square$

Lemma 12. *Suppose $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \le \frac{1}{n^{1+\Omega(1)}}$ for some constant $\alpha > 0$, and suppose $\mathbb{E}_{p\sim\mathcal{D}_n}[p] \ge \frac{1}{2} + \varepsilon_n$ with $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$ for some constant $a > 0$. Then, with probability at least $1 - \frac{1}{n^{\Omega(1)}}$ (over the random draw of $\boldsymbol{p} \sim \mathcal{D}_n$),*

$$\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] \ge \frac{1}{\sqrt{2\pi}} \cdot \frac{1-o(1)}{2\alpha\sqrt{\log k}} \cdot \frac{1}{k^{\frac{2\alpha^2}{1-o(1)}}} - \frac{2C_1}{(1-o(1))\sqrt{k}},$$

where C_1 is the constant in Berry-Esseen theorem (Lemma 4).

Proof. Given $p_{(1)}, \ldots, p_{(k)}$, each $X_{(i)}$ independently follows $\mathrm{Bern}(p_{(i)})$. We use Berry-Esseen theorem (Lemma 4) for $Y_i = X_{(i)} - p_{(i)}$, $i = 1, \ldots, k$. Noticing that $\mathbb{E}[Y_i] = p_{(i)}$, $\sigma_i^2 = \mathbb{E}[Y_i^2] = p_{(i)}(1-p_{(i)})$, and $\rho_i = \mathbb{E}[|Y_i|^3] = p_{(i)}(1-p_{(i)})[(1-p_{(i)})^2 + p_{(i)}^2] \le \sigma_i^2$, the theorem implies

$$\left|\Pr\left[\frac{\sum_{i=1}^k Y_i}{\sum_{i=1}^k \sigma_i^2} \le x\right] - \Phi(x)\right| \le \frac{C_1}{\sqrt{\sum_{i=1}^k \sigma_i^2}} \max_{1\le i\le k} \frac{\rho_i}{\sigma_i^2} \le \frac{C_1}{\sqrt{\sum_{i=1}^k \sigma_i^2}}$$

for any $x \in \mathbb{R}$, where $\Phi(x)$ is CDF of the standard normal distribution. Therefore,

$$\begin{aligned}\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] &= \Pr\left[\sum_{i=1}^{k} X_{(i)} - \sum_{i=1}^{k} p_{(i)} \le \frac{k}{2} - \sum_{i=1}^{k} p_{(i)} \,\middle|\, \boldsymbol{p}\right] \\ &= \Pr\left[\frac{\sum_{i=1}^k Y_i}{\sqrt{\sum_{i=1}^k \sigma_i^2}} \le \frac{\frac{k}{2} - \sum_{i=1}^k p_{(i)}}{\sqrt{\sum_{i=1}^k \sigma_i^2}}\right] \\ &\ge \Phi\left(\frac{\frac{k}{2} - \sum_{i=1}^k p_{(i)}}{\sqrt{\sum_{i=1}^k \sigma_i^2}}\right) - \frac{C_1}{\sqrt{\sum_{i=1}^k \sigma_i^2}}.\end{aligned}$$

We consider $\sum_{i=1}^k p_{(i)}$. By the assumption that $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) = \Pr_{p_i\sim\mathcal{D}_n}[p_i > \frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}] = \frac{1}{n^{1+\Omega(1)}}$, using a union bound we have with probability at least $1 - n\frac{1}{n^{1+\Omega(1)}} = 1 - \frac{1}{n^{\Omega(1)}}$, all p_i's (for $i = 1, \ldots, n$) satisfy $p_i \le \frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}$. Hence,

$$\sum_{i=1}^{k} p_{(i)} \le k\left(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}\right) = \frac{k}{2} + \alpha\sqrt{k\log k},$$

which implies

$$\begin{aligned}\Pr\left[\sum_{i=1}^{k} X_{(i)} \le \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] &\ge \Phi\left(\frac{\frac{k}{2} - (\frac{k}{2} + \alpha\sqrt{k\log k})}{\sqrt{\sum_{i=1}^k \sigma_i^2}}\right) - \frac{C_1}{\sqrt{\sum_{i=1}^k \sigma_i^2}} \\ &= \Phi\left(\frac{-\alpha\sqrt{k\log k}}{\sqrt{\sum_{i=1}^k \sigma_i^2}}\right) - \frac{C_1}{\sqrt{\sum_{i=1}^k \sigma_i^2}} \qquad (12)\end{aligned}$$

We then consider $\sum_{i=1}^k \sigma_i^2 = \sum_{i=1}^k p_{(i)}(1-p_{(i)}) = \sum_{i=1}^k p_{(i)} - \sum_{i=1}^k p_{(i)}^2$. We note that the p_i's (for $i = 1, \ldots, n$) are n i.i.d. random draws from distribution $\mathcal{D}_n$ whose mean is $\mathbb{E}_{p\sim\mathcal{D}_n}[p] \ge \frac{1}{2} + \varepsilon_n$, by Hoeffding's inequality, their average satisfies

$$\frac{1}{n}\sum_{i=1}^{n} p_i \ge \mathbb{E}_{p\sim\mathcal{D}_n}[p] - \varepsilon \ge \frac{1}{2} + \varepsilon_n - \varepsilon,$$

with probability at least $1 - \exp(-2n\varepsilon^2)$. We choose $\varepsilon = O(\sqrt{\frac{\log n}{n}})$ so the probability is $1 - \frac{1}{n^{\Omega(1)}}$. We also note that $\frac{1}{n}\sum_{i=1}^n p_i \le \frac{1}{k}\sum_{i=1}^k p_{(i)}$ because $p_{(1)}, \ldots, p_{(k)}$ are the k largest values in $p_1, \ldots, p_n$. Therefore,

$$\sum_{i=1}^{k} p_{(i)} \ge \frac{k}{n}\sum_{i=1}^{n} p_i \ge k\left(\frac{1}{2} + \varepsilon_n - \varepsilon\right).$$

Moreover, since previously we had $p_i \leq \frac{1}{2} + \alpha\sqrt{\frac{\log n}{n}}$ for all $i = 1, \dots, n$, it holds that

$$\sum_{i=1}^{k} p_{(i)}^2 \leq k\left(\frac{1}{2} + \alpha\sqrt{\frac{\log n}{n}}\right)^2 = k(\frac{1}{4} + o(1)).$$

Therefore,

$$\sum_{i=1}^{k} \sigma_i^2 = \sum_{i=1}^{k} p_{(i)} - \sum_{i=1}^{k} p_{(i)}^2 \geq k(\frac{1}{2} + \varepsilon_n - \varepsilon) - k(\frac{1}{4} + o(1)) = k(\frac{1}{4} - o(1)).$$

Plugging $\sum_{i=1}^{k} \sigma_i^2 \geq k(\frac{1}{4} - o(1))$ into Equation (12), we get

$$\begin{aligned}\Pr\left[\sum_{i=1}^{k} X_{(i)} \leq \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] &\geq \Phi\left(\frac{-\alpha\sqrt{k\log k}}{\sqrt{k(\frac{1}{4} - o(1))}}\right) - \frac{C_1}{\sqrt{k(\frac{1}{4} - o(1))}} \\ &= \Phi\left(\frac{-2\alpha\sqrt{\log k}}{1 - o(1)}\right) - \frac{2C_1}{(1 - o(1))\sqrt{k}}.\end{aligned}$$

Using Lemma 5 with $x = \frac{2\alpha\sqrt{\log k}}{1 - o(1)}$, we have

$$\Phi\left(\frac{-2\alpha\sqrt{\log k}}{1 - o(1)}\right) \geq \frac{1}{\sqrt{2\pi}} \frac{2\alpha\sqrt{\log k}(1 - o(1))}{4\alpha^2 \log k + 1} e^{-\frac{4\alpha^2 \log k}{2(1 - o(1))}} = \frac{1}{\sqrt{2\pi}} \frac{1 - o(1)}{2\alpha\sqrt{\log k}} \frac{1}{k^{\frac{2\alpha^2}{1 - o(1)}}}$$

which implies

$$\Pr\left[\sum_{i=1}^{k} X_{(i)} \leq \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] \geq \frac{1}{\sqrt{2\pi}} \frac{1 - o(1)}{2\alpha\sqrt{\log k}} \frac{1}{k^{\frac{2\alpha^2}{1 - o(1)}}} - \frac{2C_1}{(1 - o(1))\sqrt{k}},$$

concluding the proof. □

Proof of the second item of Theorem 4. To prove $\Gamma_n^{\boldsymbol{p}}(k) < 0$, we use Lemma 8 and Lemma 12 to get

$$\begin{aligned}\Gamma_n^{\boldsymbol{p}}(k) &= \Pr\left[\sum_{i=1}^{n} X_{(i)} \leq \frac{n}{2} \,\middle|\, \boldsymbol{p}\right] - \Pr\left[\sum_{i=1}^{k} X_{(i)} \leq \frac{k}{2} \,\middle|\, \boldsymbol{p}\right] \\ &\leq \frac{1}{n^{2(a-b)^2}} - \frac{1}{\sqrt{2\pi}} \frac{1 - o(1)}{2\alpha\sqrt{\log k}} \frac{1}{k^{\frac{2\alpha^2}{1 - o(1)}}} + \frac{2C_1}{(1 - o(1))\sqrt{k}},\end{aligned}$$

with probability at least $1 - n^{-2b^2} - n^{-\Omega(1)} = 1 - n^{-\Omega(1)}$, where $\mathbb{E}_{p\sim\mathcal{D}_n}[p] \geq \frac{1}{2} + \varepsilon_n$ with $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$ for some $a > 0$, $0 < b < a$, $1 - F_n(1 + \alpha\sqrt{\frac{\log k}{k}}) = \frac{1}{n^{1+\Omega(1)}}$ for some $\alpha > 0$, and C_1 is some constant. Since $k = n^r$, or $n = k^{\frac{1}{r}}$,

$$\Gamma_n^{\boldsymbol{p}}(k) \leq \frac{1}{k^{\frac{2(a-b)^2}{r}}} - \frac{1}{\sqrt{2\pi}} \frac{1 - o(1)}{2\alpha\sqrt{\log k}} \frac{1}{k^{\frac{2\alpha^2}{1 - o(1)}}} + \frac{2C_1}{(1 - o(1))\sqrt{k}},$$

When inequalities $\frac{2\alpha^2}{1 - o(1)} < \frac{2(a-b)^2}{r}$ and $\frac{2\alpha^2}{1 - o(1)} < \frac{1}{2}$ are satisfied, we have $\Gamma_n^{\boldsymbol{p}}(k) = -O\left(\frac{1}{\sqrt{\log k}} \frac{1}{k^{\frac{2\alpha^2}{1 - o(1)}}}\right) < 0$ for sufficiently large n. The former is satisfied when $a > \sqrt{r}\alpha$ and b is sufficiently close to 0. The latter is satisfied when $\alpha < \frac{1}{2}$. □

B Figures

Optimal Congress Size

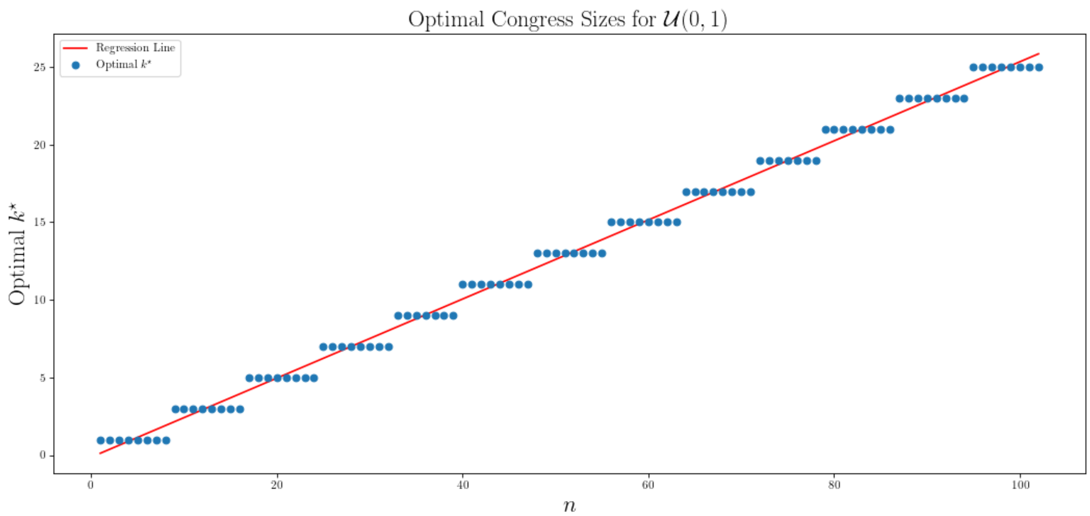

Figure 2: Optimal value of k for $\mathcal{U}(0,1)$ competence levels following their expectation. The line of best fit is very close to $n/4$.

Real-world congress sizes

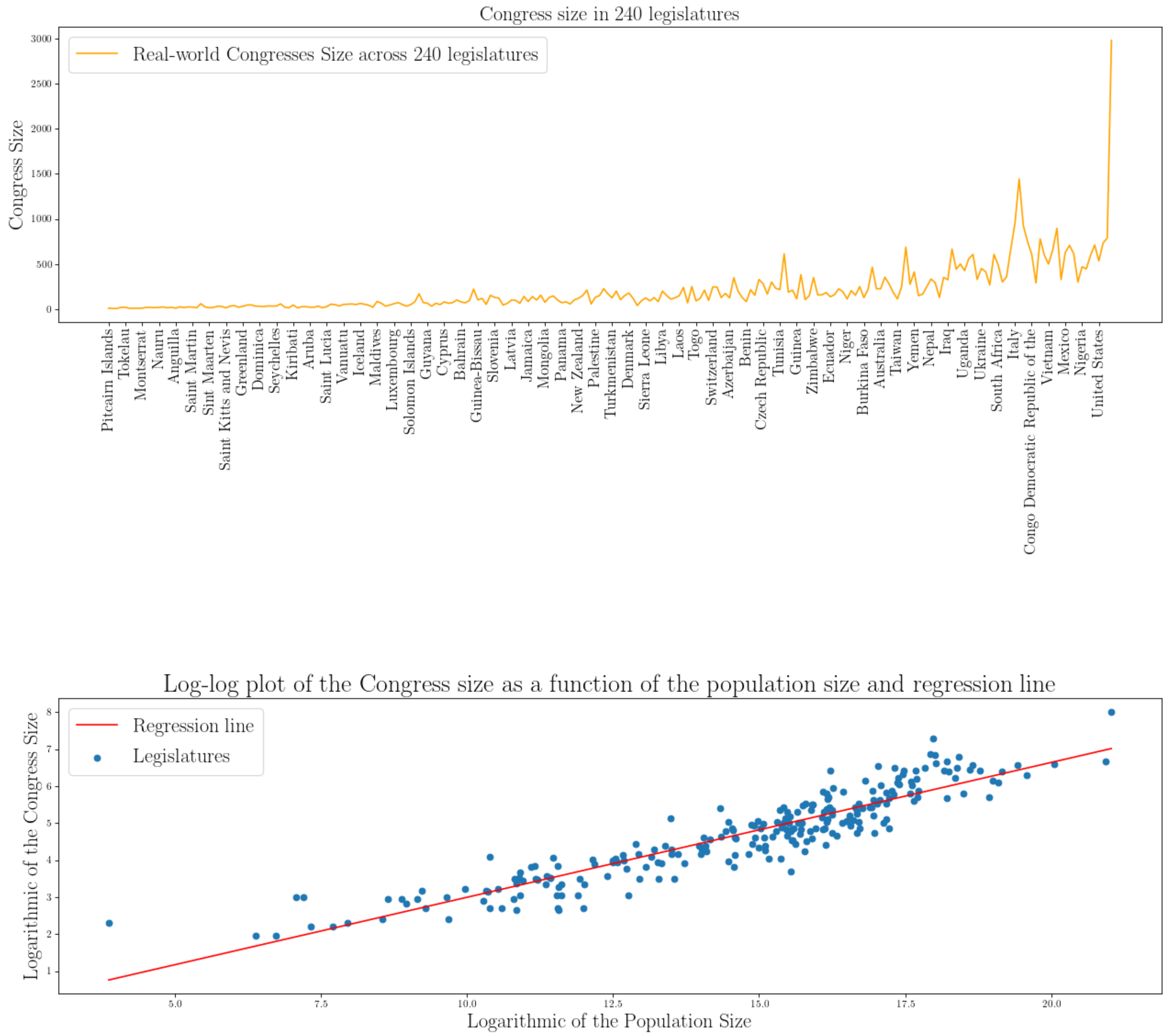

Figure 3: Congress sizes in 240 legislatures (top) and log-log plot of the Congress size as a function of the Population size (bottom). The regression line yields $\log k = 0.36 \log n - 0.65$, or $k = cn^{0.36}$, with a coefficient of determination $R^2 = 0.85$. Note that in the top plot, we only show a handful of countries for obvious space constraints. In reality, the United States is not the country with the largest congress (it has 535 congress-members per our computation, merging both chambers).

Small congresses outperform majority voting

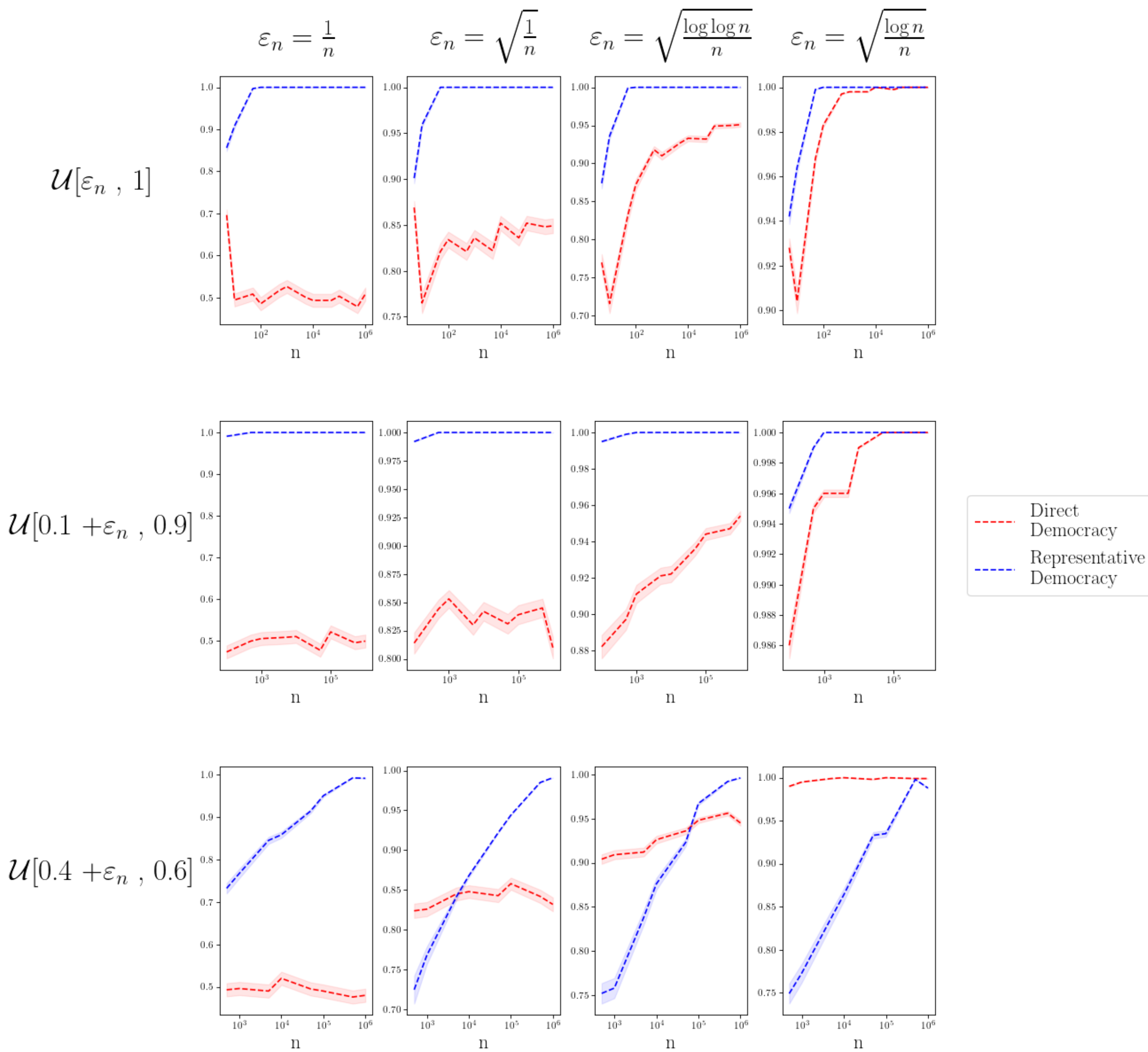

Figure 4: Estimates of $\Pr[\sum_{i=1}^{k} X_{(i)} > \frac{k}{2} \mid \boldsymbol{p}]$ (Representative Democracy) and $\Pr[\sum_{i=1}^{n} X_{(i)} > \frac{n}{2} \mid \boldsymbol{p}]$ (Direct Democracy) with 95% confidence intervals as a function of the population size for different values of ε_n, with $k = n^{0.36}$ and $\mathcal{D}_n = \mathcal{U}[0.4+\varepsilon_n, 0.6]$. For large society biases, the population size needs to reach a critical mass for the congress to outperform direct democracy. 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 $L = 0.4$.

Unsurprisingly, the larger the bias, the smaller the gain. For $L \leq 0.1$ and a bias of order $\sqrt{\log n/n}$, there is a no gain from relying on the congress, while if the bias is of order $\sqrt{\log\log n/n}$, there is positive gain. Yet, for $L = 0.4$, a bias of order $\sqrt{\log n/n}$ systematically yields a strictly negative gain for $n \leq 10^6$.

C Distribution Examples

Distributions satisfying Theorem 3

We recall the conditions on competency distributions $\mathcal{D}_n$ under which $\Gamma_n^{\boldsymbol{p}}(1) > 0$ or $\Gamma_n^{\boldsymbol{p}}(1) < 0$ in Theorem 3: for $\Gamma_n^{\boldsymbol{p}}(1) > 0$, we require $\mathbb{E}_{\mathcal{D}_n}[p] \leq \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$ and $f_n(x) \geq \underline{C}(1-x)^{\underline{\beta}-1}$ for $x \in [1-\underline{\delta}, 1]$ with constants $a, \underline{C}, \underline{\beta}, \underline{\delta} > 0$ such that $a < \sqrt{\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \cdot \min\{1, 2/\underline{\beta}\}}$; for $\Gamma_n^{\boldsymbol{p}}(1) < 0$, we require $\mathbb{E}_{\mathcal{D}_n}[p] \geq \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$ and $f_n(x) \leq \overline{C}$ for $x \in [1-\overline{\delta}, 1]$ with constants $a, \overline{C}, \overline{\delta} > 0$ such that $a > \frac{1}{\sqrt{2}}$. We give examples of beta distributions and uniform distributions satisfying those conditions:

Example 1.

- *Beta distributions: Consider* $\mathcal{D}_n = \mathrm{Beta}(\beta + \varepsilon_n, \beta)$*, where* $\mathbb{E}_{\mathcal{D}_n}[p] = \frac{\beta+\varepsilon_n}{2\beta+\varepsilon_n} = \frac{1}{2} + \frac{\varepsilon_n}{4\beta+2\varepsilon_n}$ *and* $f_n(x) = \frac{1}{\mathrm{B}(\beta+\varepsilon_n,\beta)}x^{\beta+\varepsilon_n-1}(1-x)^{\beta-1}$ *where* $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$*. Let* β *be a constant and suppose* $\varepsilon_n = 4\beta a\sqrt{\frac{\log n}{n}}$*. Since* $\varepsilon_n \approx 0$*, we have* $\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \approx \frac{1}{4} - \frac{1}{8\beta+4}$*.*
  - *For* $\Gamma_n^{\boldsymbol{p}}(1) > 0$*: First, we have* $f_n(x) \geq \underline{C}(1-x)^{\beta-1}$ *because* $\mathrm{B}(\beta+\varepsilon_n,\beta)$ *is upper bounded and* $x^{\beta+\varepsilon_n-1}$ *is lower bounded for* x *close to* 1*. In addition,* $\mathbb{E}_{\mathcal{D}_n}[p] \leq \frac{1}{2} + \frac{\varepsilon_n}{4\beta} = \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$*. When* $a < \sqrt{\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \cdot \min\{1, 2/\beta\}} \approx \sqrt{(\frac{1}{4} - \frac{1}{8\beta+4}) \cdot \min\{1, 2/\beta\}}$*, the condition is satisfied.*
  - *For* $\Gamma_n^{\boldsymbol{p}}(1) < 0$*: Clearly,* $f_n(x) \leq \frac{1}{\mathrm{B}(\beta+\varepsilon_n,\beta)} \leq \overline{C} < \infty$*. In addition,* $\mathbb{E}_{\mathcal{D}_n}[p] \approx \frac{1}{2} + \frac{\varepsilon_n}{4\beta} = \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$*. When* $a > \frac{1}{\sqrt{2}}$*, the condition is satisfied.*
- *Uniform distributions: Let* $\mathcal{D}_n = \mathcal{U}(2\varepsilon_n, 1)$*, where* $\mathbb{E}_{\mathcal{D}_n}[p] = \frac{1}{2} + \varepsilon_n$ *and* $f_n(x) = \frac{1}{1-2\varepsilon_n}$*. Let* $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$*. Since* $\varepsilon_n \approx 0$*, we have* $\underline{C} = 1 \leq f_n(x) \leq 2 = \overline{C}$*. Then*
  - *For* $\Gamma_n^{\boldsymbol{p}}(1) > 0$*: the condition is satisfied when* $a < \sqrt{\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \cdot \min\{1, 2/\underline{\beta}\}} \approx \sqrt{\frac{1}{6}}$ *(here* $\underline{\beta} = 1$*).*
  - *For* $\Gamma_n^{\boldsymbol{p}}(1) < 0$*: the condition is satisfied when* $a > \frac{1}{\sqrt{2}}$*.*

Distributions satisfying Theorem 4

We recall the conditions on competency distribution $\mathcal{D}_n$ under which $\Gamma_n^{\boldsymbol{p}}(k) > 0$ or $\Gamma_n^{\boldsymbol{p}}(k) < 0$ in Theorem 4: for $\Gamma_n^{\boldsymbol{p}}(k) > 0$, we require that its mean satisfies $\mathbb{E}_{\mathcal{D}_n}[p] \leq \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$ and CDF satisfies $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \geq \frac{k}{n} + \Omega(\sqrt{\frac{\log n}{n}})$ for constants $a, \alpha > 0$ such that $a < \sqrt{\mathbb{E}_{\mathcal{D}_n}[p(1-p)]}$ and $\alpha > \frac{a}{2\sqrt{r \cdot \mathbb{E}_{\mathcal{D}_n}[p(1-p)]}}$; for $\Gamma_n^{\boldsymbol{p}}(k) < 0$, we require that its mean satisfies $\mathbb{E}_{\mathcal{D}_n}[p] \geq \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$ and CDF satisfies $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq \frac{1}{n^{1+\Omega(1)}}$ for constants $a, \alpha > 0$ such that $\alpha < \frac{1}{2}$ and $a > \sqrt{r}\alpha$. We give examples of normal distributions and beta distributions satisfying those conditions.

Example 2. *Recall that* $k = n^r$ *for some constant* $0 < r < 1$*. In this example, we show that distributions with large variance are more likely to satisfy the condition for* $\Gamma_n^{\boldsymbol{p}}(k) > 0$ *while distributions with small variance satisfy the condition for* $\Gamma_n^{\boldsymbol{p}}(k) < 0$*. We consider normal and beta distributions.*

- *Normal distributions: Let* $\mathcal{D}_n$ *be the distribution of* $p \sim \mathcal{N}(\mu_n = \frac{1}{2} + a\sqrt{\frac{\log n}{n}}, \sigma_n^2 = \frac{\sigma^2}{k})$ *conditioning on* $p \in [0,1]$*, where* σ^2 *is a constant to be chosen. We note that for large* k *(or large* n*), the variance* $\sigma_n^2 = \frac{\sigma^2}{k}$ *is small, so* p *is centered around* $\mu_n \approx \frac{1}{2}$*, thus* $\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \approx \frac{1}{4}$*.*
  - *For* $\Gamma_n^{\boldsymbol{p}}(k) > 0$*: Let* a, α *be any constants such that* $a < \sqrt{E_{\mathcal{D}_n}[p(1-p)]} \approx \frac{1}{4}$*,* $\alpha > \frac{a}{2\sqrt{r \cdot \mathbb{E}_{\mathcal{D}_n}[p(1-p)]}} \approx \frac{a}{\sqrt{r}}$*. We claim that the CDF condition* $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \geq \frac{k}{n} + \Omega(\sqrt{\frac{\log n}{n}})$ *is satisfied when* $\sigma^2 > \frac{r\alpha^2}{2\min\{1-r, 1/2\}}$*. (A proof is given below).*
  - *For* $\Gamma_n^{\boldsymbol{p}}(k) < 0$*: Let* a, α *be any constants such that* $\alpha < \frac{1}{2}$ *and* $a > \sqrt{r}\alpha$*. We claim that the CDF condition* $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq \frac{1}{n^{1+\Omega(1)}}$ *is satisfied when* $\sigma^2 < \frac{r\alpha^2}{2(1+\Omega(1))}$*.*
- *Beta distributions: Let* $\mathcal{D}_n = \mathrm{Beta}(\beta + 4\beta\varepsilon_n, \beta)$ *where* $\beta = \gamma k$ *for some constant* γ *to be chosen, and* $\varepsilon_n = a\sqrt{\frac{\log n}{n}}$*. For simplicity we suppose* $r < \frac{1}{2}$*, so* $\beta\varepsilon_n = \gamma k a\sqrt{\frac{\log n}{n}} = \gamma a\frac{\sqrt{\log n}}{n^{1/2-r}} \to 0$ *as* n *grows. Then the mean satisfies* $\mathbb{E}_{\mathcal{D}_n}[p] =$

$\frac{\beta+4\beta\varepsilon_n}{2\beta+4\beta\varepsilon_n} = \frac{1}{2} + \frac{\varepsilon_n}{1+2\beta\varepsilon_n} \approx \frac{1}{2} + \varepsilon_n = \frac{1}{2} + a\sqrt{\frac{\log n}{n}}$. *The variance of* $\mathcal{D}_n = \mathrm{Beta}(\beta + 4\beta\varepsilon_n, \beta)$ *is of the order* $\frac{1}{8\beta} = \frac{1}{8\gamma k}$, *which is larger when* γ *is smaller. Since the variance is small when* k *is large,* $p \sim \mathcal{D}_n$ *is centered around* $\frac{1}{2}$ *and hence* $\mathbb{E}_{\mathcal{D}_n}[p(1-p)] \approx \frac{1}{4}$.

- *For* $\Gamma_n^{\boldsymbol{p}}(k) > 0$*: Let* a, α *be any constants such that* $a < \sqrt{E_{\mathcal{D}_n}[p(1-p)]} \approx \frac{1}{4}$, $\alpha > \frac{a}{2\sqrt{r \cdot \mathbb{E}_{\mathcal{D}_n}[p(1-p)]}} \approx \frac{a}{\sqrt{r}}$. *We claim that the CDF condition* $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \geq \frac{k}{n} + \Omega(\sqrt{\frac{\log n}{n}})$ *is satisfied when* $\gamma < \frac{1}{4\alpha^2}\left(\frac{1}{2r} - 1\right)$.
- *For* $\Gamma_n^{\boldsymbol{p}}(k) < 0$*: Let* a, α *be any constants such that* $\alpha < \frac{1}{2}$ *and* $a > \sqrt{r}\alpha$. *We claim that the CDF condition* $1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq \frac{1}{n^{1+\Omega(1)}}$ *is satisfied when* $\gamma > \frac{1}{4\alpha^2}\left(\frac{1+\Omega(1)}{r} + 1\right)$.

The rest of this section proves the above claims.

Proof for normal distributions. Since the random variable $p \sim \mathcal{N}(\mu_n = \frac{1}{2} + a\sqrt{\frac{\log n}{n}}, \sigma_n^2 = \frac{\sigma^2}{k})$ is below 0 or above 1 with exponentially small probability, we can approximate the PDF or CDF of $\mathcal{D}_n$ by the PDF and CDF of $\mathcal{N}(\mu_n = \frac{1}{2} + a\sqrt{\frac{\log n}{n}}, \sigma_n^2 = \frac{\sigma^2}{k})$, so

$$
\begin{aligned}
1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \approx \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(x-\mu_n)^2}{2\sigma_n^2}} \mathrm{d}x &= \int_{\frac{\alpha\sqrt{\frac{\log k}{k}} - \mu_n}{\sigma_n}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= \int_{\frac{\sqrt{k}(\alpha\sqrt{\frac{\log k}{k}} - a\sqrt{\frac{\log n}{n}})}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= \int_{\frac{\alpha}{\sigma}\sqrt{\log k}(1-o(1))}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t
\end{aligned}
$$

Using $\frac{1}{\sqrt{2\pi}} \frac{x}{x^2+1} e^{-\frac{x^2}{2}} \leq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}$ (Lemma 5), we get

$$
\frac{1}{\sqrt{2\pi}} \frac{\frac{\alpha}{\sigma}\sqrt{\log k}}{(\frac{\alpha}{\sigma}\sqrt{\log k})^2 + 1} e^{-\frac{(\frac{\alpha}{\sigma}\sqrt{\log k})^2}{2}} \leq 1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{\frac{\alpha}{\sigma}\sqrt{\log k}(1-o(1))} e^{-\frac{(\frac{\alpha}{\sigma}\sqrt{\log k}(1-o(1)))^2}{2}},
$$

or asymptotically

$$
\Omega\left(\frac{1}{\sqrt{\log k}} k^{-\frac{\alpha^2}{2\sigma^2}}\right) \leq 1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq O\left(\frac{1}{\sqrt{\log k}} k^{-\frac{\alpha^2}{2\sigma^2}(1-o(1))}\right).
$$

Plugging in $k = n^r$,

$$
\Omega\left(\frac{1}{\sqrt{\log n}} \frac{1}{n^{r\frac{\alpha^2}{2\sigma^2}}}\right) \leq 1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq O\left(\frac{1}{\sqrt{\log n}} \frac{1}{n^{r\frac{\alpha^2}{2\sigma^2}(1-o(1))}}\right).
$$

To satisfy the condition for $\Gamma_n^{\boldsymbol{p}}(k) > 0$, it suffices to require

$$
1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \geq \Omega\left(\frac{1}{\sqrt{\log n}} \frac{1}{n^{r\frac{\alpha^2}{2\sigma^2}}}\right) \geq \frac{k}{n} + \Omega\left(\sqrt{\frac{\log n}{n}}\right) = \frac{1}{n^{1-r}} + \Omega\left(\frac{\sqrt{\log n}}{n^{1/2}}\right),
$$

which is satisfied when

$$
r\frac{\alpha^2}{2\sigma^2} < \min\{1-r, 1/2\},
$$

i.e., $\sigma^2 > \frac{r\alpha^2}{2\min\{1-r,1/2\}}$.

For $\Gamma_n^{\boldsymbol{p}}(k) < 0$, it suffices to require

$$
1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \leq O\left(\frac{1}{\sqrt{\log n}} \frac{1}{n^{r\frac{\alpha^2}{2\sigma^2}(1-o(1))}}\right) \leq \frac{1}{n^{1+\Omega(1)}},
$$

which is satisfied when

$$
r\frac{\alpha^2}{2\sigma^2} > 1 + \Omega(1),
$$

i.e., $\sigma^2 < \frac{r\alpha^2}{2(1+\Omega(1))}$. □

We then prove the claims for beta distributions.

Proof for beta distributions. For $\mathcal{D}_n = \mathrm{Beta}(\beta + 4\beta\varepsilon_n, \beta)$, we have

$$
\begin{aligned}
1 - F_n\Big(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}\Big) &= \int_{\frac{1}{2}+\alpha\sqrt{\frac{\log k}{k}}}^{1} \frac{1}{\mathrm{B}(\beta + 4\beta\varepsilon_n, \beta)} x^{\beta+4\beta\varepsilon_n-1}(1-x)^{\beta-1}\mathrm{d}x \\
&= \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{1}{\mathrm{B}(\beta + 4\beta\varepsilon_n, \beta)} \left(\frac{1}{2}+t\right)^{\beta+4\beta\varepsilon_n-1} \left(\frac{1}{2}-t\right)^{\beta-1} \mathrm{d}t \qquad (13)
\end{aligned}
$$

where $\mathrm{B}(\beta + 4\beta\varepsilon_n, \beta) = \frac{\Gamma(\beta+4\beta\varepsilon_n)\Gamma(\beta)}{\Gamma(2\beta+4\beta\varepsilon_n)}$, and $\beta = \gamma k$. We note that since $r < \frac{1}{2}$, $4\beta\varepsilon_n = 4\gamma(n^r)a\frac{\sqrt{\log n}}{n^{1/2}} = o(1) < 1$ as n grows large.

The case of $\Gamma_n^{\boldsymbol{p}}(k) < 0$. We first consider the case of $\Gamma_n^{\boldsymbol{p}}(k) < 0$. We note that by monotonicity of $\Gamma(\cdot)$, assuming $\beta = \gamma k$ is an integer,

$$
\begin{aligned}
\mathrm{B}(\beta + 4\beta\varepsilon_n, \beta) = \frac{\Gamma(\beta+4\beta\varepsilon_n)\Gamma(\beta)}{\Gamma(2\beta+4\beta\varepsilon_n)} &\geq \frac{\Gamma(\beta)\Gamma(\beta)}{\Gamma(2\beta+1)} \\
&= \frac{(\beta-1)!(\beta-1)!}{(2\beta)!} \\
&= \frac{\beta!\beta!}{(2\beta)!\beta^2}.
\end{aligned}
$$

By Stirling's approximation, $\frac{n!n!}{(2n)!} \geq \frac{\sqrt{\pi n}}{4^n}$, hence

$$
\mathrm{B}(\beta + 4\beta\varepsilon_n, \beta) \geq \frac{\sqrt{\pi\beta}}{4^\beta\beta^2}.
$$

Plugging into Equation (13),

$$
\begin{aligned}
1 - F_n\Big(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}\Big) &\leq \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4^\beta\beta^2}{\sqrt{\pi\beta}} \left(\frac{1}{2}+t\right)^{\beta+4\beta\varepsilon_n-1} \left(\frac{1}{2}-t\right)^{\beta-1} \mathrm{d}t \\
\text{(because } \frac{1}{2}+t \leq 1\text{)} \quad &\leq \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4^\beta\beta^2}{\sqrt{\pi\beta}} \left(\frac{1}{2}+t\right)^{\beta-1} \left(\frac{1}{2}-t\right)^{\beta-1} \mathrm{d}t \\
&= \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4\beta^2}{\sqrt{\pi\beta}} (1+2t)^{\beta-1} (1-2t)^{\beta-1} \mathrm{d}t \\
&= \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4\beta^2}{\sqrt{\pi\beta}} \left(1-4t^2\right)^{\beta-1} \mathrm{d}t \\
\text{(using } 1-x \leq e^{-x}\text{)} \quad &\leq \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4\beta^2}{\sqrt{\pi\beta}} e^{-4t^2(\beta-1)}\mathrm{d}t \\
&\leq \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4e\beta^2}{\sqrt{\pi\beta}} e^{-4t^2\beta}\mathrm{d}t \\
\text{(let } u = \sqrt{8\beta}t\text{)} \quad &= \int_{\alpha\sqrt{8\gamma\log k}}^{\frac{1}{2}\sqrt{8\gamma k}} \frac{4e\beta}{\sqrt{8\pi}} e^{-\frac{u^2}{2}}\mathrm{d}u
\end{aligned}
$$

Using $\int_x^\infty e^{-\frac{u^2}{2}}\mathrm{d}u \le \frac{1}{x}e^{-\frac{x^2}{2}}$ (Lemma 5), we get

$$\begin{aligned}
1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \le \int_{\alpha\sqrt{8\gamma\log k}}^\infty \frac{4e\beta}{\sqrt{8\pi}} e^{-\frac{u^2}{2}}\mathrm{d}u &\le \frac{4e\beta}{\sqrt{8\pi}} \frac{1}{\alpha\sqrt{8\gamma\log k}} e^{-\frac{(\alpha\sqrt{8\gamma\log k})^2}{2}} \\
&= \frac{e\gamma k}{2\alpha\sqrt{\pi\gamma\log k}} k^{-4\alpha^2\gamma} \\
&= O\left(\frac{1}{\sqrt{\log k}} \frac{1}{k^{4\alpha^2\gamma-1}}\right) \\
&= O\left(\frac{1}{\sqrt{\log n}} \frac{1}{n^{r(4\alpha^2\gamma-1)}}\right).
\end{aligned}$$

To satisfy the CDF condition, it suffices to require

$$1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \le O\left(\frac{1}{\sqrt{\log n}} \frac{1}{n^{r(4\alpha^2\gamma-1)}}\right) \le \frac{1}{n^{1+\Omega(1)}},$$

which is satisfied when

$$r(4\alpha^2\gamma - 1) > 1 + \Omega(1),$$

i.e., $\gamma > \frac{1}{4\alpha^2}\left(\frac{1+\Omega(1)}{r} + 1\right)$.

The case of $\Gamma_n^{\boldsymbol{p}}(k) > 0$. Now we consider the case of $\Gamma_n^{\boldsymbol{p}}(k) > 0$. We note that by monotonicity of $\Gamma(\cdot)$, assuming $\beta = \gamma k$ is an integer,

$$\begin{aligned}
\mathrm{B}(\beta + 4\beta\varepsilon_n, \beta) = \frac{\Gamma(\beta + 4\beta\varepsilon_n)\Gamma(\beta)}{\Gamma(2\beta + 4\beta\varepsilon_n)} &\le \frac{\Gamma(\beta+1)\Gamma(\beta)}{\Gamma(2\beta)} \\
&= \frac{\beta!(\beta-1)!}{(2\beta-1)!} \\
&= \frac{\beta!\beta!}{(2\beta)!} \frac{2\beta}{\beta}.
\end{aligned}$$

By Stirling's approximation, $\frac{n!n!}{(2n)!} \le \frac{\sqrt{\pi n}}{4^n(1-1/8n)} \le \frac{3}{2}\frac{\sqrt{\pi n}}{4^n}$, hence

$$\mathrm{B}(\beta + 4\beta\varepsilon_n, \beta) \le \frac{3\sqrt{\pi\beta}}{4^\beta}.$$

Plugging into Equation (13),

$$\begin{aligned}
1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) &\ge \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4^\beta}{3\sqrt{\pi\beta}} \left(\frac{1}{2}+t\right)^{\beta+4\beta\varepsilon_n-1} \left(\frac{1}{2}-t\right)^{\beta-1} \mathrm{d}t \\
(4\beta\varepsilon \le 1) \quad &\ge \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{4^\beta}{3\sqrt{\pi\beta}} \left(\frac{1}{2}+t\right)^{\beta} \left(\frac{1}{2}-t\right)^{\beta} \mathrm{d}t \\
&= \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{1}{3\sqrt{\pi\beta}} (1+2t)^\beta (1-2t)^\beta \mathrm{d}t \\
&= \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{1}{3\sqrt{\pi\beta}} \left(1-4t^2\right)^\beta \mathrm{d}t \\
(\text{using } (1-\frac{x}{n})^n \ge e^{-x}(1-\frac{x^2}{n}) \text{ for } x \le n) \quad &\ge \int_{\alpha\sqrt{\frac{\log k}{k}}}^{\frac{1}{2}} \frac{1}{3\sqrt{\pi\beta}} e^{-4\beta t^2}(1-16\beta t^4)\mathrm{d}t \\
(\text{let } u = \sqrt{8\beta}t) \quad &= \int_{\alpha\sqrt{8\gamma\log k}}^{\frac{1}{2}\sqrt{8\gamma k}} \frac{1}{3\sqrt{8\pi}\beta} e^{-\frac{u^2}{2}}(1-\frac{u^4}{4\beta})\mathrm{d}u \\
(1-\frac{u^4}{4\beta} \ge \frac{3}{4} \text{ for } u \le \beta^{1/4}) \quad &\ge \int_{\alpha\sqrt{8\gamma\log k}}^{(\gamma k)^{1/4}} \frac{1}{3\sqrt{8\pi}\beta} e^{-\frac{u^2}{2}} \frac{3}{4}\mathrm{d}u \\
&= \frac{1}{4\sqrt{8\pi}\beta} \int_{\alpha\sqrt{8\gamma\log k}}^{(\gamma k)^{1/4}} e^{-\frac{u^2}{2}}\mathrm{d}u
\end{aligned}$$

Using $\int_x^y e^{-\frac{u^2}{2}}\,\mathrm{d}u \geq (-\frac{u}{u^2+1})e^{-\frac{u^2}{2}}\Big|_x^y$ (see the proof of Lemma 5), we get

$$\begin{aligned}
1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) &\geq \frac{1}{4\sqrt{8\pi}\beta}\int_{\alpha\sqrt{8\gamma\log k}}^{(\gamma k)^{1/4}} e^{-\frac{u^2}{2}}\,\mathrm{d}u \\
&\geq \frac{1}{4\sqrt{8\pi}\beta}\left(\frac{\alpha\sqrt{8\gamma\log k}}{\alpha^2 8\gamma\log k+1}e^{-\frac{\alpha^2 8\gamma\log k}{2}} - \frac{(\gamma k)^{1/4}}{\sqrt{\gamma k}+1}e^{-\frac{\sqrt{\gamma k}}{2}}\right) \\
&= \frac{1}{4\sqrt{8\pi}\gamma k}\left(\frac{\alpha\sqrt{8\gamma\log k}}{\alpha^2 8\gamma\log k+1}k^{-4\alpha^2\gamma} - o\left(e^{-\frac{\sqrt{\gamma k}}{2}}\right)\right) \\
&= \Omega\left(\frac{1}{\sqrt{\log k}}\frac{1}{k^{4\alpha^2\gamma+1}}\right) \\
&= \Omega\left(\frac{1}{\sqrt{\log n}}\frac{1}{n^{r(4\alpha^2\gamma+1)}}\right)
\end{aligned}$$

To satisfy the CDF condition, it suffices to require

$$1 - F_n(\frac{1}{2} + \alpha\sqrt{\frac{\log k}{k}}) \geq \Omega\left(\frac{1}{\sqrt{\log n}}\frac{1}{n^{r(4\alpha^2\gamma+1)}}\right) \geq \frac{k}{n} + \Omega\left(\sqrt{\frac{\log n}{n}}\right) = \frac{1}{n^{1-r}} + \Omega\left(\frac{\sqrt{\log n}}{n^{1/2}}\right),$$

which is satisfied when

$$r(4\alpha^2\gamma + 1) < \min\{1 - r, 1/2\} = 1/2$$

i.e., $\gamma < \frac{1}{4\alpha^2}\left(\frac{1}{2r} - 1\right)$. $\square$