Multiwinner Voting with Interval Preferences under Incomplete Information

Drew Springham[1] **Edith Elkind**[2] **Bart de Keijzer**[1] **Maria Polukarov**[1]

[1]King's College London, London, UK
[2]Northwestern University, Evanston, IL, USA

Abstract

In multiwinner approval elections with many candidates, voters may struggle to determine their preferences over the entire slate of candidates. It is therefore of interest to explore which (if any) fairness guarantees can be provided under reduced communication. In this paper, we consider voters with one-dimensional preferences: voters and candidates are associated with points in $\mathbb{R}$, and each voter's approval set forms an interval of $\mathbb{R}$. We put forward a probabilistic preference model, where the voter set consists of different groups; each group is associated with a distribution over an interval of $\mathbb{R}$, so that each voter draws the endpoints of her approval interval from the distribution associated with her group. We present an algorithm for computing committees that provide *Proportional Justified Representation + (PJR+)*, which queries voters' preferences, and show that, in expectation, it makes $\mathcal{O}(\log k)$ queries per voter, where k is the target committee size.

1 INTRODUCTION

The problem of selecting a set of winning candidates from a wider pool, given the preferences of many agents, arises in a wide variety of settings. This problem is formalized as *Multiwinner Voting (MWV)*. Real-world examples include, but are not limited to, committee selection, parliamentary elections, group recommendations, and opinion summarization [Lackner and Skowron, 2023, Halpern et al., 2024, Revel et al., 2025]. Often, we wish to select a committee of size k in a way that is "fair"; this is captured by the concept of *Justified Representation (JR)* and its extensions, such as PJR, PJR+, EJR, EJR+, FPJR, and FJR [Sánchez-Fernández et al., 2026, Brill and Peters, 2023, Aziz et al., 2017, Peters and Skowron, 2020, Kalayci et al., 2025]. These extensions capture the idea that groups making up at least an ℓ/k fraction of the voter population should be represented by at least ℓ of the k candidates in the winning set. E.g., for the winning set W to provide PJR+—the axiom we focus on in this work—for each group of voters S that forms an ℓ/k fraction of the population and approves of a candidate not in W, the set W must contain ℓ candidates approved by some voter in S.

Importantly, some multiwinner elections have a large number of candidates in consideration. E.g., in the well-known *participatory budgeting* setting, in which residents of a city can propose projects to be completed, and then vote on the proposals, there may be dozens of projects to choose from: indeed, there are real-life examples of participatory budgeting elections with more than 150 candidate projects [Faliszewski et al., 2023]. Another example is debates on Pol.is [Small et al., 2021], an online platform for civic participation. Here, citizens vote on comments (candidates) that are made by others, and we wish to select a representative slate of opinions. In some debates, e.g., during an online citizens' assembly meeting on climate law in Austria, over 1000 comments were made [The Computational Democracy Project, 2022]. In such circumstances, one cannot realistically expect the voters to accurately evaluate all candidates. Therefore, it is of interest to consider multiwinner voting mechanisms that operate when some preference information is unknown or uncertain, but still provide guarantees on the social desirability of the outcome.

In this paper, we introduce the general problem of multiwinner voting with incomplete information in a spatial preferences setting, where voters and candidates are associated with points in $\mathbb{R}^d$, and each voter approves candidates within a certain distance from her. We focus on the one-dimensional case of this general problem, where each voter's approval set forms an interval of $\mathbb{R}$ (her *approval interval*). Such a model may be appropriate, e.g., in a political setting, where candidates are positioned on the left-right political spectrum; then, each voter only disapproves of candidates they find too extreme (to the left or to the right).

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:6512–6536.

We put forward a rich probabilistic model to represent the voters' preferences in this setting: the *Random Interval Voter model (RIV)*. In RIV (Definition 6), we assume that voters are divided into groups, so that each group is associated with a subinterval of $\mathbb{R}$ and a voter's approval interval is obtained by independently sampling two points from a given distribution over the interval associated with their group. In particular, the intervals associated with different groups may overlap, so a candidate can be approved by voters from several groups whenever their intervals intersect.

When the voter population is large, a probabilistic model can be thought of as providing aggregated preference information, based, e.g., on historical data. In particular, an attractive feature of RIV is that it can express both large-scale trends and small-scale behavior. Indeed, by associating each voter with a subinterval of $\mathbb{R}$, we can capture settings where voters are partitioned into well-defined groups (parties), so that voters' preferences are consistent with this partition; in this setting, each party naturally corresponds to a subinterval of $\mathbb{R}$. Thus, the RIV model can be thought of as a generalized probabilistic variant of open party-list elections, where candidates may appear within *multiple parties' lists*, so long as their appearance is consistent with the overall party ordering. By sampling voters' endpoints from a party-specific cumulative distribution function, we capture voters' within-party preferences. Together, both factors give RIV sufficient depth to be a nuanced model of how agents might vote in practice.

Our contributions. We develop a preference elicitation framework that is tailored to the RIV model. Our framework allows for *interval queries*, where a voter is asked if her set of approved candidates is contained within a given interval. We show that, in this model, we can elicit the preferences of each voter using $\mathcal{O}(\log m)$ queries per voter in expectation, where m is the number of candidates; this enables us to construct an outcome in the core [Aziz et al., 2017, Pierczyński and Skowron, 2022]. Then, our main technical contribution is an algorithm that outputs committees providing a notion of proportionality that is slightly weaker than core stability, namely, PJR+ [Brill and Peters, 2023], and only requires $\mathcal{O}(\log k)$ queries per voter in expectation, where k is the target committee size. Notably, unlike for the core, our bound for PJR+ does not depend on the total number of candidates. Moreover, while the bound on the running time holds in expectation (and also with high probability), the PJR+ guarantee holds with probability 1.

Related work. There is substantial literature on approval-based multiwinner voting (see [Lackner and Skowron, 2023] for an overview). The preference elicitation aspect of voting has also been studied extensively, especially in the single-winner context (e.g. [Conitzer and Sandholm, 2002, Meir et al., 2014, Dey and Bhattacharyya, 2015, Dery et al., 2016, Zhao et al., 2018]).

A number of papers on preference elicitation in multiwinner voting optimize objectives other than proportionality, e.g., minimax regret [Lu and Boutilier, 2013], diversity [Lindeboom et al., 2025], or social welfare [Mandal et al., 2020].

Imber et al. [2022] consider winner determination with incomplete preferences, under several voting rules that guarantee forms of justified representation. However, their model does not allow for preference elicitation: In their setting, each voter reports a subset of her approved candidates and a subset of her disapproved candidates, with the remaining candidates being unknown, and the goal is to decide which candidates are possibly or necessarily in the winning set. Subsequently, Imber et al. [2024] considered the same problem under the assumption that voters and candidate lie in a d-dimensional Euclidean space. There are two key differences between this line of work and our contribution: (1) we allow the algorithm to query the voters, and (2) we take a probabilistic rather than a nondeterministic perspective.

The closest in spirit to our approach is the work of Halpern et al. [2023], who also consider querying voters in order to achieve a form of justified representation. They formalize an idea of approximate justified representation, by scaling up the size of a voter group that "deserves" representation by a factor of $(1 + \varepsilon)$, and provide adaptive algorithms to achieve approximate EJR (a notion that is similar to PJR+, but incomparable with it). They also provide a lower bound on the number of voters that need to be sampled to achieve EJR with probability of at least $1-\delta$. Brill and Peters [2023] discuss a similar technique for achieving EJR+ with high probability in the general preference setting. Our work differs from Halpern et al. [2023] and Brill and Peters [2023] in two main ways: (1) while their model allows for unrestricted approval preferences, we assume that voters and candidates are located in a metric space; (2) our model is Bayesian, i.e., our algorithm makes use of distributional information and our bounds on the number of queries hold with high probability and in expectation. In particular, our query model allows for queries that ask a voter if her approval set is fully contained within a given interval; such queries are very natural in the metric setting, but their non-metric analogue (asking a voter whether all her approved candidates belong to some subset of candidates) has a higher cognitive and communication cost. The metric and distributional assumptions, together with the richer query model, enable us to obtain stronger results: while the number of queries in the work of Halpern et al. [2023] scales linearly with the number of candidates m, our algorithm achieves PJR+ using a number of queries that, in expectation, does not depend on m. Moreover, we achieve proportionality with probability 1, whereas Halpern et al. [2023] and Brill and Peters [2023] aim for EJR+ with probability $1 - \varepsilon$.

Since this paper was accepted, we have extended these results to a more general setting [Springham et al., 2026], encompassing general joint distributions over voters' approval regions in d dimensions and attaining exact EJR+; see Appendix G for further discussion.

Outline. In Section 2 we introduce the preliminary definitions and results. In Section 3, we describe our probabilistic model, and formulate the general problem of multiwinner voting with incomplete information in a spatial setting. Section 3.2 discusses our preference elicitation framework. We also formulate accompanying lemmas, which we put to use in Section 4, where we prove that we can construct a committee that provides PJR+ using limited queries per voter in expectation. Section 5 concludes.

2 PRELIMINARIES

All our indexing of ordered sets (or lists) starts at 0. We write $X[a : b] := \{X[i] : a \leq i < b\}$. We also use the notation $[t] := \{i \in \mathbb{N} : 1 \leq i \leq t\}$ and $\mathbb{1}[P]$ as the indicator function that event P has occurred.

Definition 1. *(Approval-based MWV election) An* MWV election *is a tuple $E = (V, C, k, A)$, where V is a set of n voters, C is a set of $m \geq 2$ candidates, $k \in \mathbb{N}$ is a total number of candidates to elect, and $A : V \to 2^C$ is a function that maps each voter to the set of candidates she approves. For each $G \subseteq V$, we write $A(G) = \bigcap_{v \in G} A(v)$. The primary task associated with an MWV election is to select a set, or committee, of* winners *$W \subseteq C$ with $|W| = k$.*

The literature on multiwinner voting with approval preferences defines a number of fairness axioms, ranging from Justified Representation (JR) to core stability [Lackner and Skowron, 2023]. All these axioms aim to capture the idea that large groups of voters with similar preferences should be represented in the committee in proportion to the group size; they differ in how they define which groups of voters are considered to have similar preferences and what it means to represent a group. Below, we define three axioms from this hierarchy, namely, PJR+, EJR+ [Brill and Peters, 2023], and core stability [Aziz et al., 2017]. We will show that our method always selects PJR+ committees while asking a small number of queries in expectation and with high probability; obtaining a similar result for more demanding axioms such as EJR+ remains a direction for future work, discussed in Section 5.

Definition 2. *(α-PJR+) A committee W satisfies α-PJR+ [Brill and Peters, 2023][1] for $\alpha \geq 1$ if $|W| = k$ and for every $\ell \in [k]$ and every group $G \subseteq V$ that satisfies $|G| \geq \alpha \cdot n\ell/k$, $A(G) \setminus W \neq \varnothing$ it holds that $|W \cap \bigcup_{v \in G} A(v)| \geq \ell$. If W satisfies 1-PJR+, we say that it satisfies PJR+.*

Definition 3. *(α-EJR+) A committee W satisfies α-EJR+ [Brill and Peters, 2023] for $\alpha \geq 1$ if $|W| = k$ and for every $\ell \in [k]$ and every group $G \subseteq V$ that satisfies $|G| \geq \alpha \cdot n\ell/k$, $A(G) \setminus W \neq \varnothing$ it holds that some voter $v \in G$ has $|W \cap A(v)| \geq \ell$. If W satisfies 1-EJR+, we say that it satisfies EJR+.*

Remark 1. *A committee W satisfies PJR+ if there is no group of voters who (i) deserve ℓ representatives, (ii) can all agree on a candidate not in W, but (iii) collectively approve fewer than ℓ members of W. For $\alpha > 1$, α-PJR+ (resp. α-EJR+) is a relaxation of PJR+ (resp. EJR+): if W satisfies PJR+, then it satisfies α-PJR+ for $\alpha > 1$. For $\alpha \geq 1$, α-PJR+ is also a relaxation of α-EJR+.*

Definition 4. *(The core) A committee W satisfies* core stability *[Aziz et al., 2017, Lackner and Skowron, 2023], or is said to be* in the core, *if $|W| = k$ and for every $\ell \in [k]$, every $T \subseteq C$ with $|T| \leq \ell$, and every group $G \subseteq V$ with $|G| \geq n\ell/k$ there is a voter $v \in G$ with $|W \cap A(v)| \geq |T \cap A(v)|$.*

Core stability is more demanding than PJR+: if a committee fails PJR+ then it is not in the core, but the converse is not true[2]. It is a prominent open problem[3] whether every approval-based MWV election has a core stable committee [Aziz et al., 2017, Lackner and Skowron, 2023].

Every MWV election has a committee that satisfies EJR+. In particular, the Method of Equal Shares (MES) [Peters and Skowron, 2020] runs in polynomial time and outputs committees that satisfy EJR+, and therefore PJR+. In the algorithms we propose later, we will use MES as a subroutine, serving as a "fallback" method to find a PJR+ committee (Algorithm 2).

We will consider one-dimensional approval preferences.

Definition 5. *(Candidate-Interval (CI) preferences) Given a set of candidates C and a linear order $\lhd$ over C, we say that a subset of candidates $T \subseteq C$ is* consecutive *if for every $x, z \in T$ and $y \in C$ with $x \lhd y \lhd z$ it holds that $y \in T$. An election E has* Candidate Interval (CI) preferences *if there exists a linear order $\lhd$ over C where the set $A(v)$ is consecutive for every voter v [Elkind and Lackner, 2015].*

If $C \subseteq \mathbb{R}$, each voter approves of an interval of candidates: $T \subseteq C$ is consecutive if $T = C \cap I$ for some interval $I = [a, b]$. CI preferences are single-peaked preferences in the dichotomous setting; this class of preferences has been explored in a number of papers, e.g., [Peters and Lackner, 2020, Godziszewski et al., 2021, Terzopoulou et al., 2021, Pierczyński and Skowron, 2022].

We say an event occurs *with high probability (w.h.p.)* when its failure probability tends to 0 as $m \to \infty$ under the

[1] PJR+ is also known as IPSC [Aziz and Lee, 2021].

[2] This is claimed by Kalayci et al. [2025] without proof; for completeness, we provide a proof in Appendix A.

[3] However, the core can be empty in participatory budgeting elections with approval preferences [Maly, 2025].

sample-size hypothesis of the relevant theorem; the precise bound is stated in each result.

We use the following concentration inequality in our proofs.

Fact 2 (Bernstein inequality [Boucheron et al., 2013]). *Let $X_1, \dots, X_n$ be independent random variables with $|X_i| \le 1$ and $\mathrm{Var}(X_i) \le \sigma^2$. Then for any $t > 0$, $\Pr\left[\sum_{i=1}^n X_i \ge t\right] \le \exp\left(-\frac{t^2/2}{n\sigma^2 + t/3}\right)$.*

3 RANDOM INTERVAL VOTER MODEL (RIV)

In this section, we introduce the probabilistic model for CI preferences that will be used throughout this paper; we briefly discuss an extension of this model to more than one dimension in Section 5.

Definition 6 (RIV). *An* RIV model *is a σ-length list of pairs $\mathcal{M} = (F_t, p_t)_{t\in[\sigma]}$, where $\sigma \in \mathbb{N}$, $(p_t)_{t\in[\sigma]}$ is a probability vector, and for each $t \in [\sigma]$ it holds that F_t is an invertible cumulative distribution function*[4] *on $[0,1]$. Given a candidate set $C \subset [0,1]$, we say that a voter v is* sampled according to $\mathcal{M}$ *if her ballot is obtained as follows: A single distribution index t is drawn, where t is chosen with probability p_t; two positions $Z_1, Z_2 \in [0,1]$ are drawn independently from distribution F_t; we set $a_v = \min(Z_1, Z_2)$, $b_v = \max(Z_1, Z_2)$, and the voter's approval ballot is defined as $A^*(v) = [a_v, b_v]$, $A(v) = A^*(v) \cap C$.*

The independence of Z_1 and Z_2 in the RIV model is a substantive assumption: when relaxed to allow the two endpoints to be drawn jointly (the Generalized Interval Voter model, Appendix E), the proportionality guarantee degrades from PJR+ to 2-PJR+ (Theorem 13).

We now define some notation that will be used throughout, and state a useful lemma.

Definition 7. *We abuse notation and say that $v \in F_t$ if distribution t is chosen for v. We have that $\Pr[Z_1 \le x \mid v \in F_t] = F_t(x)$, and so we define the combined cumulative distribution function $G(x) \coloneqq \Pr[Z_1 \le x] = \sum_{t\in[\sigma]} p_t F_t(x)$. We consider $\mathbf{p} = (p_t)_{t\in[\sigma]}, \mathbf{F}(c) = (F_t(c))_{t\in[\sigma]}$ as vectors, so $G(c) = \mathbf{p}\cdot\mathbf{F}(c)$. For each $T \subseteq C$ and $x \in [0,1]$, let $x_T^- = \max(\{c \in T : c \le x\}), x_T^+ = \min(\{c \in T : c \ge x\})$. where $\min\varnothing = 1$ and $\max\varnothing = 0$.*

x_T^- (resp., x_T^+) is the closest candidate in T to the left (resp., to the right) of x, or 0 (resp., 1) if no such candidate exists. The following lemma gives a useful expression for the probability of approving and disapproving candidates.

[4]This means F_t is continuous: it assigns no point mass for any $x \in [0,1]$, i.e., $\Pr(X = x) = 0$ for all $x \in [0,1]$ where $X \sim F_t$.

Lemma 3. *For an RIV instance, a subset of candidates $S \subseteq C$, and a candidate $c \in C \setminus S$ we have*

$$\Pr[c \in A(v) \wedge A(v) \cap S = \varnothing] = \sum_{t\in[\sigma]} 2p_t \left(F_t(c) - F_t(c_S^-)\right)\left(F_t(c_S^+) - F_t(c)\right).$$

Proof. Let X be the event that $c \in A(v) \wedge A(v) \cap S = \varnothing$. For $c \in A(v)$, we need $a_v \le c \le b_v$. Then if $c \in A(v)$, in order for $A(v) \cap S = \varnothing$, we also need $c_S^- \le a_v, b_v \le c_S^+$. Note that there are two ways in which this can occur: with $Z_1 = a_v, Z_2 = b_v$, and $Z_2 = a_v, Z_1 = b_v$. So, we have

$$\begin{aligned}\Pr[X] &= \sum_{t\in[\sigma]} \Pr[X \mid v \in F_t]\Pr[v \in F_t] \\ &= \sum_{t\in[\sigma]} p_t \Pr[c_S^- \le a_v \le c \le b_v \le c_S^+ \mid v \in F_t] \\ &= \sum_{t\in[\sigma]} 2p_t \left(F_t(c) - F_t(c_S^-)\right)\left(F_t(c_S^+) - F_t(c)\right),\end{aligned}$$

our desired result. □

3.1 QUERIES

We use a type of query called an *interval query* $\textsc{Int}(x, y, v)$. We communicate two positions $x, y \in C \cup \{0,1\}$ to v, who responds with $\mathbb{1}[A^*(v) \subseteq [x,y]]$. Note that the response is a single bit, and so the number of bits communicated by the voter is the same as the number of queries made to the voter. We refer to a sequence of queries, together with voters' responses, as a *dialogue*; e.g., a dialogue can take the form $\mathcal{I} = (A^*(v) \subseteq [c_1, c_2], A^*(v') \not\subseteq [c_3, c_4])$. INTERVAL queries allow us to effectively ask a voter "Would you only approve of some compromise between x and y?"

While one could alternatively use *point queries* (asking whether a voter approves a specific position $x \in [0,1]$), interval queries offer significant advantages. First, interval queries achieve $\mathcal{O}(\log|P|)$ queries in the *worst case* via the binary search in Algorithm 1, whereas point queries only guarantee this in expectation. Second, using only point queries requires asking voters about arbitrary positions throughout $[0,1]$, not just at candidate positions, which imposes cognitive demands on voters. In contrast, interval queries ask naturally about ranges of actual candidates, which we believe aligns with how voters would naturally reason about preferences in spatial settings.

We note that under CI preferences each voter's ballot is fully determined by her leftmost and rightmost approved candidates; thus, there are at most m^2 distinct ballots. Consequently, if we consider a richer elicitation framework in which each voter can directly communicate which of the possible m^2 ballots she holds, we could elicit full information regarding the voters' preferences with $\log\left(m^2\right) =$

Algorithm 1: Algorithm for resolving a voter v's preferences over a set of candidates P

Function `resolve`(v, P)**:**
  $((a_v)^-_P, (a_v)^+_P) \leftarrow$`search`$(v, P \cup \{0,1\}, \textit{left})$;
  $((b_v)^-_P, (b_v)^+_P) \leftarrow$`search`$(v, P \cup \{0,1\}, \textit{right})$;
  return $P \cap [(a_v)^+_P, (b_v)^-_P]$
Function `search`(v, P, dir)**:**
  $(\check{x}, \hat{x}) = (0, 1)$;
  while $P \neq \varnothing$ **do**
    $p \leftarrow$ median of P;
    if $\mathrm{dir} = \textit{left}$ **:**
      $goRight \leftarrow$ response to $\textsc{Int}(p, 1, v)$;
    else
      $goRight \leftarrow \neg$(response to $\textsc{Int}(0, p, v)$);
    if $goRight$ **:**
      $\check{x} \leftarrow p$; $P \leftarrow \{y \in P : y > p\}$;
    else
      $\hat{x} \leftarrow p$; $P \leftarrow \{y \in P : y < p\}$;
  return $(\check{x}, \hat{x})$

$\mathcal{O}(\log m)$ communication complexity per voter. In what follows, we show that we can elicit full preferences with $\mathcal{O}(\log m)$ communication complexity per voter in our model as well (Section 3.2); moreover, it turns out that, with a much smaller number of queries (which, crucially, is independent of m), we can usually find a PJR+ outcome (Section 4).

3.2 ELICITING PREFERENCES

We shall now show that we can elicit a voter's full preferences using $\mathcal{O}(\log m)$ queries. We formalize the problem of eliciting a voter's preferences regarding a candidate subset $P \subseteq C$ as follows.

Definition 8. *A voter v is* resolved for *$P \subseteq C$ given a dialogue $\mathcal{I}$ if $\Pr[x \in A(v) \mid \mathcal{I}] \in \{0, 1\}$ for all $x \in P$. We drop $\mathcal{I}$ from the phrase when it is clear from the context.*

Our algorithm for resolving a given voter v (Algorithm 1) uses $\textsc{Int}$ queries to perform two binary searches over set P. These searches determine the consecutive elements $(a_v)^-_P, (a_v)^+_P \in P$ and $(b_v)^-_P, (b_v)^+_P \in P$ with $a_v \in ((a_v)^-_P, (a_v)^+_P)$ and $b_v \in ((b_v)^-_P, (b_v)^+_P)$. From there, we know that $A(v) \cap P = [(a_v)^+_P, (b_v)^-_P] \cap P$. For each endpoint a_v, b_v, we are using $\textsc{Int}$ queries as a comparison function in a binary search, so we need to use only $\mathcal{O}(\log |P|)$ $\textsc{Int}$ queries in total to resolve v on P.

By executing Algorithm 1 with $P = C$, we can learn all voters' complete preferences using $\mathcal{O}(\log m)$ queries per voter. This enables us to find proportional committees. In particular, since for CI preferences the core is always nonempty, and a committee in the core can be computed in polynomial time given the voters' preferences [Pierczyński and Skowron, 2022], we obtain the following corollary.

Corollary 4. *Let E be an election with candidate set C, $|C| = m$, obtained by sampling the voters' preferences from an RIV. One can compute a committee in the core of E using $\mathcal{O}(\log m)$ queries per voter.*

4 FINDING PJR+ COMMITTEES FOR RIV

We have seen that if voters' preferences are sampled from an RIV model, one can find a core stable outcome using $\mathcal{O}(\log m)$ queries per voter. We will now present our main result: for a slightly less ambitious notion of proportionality, namely, PJR+, one can find proportional outcomes using $\mathcal{O}(\log k)$ queries per voter in expectation; notably, the expected number of queries does not depend on m. In particular, this means that our algorithm for PJR+ does not elicit voters' full preferences; in that sense, it follows the line of work on learning solution concepts in games [Balcan et al., 2015, Jha and Zick, 2023].

Theorem 5. *For an RIV, Algorithm 2 finds a PJR+ committee. When the number of voters n satisfies $n = \Omega\left(k^3 \log\left(k \log m\right)\right)$,[5] the algorithm uses $\mathcal{O}(\log k)$ queries per voter in expectation and with high probability.*

The proof of Theorem 5 is based on a sequence of algorithms (Algorithm 2). Briefly, our algorithm first "guesses" a committee $\widehat{W}$; we will argue that $\widehat{W}$ provides PJR+ with high probability. It then elicits preference information from each voter to obtain an approximation to the voter's true approval window $[a_v, b_v]$. Given this information, the algorithm then checks whether $\widehat{W}$ necessarily provides PJR+. If the check fails (which happens with low probability), the algorithm elicits voters' full preferences, and computes a PJR+ committee using MES.

The stage that computes $\widehat{W}$ proceeds as follows. For each $i \in [k]$, we find $w_i = \arg\min_{c \in C \setminus \widehat{W}} |G(c) - \frac{i}{k+1}|$ and we add w_i to $\widehat{W}$. Recall that $G(c) = \mathbf{p} \cdot \mathbf{F}(c)$, the combined cumulative distribution function. We shall refer to the points $\{\frac{i}{k+1} : i \in [k]\}$ as *marked points*. So, we select the candidates that are closest (with respect to G) to these equally spread marked points. Intuitively, in this way we ensure that the selected candidates are approximately uniformly distributed with respect to G.

Example 6. *We provide an example of how we construct $\widehat{W}$. Let $\sigma = 1$ (an extension to multiple distributions is*

[5]When $\sigma = 1$ (all voters share a single CDF), this bound improves to $n = \Omega(k^2 \log(k \log m))$: using AM-GM gives $\Pr[Y] \leq 2\ell^2/(k+1)^2$, and taking $\Delta = (4k)^{-1}$ yields $\Pr[X_v] \leq \ell/k - \Omega(1/k)$, so Bernstein gives $\Pr[s_{\ell,c,j} \geq \ell n/k] \leq \exp(-\Omega(n/k^2))$.

Algorithm 2: Finding a PJR+ committee

Function `RIVPJR+` *(An RIV $\mathcal{M}$, candidates C, k)***:**
- $\Delta = (4k)^{-2}$; $G = \mathbf{p} \cdot \mathbf{F}$; $\widehat{W} = \varnothing$;
- **for** $i \in [k]$ **:**
  - $c \leftarrow \arg\min_{c \in C \setminus \widehat{W}} \left|G(c) - \frac{i}{k+1}\right|$;
  - $\widehat{W} \leftarrow \widehat{W} \cup \{c\}$;
- Sort $\widehat{W}$;
- $P \leftarrow \widehat{W} \cup \left\{ \left(G^{-1}(r\Delta)\right)^{+}_{C} : r \in [1/\Delta] \right\}$;
- $P \leftarrow P \cup \left\{ \left(G^{-1}(r\Delta)\right)^{-}_{C} : r \in [1/\Delta] \right\}$;
- **for** $v \in V$ **:** $\widehat{A}(v) \leftarrow$ `resolve`(v, P) ;
- $\widehat{E} \leftarrow (V, P, k, \widehat{A})$;
- **if** `validate` $(\widehat{W}, \widehat{E}, C)$ **: return** $\widehat{W}$;
- **for** $v \in V$ **:** $A(v) \leftarrow$ `resolve`(v, C) ;
- **return** $W^* = \mathit{MES}(V, C, k, A)$;

Function `validate` $(\widehat{W}, \widehat{E}, C)$ **:**
- **for** $\ell \in [k]$, $c \in C \setminus \widehat{W}$ **:**
  - $\rho \leftarrow |\{x \in \widehat{W} : x < c\}|$;
  - **for** $j \in [\ell - 1] \cup \{0\}$ **:**
    - $R \leftarrow \widehat{W}[\rho - j : \rho + \ell - 1 - j]$;
    - $s_{\ell,c,j} \leftarrow |\{v \in V : \mathtt{poss}(v, c, \widehat{W} \setminus R)\}|$;
    - **if** $s_{\ell,c,j} \geq \frac{n\ell}{k}$ **: return** *False* ;
- **return** *True*;

Function `poss` (v, c, S) **:**
- $H_1, H_2 \leftarrow ((a_v)^{-}_{P}, (a_v)^{+}_{P}], [(b_v)^{-}_{P}, (b_v)^{+}_{P})$ (found previously using `resolve`);
- `/* Do there exist positions for` a_v, b_v `that are compatible with query info AND approving` c `but none of` S`? */`
- **return** $H_1 \cap (c^{-}_{S}, c] \neq \varnothing \wedge H_2 \cap [c, c^{+}_{S}) \neq \varnothing$;

straightforward), assume $G = F_1$ is the uniform distribution here, and consider candidate positions in Figure 1 with $k = 8$. We iterate over marked points from left to right[6], selecting the candidate that is closest (with respect to G) to the marked point. Among all candidates, c_1 has the minimum distance to marked point 1, so we select it. We then pick c_2 for marked point 2 and c_3 for marked point 3. For marked point 4, note that c_3 was already selected, so the closest unselected candidate is c_4, which is further from marked point 4 than c_3. Eventually, we obtain $\widehat{W} = \{c_1, \ldots, c_8\}$, $|\widehat{W}| = k = 8$. We see that the k candidates are spread across the interval as uniformly (w.r.t. G) as possible.

Given our guessed committee $\widehat{W}$, we verify whether it provides PJR+. With $\Delta = (4k)^{-2}$, we pick the two candidates closest to the left and right of $G^{-1}(j\Delta)$ for each

[6]Iteration over marked points in this step can be performed in arbitrary order, but we choose left to right for simplicity.

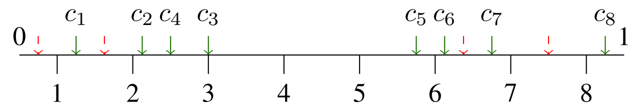

Figure 1: The algorithm selects candidates c for the committee $\widehat{W}$, in order $c_1, \ldots, c_8$. Unelected candidates are indicated by dashed red arrows.

$j \in [1/\Delta] = [16k^2]$, along with $\widehat{W}$, and use Algorithm 1 to resolve each voter on these candidates. This provides an approximation for the voter's approval interval "up to P", in that we discover the candidates in P that are closest either side of a_v and b_v. In particular, we discover with certainty which of the $\widehat{W}$ candidates each voter approves.

For every $\ell \in [k]$, every candidate $c \in C \setminus \widehat{W}$ and for every $0 \leq j < \ell$, we count the number of voters for which it could be possible (given the information we have elicited from the voter) that they approve of: (1) c, (2) at most j candidates in $\widehat{W}$ to the left of c, and (3) fewer than ℓ candidates in $\widehat{W}$. If there are at least $n\ell/k$ such voters for a given choice of ℓ, c, and j, then $\widehat{W}$ possibly violates PJR+ (given what we know about those voters' preferences), as this set of voters possibly satisfies conditions (i-iii) stated in Remark 1. This count is performed by the algorithm through the calls to `poss`. Each voter for which the above three properties hold, contributes to $s_{\ell,c,j}$, and if $s_{\ell,c,j} < n\ell/k$ for all $\ell \in [k]$, $c \in C \setminus \widehat{W}$, $0 \leq j < \ell$, we output $\widehat{W}$. Otherwise, we query all voters in V about all candidates in C to get full information about E, and then run MES to output a committee W^* that provides PJR+.

In Appendix C we show that if $s_{\ell,c,j} < n\ell/k$ for all $\ell \in [k]$, $c \in C \setminus \widehat{W}$, $0 \leq j < \ell$, then $\widehat{W}$ is guaranteed to provide PJR+. The main insight we use for proving this is that if a group of voters violates PJR+, then the (at most ℓ) members of $\widehat{W}$ they support are consecutive, which corresponds to the sets that are assigned to R in Algorithm 2.

If on the other hand $s_{\ell,c,j} \geq n\ell/k$ for some ℓ, c, j, then we use $\mathcal{O}(\log m)$ queries per voter in order to obtain full information about $A(v)$ for each $v \in V$. However, we show that the latter event is rare, so with high probability and in expectation we require $\mathcal{O}(\log k)$ queries. The time complexity of Algorithm 2 (without MES, which runs in $\mathcal{O}(m^2 n \log n)$ [Peters et al., 2021]) is $\mathcal{O}(nmk^2)$.

To prove Theorem 5, we require additional notation; we first discuss it informally, and then provide rigorous definitions. The quantities $\widehat{W}, s_{\ell,c,j}$ are defined in Algorithm 2 in the `RIVPJR+` and `validate` functions (in what follows, we consider the values of these variables at the termination of the respective algorithms). We will consider a fixed $0 \leq j < \ell \leq k$ and $c \in C \setminus \widehat{W}$ for convenience.

Assume that $\widehat{W}$ is ordered from left to right. The function ρ counts the candidates in $\widehat{W}$ to the left of c, so that $\widehat{W}[\rho - r]$ is the r-th candidate in $\widehat{W}$ to the left of c. If $r > \rho$, we say that $\widehat{W}[\rho - r] = 0$. Y is the event that an arbitrary voter v drawn from the RIV model approves candidate c, at most j candidates in $\widehat{W}$ to the left of c, and strictly fewer than ℓ candidates in $\widehat{W}$ in total.

The $j+1$'th candidate in $\widehat{W}$ to the left of c we define as w^-, and w^+ is the $\ell - j$'th candidate in $\widehat{W}$ to the right of c: if v can only approve at most j candidates to the left of c, and $< \ell - j$ candidates in $\widehat{W}$ to the right of c, then w^- is the first candidate in $\widehat{W}$ to the left of c that v cannot approve, and w^+ is the first candidate in $\widehat{W}$ to the right of c that v cannot approve. We also define events Y^- and Y^+: Y^- is defined as the event that the voter's left endpoint lies between c and w^- and Y^+ is defined as the event that the voter's right endpoint lies between c and w^+. This means that the event Y occurring is equivalent to both Y^- and Y^+ occurring. Finally, the vector $\mathbf{d}^-$ is the vector of probabilities where element d_t^- is the probability, conditioned on $v \in F_t$, that a_v will sit between c and w^-. $\mathbf{d}^+$ is defined similarly.

Definition 9. *We define $\widehat{W}[s] = 0$ for $s < 0$, and $\widehat{W}[s] = 1$ for $s \geq k$. Set $\rho = |\{x \in \widehat{W} : x < c\}|$. For $0 \leq j < \ell \leq k$ and $c \in C \setminus \widehat{W}$, define random event Y as*

$$c \in A(v) \text{ and } A(v) \cap \widehat{W} \subseteq \widehat{W}[\rho - j : \rho + \ell - 1 - j].$$

Let $w^- = \widehat{W}[\rho - j - 1]$ and $w^+ = \widehat{W}[\rho + \ell - 1 - j]$. Let Y^- be the event that $a_v \in (w^-, c]$ and Y^+ the event that $b_v \in [c, w^+)$. Let $\mathbf{d}^+ = \mathbf{F}(w^+) - \mathbf{F}(c)$, $\mathbf{d}^- = \mathbf{F}(c) - \mathbf{F}(w^-)$.

The proof of Theorem 5 proceeds in five steps: (1) Lemma 7 shows that the G-probability mass between the boundary committee members w^- and w^+ is at most $2\ell/(k+1)$, exploiting the greedy placement of $\widehat{W}$ at G-quantiles $i/(k+1)$. (2) Lemma 8 applies this to show $\Pr[Y] \leq \ell/(k+1)$, where Y is the event that a voter belongs to a group potentially violating PJR+. (3) Lemma 9 bounds the probability that enough voters satisfy Y, accounting for query uncertainty via Δ; Fact 2 gives $\Pr[s_{\ell,c,j} \geq n\ell/k] \leq \exp(-n\ell/24k^3)$. (4) By Lemma 10, for each (ℓ, j), the union over $c \in C \setminus \widehat{W}$ reduces to at most $68k^2$ "cells" (one per open query interval in P, plus one per candidate in $P \setminus \widehat{W}$). There are $k(k+1)/2 \leq k^2$ pairs ($\ell \in [k]$, $0 \leq j < \ell$), so the total union has at most $68k^4$ terms, giving a validation failure probability of at most $68k^4 \exp(-n/24k^3)$ (Appendix C). (5) When $n = \Omega\left(k^3 \log\left(k \log m\right)\right)$, this failure probability is $\mathcal{O}(1/\log m)$, so the expected queries are $\mathcal{O}(\log k)$ per voter.

We now find a bound for $\Pr[Y]$ in terms of ℓ and k. First, we shall show that $G(c) - G(w^-) = \mathbf{p} \cdot \mathbf{d}^- \leq \frac{2j}{k+1} + 2\varepsilon$ and $G(w^+) - G(c) = \mathbf{p} \cdot \mathbf{d}^+ \leq \frac{2(\ell-j)}{k+1} - 2\varepsilon$ for some ε we shall define later, and so $\mathbf{p} \cdot (\mathbf{d}^+ + \mathbf{d}^-) \leq \frac{2\ell}{k+1}$.

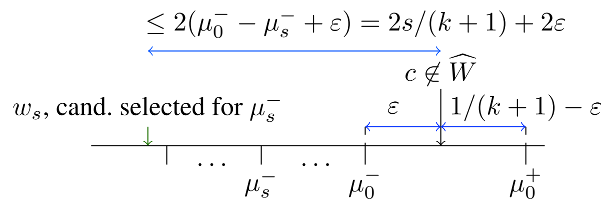

Figure 2: Intuition for Lemma 7. The w_s cannot be further in probability from μ_s^- than a candidate $c \notin \widehat{W}$, as otherwise c would have been chosen by μ_s^- instead.

Lemma 7. $G(w^+) - G(w^-) = \mathbf{p} \cdot (\mathbf{d}^+ + \mathbf{d}^-) \leq \frac{2\ell}{k+1}$.

Proof. A diagram of the argument is provided in Figure 2. Let μ_0^- be the marked point closest to the left of c. That is, $\mu_0^- = \frac{\alpha}{k+1}$ for $\alpha \in [k]$ such that $\frac{\alpha}{k+1} < G(c) < \frac{\alpha+1}{k+1}$ for some $\alpha \in [k]$: these inequalities are strict since otherwise c would be selected by the marked point. Let $\mu_s^- = \frac{\alpha - s}{k+1}$ be the $(s+1)$'th closest marked point to the left of c. First, we shall show that $\mathbf{p} \cdot \mathbf{d}^- \leq \frac{2j}{k+1} + 2\varepsilon$ where $\varepsilon = G(c) - \mu_0^-$. We have that $\mathbf{p} \cdot \mathbf{d}^- = G(c) - G(w^-)$. Consider the selection process for $\widehat{W}$: for each marked point μ_s^-, we select a candidate $w_s \in \widehat{W}$ for μ_s^-, and in particular since $c \notin \widehat{W}$, we know that $|G(w_s) - \mu_s^-| \leq |G(c) - \mu_s^-|$. From this, we know that for any candidate $w \in \widehat{W}$, $w < c$ if and only if it is selected for some marked point $\mu_s^- < c$. This means that the number of $\widehat{W}$ candidates to the left of c is $|\widehat{W} \cap [0, c)| = \alpha$. If $j \geq \alpha$, then we have that $w^- = 0$ and so we have $\mathbf{p} \cdot \mathbf{d}^- = G(c) = \frac{\alpha}{k+1} + \varepsilon \leq \frac{2j}{k+1} + 2\varepsilon$.

Now, if $j < \alpha$, then since $\mu_s^- - G(w_s) \leq G(c) - \mu_s^-$, we can say that $G(w_s) \in [2\mu_s^- - G(c), G(c))$. Therefore we know that there exist at least $j+1$ candidates from $\widehat{W}$ in the interval $[2\mu_j^- - G(c), G(c))$, and thus $G(w^-) = G(\widehat{W}[\rho - j - 1]) \in [2\mu_j^- - G(c), G(c))$. From this, we can say that $\mathbf{p} \cdot \mathbf{d}^- = G(c) - G(w^-) \leq 2(G(c) - \mu_j^-) = 2(\varepsilon + \frac{j}{k+1})$. We can argue similarly that $\mathbf{p} \cdot \mathbf{d}^+ \leq \frac{2(\ell - j)}{k+1} - 2\varepsilon$, and thus $\mathbf{p} \cdot (\mathbf{d}^+ + \mathbf{d}^-) \leq \frac{2\ell}{k+1}$. □

We apply this bound $\mathbf{p} \cdot (\mathbf{d}^+ + \mathbf{d}^-) \leq \frac{2\ell}{k+1}$ to $\Pr[Y]$.

Lemma 8. $\Pr[Y] \leq \frac{\ell}{k+1}$.

Proof. Applying Lemmas 3 and 7, $\Pr[Y^+ \wedge Y^-] = \sum_{t \in [\sigma]} 2 p_t d_t^+ d_t^- \leq \frac{1}{2} \sum_{t \in [\sigma]} p_t \left(d_t^+ + d_t^-\right) \leq \frac{\ell}{k+1}$, where $0 \leq d_t^+ + d_t^- \leq 1$ implies that $d_t^+ d_t^- \leq \frac{d_t^+ + d_t^-}{4}$. □

Using Lemma 8, we can obtain a probabilistic bound on the event $s_{\ell,c,j} \geq n\ell/k$.

Lemma 9. *For $\ell \in [k], c \in C \setminus \widehat{W}, 0 \leq j < \ell$, we have* $\Pr\left[s_{\ell,c,j} \geq \frac{\ell n}{k}\right] \leq \exp\left(-n\ell/24k^3\right)$.

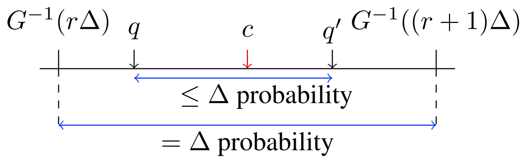

Figure 3: Intuition for Lemma 9. Since $c \notin P$, q, q' exist. The probability between q, q' is at most Δ, so the probability of $\Pr[X_v \wedge c \notin A(v)] \leq 2\Delta$.

The full proof can be found in Appendix B. Briefly, for a given voter v, let X_v be the event that $\texttt{poss}(v, c, \widehat{W} \setminus R)$ returns true. Then $\Pr[X_v] \leq \Pr[Y] + \Pr[X_v \wedge \neg Y]$. The event $X_v \wedge \neg Y$ can be considered a "false positive" event. Since all of $\widehat{W}$ is part of the query set P, we cannot return a false positive if the voter approves of some candidate in $S = \widehat{W} \setminus R$. Instead, we can only return a false positive if $c \notin A(v)$. In this case, at least one of the voters endpoints a_v, b_v sits between the two closest query points q, q' left and right of c; otherwise either $c \in A(v)$ or we can detect from the queries that $c \notin A(v)$, both contradictions. See Figure 3 for a diagram. Then $\Pr[X_v \wedge \neg Y] = \Pr[|\{a_v, b_v\} \cap [q, q']| \geq 1]$ is upper bounded by $\leq 2\Delta$, and so using Lemma 8, $\Pr[X_v] \leq \Pr[Y] + 2\Delta \leq \ell/(k+1) + 2\Delta$. We then apply Fact 2 to show that $\Pr\left[s_{\ell,c,j} \geq \frac{\ell n}{k}\right] \leq \exp\left(-n\ell/24k^3\right)$ (full proof in Appendix B).

Lemma 10. *For any consecutive query points $q, q' \in P$, for any $\ell \in [k], 0 \leq j < \ell$ and $c, c' \in C \cap (q, q')$, we have $s_{\ell,c,j} = s_{\ell,c',j}$.*

Proof. Consider a voter v, and H_1, H_2 computed in $\texttt{poss}$. The values H_1, H_2 are identical in $\texttt{poss}(v, c, S)$ and $\texttt{poss}(v, c', S)$. Also, since $\widehat{W} \subseteq P$ and $c, c' \in C \cap (q, q')$, for a given ℓ, j, the set R computed in $\texttt{validate}$, and thus the S parameter in $\texttt{poss}$, is the same for c and c'. Therefore $H_1 \cap (c_S^-, c] \neq \varnothing$ iff $H_1 \cap ((c')_S^-, c'] \neq \varnothing$, and $H_2 \cap [c, c_S^+) \neq \varnothing$ iff $H_2 \cap [c', (c')_S^+) \neq \varnothing$, so $\texttt{poss}(v, c, S)$=$\texttt{poss}(v, c', S)$, and $s_{\ell,c,j} = s_{\ell,c',j}$. $\square$

We can now prove Theorem 5.

Proof of Theorem 5. First, the output of Algorithm 2 always provides PJR+[7]: if $\widehat{W}$ is output, the voters in any group satisfying (i-iii) of Remark 1 are counted by some $s_{\ell,c,j}$, and otherwise W^* satisfies PJR+.

We bound $s_{\ell,c,j}$ with Lemma 9. Let $Q_i = (p_i, p_{i+1})$ be the open intervals between consecutive elements $p_i \in P$. For $\ell \in [k], 0 \leq j < \ell, c \in C \setminus \widehat{W}$, let $X_{\ell,c,j}$ be the event that $s_{\ell,c,j} \geq n\ell/k$. By Lemma 10, for any two $c, c' \in Q_i$, $X_{\ell,c,j} = X_{\ell,c',j}$; thus the union over c in any open interval reduces to one term per interval. Candidates in $P \setminus \widehat{W}$ (at most $|P| \leq 34k^2$ boundary points) each contribute an individual term, also bounded by Lemma 9. Since $|P| \leq k + 32k^2$, there are at most $34k^2$ open intervals plus at most $34k^2$ boundary candidates, giving at most $68k^2$ cells in total, and

$$\Pr\left[\bigcup_{\ell,j,c} X_{\ell,c,j}\right] \leq 68k^4 \exp\left(-\frac{n}{24k^3}\right).$$

Therefore, the expected number of queries per voter is

$$\mathbf{E}(Q) \leq 68k^4 \log m \exp\left(-\frac{n}{24k^3}\right) + \mathcal{O}(\log k) \leq \mathcal{O}(\log k),$$

if $n \in \Omega\left(k^3 \log(k^4 \log m)\right) = \Omega\left(k^3 \log\left(k \log m\right)\right)$. $\square$

Note that PJR+ still holds if $n = o(k^3 \log(k \log m))$, but we may need $\omega(\log k)$ expected queries per voter (though still $\mathcal{O}(\log m)$ per voter).

Remark 11. *Algorithm 2 requires access to $G = \mathbf{p} \cdot \mathbf{F}$. If only a strictly increasing approximation $\hat{G}$ satisfying $\|\hat{G} - G\|_\infty \leq \varepsilon_G \leq \Delta/2 = \left(32k^2\right)^{-1}$ is available, the algorithm remains correct with the same sample bound $n = \Omega\left(k^3 \log\left(k \log m\right)\right)$: the quantile-position error of at most $2\varepsilon_G$ per winner increases $\Pr[Y]$ by $2\varepsilon_G$ and the false-positive probability by $4\varepsilon_G$, while the validation gap shrinks from $\ell/(k(k+1)) - 2\Delta$ to $\ell/(k(k+1)) - 5\Delta > 0$, so Bernstein still applies. See Appendix F.*

We now briefly discuss EJR+ in this setting. For a committee W, define $N_{\ell,c,j} = \{v \in V : c \in A(v), A(v) \cap W \subseteq R\}$ where $R = \widehat{W}[\rho - j : \rho + \ell - 1 - j]$, and let w_i^+ be the i'th candidate in W to the right of $c \notin W$ and w_i^- be the i'th candidate in W to the left of c. We have seen that a committee W satisfying PJR+ is equivalent to all $|N_{\ell,c,j}| < \frac{n\ell}{k}$. For W to satisfy EJR+, we require that all $N_{\ell,c} = \cup_{j=0}^{\ell-1} N_{\ell,c,j}$ have size at most $|N_{\ell,c}| < \frac{n\ell}{k}$. The conditions for a voter to exist in $N_{\ell,c,j}$ simplify to their endpoints satisfying $a_v \in (w_{j+1}^-, c], b_v \in [c, w_{\ell-j}^+)$. However, considering disjoint conditions for a voter to exist in $N_{\ell,c}$, we need the voter to approve of at most j candidates in W to the left of c and exactly $\ell - j - 1$ candidates in W to the right of c for some j, i.e. the event $\bigsqcup_{j=0}^{\ell-1}\{a_v \in (w_{j+1}^-, c], b_v \in [w_{\ell-j-1}^+, w_{\ell-j}^+)\}$; See Figure 4 for an illustration. In Lemma 7, we present a straightforward technique for bounding the probability between c and e.g. $w_{\ell-j}^+$, however a technique for bounding the probability between $[w_{\ell-j-1}^+, w_{\ell-j}^+)$ is significantly more complex, hence why we have primarily discussed PJR+. While guaranteeing EJR+ is a direction for future work, we believe that $\widehat{W}$ constructed in Algorithm 2 provides EJR+ with high probability, and we note that we can adapt Algorithm 2 and the proof to show that $\widehat{W}$ provides at least 2-EJR+ with high probability, and 4/3-EJR+ when $\sigma = 1$: the proof of this can be found in Appendix D.

[7]Details can be found in Appendix C.

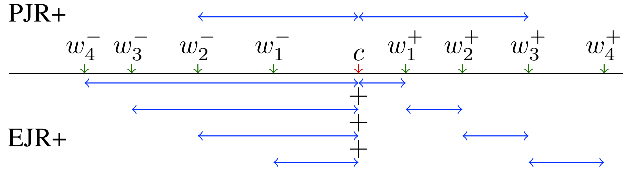

Figure 4: An example of the difference between PJR+ and EJR+ with $\ell = 4, j = 1$. For a voter to be in $N_{\ell,c,j}$, the voters endpoints must sit in the two blue regions above the line. For a voter to be in $N_{\ell,c}$, the voter's endpoints must sit in one of the 4 blue configurations below the line.

Theorem 12. *For an RIV, an adaptation of Algorithm 2 finds a 2-EJR+ committee. When the number of voters n satisfies $n = \Omega\left(k^3 \log\left(k \log m\right)\right)$, the algorithm uses $\mathcal{O}(\log k)$ queries per voter in expectation and with high probability. When $\sigma = 1$ (all voters share a single CDF), we can achieve $4/3$-EJR+ with $n = \Omega\left(k^4 \log\left(k \log m\right)\right)$.*

When $\sigma = 1$ (all voters share a single CDF), we can achieve $4/3$-EJR+. The bound on the probability that a voter approves c but few winners decomposes as a sum over types $t \in [\sigma]$, each contributing a product of two probability terms. For general $\sigma > 1$, we have only linear bounds on each term averaged over types, and converting these into a bound on the averaged product requires a product-to-sum inequality that is loose. When $\sigma = 1$, there is no averaging step, and the product can be bounded sharply, yielding the tighter $4/3$-EJR+ bound.

We also note that Algorithm 2 can be adapted to work with any random distribution over *intervals* to guarantee 2-PJR+. We use $G(x)$ as half the sum of the two cumulative marginal distributions over the two voter endpoints a_v, b_v and by changing `validate` to use a quota of $\frac{2n\ell}{k}$ instead of $\frac{n\ell}{k}$. The full proof can be found in Appendix E.

Theorem 13 (Informal)**.** *For a random interval voter model where voters' endpoints are drawn jointly from an arbitrary distribution over pairs, a variant of Algorithm 2 finds a 2-PJR+ committee using $\mathcal{O}(\log k)$ queries per voter in expectation and w.h.p. if $n = \Omega(k^3 \log(k \log m))$. However, there is a joint distribution such that for any $\varepsilon > 0$, $\widehat{W}$ fails $(2-\varepsilon)$-PJR+ w.h.p.*

Proof idea. We use G as the mean of the two marginal cumulative distributions F_{Z_1}, F_{Z_2} of the two voter endpoints. The result that $G(w^+) - G(w^-) \leq \frac{2\ell}{k+1}$ from Lemma 7 still holds, and so we wish to show that $\Pr[Y] \leq \frac{2\ell}{k+1}$. We have that $\Pr[Y] \leq \min(\Pr[Y^+], \Pr[Y^-])$. We also have that $\Pr[Y^+] = F_{Z_2}(w^+) - F_{Z_2}(c)$ and $\Pr[Y^-] = F_{Z_1}(c) - F_{Z_1}(w^-)$. Now, $\Pr[Y^+] + \Pr[Y^-] \leq 2(G(w^+) - G(w^-)) \leq \frac{4\ell}{k+1}$, and thus $\Pr[Y] \leq \frac{1}{2}\frac{4\ell}{k+1} = \frac{2\ell}{k+1}$. We can apply a technique as in Lemma 9 to bound $\Pr[s_{\ell,c,j} \geq 2n\ell/k]$, and show that $\widehat{W}$ is 2-PJR+ w.h.p. $\square$

5 CONCLUSION AND FUTURE WORK

We introduced a new random voter model for multiwinner voting with one-dimensional approval preferences called the Random Interval Voter model (RIV), as well as a query framework for this model. Given an RIV election, we can find a committee that satisfies PJR+ using $\mathcal{O}(\log k)$ queries per voter in expectation.

The query complexity, as defined in our work, can be seen as a crude heuristic for *cognitive burden*. We believe that $\mathcal{O}(\log k)$ is a reasonable measure of what a voter can evaluate; however, the constant in the analysis may still be too large. We note that we can reduce $|P|$ used in Algorithm 2 by a constant factor, at the expense of requiring n to be larger by a constant factor.

As discussed in Section 4, an immediate open question is whether our approach can be extended to EJR+. To adapt our analysis to EJR+, we would have to bound $\Pr[\bigvee_{j=0}^{\ell-1} Y_{c,\ell,j}] < \frac{\ell}{k} - \frac{1}{\text{poly(k)}}$. To use a technique similar to in Section 4, we would need to obtain a bound for the probability between consecutive candidates in $\widehat{W}$. We can obtain a bound on $G(\widehat{W}[r]) - G(\widehat{W}[r-1])$ that is tight for each individual value of r, in that there exists an election where the bound matches for a given r; however, this bound is loose in that we cannot construct an election such that the bound is tight for all r, and it is too loose to prove EJR+. Indeed, the elections that are tight for a given r are often very loose for other values of r, so we would need a more sophisticated bound that simultaneously considers multiple values of r.

Another important direction is to establish lower bounds on the amount of information required in this model to guarantee forms of justified representation. The fact that we focus on *expected* rather than maximum number of queries makes it more challenging to establish lower bounds.

It would also be interesting to explore an alternative probabilistic model where each voter has an ideal position and a radius around that position they are willing to approve, with both the ideal position and the radius drawn from given distributions [Bredereck et al., 2019, Elkind et al., 2023, Szufa et al., 2022].

Finally, it is natural to ask if our results can be extended beyond one dimension. Indeed, spatial models with $d > 1$ are popular in the literature [Miller, 2018, Godziszewski et al., 2021, Imber et al., 2024]: each dimension corresponds to an issue, and a point in this space (a candidate) represents a specific stance on all of the issues simultaneously. Lewenberg et al. [2016] suggest that political opinions in the UK lie in a 10-dimensional space, so expanding our model to higher dimensions may be necessary to accurately model real-world political elections. We refer the reader to Appendix G for further discussion, and to Springham et al. [2026] for a full treatment of this d-dimensional generalization.

Acknowledgements

Drew Springham was supported by UK Research and Innovation [grant number EP/S023356/1], in the UKRI Centre for Doctoral Training in Safe and Trusted Artificial Intelligence (www.safeandtrustedai.org). Bart de Keijzer was supported by EPSRC grant EP/X021696/1. The authors thank the anonymous reviewers for constructive feedback and helpful suggestions.

References

Haris Aziz and Barton E Lee. Proportionally representative participatory budgeting with ordinal preferences. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 5110–5118, 2021.

Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2):461–485, 2017.

Maria-Florina Balcan, Ariel D. Procaccia, and Yair Zick. Learning cooperative games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI 2015*, pages 475–481, 2015.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities : a nonasymptotic theory of independence*. Oxford University Press, Oxford, United Kingdom, first edition, 2013.

Robert Bredereck, Piotr Faliszewski, Andrzej Kaczmarczyk, and Rolf Niedermeier. An experimental view on committees providing justified representation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI 2019*, pages 109–115, 2019.

Markus Brill and Jannik Peters. Robust and verifiable proportionality axioms for multiwinner voting. In *Proceedings of the 24th ACM Conference on Economics and Computation, ACM EC 2023*, page 301, 2023.

Vincent Conitzer and Tuomas Sandholm. Vote elicitation: Complexity and strategy-proofness. In *Proceedings of the 18th National Conference on Artificial Intelligence, AAAI 2002*, pages 392–397, 2002.

Lihi Naamani Dery, Meir Kalech, Lior Rokach, and Bracha Shapira. Reducing preference elicitation in group decision making. *Expert Syst. Appl.*, 61:246–261, 2016.

Palash Dey and Arnab Bhattacharyya. Sample complexity for winner prediction in elections. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015*, pages 1421–1430, 2015.

Edith Elkind and Martin Lackner. Structure in dichotomous preferences. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI 2015*, pages 2019–2025, 2015.

Edith Elkind, Piotr Faliszewski, Ayumi Igarashi, Pasin Manurangsi, Ulrike Schmidt-Kraepelin, and Warut Suksompong. Justifying groups in multiwinner approval voting. *Theor. Comput. Sci.*, 969:114039, 2023.

Piotr Faliszewski, Jarosaw Flis, Dominik Peters, Grzegorz Pierczyński, Piotr Skowron, Dariusz Stolicki, Stanisaw Szufa, and Nimrod Talmon. Participatory budgeting: Data, tools and analysis. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence, IJCAI 2023*, pages 2667–2674, 2023.

Micha Tomasz Godziszewski, Pawe Batko, Piotr Skowron, and Piotr Faliszewski. An analysis of approval-based committee rules for 2D-Euclidean elections. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 5448–5455, 2021.

Daniel Halpern, Gregory Kehne, Ariel D Procaccia, Jamie Tucker-Foltz, and Manuel Wüthrich. Representation with incomplete votes. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, pages 5657–5664, 2023.

Daniel Halpern, Safwan Hossain, and Jamie Tucker-Foltz. Computing voting rules with elicited incomplete votes. In *Proceedings of the 25th ACM Conference on Economics and Computation, ACM EC 2024*, pages 941–963, 2024.

Aviram Imber, Jonas Israel, Markus Brill, and Benny Kimelfeld. Approval-based committee voting under incomplete information. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, pages 5076–5083, 2022.

Aviram Imber, Jonas Israel, Markus Brill, Hadas Shachnai, and Benny Kimelfeld. Spatial voting with incomplete voter information. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence, AAAI 2024*, pages 9790–9797, 2024.

Tushant Jha and Yair Zick. A learning framework for distribution-based game-theoretic solution concepts. *ACM Transactions on Economics and Computation*, 11: 5:1–5:23, 2023.

Yusuf Hakan Kalayci, Jiasen Liu, and David Kempe. Full proportional justified representation. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2025*, pages 1070–1078, 2025.

Martin Lackner and Piotr Skowron. *Multi-Winner Voting with Approval Preferences*. Springer International Publishing, Cham, 2023.

Yoad Lewenberg, Yoram Bachrach, Lucas Bordeaux, and Pushmeet Kohli. Political dimensionality estimation using a probabilistic graphical model. In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence, UAI 2016*, pages 447–456, 2016.

Feline Lindeboom, Martijn Brehm, Davide Grossi, and Pradeep K. Murukannaiah. A voice for minorities: diversity in approval-based committee elections under incomplete or inaccurate information. *CoRR*, abs/2506.10843, 2025.

Tyler Lu and Craig Boutilier. Multi-winner social choice with incomplete preferences. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence, IJCAI 2013*, pages 263–270, 2013.

Jan Maly. The core in participatory budgeting can be empty. *Economics Letters*, 254:112472, 2025.

Debmalya Mandal, Nisarg Shah, and David P Woodruff. Optimal communication-distortion tradeoff in voting. In *Proceedings of the 21st ACM Conference on Economics and Computation, ACM EC 2020*, pages 795–813, 2020.

Reshef Meir, Omer Lev, and Jeffrey S Rosenschein. A local-dominance theory of voting equilibria. In *Proceedings of the ACM Conference on Economics and Computation, ACM EC 2014*, pages 313–330. ACM, 2014.

Nicholas R. Miller. The spatial model of social choice and voting. In *Handbook of Social Choice and Voting*, chapter 10, pages 163–181. Edward Elgar Publishing Limited, 2018.

Dominik Peters and Martin Lackner. Preferences single-peaked on a circle. *Journal of Artificial Intelligence Research*, 68:463–502, 2020.

Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation, ACM EC 2020*, pages 793–794, 2020.

Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. Proportional participatory budgeting with additive utilities. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems, NeurIPS 2021*, pages 12726–12737, 2021.

Grzegorz Pierczyński and Piotr Skowron. Core-stable committees under restricted domains. In *Proceedings of the 18th International Conference on Web and Internet Economics, WINE 2022*, pages 311–329, 2022.

Manon Revel, Smitha Milli, Tyler Lu, Jamelle Watson-Daniels, and Max Nickel. Representative ranking for deliberation in the public sphere. In *Proceedings of the 42nd International Conference on Machine Learning, ICML 2025*, 2025.

Christopher Small, Michael Bjorkegren, Timo Erkkilä, Lynette Shaw, and Colin Megill. Polis: Scaling deliberation by mapping high dimensional opinion spaces. *Recerca: revista de pensament i anàlisi*, 26(2), 2021.

Drew Springham, Edith Elkind, Bart de Keijzer, and Maria Polukarov. Multiwinner voting with spatial preferences under incomplete information. *CoRR*, abs/2607.01036, 2026.

Stanisaw Szufa, Piotr Faliszewski, Łukasz Janeczko, Martin Lackner, Arkadii Slinko, Krzysztof Sornat, and Nimrod Talmon. How to sample approval elections? In *Proceedings of the 31st International Joint Conference on Artificial Intelligence, IJCAI 2022*, pages 496–502, 2022.

Luis Sánchez-Fernández, Edith Elkind, Martin Lackner, Norberto Fernández García, Jesús A. Fisteus, Pablo Basanta Val, and Piotr Skowron. Proportional justified representation. *Artificial Intelligence*, page 104503, 2026.

Zoi Terzopoulou, Alexander Karpov, and Svetlana Obraztsova. Restricted domains of dichotomous preferences with possibly incomplete information. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 5726–5733, 2021.

The Computational Democracy Project. The Klimarat in Austria, 2022. URL `compdemocracy.org/Case-studies/2022-Austria-Klimarat/`.

Zhibing Zhao, Haoming Li, Junming Wang, Jeffrey O Kephart, Nicholas Mattei, Hui Su, and Lirong Xia. A cost-effective framework for preference elicitation and aggregation. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence, UAI 2018*, pages 446–456, 2018.

Appendix

Drew Springham[1] **Edith Elkind**[2] **Bart de Keijzer**[1] **Maria Polukarov**[1]

[1]King's College London, London, UK
[2]Northwestern University, Evanston, IL, USA

A CORE IMPLIES PJR+

Proposition 14. *If W fails PJR+, then W is not in the core.*

Proof. Suppose that committee W fails PJR+. Then there exists some $\ell \in [k], G \subseteq V$ such that $|G| \geq n\ell/k$ and $A(G) \setminus W \neq \varnothing$, but $|W \cap \bigcup_{v \in G} A(v)| < \ell$. Let $c \in A(G) \setminus W$, let $T' = W \cap \bigcup_{v \in G} A(v)$ and $T = \{c\} \cup T'$. Then $|T| \leq \ell$, but for all $v \in G$ it holds that $|W \cap A(v)| = |T' \cap A(v)| = |T \cap A(v)| - 1 < |T \cap A(v)|$. Hence, W is not in the core. $\square$

B PROOF OF LEMMA 9

Lemma 9. *For $\ell \in [k], c \in C \setminus \widehat{W}, 0 \leq j < \ell$, we have* $\Pr\left[s_{\ell,c,j} \geq \frac{\ell n}{k}\right] \leq \exp\left(-n\ell/24k^3\right)$.

Proof of Lemma 9. Fix $\ell \in [k], c \in C \setminus \widehat{W}, 0 \leq j < \ell$. Let us bound $s_{\ell,c,j}$. Let R and $S = \widehat{W} \setminus R$ be as defined in the algorithm. Let X_v be the event that $\texttt{poss}(v, c, \widehat{W} \setminus R)$ evaluates to true. This is the event that the voter's responses to the queries are consistent with the voter approving c and none of S. Recall that Y is the event that the voter *actually* approves of c and none of S. We decompose $\Pr[X_v] \leq \Pr[Y] + \Pr[X_v \wedge \neg Y]$ and we look to upper bound $\Pr[X_v \wedge \neg Y]$. First, see that event Y occurring is equivalent to $c \in A(v)$ and $A(v) \cap S = \varnothing$. So, $\Pr[X_v \wedge \neg Y] \leq \Pr[X_v \wedge c \notin A(v)] + \Pr[X_v \wedge A(v) \cap S \neq \varnothing]$. If a voter approves of some candidate $w \in S$, then it cannot be the case that X_v occurs, since w is queried, and so $\texttt{poss}$ would return false. Therefore $\Pr[X_v \wedge A(v) \cap S \neq \varnothing] = 0$ and $\Pr[X_v \wedge \neg Y] \leq \Pr[X_v \wedge c \notin A(v)]$.

If $c \in P$, then $\Pr[X_v \wedge c \notin A(v)] = 0$, so suppose $c \notin P$. For some r, we have that $G(c) \in (r\Delta, (r+1)\Delta)$. There exist queried positions q, q' either side of c, such that $q \in [G^{-1}(r\Delta), c)$ and $q' \in (c, G^{-1}((r+1)\Delta)]$; if no queried position exists in $[G^{-1}(r\Delta), c)$, then $(G^{-1}(r\Delta))^+_C = c$ and then $c \in P$, a contradiction, so $q \in [G^{-1}(r\Delta), c)$ exists (and symmetrically for q').

Suppose that both $a_v, b_v \notin [q, q']$. Then we have 2 cases: 1) $a_v, b_v < q$ or $a_v, b_v > q'$, in which case X_v cannot have occurred, as it is clear based on the query information that c is not approved, or 2) $a_v < q < c < q' < b_v$, in which case $c \in A(v)$. Contrapositively, if X_v and $c \notin A(v)$ both occur, then at least one of a_v, b_v sits in $[q, q']$ and the probability of this is $\leq 2\,(G(q') - G(q))$. So $\Pr[X_v \wedge c \notin A(v)] \leq 2\,(G(q') - G(q)) \leq 2\Delta$ since $G(q') \leq (r+1)\Delta$ and $G(q) \geq r\Delta$.

Combining, $\Pr[X_v] \leq p$ where $p = \frac{\ell}{k+1} + 2\Delta$. Let $X = \sum_{i=1}^{n} X_{v_i}$ and set $a = \frac{\ell}{k(k+1)} - 2\Delta$ (positive since $8\ell k > k+1$ for all $\ell, k \geq 1$). The bounds $k+1 \leq 2k$ and $2\Delta = \frac{1}{8k^2} \leq \frac{\ell}{8k^2}$ give $a \geq \frac{3\ell}{8k^2}$ and $p \leq \frac{9\ell}{8k}$. Applying Fact 2 to $X_{v_i} - p$ with $\sigma^2 = p(1-p) \leq p$, deviation $t = \frac{n\ell}{k} - np = na$, and bound $n\sigma^2 + \frac{t}{3} \leq \frac{9n\ell}{8k} + \frac{n\ell}{3k} = \frac{35n\ell}{24k}$:

$$\Pr\left[X \geq \frac{n\ell}{k}\right] \leq \exp\left(-\frac{t^2/2}{n\sigma^2 + t/3}\right) \leq \exp\left(-\frac{9n^2\ell^2/(128k^4)}{35n\ell/(24k)}\right) = \exp\left(-\frac{27n\ell}{560k^3}\right) \leq \exp\left(-\frac{n\ell}{24k^3}\right),$$

since $27 \cdot 24 = 648 > 560$. $\square$

Algorithm 3: Finding a 2-EJR+ committee

```
Function validate(Ŵ, Ê, C):
    for ℓ = 1 to k, c ∈ C \ Ŵ :
        ρ ← |{x ∈ Ŵ : x < c}|;
        for v ∈ V :
            for j = 0 to ℓ − 1 :
                R ← Ŵ[ρ − j : ρ + ℓ − 1 − j];
                if poss(v, c, Ŵ \ R) :
                    s_{ℓ,c} ← s_{ℓ,c} + 1;
                    if s_{ℓ,c} ≥ 2nℓ/k : return False ;
                    continue to next v;
    return True;
```

C PROOF THAT $\widehat{W}$ PROVIDES PJR+

Lemma 15. *If all $s_{\ell,c,j} < \frac{n\ell}{k}$, then $\widehat{W}$ provides PJR+ with certainty.*

Proof. Suppose that there exists an $\ell \in [k]$ and a subset of voters $G \subseteq V$ with $|G| \geq n\ell/k$ such that $|\widehat{W} \cap \bigcup_{v\in G} A(v)| < \ell$, but there exists a candidate $c \in A(G) \setminus \widehat{W}$. Given that such a G, c, ℓ exists, assume that ℓ is the minimum such value, so $|\widehat{W} \cap \bigcup_{v\in G} A(v)| = \ell - 1$. We will show that this implies $s_{\ell,c,j} \geq n\ell/k$ for some $j \in \{0\} \cup [\ell - 1]$, i.e., Algorithm 2 does not output $\widehat{W}$. Let

$$H = \widehat{W} \cap \bigcup_{v\in G} A(v), \quad S = \widehat{W} \setminus H, \quad L = \{t \in H : t < c\}, \quad j = |L|.$$

Since $c \in A(G)$, we know that all $A(v)$ for $v \in G$ overlap at c, and so therefore $\bigcup_{v\in G} A^*(v) = [\alpha, \beta]$ for some $0 \leq \alpha \leq \beta \leq 1$. Therefore, $H = \widehat{W} \cap [\alpha, \beta]$. We also have $|H| = \ell - 1$ since ℓ was minimal.

In the iteration of the algorithm using the above ℓ, c, j, we use $R = \widehat{W}[\rho - j : \rho + \ell - 1 - j]$. We shall show that $H = R$. First, $L = \widehat{W}[\rho - j : \rho]$; L is the closest j candidates of H that are to the left of c, and $\widehat{W}[\rho - j : \rho]$ is the closest j candidates in $\widehat{W}$ to the left of c. Since $H = \widehat{W} \cap [\alpha, \beta]$, these two sets are the same.

We also have $H \setminus L = \widehat{W}[\rho : \rho + \ell - 1 - j]$. $\widehat{W}[\rho : \rho + \ell - 1 - j]$ is the $\ell - 1 - j$ candidates of $\widehat{W}$ to the right of c and $H \setminus L$ is the remaining $|H| - j = \ell - 1 - j$ candidates in H. Again, since $H = \widehat{W} \cap [\alpha, \beta]$, these two sets are the same, and so $H = R$, and therefore in this iteration of the algorithm, we call $\mathtt{poss}(v, c, S)$ for each voter, where S is the set of $\widehat{W}$ candidates that are not approved by any voter in the group.

Now, for each $v \in G$, we have $\mathtt{poss}(v, c, S)$ returns true: Since v approves of c and none of S, we know that $a_v \in (c_S^-, c]$ and $b_v \in [c, c_S^+)$. We also have $a_v \in ((a_v)_P^-, (a_v)_P^+]$ and $b_v \in [(b_v)_P^-, (b_v)_P^+)$, where the quantities $(a_v)_P^-, (a_v)_P^+, (b_v)_P^-, (b_v)_P^+$ are found in Algorithm 1. Therefore $((a_v)_P^-, (a_v)_P^+] \cap (c_S^-, c] \neq \varnothing$ and $[(b_v)_P^-, (b_v)_P^+) \cap [c, c_S^+) \neq \varnothing$, and we have $\mathtt{poss}(v, c, S)$ for all $v \in G$. Hence $s_{\ell,c,j} \geq |G| \geq n\ell/k$. Therefore, if $\widehat{W}$ is output by Algorithm 2, then it provides PJR+ for the election. □

D 2 APPROXIMATION TO EJR+

Theorem 12. *For an RIV, an adaptation of Algorithm 2 finds a 2-EJR+ committee. When the number of voters n satisfies $n = \Omega\left(k^3 \log\left(k \log m\right)\right)$, the algorithm uses $\mathcal{O}(\log k)$ queries per voter in expectation and with high probability. When $\sigma = 1$ (all voters share a single CDF), we can achieve $4/3$-EJR+ with $n = \Omega\left(k^4 \log\left(k \log m\right)\right)$.*

Proof roadmap. The proof centers on bounding $\pi(C) = \Pr[c \in A(v), |A(v) \cap \widehat{W}| < \ell]$, the probability that a random voter approves the non-winner c but fewer than ℓ winners.

Step 1: structural reductions. Lemma 16 shows that it suffices to consider $|C| = k+1$. Lemma 17 establishes that relocating winners away from c cannot decrease $\pi(C)$. Using these, Lemma 18 shows WLOG each marked point selects a winner on the far side of itself from c.

Step 2: geometric analysis. Definition 11 introduces variables measuring the fractional offsets of c from its surrounding marked points and the alignment between winners and marked points on each side of c. Lemmas 20 and 21 show that at least $\ell/2$ of the first ℓ winners on each side lie in the region where winners and marked points are in one-to-one correspondence, bounding the cumulative misalignment.

Step 3: decomposing $\pi(C)$. Lemma 22 writes $\pi(C)/2$ as a sum of products of probability gaps $d_i^{\pm}$. Applying $d_i^+ d_j^- \leq (d_i^+ + d_j^-)/4$ and Lemmas 23 to 25 yields Theorem 26: $\pi(C) \leq 2\ell/(k+1)$.

Step 4: concentration. Lemma 27 gives $\Pr[X_v] \leq 2\ell/(k+1) + 4\Delta$; a Bernstein bound yields $\Pr[s_{\ell,c} \geq 2\ell n/k] \leq \exp(-n/12k^3)$. A union bound over (ℓ, Q) pairs (where Q ranges over intervals between consecutive query points) completes the proof.

Sharpening to $4/3$*-EJR+ when* $\sigma = 1$. The step $d_i^+ d_j^- \leq (d_i^+ + d_j^-)/4$ in Lemma 22 is lossy. When $\sigma = 1$, $\pi(C)/2$ becomes a polynomial in five real variables $(\varepsilon, \gamma, \zeta, \alpha^+, \alpha^-)$ which is maximized directly via sequential one-variable maximization, yielding the sharper bound $\pi(C) \leq 4\ell/(3k) - 1/(2k^2)$ and the $4/3$-EJR+ guarantee.

We note a couple of differences between `validate` in Algorithm 2 and Algorithm 3. We add a voter to counter $s_{\ell,c}$ if $\texttt{poss}(v, c, \widehat{W} \setminus R)$ for *any* $0 \leq j < \ell$. We also change the quota size to $\frac{2n\ell}{k}$ instead of $\frac{n\ell}{k}$. Now, we look to bound the probability that a voter will approve of a candidate c and less than ℓ candidates in $\widehat{W}$. We first show that we can make some assumptions WLOG about the structure of candidates.

Definition 10. *Fix some $c \in [0,1]$, some $k \in \mathbb{N}$, and $\ell \in [k]$. We aim to find C such that $c \in C$, $c \notin \widehat{W}(C)$ and where $\pi(C) = \Pr[c \in A(v), |A(v) \cap \widehat{W}(C)| < \ell]$ is maximized over all $C \subset [0,1]$. We consider a candidate set* valid *if $c \in C$, but $c \notin \widehat{W}(C)$, where $\widehat{W}(C)$ is the set of winners that Algorithm 3 would find on the candidate set.*

Lemma 16. *WLOG, we only need to consider C where $|C| = k+1$, so C is comprised of $c \notin \widehat{W}$, and some set $\widehat{W}$.*

Proof. If C contained other unelected candidates, then $\pi(C) = \pi(\widehat{W}(C) \cup \{c\})$ remains the same. ☐

Lemma 17. *Let C be a valid candidate set. Suppose there exists an injective function $\phi : C \to [0,1]$ such that $\phi(c) = c$, and for every $x \in C$, $|G(c) - G(\phi(x))| \geq |G(c) - G(x)|$. Let $C' = \phi(C)$. If C' is still valid, then $\pi(C') \geq \pi(C)$.*

What we are saying with this lemma is that if we can move the winners away from c without c becoming elected, then this cannot decrease π.

Proof. Let $\widehat{W}^+ = \{w \in \widehat{W}(c) : w > c\}$ and $\widehat{W}^- = \{w \in \widehat{W}(c) : w < c\}$. Consider breaking ϕ into two functions, where ϕ_1 moves only the candidates to the right of c, and ϕ_2 moves only the candidates to the left of c: $\phi_1(\widehat{W}^+) = \widehat{W}^+$ and $\phi_1(\widehat{W}^-) = \phi(\widehat{W}^-)$ and $\phi_2(\phi_1(\widehat{W}^+)) = \phi_2(\widehat{W}^+) = \phi(\widehat{W}^+)$ and $\phi_2(\phi_1(\widehat{W}^-)) = \phi_2(\phi(\widehat{W}^-)) = \phi(\widehat{W}^-)$. $\pi(C) = \sum_{i=0}^{\ell-1} \Pr[c \in A(v), |A(v) \cap W^-| = i, |A(v) \cap W^+| < \ell - i] = \sum_{j=0}^{\ell-1} \Pr[c \in A(v), |A(v) \cap W^-| < \ell - j, |A(v) \cap W^+| = j]$. By moving the winners on the right of c away from c, then $\Pr[c \in A(v), |A(v) \cap W^-| = i, |A(v) \cap W^+| < \ell - i]$ cannot decrease. Similarly when we move only the winners on the left of c away from c, $\Pr[c \in A(v), |A(v) \cap W^-| < \ell - j, |A(v) \cap W^+| = j]$ cannot decrease. Now, we see that by applying ϕ_1, $\pi(\phi_1(C)) \geq \pi(C)$. We can then apply ϕ_2 to get $\pi(\phi(C)) = \pi(\phi_2(\phi_1(C))) \geq \pi(\phi_1(C)) \geq \pi(C)$. ☐

Lemma 18 (No backpicking)**.** *WLOG, we can assume that marked points μ must select their corresponding candidate w such that $w \geq \mu$ if $\mu \geq c$, or $w \leq \mu$ if $\mu \leq c$.*

Marked points must select candidates further away from c than μ.

Proof. Suppose that we have $c < w < \mu$ where μ selects w. Consider the alternate candidate set where we just move w to μ. We assume Lemma 16, so there does not currently exist a candidate at μ. Then μ must still pick w, no other winner selection

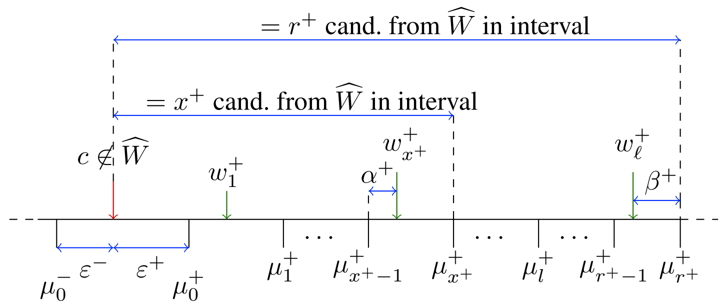

Figure 5: A diagram of the variables introduced in Definition 11.

is changed, but we have moved w away from c, and so by Lemma 17, π cannot have decreased. We can think similarly if $\mu < w < c$. $\square$

Now, taking Lemma 16 and Lemma 18 as assumptions, we will analyze $\pi(C)$.

We shall act as though there exist at least ℓ marked points, and therefore $\widehat{W}$ candidates, to the left and right of c. If this is not the case, we note that Lemmas 23 to 25 still hold, since in this case we are acting as though the marked points and $\widehat{W}$ candidates extend beyond $[0, 1]$, only increasing $\pi(C)$, and so for deriving bounds on $\pi(C)/2$ and for proving Theorem 13, we make this assumption for the simplicity of presentation.

We first make some definitions. Before the formal definition, we shall give some intuition. We shall discuss only the definitions with a $+$ superscript, as the $-$ superscript definitions are analogous. See Figure 5 for a diagram. For a C valid candidate set, we define μ_t^+ as the $(t+1)$'th marked points to the right of c respectively. x^+ is the largest index such that between c and the $x^+ + 1$'th marked point $\mu_{x^+}^+$ to the right of c (non-inclusive), there exists exactly x^+ $\widehat{W}$ candidates. In other words, between these points there exists x^+ marked points and x^+ $\widehat{W}$ candidates. These candidates must have been picked by these marked points, since Lemma 18 does not allow any other marked point to have selected any of these candidates.

ε^+ is the probability between c and the marked point directly to the right of c. α^+ is the probability between the x^+'th winner to the right of c, and the x^+'th marked point. Finally, β^+ is the probability between μ_r^+ and the ℓ'th winner to the right of c.

Definition 11. *For a C valid candidate set, we define μ_t^+ as the $(t+1)$'th marked points to the right of c respectively. We define μ_t^- as the $(t+1)$'th marked points to the left of c respectively. Formally, $\mu_t^+ = \frac{\lceil G(c)(k+1) \rceil + t}{(k+1)}$, and $\mu_t^- = \frac{\lceil G(c)(k+1) \rceil - t - 1}{(k+1)}$. We define $x^+ \in [\ell]$ as the largest number such that $|\widehat{W} \cap (G(c), \mu_{x^+}^+)| = x^+$, and $x^- \in [\ell]$ as the largest number such that $|\widehat{W} \cap (\mu_{x^-}^-, G(c))| = x^-$. We define $\varepsilon^+ = \mu_0^+ - G(c)$ and $\varepsilon^- = G(c) - \mu_0^-$. Let w_t^+ be the t'th candidate in $\widehat{W}$ to the right of c and w_t^- be the t'th candidate in $\widehat{W}$ to the left of c with $w_0^+ = w_0^- = c$. We define $\alpha^+ = |G(w_{x^+}^+) - \mu_{x^+-1}^+|$ and $\alpha^- = |G(w_{x^-}^-) - \mu_{x^--1}^-|$. We define $r^+ > x^+$ to be smallest such that $|\widehat{W} \cap (G(c), \mu_{r^+}^+)| = r^+$ and r^- to be smallest such that $|\widehat{W} \cap (\mu_{r^-}^-, G(c))| = r^-$. We define $\beta^+ = |\mu_{r^+}^+ - G(w_\ell^+)|$ and $\beta^- = |\mu_{r^-}^- - G(w_\ell^-)|$.*

Note that we can never have $|\widehat{W} \cap (G(c), \mu_t^+)| > t$ and similarly on the left: If so, then $(G(c), \mu_t^+)$ contains exactly t marked points but $\geq t+1$ $\widehat{W}$ candidates, implying that some marked point has violated Lemma 18.

We establish some relationships between these variables.

Lemma 19. *$\varepsilon^+ = \frac{1}{k+1} - \varepsilon^-$, and $\varepsilon^+, \varepsilon^- \in (0, 1/(k+1))$*

This can be seen easily from Figure 5.

Proof. Follows from the fact that $\mu_0^+ - \mu_0^- = \frac{1}{k+1}$. We also have $\varepsilon^+, \varepsilon^- \in (0, 1/(k+1))$ since if $\varepsilon^+ = 0$ or $\varepsilon^+ = 1/(k+1)$, then c must be elected. $\square$

Lemma 20. *x^+ and x^- exist. $\mu_{x^+}^+$ selects a candidate from $(\mu_{r^+-1}^+, \mu_{r^+}^+)$ and similarly for μ^-*

Since r^+ is the first index after x^+ that we have $|\widehat{W} \cap (\mu_i^-, G(c))| = i$, it must be that the marked point $\mu_{x^+}^+$ selects a candidate from $(\mu_{r^+-1}^+, \mu_{r^+}^+)$.

Proof. First, x^+ exists: $|\widehat{W} \cap (G(c), \mu_1^+)| = 1$, since otherwise if $|\widehat{W} \cap (G(c), \mu_1^+)| = 0$, then μ_0^+ would select c instead, so $1 \leq x^+ < \ell$. This can be seen from Figure 5.

We know that $|\widehat{W} \cap (G(c), \mu_{x^+}^+)| = x^+$, and r^+ is the smallest $> x^+$ such that $|\widehat{W} \cap (G(c), \mu_{r^+}^+)| = r^+$, so for any $x^+ < t < r^+$, we can say $|\widehat{W} \cap (G(c), \mu_t^+)| < t$, and in particular $|\widehat{W} \cap (\mu_{x^+}^+, \mu_t^+)| \leq t - x^+ - 1$.

We also have that $\mu_{x^+}^+$ cannot select its candidate $> \mu_{r^+}^+$: since $|\widehat{W} \cap (G(c), \mu_{r^+}^+)| = r^+$, and marked points always select away from c, there are r^+ marked points in $(G(c), \mu_{r^+}^+)$ and r^+ candidates in that interval, $\mu_{x^+}^+$ must elect from $(G(c), \mu_{r^+}^+)$.

We now prove that $\mu_{x^+}^+$ cannot select it's candidate from $(\mu_{x^+}^+, \mu_t^+)$ for any $x^+ < t < r^+$ by induction. If we consider $t = x^+ + 1$, we know $|\widehat{W} \cap (\mu_{x^+}^+, \mu_t^+)| = 0$, so clearly no candidate is selected in $(\mu_{x^+}^+, \mu_t^+)$. Suppose for our inductive hypothesis that $\mu_{x^+}^+$ does not select it's candidate from $(\mu_{x^+}^+, \mu_t^+)$ for some $x^+ < t < r^+ - 1$. Can $\mu_{x^+}^+$ select its candidate from $(\mu_{x^+}^+, \mu_{t+1}^+)$? By the inductive hypothesis, if it does, then it must select from (μ_t^+, μ_{t+1}^+). But $|\widehat{W} \cap (\mu_{x^+}^+, \mu_{t+1}^+)| \leq t - x^+$: there are $t + 1 - x^+$ marked points in $[\mu_{x^+}^+, \mu_{t+1}^+)$ but only $\leq t - x^+$ elected candidates in $(\mu_{x^+}^+, \mu_{t+1}^+)$. Therefore $\mu_{x^+}^+$ cannot select its candidate w from $(\mu_{x^+}^+, \mu_{t+1}^+)$, since w would otherwise be closer to some other marked point that has *not* selected a candidate.

Therefore it must be that $\mu_{x^+}^+$ selects a candidate from $(\mu_{r^+-1}^+, \mu_{r^+}^+)$ and similarly for μ^-. $\square$

Lemma 21. *$\frac{r^+}{k+1} \leq \frac{2x^+}{k+1} + \varepsilon^+ + \beta^+$ and so $x^+ \geq \ell/2$. $\frac{r^-}{k+1} \leq \frac{2x^-}{k+1} + \varepsilon^- + \beta^-$ and so $x^- \geq \ell/2$.*

Since $\mu_{x^+}^+$ selects a candidate from $(\mu_{r^+-1}^+, \mu_{r^+}^+)$, this allows us to bound the probability between $\mu_{r^+}^+$ and $\mu_{x^+}^+$, since we need to avoid c being selected for $\mu_{x^+}^+$ instead.

Proof. By Lemma 20, we know that $\mu_{x^+}^+ = \mu$ selects its candidate w from $(\mu_{r^+-1}^+, \mu_{r^+}^+)$, so $G(w) \in (\mu_{r^+-1}^+, \mu_{r^+}^+)$. We also know that $\mu - G(c) \geq G(w) - \mu$ (recall that $G(w) \geq \mu$ WLOG) since otherwise μ would have selected c instead. We have $\frac{x^+}{k+1} + \varepsilon^+ = \mu - \mu_0^+ + \varepsilon^+ = \mu - G(c) \geq \mu_{r^+}^+ - \beta^+ - \mu = \frac{r^+ - x^+}{k+1} - \beta^+$ and thus $\frac{2x^+}{k+1} + \varepsilon^+ + \beta^+ \geq \frac{r^+}{k+1}$.

Suppose to the contrary that $x^+ < \ell/2$. Then $x^+ \leq (\ell-1)/2$. Then we have $\frac{\ell+1}{k+1} \leq \frac{r^+}{k+1} \leq \frac{2x^+ + 1}{k+1} + \varepsilon^+ \leq \frac{\ell}{k+1} + \varepsilon^+ < \frac{\ell+1}{k+1}$, a contradiction. We get an analogous result for $x^- \geq \ell/2$ $\square$

We now make some definitions that will be used to bound $\pi(C)$.

Definition 12. *Define $\delta_{t,i}^+ = |F_t(w_{i+1}^+) - F_t(w_i^+)|$, and $d_{t,i}^+ = |F_t(w_{i+1}^+) - F_t(c)|$. Define $\delta_{t,i}^- = |F_t(w_{i+1}^-) - F_t(w_i^-)|$, and $d_{t,i}^- = |F_t(w_{i+1}^-) - F_t(c)|$. Define $d_i^+ = \sum_{t \in [\sigma]} p_t d_{t,i}^+$. Define $d_i^- = \sum_{t \in [\sigma]} p_t d_{t,i}^-$*

Using these definitions, we can bound $\pi(C)$.

Lemma 22. *$\pi(C)/2 \leq \frac{1}{4}\left(d_{x^+-1}^+ + d_{x^--1}^- + (d_{\ell-1}^+ - d_{x^+-1}^+) + d_{\ell-1-x^+}^- + d_{\ell-1-x^-}^+ + (d_{\ell-1}^- - d_{x^--1}^-)\right)$*

Proof. Let $\widehat{W}^+ = \{w \in \widehat{W} : w > c\}$ and $\widehat{W}^- = \{w \in \widehat{W} : w < c\}$. Consider $\pi(C) = \Pr[c \in A(v), |A(v) \cap \widehat{W}(C)| < \ell]$. We can break this down as

$$z_{i,j} = \Pr[c \in A(v), |A(v) \cap \widehat{W}^-| = i, |A(v) \cap \widehat{W}^+| = j],$$

so $\pi(C) = \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1-i} z_{i,j}$. We further have

$$z_{i,j} = \sum_{t \in [\sigma]} \Pr[c \in A(v), |A(v) \cap \widehat{W}^-| = i, |A(v) \cap \widehat{W}^+| = j \mid v \in F_t] p_t.$$

We can then further break this down as

$$\Pr[c \in A(v), |A(v) \cap \widehat{W}^-| = i, |A(v) \cap \widehat{W}^+| = j \mid v \in F_t] = 2\delta_{t,i}^- \delta_{t,j}^+$$

and thus

$$\pi(C)/2 \leq \sum_{t\in[\sigma]} \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1-i} p_t \delta^+_{t,j} \delta^-_{t,i}.$$

Now, since $x^+, x^- \geq \ell/2$, we have $x^+ + x^- \geq \ell$, and so we have that

$$\begin{aligned}\sum_{t\in[\sigma]} p_t \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1-i} \delta^+_{t,i}\delta^-_{t,i} &\leq \sum_{t\in[\sigma]} p_t \left(\sum_{i=0}^{x^+-1} \sum_{j=0}^{x^--1} \delta^+_{t,i}\delta^-_{t,j} + \sum_{i=x^+}^{\ell-1} \sum_{j=0}^{\ell-1-x^+} \delta^+_{t,i}\delta^-_{t,j} + \sum_{i=0}^{\ell-x^--1} \sum_{j=x^-}^{\ell-1} \delta^+_{t,i}\delta^-_{t,j} \right) \\ &= \sum_{t\in[\sigma]} p_t \left(d^+_{t,x^+-1} d^-_{t,x^--1} + (d^+_{t,\ell-1} - d^+_{t,x^+-1}) d^-_{t,\ell-1-x^+} + d^+_{t,\ell-1-x^-} (d^-_{t,\ell-1} - d^-_{t,x^--1}) \right)\end{aligned}$$

Recalling Lemma 8, we have that

$$\begin{aligned} d^+_{t,x^+-1} d^-_{t,x^--1} &\leq \frac{d^+_{t,x^+-1} + d^-_{t,x^--1}}{4}, \\ (d^+_{t,\ell-1} - d^+_{t,x^+-1}) d^-_{t,\ell-1-x^+} &\leq \frac{(d^+_{t,\ell-1} - d^+_{t,x^+-1}) + d^-_{t,\ell-1-x^+}}{4}, \\ d^+_{t,\ell-1-x^-} (d^-_{t,\ell-1} - d^-_{t,x^--1}) &\leq \frac{d^+_{t,\ell-1-x^-} + (d^-_{t,\ell-1} - d^-_{t,x^--1})}{4}. \end{aligned}$$

Hence we get

$$\begin{aligned} \pi(C)/2 &\leq \sum_{t\in[\sigma]} p_t \left(\frac{d^+_{t,x^+-1} + d^-_{t,x^--1} + (d^+_{t,\ell-1} - d^+_{t,x^+-1}) + d^-_{t,\ell-1-x^+} + d^+_{t,\ell-1-x^-} + (d^-_{t,\ell-1} - d^-_{t,x^--1})}{4} \right) \\ &\leq \frac{1}{4} \left(d^+_{x^+-1} + d^-_{x^--1} + (d^+_{\ell-1} - d^+_{x^+-1}) + d^-_{\ell-1-x^+} + d^+_{\ell-1-x^-} + (d^-_{\ell-1} - d^-_{x^--1}) \right) \end{aligned}$$

□

Now, we shall bound the components of this sum.

Lemma 23. *$d^+_{x^+-1} = \frac{x^+-1}{k+1} + \varepsilon^+ + \alpha^+$ and $d^-_{x^--1} = \frac{x^--1}{k+1} + \varepsilon^- + \alpha^-$*

Proof. First, we know that $|\widehat{W} \cap (G(c), \mu^+_{x^+})| = x^+$. There are x^+ marked points in $(G(c), \mu^+_{x^+})$, and x^+ elected winners in $(G(c), \mu^+_{x^+})$, so we know that the x^+'th winner to the right of c exists in $(G(c), \mu^+_{x^+})$. Since $\mu^+_{x^+-1}$ cannot select a winner towards c, it must select a winner in $(\mu^+_{x^+-1}, \mu^+_{x^+})$.

So, $d^+_{x^+-1} = G(w^+_{x^+}) - G(c) = \alpha^+ + \mu^+_{x^+-1} - \mu^+_0 + \varepsilon^+ = \frac{x^+-1}{k+1} + \varepsilon^+ + \alpha^+$. □

Lemma 24. *$d^+_{\ell-1} - d^+_{x^+-1} \leq \frac{x^++1}{k+1} + \varepsilon^+ - \alpha^+$ and $d^-_{\ell-1} - d^-_{x^--1} \leq \frac{x^-+1}{k+1} + \varepsilon^- - \alpha^-$*

Proof.

$$d^+_{\ell-1} - d^+_{x^+-1} = G(w^+_\ell) - G(w^+_{x^+}) \leq \mu^+_{r^+} - \beta^+ - \mu^+_{x^+-1} - \alpha^+ = \frac{r^+ - (x^+-1)}{k+1} - \alpha^+ - \beta^+ \leq \frac{2x^+ - (x^+-1)}{k+1} + \varepsilon^+ - \alpha^+$$

□

Lemma 25. *$d^+_{\ell-1-x^-} \leq 2\left(\frac{\ell-x^--1}{k+1} + \varepsilon^+\right)$ and $d^-_{\ell-1-x^+} \leq 2\left(\frac{\ell-x^+-1}{k+1} + \varepsilon^-\right)$.*

Proof. By a similar reasoning to Lemma 7, there must exist $\ell - x^-$ $\widehat{W}$ candidates in $(G(c), 2\mu^+_{\ell-x^--1} - G(c)))$. Therefore $d^+_{\ell-1-x^-} \leq 2(\mu^+_{\ell-x^--1} - G(c)) = 2\left(\frac{\ell-x^--1}{k+1} + \varepsilon^+\right)$. □

Theorem 26. $\pi(C) \le \frac{2\ell}{k+1}$

Proof. Recall from Lemma 22 that

$$\pi(C)/2 \le \frac{1}{4}\left(d^+_{x^+-1} + d^-_{x^--1} + (d^+_{\ell-1} - d^+_{x^+-1}) + d^-_{\ell-1-x^+} + d^+_{\ell-1-x^-} + (d^-_{\ell-1} - d^-_{x^--1})\right).$$

With Lemmas 23 to 25, we get

$$\begin{aligned}\pi(C)/2 &\le \frac{1}{4}\left(d^+_{x^+-1} + d^-_{x^--1} + (d^+_{\ell-1} - d^+_{x^+-1}) + d^-_{\ell-1-x^+} + d^+_{\ell-1-x^-} + (d^-_{\ell-1} - d^-_{x^--1})\right)\\ &\le \frac{1}{4}\left(4\varepsilon^+ + 4\varepsilon^- + 4\left(\frac{\ell-1}{k+1}\right)\right)\\ &= \frac{1}{4}\left(4\varepsilon^+ + 4\left(\frac{1}{k+1} - \varepsilon^+\right) + 4\left(\frac{\ell-1}{k+1}\right)\right)\\ &= \frac{\ell}{k+1}\end{aligned}$$

so hence $\pi(C) \le \frac{2\ell}{k+1}$ □

Lemma 27. *For $\Delta = (4k)^{-2}$, and for every $\ell \in [k], c \in C \setminus \widehat{W}$, it holds that* $\Pr\left[s_{\ell,c} \ge \frac{2\ell n}{k}\right] \le \exp\left(-\frac{n}{12k^3}\right)$.

Proof. Fix $\ell \in [k], c \in C \setminus \widehat{W}$. Let Y_v be the event that voter v approves of c and less than ℓ candidates in $\widehat{W}$. We have $\Pr[Y_v] = \pi(C)$. Let X_v be the event that $\texttt{poss}(v, c, \widehat{W} \setminus R)$ evaluates to true for some $R = \widehat{W}[\rho - j : \rho + \ell - 1 - j]$. We want to find $\Pr[X_v] = \Pr[X_v \wedge Y_v] + \Pr[X_v \wedge \neg Y_v]$.

First, we have $\Pr[X_v \wedge Y_v] = \Pr[Y_v] \le \pi(C)$, since $\Pr[X_v \mid Y_v] = 1$: if the voter approves of c and less than ℓ candidates in $\widehat{W}$, then certainly for some R, it is true that $\texttt{poss}(v, c, \widehat{W} \setminus R)$. In particular, we can choose any $R \supseteq A(v) \cap \widehat{W}$.

Let us consider $\Pr[X_v \wedge \neg Y_v] \le \Pr[X_v \mid \neg Y_v]$, the probability that $\texttt{poss}(v, c, \widehat{W} \setminus R)$ for some R despite the fact that the voter either does not approve of c, or approves of $\ge \ell$ candidates of $\widehat{W}$. If $|A(v) \cap \widehat{W}| \ge \ell$, then there cannot exist any R such that $\texttt{poss}(v, c, \widehat{W} \setminus R)$, so it must be the case that the voter does not approve of c, but we still have $\texttt{poss}(v, c, \widehat{W} \setminus R)$ for some R. It must be the case that $c \notin P$.

Let $q_r = G^{-1}(r\Delta)$ be the r-th position around which we query. Then consider r with $q_r \le c < q_{r+1}$. Since $c \notin P$, we know that $q_r \le (q_r)^+_C < c < (q_{r+1})^-_C \le q_{r+1}$, and so there exist queried positions either side of c with probability (with respect to G) at most Δ between them. With a_v, b_v as the endpoints of v's approval interval, then if X_v occurs, then $a_v \in (c^-_S, (q_{r+1})^-_C]$ and $b_v \in [(q_r)^+_C, c^+_S)$ for some $S = \widehat{W} \setminus R$. Indeed, if $a_v \le c^-_S$ or $a_v > (q_{r+1})^-_C$, then given the elicited information, it would be impossible for both c to be approved and c^-_S to be disapproved, and analogously for b_v. Since $\neg Y_v$ has occurred, we know that either $a_v > c$ or $b_v < c$. Let δ^- be the event that $c < a_v < (q_{r+1})^-_C$ and let δ^+ be the event that $(q_r)^+_C < b_v < c$. We then have

$$\Pr[X_v] \le \Pr[Y_v] + \Pr[\delta^+ \vee \delta^-] \le \frac{2\ell}{k+1} + 4\Delta.$$

Let $X = \sum_{i=1}^n X_{v_i}$ with $p = \frac{2\ell}{k+1} + 4\Delta$ and set $a = \frac{2\ell}{k(k+1)} - 4\Delta$ (positive since $8\ell k > k+1$ for all $\ell, k \ge 1$). The bounds $k+1 \le 2k$ and $4\Delta = \frac{1}{4k^2} \le \frac{\ell}{4k^2}$ give $a \ge \frac{3\ell}{4k^2}$ and $p \le \frac{9\ell}{4k}$. Applying Fact 2 to $X_{v_i} - p$ with $\sigma^2 = p(1-p) \le p$, deviation $t = \frac{2\ell n}{k} - np = na$, and bound $n\sigma^2 + \frac{t}{3} \le \frac{9n\ell}{4k} + \frac{2n\ell}{3k} = \frac{35n\ell}{12k}$:

$$\Pr\left[X \ge \frac{2\ell n}{k}\right] \le \exp\left(-\frac{t^2/2}{n\sigma^2 + t/3}\right) \le \exp\left(-\frac{9n^2\ell^2/(32k^4)}{35n\ell/(12k)}\right) = \exp\left(-\frac{27n\ell}{280k^3}\right) \le \exp\left(-\frac{n}{12k^3}\right),$$

since $27 \cdot 12 = 324 > 280$. □

We can now prove Theorem 12.

Proof of Theorem 12, 2-EJR+ claim. First, the output of Algorithm 3 always provides 2-EJR+: if $\widehat{W}$ is output by the algorithm, then for any group $G_{\ell,c} = \{v \in V : c \in A(v), |A(v) \cap W| < \ell\}$, the voters of $G_{\ell,c}$ are counted in $s_{\ell,c}$, so $s_{\ell,c} \geq |G_{\ell,c}|$. $\widehat{W}$ is output if $s_{\ell,c} < \frac{2n\ell}{k}$ for all ℓ, c, giving $|G_{\ell,c}| < \frac{2n\ell}{k}$, so 2-EJR+ holds. Otherwise W^* is output and satisfies EJR+.

We bound $s_{\ell,c}$ with Lemma 27. Let $Q_i = (p_i, p_{i+1})$ be the intervals between consecutive elements $p_i \in P$, with $|P| \leq k + 32k^2 \leq 34k^2$. By an argument analogous to Lemma 10, all $c, c' \in Q_i$ satisfy $X_{\ell,c} = X_{\ell,c'}$ (where $X_{\ell,c}$ is the event $s_{\ell,c} \geq 2\ell n/k$), so the union over $c \in C \setminus \widehat{W}$ reduces to a union over intervals Q_i:

$$\Pr\left[\bigcup_{\ell \in [k]} \bigcup_{c \in C \setminus \widehat{W}} X_{\ell,c}\right] \leq (|P|+1) \sum_{\ell=1}^{k} \exp\left(-\frac{n}{12k^3}\right) \leq O(k^3) \exp\left(-\frac{n}{12k^3}\right).$$

Therefore, since $|P| = \mathcal{O}(k^2)$, the expected number of queries is

$$\mathbf{E}(Q) \leq O(k^3) \log m \exp\left(-\frac{n}{12k^3}\right) + \mathcal{O}(\log k) \leq \mathcal{O}(\log k),$$

if $n \in \Omega(k^3 \log(k^3 \log m)) = \Omega\left(k^3 \log\left(k \log m\right)\right)$. □

In fact, if $\sigma = 1$, we can achieve $4/3$-EJR+.

Proof of Theorem 12, 4/3-EJR+ when $\sigma = 1$ claim. We use Algorithm 3, with the quota changed from $\frac{2\ell n}{k}$ to $\frac{4\ell n}{3k}$. From Lemma 22, we have that

$$\begin{aligned}
\pi(C)/2 &\leq \sum_{t \in [\sigma]} p_t \left(d^+_{t,x^+-1} d^-_{t,x^--1} + (d^+_{t,\ell-1} - d^+_{t,x^+-1}) d^-_{t,\ell-1-x^+} + d^+_{t,\ell-1-x^-}(d^-_{t,\ell-1} - d^-_{t,x^--1})\right) \\
&= d^+_{x^+-1} d^-_{x^--1} + (d^+_{\ell-1} - d^+_{x^+-1}) d^-_{\ell-1-x^+} + d^+_{\ell-1-x^-}(d^-_{\ell-1} - d^-_{x^--1}).
\end{aligned}$$

With Lemmas 23 to 25, we get

$$\begin{aligned}
\pi(C)/2 \leq &\left(\frac{x^+-1}{k+1} + \varepsilon^+ + \alpha^+\right)\left(\frac{x^-}{k+1} - \varepsilon^+ + \alpha^-\right) + 2\left(\frac{x^++1}{k+1} + \varepsilon^+ - \alpha^+\right)\left(\frac{\ell - x^+}{k+1} - \varepsilon^+\right) \\
&+ 2\left(\frac{\ell - x^- - 1}{k+1} + \varepsilon^+\right)\left(\frac{x^-+2}{k+1} - \varepsilon^+ - \alpha^-\right).
\end{aligned}$$

Let $h = \frac{1}{k+1}$. We now look to maximise the function

$$\begin{aligned}
f(x^+, x^-, \alpha^+, \alpha^-, \varepsilon) = &\left(\frac{x^+-1}{k+1} + \varepsilon^+ + \alpha^+\right)\left(\frac{x^-}{k+1} - \varepsilon^+ + \alpha^-\right) + 2\left(\frac{x^++1}{k+1} + \varepsilon^+ - \alpha^+\right)\left(\frac{\ell - x^+}{k+1} - \varepsilon^+\right) \\
&+ 2\left(\frac{\ell - x^- - 1}{k+1} + \varepsilon^+\right)\left(\frac{x^-+2}{k+1} - \varepsilon^+ - \alpha^-\right)
\end{aligned}$$

Let $\gamma = x^+ - x^-$ and $\zeta = x^+ + x^-$. We can rewrite f as

$$\begin{aligned}
f(\gamma, \zeta, \alpha^+, \alpha^-, \varepsilon) = &\frac{1}{2}\zeta h\left(3\alpha^- + 3\alpha^+ + h(4\ell - 9)\right) + \alpha^-\left(\alpha^+ - 2h\ell + h - \varepsilon\right) + \frac{1}{2}\gamma h\left(\alpha^+ - \alpha^- + 5h - 10\varepsilon\right) \\
&+ \alpha^+(\varepsilon - 2h\ell) - \frac{5}{4}\gamma^2 h^2 - \frac{3\zeta^2 h^2}{4} + 6h^2\ell - 4h^2 + 5h\varepsilon - 5\varepsilon^2
\end{aligned}$$

We now treat f as a continuous function over $\mathbb{R}^5$, with only the constraints that $0 \leq \alpha^+, \alpha^- \leq h$. Since we are removing constraints, the maximum of f cannot be less than the max of $\pi(C)$. f is quadratic in its 5 variables, and so we proceed by sequential maximization.

Maximization over ε. For any fixed $(\zeta, \gamma, \alpha^+, \alpha^-)$, the function is quadratic in ε and $\frac{\partial^2 f}{\partial \varepsilon^2} = -10 < 0$. Hence f is strictly concave in ε. The first-order condition $\partial f / \partial \varepsilon = 0$ therefore yields the unique global maximizer $\varepsilon^*(\zeta, \gamma, \alpha^+, \alpha^-) = \frac{1}{10}\left(-\alpha^- + \alpha^+ + 5h(1-\gamma)\right)$.

Substituting $\varepsilon = \varepsilon^*$ into f yields the reduced objective, which is now independent of γ

$$
\begin{aligned}
f_1(\zeta, \alpha^+, \alpha^-) &:= f(\zeta, \gamma, \alpha^+, \alpha^-, \varepsilon^*) \\
&= \frac{1}{20}\Big((\alpha^-)^2 + (\alpha^+)^2 - 15\zeta^2 h^2 + \zeta h\left(10h(4\ell - 9)\right) + 18\alpha^+\alpha^- \\
&\quad + 10h(1 - 4\ell + 3\zeta)(\alpha^+ + \alpha^-) + 5h^2(24\ell - 11)\Big),
\end{aligned}
$$

Maximization of f_1 over ζ. The second derivative of f_1 with respect to ζ is

$$\frac{\partial^2 f_1}{\partial \zeta^2} = -\frac{3h^2}{2} < 0.$$

Since $h = \frac{1}{1+k} > 0$, this quantity is strictly negative. Therefore f_1 is strictly concave in ζ for every fixed $(\gamma, \alpha^+, \alpha^-)$.

The unique maximizer in ζ satisfies

$$\frac{\partial f_1}{\partial \zeta} = 0,$$

which yields

$$\zeta^*(\alpha^+, \alpha^-) = \frac{\alpha^- + \alpha^+}{h} + \frac{4\ell}{3} - 3.$$

Substituting $\zeta = \zeta^*$ into f_1 gives

$$f_2(\alpha^+, \alpha^-) := f_1(\zeta^*, \alpha^+, \alpha^-),$$

where

$$f_2(\alpha^+, \alpha^-) = \frac{4}{15}\left(-15h\left(\alpha^- + \alpha^+\right) + 9\alpha^-\alpha^+ + 3\left(\alpha^-\right)^2 + 3\left(\alpha^+\right)^2 + 5h^2\left(\ell^2 + 3\right)\right).$$

Thus the problem reduces to maximizing f_2 over (α^+, α^-) subject to the box constraints that $0 \le \alpha^+, \alpha^- \le h$.

Maximization of f_2 with respect to α^+ We have that $0 \le \alpha^+ \le h$ and $0 \le \alpha^- \le h$.

Since the constant factor $\frac{4}{15}$ does not affect maximization, we focus on the terms involving α^+:

$$g(\alpha^+) = 3(\alpha^+)^2 + (9\alpha^- - 15h)\alpha^+ + (\text{terms independent of } \alpha^+).$$

The second derivative is $\frac{\partial^2 g}{\partial (\alpha^+)^2} = 6 > 0$. Hence g is strictly convex in α^+. Therefore, there is no interior maximum; any maximizer must lie on the boundary of $[0, h]$. We compare f_2 at $\alpha^+ = 0$ and $\alpha^+ = h$. The difference is

$$f_2(h, \alpha^-) - f_2(0, \alpha^-) = \frac{4}{15}\Big(3h^2 + (9\alpha^- - 15h)h\Big) = \frac{4}{15}\left(9h\alpha^- - 12h^2\right) = \frac{12h}{15}\left(3\alpha^- - 4h\right) < 0,$$

since $0 \le \alpha^- \le h$, so that $3\alpha^- - 4h < 0$.

For every admissible $\alpha^- \in [0, h]$,

$$f_2(h, \alpha^-) < f_2(0, \alpha^-).$$

Hence the maximizer satisfies $\alpha^{+*} = 0$, and symmetrically, we find $\alpha^{-*} = 0$. So $f_2(0, 0) = \frac{4(\ell^2+3)}{3(k+1)^2}$

We now know that $\pi(C)/2 \le \frac{4(\ell^2+3)}{3(k+1)^2}$ and thus $\pi(C) \le \frac{8(\ell^2+3)}{3(k+1)^2}$.

We now proceed by cases:

Case 1: $\ell = 1$. We can use the result from Lemma 8, since we need the probability that a voter approves of c and no candidates from $\widehat{W}$ which is the same as when working with PJR+ with $\ell = 1$. We get $\pi(C) \le \frac{\ell}{k+1} \le \frac{4\ell}{3k} - \frac{1}{2k^2}$.

Case 2: $2 \le \ell (\le k)$ **and** $\ell \ge \frac{15k}{32}$**.** Then we have $\Pr[c \in A(v)] \le \frac{1}{2} \le \frac{4\ell}{3k} - \frac{1}{2k^2}$

Case 3: $2 \le k \le 5$ **and** $2 \le \ell < \frac{15k}{32}$**.** Note that we only have $2 < \frac{15k}{32}$ for $\ell \in [k]$ when $k \ge 5$, so we only need to focus on $k = 5$ and $\ell = 2$. Again, we have that $\Pr[c \in A(v)] \le \frac{1}{2} \le \frac{4\ell}{3k} - \frac{1}{2k^2}$.

Case 4: $k \ge 6$ **and** $2 \le \ell < \frac{15k}{32}$**.** Define $f(\ell) = \frac{4\ell}{3k} - \frac{1}{2k^2} - \frac{8(\ell^2+3)}{3(k+1)^2}$. f is quadratic in ℓ, and is concave. We'll show that $f(\ell) \ge 0$ at $\ell = 2$ and $\ell = \frac{15k}{32}$ to show that it is nonnegative on the interval $[2, \frac{15k}{32})$. We have $f(2) = \frac{8}{3k} - \frac{1}{2k^2} - \frac{56}{3(k+1)^2} \ge 0$ for $k \ge 6$, and $f(\frac{15k}{32}) = \frac{15}{24} - \frac{1}{2k^2} - \frac{225k^2+3072}{384(k+1)^2} \ge 0$ for $k \ge 6$. Hence $\frac{4\ell}{3k} - \frac{1}{2k^2} \ge \frac{8(\ell^2+3)}{3(k+1)^2}$.

So in any case, we get that $\pi(C) \le \frac{4\ell}{3k} - \frac{1}{2k^2}$.

We apply Fact 2 to $X = \sum_i X_{v_i}$ with threshold $\frac{4\ell n}{3k}$. As in Lemma 27, $\Pr[X_v] \le \pi(C) + 4\Delta \le \frac{4\ell}{3k} - \frac{1}{2k^2} + \frac{1}{4k^2} = \frac{4\ell}{3k} - \frac{1}{4k^2} =: p$. For $\ell > \frac{3k}{4}$ the threshold exceeds n, so $\Pr[s_{\ell,c} \ge \frac{4\ell n}{3k}] = 0$ trivially. For $\ell \le \frac{3k}{4}$ the gap $a = \frac{4\ell}{3k} - p \ge \frac{1}{4k^2}$ gives $\varepsilon^2 = n^2 a^2 \ge \frac{n^2}{16k^4}$. For the denominator, $2p(1-p) \le 2p \le \frac{8\ell}{3k}$ and $\frac{2}{3}a \le \frac{8\ell}{9k}$, so $2p(1-p) + \frac{2}{3}a \le \frac{32\ell}{9k} \le \frac{8}{3}$ (using $\ell \le \frac{3k}{4}$). Therefore

$$\frac{\varepsilon^2}{2np(1-p) + \frac{2}{3}\varepsilon} \ge \frac{n^2/(16k^4)}{n \cdot 8/3} = \frac{3n}{128k^4},$$

and $\Pr\left[s_{\ell,c} \ge \frac{4\ell n}{3k}\right] \le \exp\left(-\frac{3n}{128k^4}\right)$. By the union bound argument of Theorem 12, when $n = \Omega\left(k^4 \log\left(k \log m\right)\right)$, the expected number of queries is $\mathcal{O}(\log k)$. □

E PROOF OF THEOREM 13

Definition 13 (Generalized Interval Voter (GIV) model)**.** *A Generalized Interval Voter (GIV) model is a joint probability distribution $F_{Z_1,Z_2}(x,y)$ over two random variables $Z_1 \le Z_2 \in [0,1]$. Given a candidate set $C \subset [0,1]$, we say a voter v is sampled according to the model if her ballot is obtained by sampling Z_1, Z_2 jointly (possibly dependently) from F_{Z_1,Z_2}, and so that $A^*(v) = [Z_1, Z_2]$, $A(v) = A^*(v) \cap C$. We call $a_v = Z_1, b_v = Z_2$ for convenience, so $A^* = [a_v, b_v]$.*

Note that GIV generalises RIV: Given an RIV $\mathcal{M} = (F_t, p_t)_{t\in[\sigma]}$, we can define (for $x \le y$)

$$F_{Z_1,Z_2}(x,y) = \sum_{t\in[\sigma]} p_t \left(2F_t(x)F_t(y) - F_t(x)^2\right)$$

Theorem 13. *For a GIV, Algorithm 4 finds a 2-PJR+ committee. When the number of voters n satisfies $n = \Omega(k^3 \log(k \log m))$, the algorithm uses $\mathcal{O}(\log k)$ queries per voter in expectation and with high probability.*

The proof of Theorem 13 is based on a sequence of algorithms (Algorithms 4 and 5). Our guessing stage is the same as before, using $G(x) = \frac{F_{Z_1}(x)+F_{Z_2}(x)}{2}$, where F_{Z_1}, F_{Z_2} are the marginal cumulative distribution functions of the left and right endpoints respectively:

$$F_{Z_1}(x) = F_{Z_1,Z_2}(x,1)$$
$$F_{Z_2}(y) = F_{Z_1,Z_2}(1,y).$$

Given our guessed committee $\widehat{W}$, we verify whether it provides 2-PJR+. We perform the same queries as before. We check for every $\ell \in [k]$, every candidate $c \in C \setminus \widehat{W}$ and for every $0 \le j < \ell$ whether it could be consistent, given the information we have elicited from the voter, for the voter to approve of: (1) c, (2) at most j candidates in $\widehat{W}$ to the left of c, and (3) fewer than ℓ candidates in $\widehat{W}$. If this event has positive probability, captured by the `poss` function, we add the voter to the counter $s_{\ell,c,j}$. Finally, if $s_{\ell,c,j} < 2n\ell/k$ for all $\ell \in [k]$, $c \in C \setminus \widehat{W}$, $0 \le j < \ell$, we output $\widehat{W}$. Otherwise, we query all voters in V about all candidates in C to get full information about E, and then run MES to output a committee W^* that provides (2-)PJR+.

To prove Theorem 13, we use the same definitions as before. The quantities $\widehat{W}, s_{\ell,c,j}$ are defined in Algorithms 4 and 5 (in what follows, we consider the values of these variables at the termination of the respective algorithms). We will consider a fixed $0 \le j < \ell \le k$ and $c \in C \setminus \widehat{W}$ for convenience.

Algorithm 4: Finding a 2-PJR+ committee

input : A GIV F_{Z_1,Z_2}
$\Delta = (4k)^{-2}$ $G(x) = \frac{1}{2}(F_{Z_1}(x) + F_{Z_2}(x))$
for $i \in [k]$ **:**
  $c \leftarrow \arg\min_{c \in C \setminus \widehat{W}} \left|G(c) - \frac{i}{k+1}\right|$;
  $\widehat{W} \leftarrow \widehat{W} \cup \{c\}$;
Sort $\widehat{W}$;
$P \leftarrow \widehat{W} \cup \left\{ \left(G^{-1}(j\Delta)\right)^{+}_{C} : j \in [1/\Delta] \right\}$;
$P \leftarrow P \cup \left\{ \left(G^{-1}(j\Delta)\right)^{-}_{C} : j \in [1/\Delta] \right\}$;
for $v \in V$ **:** $\widehat{A}(v) \leftarrow \texttt{resolve}(v, C, P)$;
$\widehat{E} \leftarrow (V, P, k, \widehat{A})$;
if 2-validate *($\widehat{W}, \widehat{E}, C$) (Algorithm 5)* **: output :** $\widehat{W}$
for $v \in V$ **:** $A(v) \leftarrow \texttt{resolve}(v, C)$;
output : $W^* = \texttt{MES}((V, C, k, A))$

Algorithm 5: Validating if a committee $\widehat{W}$ provides 2-PJR+

Function 2-validate($\widehat{W}$, $\widehat{E}$, C)**:**
  for $\ell = 1$ *to* k, $c \in C \setminus \widehat{W}$ **:**
    $\rho \leftarrow |\{x \in \widehat{W} : x < c\}|$;
    for $j = 0$ *to* $\ell - 1$ **:**
      $R \leftarrow \widehat{W}[\rho - j : \rho + \ell - 1 - j]$;
      $s_{\ell,c,j} \leftarrow |\{v \in V : \texttt{poss}(v, c, \widehat{W} \setminus R)\}|$;
      if $s_{\ell,c,j} \geq \frac{2n\ell}{k}$ **: return** *False* ;
  return *True*;

Definition 14. *We shall interpret $\widehat{W}[s] = 0$ for $s < 0$, and $\widehat{W}[s] = 1$ for $s \geq k$. Set $\rho = |\{c' \in \widehat{W} : c' < c\}|$. For $0 \leq j < \ell \leq k$ and $c \in C \setminus \widehat{W}$, define the random event Y as*

$$c \in A(v) \text{ and } A(v) \cap \widehat{W} \subseteq \widehat{W}[\rho - j : \rho + \ell - 1 - j].$$

Let $w^- = \widehat{W}[\rho - j - 1]$ and $w^+ = \widehat{W}[\rho + \ell - 1 - j]$. Let Y^- be the event that $a_v \in (w^-, c]$ and Y^+ be the event such that $b_v \in [c, w^+)$, so that $Y = Y^+ \wedge Y^-$.

We now find a bound for $\Pr[Y]$ in terms of ℓ and k. Consider a fixed $\ell \in [k], c \in C \setminus \widehat{W}, 0 \leq j < \ell$. First, we shall show that $G(c) - G(w^-) \leq \frac{2j}{k+1} + 2\varepsilon$ and $G(w^+) - G(c) \leq \frac{2(\ell-j)}{k+1} - 2\varepsilon$ for some ε we shall define later, and so $G(w^+) - G(w^-) \leq \frac{2\ell}{k+1}$.

Lemma 28. $G(w^+) - G(w^-) \leq \frac{2\ell}{k+1}$.

Proof. Let μ_0^- be the marked point closest to the left of c. That is, $\mu_0^- = \frac{\alpha}{k+1}$ for $\alpha \in [k]$ such that $\frac{\alpha}{k+1} < G(c) < \frac{\alpha+1}{k+1}$ for some $\alpha \in [k]$. Let $\mu_s^- = \frac{\alpha - s}{k+1}$ be the $(s+1)$'th closest marked point to the left of c. First, we shall show that $G(c) - G(w^-) \leq \frac{2j}{k+1} + 2\varepsilon$ where $\varepsilon = G(c) - \mu_0^-$. Consider the selection process for $\widehat{W}$: for each marked point μ_s^-, we select a candidate $w_s \in \widehat{W}$ for μ_s^-, and in particular since $c \notin \widehat{W}$, we know that $|G(w_s) - \mu_s^-| \leq |G(c) - \mu_s^-|$. From this, we know that for any candidate $w \in \widehat{W}$, $w < c$ if and only if it is selected for some marked point $\mu_s^- \leq \mu_0^- < G(c)$. This means that the number of candidates of $\widehat{W}$ to the left of c is $|\{w \in \widehat{W} : w < c\}| = \alpha$. If $j \geq \alpha$, then we have that $w^- = 0$ and so we have $G(c) - G(w^-) = G(c) = \frac{\alpha}{k+1} + \varepsilon \leq \frac{2j}{k+1} + 2\varepsilon$.

Now, if $j < \alpha$, then since $|G(w_s) - \mu_s^-| \leq |G(c) - \mu_s^-|$, we can say that $G(w_s) \in [2\mu_s^- - G(c), G(c))$. Therefore we know that there exist at least $j+1$ candidates from $\widehat{W}$ in the interval $[2\mu_j^- - G(c), G(c))$, and thus $G(w^-) = G(\widehat{W}[\rho - j - 1]) \in$

$[2\mu_j^- - G(c), G(c))$. From this, we can say that $G(c) - G(w^-) \leq 2(G(c) - \mu_j^-) = 2(G(c) - \mu_0^- + \mu_0^- - \mu_j^-) = 2(\varepsilon + \frac{j}{k+1})$. We can argue similarly that $G(w^+) - G(c) \leq \frac{2(\ell-j-1)}{k+1} + 2(\frac{1}{k+1} - \varepsilon) = \frac{2(\ell-j)}{k+1} - 2\varepsilon$, and thus $G(w^+) - G(w^-) \leq \frac{2\ell}{k+1}$. $\square$

We apply this bound $G(w^+) - G(w^-) \leq \frac{2\ell}{k+1}$ to $\Pr[Y]$.

Lemma 29. $\Pr[Y] \leq \frac{2\ell}{k+1}$.

Proof. We know that $\Pr[Y] = \Pr[Y^+ \wedge Y^-] \leq \min(\Pr[Y^+], \Pr[Y^-]) \leq \frac{\Pr[Y^+] + \Pr[Y^-]}{2}$. Now, we have

$$\begin{aligned}
\frac{2\ell}{k+1} &\geq G(w^+) - G(w^-) \\
&= \frac{1}{2}\left(F_{Z_1}(w^+) - F_{Z_1}(w^-) + F_{Z_2}(w^+) - F_{Z_2}(w^-)\right) \\
&\geq \frac{1}{2}\left(F_{Z_1}(c) - F_{Z_1}(w^-) + F_{Z_2}(w^+) - F_{Z_2}(c)\right) \\
&= \frac{1}{2}(\Pr[Y^-] + \Pr[Y^+]) \\
&\geq \Pr[Y]
\end{aligned}$$

$\square$

Lemma 30. *For $\ell \in [k], c \in C \setminus \widehat{W}, 0 \leq j < \ell$, we have* $\Pr\left[s_{\ell,c,j} \geq \frac{2\ell n}{k}\right] \leq \exp\left(-\frac{27n\ell}{280k^3}\right)$.

Proof. Fix $\ell \in [k], c \in C \setminus \widehat{W}, 0 \leq j < \ell$. Let X_v be the event that $\texttt{poss}(v, c, \widehat{W} \setminus R)$ evaluates to true. By the same argument as Lemma 9, since $G = (F_{Z_1} + F_{Z_2})/2$ and F_{Z_2} is non-decreasing, $F_{Z_1}(q') - F_{Z_1}(c) \leq 2(G(q') - G(c)) \leq 2\Delta$ and symmetrically $F_{Z_2}(c) - F_{Z_2}(q) \leq 2\Delta$, giving $\Pr[X_v] \leq \frac{2\ell}{k+1} + 4\Delta$. Let $X = \sum_{i=1}^{n} X_{v_i}$. The same argument as Lemma 27 applies with $p = \frac{2\ell}{k+1} + 4\Delta$, $a = \frac{2\ell}{k(k+1)} - 4\Delta \geq \frac{3\ell}{4k^2}$, $\sigma^2 \leq p \leq \frac{9\ell}{4k}$, and $t = na$. By Fact 2:

$$\Pr\left[X \geq \frac{2\ell n}{k}\right] \leq \exp\left(-\frac{t^2/2}{n\sigma^2 + t/3}\right) \leq \exp\left(-\frac{27n\ell}{280k^3}\right).$$

$\square$

We can now prove Theorem 13.

Proof of Theorem 13. First, the output of Algorithm 4 always provides 2-PJR+: if $\widehat{W}$ is output by the algorithm, the voters in any 2-large cohesive group would be counted by some $s_{\ell,c,j}$, and otherwise W^* is output and satisfies PJR+.

We bound $s_{\ell,c,j}$ with Lemma 30. By Lemma 10 (which holds for the GIV by the same argument, since its proof uses only the structure of $\texttt{poss}$ and $\widehat{W} \subseteq P$), for any two $c, c' \in Q_i = (p_i, p_{i+1})$ between consecutive query points we have $X_{\ell,c,j} = X_{\ell,c',j}$; thus the union over c reduces to one term per interval. With $|P| \leq k + 32k^2 \leq 34k^2$, there are at most $68k^2$ cells (open intervals plus boundary candidates), giving at most $68k^4$ terms after a union over k^2 (ℓ, j) pairs, and

$$\Pr\left[\bigcup_{\ell,j,c} X_{\ell,c,j}\right] \leq 68k^4 \exp\left(-\frac{27n}{280k^3}\right).$$

Hence, if $n = \Omega\left(k^3 \log\left(k^4 \log m\right)\right) = \Omega\left(k^3 \log(k \log m)\right)$, and since $|P| = \mathcal{O}(k^2)$, the expected number of queries is $\mathbf{E}(Q) \leq 68k^4 \log m \exp\left(-\frac{27n}{280k^3}\right) + \mathcal{O}(\log k) \leq \mathcal{O}(\log k)$. $\square$

Unfortunately, we cannot achieve a stronger form of PJR+ for arbitrary joint distributions.

Proposition 31. *For $\varepsilon > 0$, $\widehat{W}$ from Algorithm 4 does not provide $(2 - \varepsilon)$-PJR+ on some distribution and choice of C, k.*

Proof. Consider the cumulative probability distribution given by $F_{Z_1,Z_2}(x,y) = \min(2x,1)\max(2y-1,0)$. In words, Z_1 is drawn uniformly at random from $[0,1/2]$, and Z_2 is drawn uniformly at random, and independently of Z_1, from $[1/2,1]$. Then $F_{Z_1}(x) = \min(2x,1), F_{Z_2}(y) = \max(2y-1,0)$, and $G(x) = \frac{1}{2}(F_{Z_1}(x) + F_{Z_2}(x)) = x$ for $x \in [0,1]$. Let $k = 4t$ for some $t \in \mathbb{N}$ with $t \geq \varepsilon^{-1}$. The marked points are then $\{i/(4t+1) : i \in [k]\}$.

We construct our candidate set. We have a single candidate c^* at $1/2$. Then for each $t+1 \leq i \leq 3t$, set a candidate at the i'th marked point $i/(k+1)$. Finally, place t candidates arbitrarily within $X_0 = [0,\delta)$ and t candidates within $X_1 = (1-\delta,1]$, with $\delta = \varepsilon/100$. For each marked point $i/(k+1)$ with $t+1 \leq i \leq 3t$, it will select the candidate that sits at its position. For each other marked point, there will exist a candidate either within X_0 or X_1 which is closer to it than $c^* = 1/2$. Therefore c^* is not selected. Take $\ell = 2t+1$ and $j = t$. Then the $j+1$'th candidate left of c is in X_0, and then $\ell - j$'th candidate right of c is in X_1. Let Y be the event that the voter approves of c^*, $\leq j$ candidates in $\widehat{W}$ to the right of c, and $< \ell$ candidates in $\widehat{W}$ in total. We have $\Pr[Y] \geq (F_{Z_1}(1/2) - F_{Z_1}(\delta))(F_{Z_2}(1-\delta) - F_{Z_2}(1/2)) = (1-2\delta)^2 \geq 1-4\delta$. However, we have that $\frac{(2-\varepsilon)\ell}{k} = \frac{(2-\varepsilon)(2t+1)}{4t+1} \leq 1 - \varepsilon/4 < 1 - \varepsilon/25 = 1 - 4\delta \leq \Pr[Y]$.

Thus the probability that $s_{\ell,c,j} < \frac{(2-\varepsilon)n\ell}{k}$ is, by Fact 2 applied to the lower tail, with variance $p(1-p) \leq 4\delta = \varepsilon/25$ and gap $q = (1-4\delta) - (2-\varepsilon)\ell/k \geq \varepsilon/5$,

$$\Pr\left(\mathrm{Binom}\left(n, 1-4\delta\right) < (2-\varepsilon)n\ell/k\right) \leq \exp\left(-\frac{3n\varepsilon}{16}\right),$$

and thus for large enough n, $s_{\ell,c,j} \geq \frac{(2-\varepsilon)n\ell}{k}$ with high probability and we use $\mathcal{O}(\log m)$ queries in expectation. □

F ROBUSTNESS TO APPROXIMATE KNOWLEDGE OF G

We verify the claim in Remark 11: if the algorithm uses $\hat{G}$ with $\|\hat{G} - G\|_\infty \leq \varepsilon_G \leq \Delta/2$, then Theorem 5 holds with the same sample bound $n = \Omega\left(k^3 \log\left(k \log m\right)\right)$.

When $\hat{G}$ replaces G, the algorithm selects winners at $\hat{G}$-quantiles and places query points at $\hat{G}^{-1}(r\Delta)$.

Perturbed Lemma 7. Assume $\hat{G}$ is strictly increasing (i.e. a strictly monotone CDF approximation). Since $\mu_s^- < \hat{G}(c)$, the selection criterion $|\hat{G}(w_s) - \mu_s^-| \leq |\hat{G}(c) - \mu_s^-|$ gives $\hat{G}(w_s) \leq \hat{G}(c)$, hence $w_s \leq c$ (by monotonicity of $\hat{G}$), hence $G(w_s) \leq G(c)$ (by monotonicity of G). Moreover, $|G(w_s) - \mu_s^-| \leq |G(c) - \mu_s^-| + 2\varepsilon_G$, so $G(w_s) \geq 2\mu_s^- - G(c) - 2\varepsilon_G$. Thus there are at least $j+1$ winners in the G-interval $[2\mu_j^- - G(c) - 2\varepsilon_G,\, G(c))$, giving $G(c) - G(w^-) \leq 2\big(G(c) - \mu_j^-\big) + 2\varepsilon_G$. Analogously, $G(w^+) - G(c) \leq 2\big(\mu_{\ell-j}^+ - G(c)\big) + 2\varepsilon_G$, and summing: $G(w^+) - G(w^-) \leq 2\big(\mu_{\ell-j}^+ - \mu_j^-\big) + 4\varepsilon_G = 2\ell/(k+1) + 4\varepsilon_G$.

Perturbed Lemma 8. By the AM-GM step ($d_t^+ d_t^- \leq (d_t^+ + d_t^-)/4$), $\Pr[Y] \leq \frac{1}{2}(G(w^+) - G(w^-)) \leq \ell/(k+1) + 2\varepsilon_G$.

Perturbed Lemma 9. Consecutive query points $\hat{G}^{-1}(r\Delta)$ and $\hat{G}^{-1}((r+1)\Delta)$ enclose a G-interval of mass at most $\Delta + 2\varepsilon_G$. Hence $\Pr[X_v \wedge \neg Y] \leq 2(\Delta + 2\varepsilon_G)$, giving $\Pr[X_v] \leq \ell/(k+1) + 2\Delta + 6\varepsilon_G$. The validation gap is $\ell/k - \Pr[X_v] \geq \ell/\big(k(k+1)\big) - 2\Delta - 6\varepsilon_G \geq \ell/\big(k(k+1)\big) - 5\Delta > 0$, where the last inequality uses $\varepsilon_G \leq \Delta/2$ and the positivity follows for all $\ell \geq 1, k \geq 1$ since $\ell/(k(k+1)) > 5/(16k^2)$.

The Bernstein bound of Lemma 9 and the union bound of Theorem 5 then apply unchanged, yielding the same failure probability $O(k^4 \exp(-\Omega(n/k^3)))$ and sample bound $n = \Omega\left(k^3 \log\left(k \log m\right)\right)$.

G DISCUSSION OF MULTIDIMENSIONAL MODELS

It is natural to ask if our results can be extended beyond one dimension. Towards this goal, we propose the d-dimensional Random Euclidean Voter Model, which is a generalization of RIV.

Definition 15. *Every bounded interval I of $\mathbb{R}$ together with a finite set $C \subset I^d$, a probability distribution $\mathcal{D}$ over convex subsets of I^d, and $n \in \mathbb{N}$ define a d-*dimensional Random Euclidean Voter Model (d-REV) *as follows: we draw n samples $A_1^*, \ldots, A_n^* \sim \mathcal{D}$, let $V = [n]$, and define the approval set of voter $v \in V$ as $A(v) = A_v^* \cap C$. We define a new type of query called a* PLANAR *query: Given a candidate $\mathbf{c} \in C$, a vector $\mathbf{p} \in \mathbb{R}^d$, a direction $\preceq \in \{\leq, \geq\}$ and a voter $v \in V$, the query* PLANAR$(c, \mathbf{x}, \preceq, v)$ *computes the half space defined by $H = \{\mathbf{x} \in I^d : \mathbf{p} \cdot \mathbf{x} \preceq \mathbf{p} \cdot \mathbf{c}\}$ and returns $\mathbb{1}[A^*(v) \subseteq H]$.*

We note that PLANAR queries in the d-REV model correspond exactly to INT queries in the RIV model. In general, voters' approval sets may have complex geometries, so that communicating (a description of) a voter's approval set may require a significant amount of information. However, PLANAR queries enable efficient communication; each query consists of two d-dimensional vectors and a direction, and the voter's response is always a single bit.

The spatial model with $d > 1$ is more expressive than the RIV model; in particular, it can capture important aspects of participatory budgeting. Indeed, consider a 2-REV model representing the geographical layout of a city, where each project is associated with a particular location, and voters approve of all projects that lie within some area around their residence. We are able to query voters using PLANAR queries to ask a voter whether all of their approved projects lie half space on the map. We hope that our investigation of the 1-dimensional model may provide useful insights into how to tackle higher-dimensional settings.

Springham et al. [2026] pursue exactly this direction, developing a general theory of committee-selection rules ("W-selection modules") for the d-REV model that subsumes the present results and closes several of their gaps. Relative to the present paper, that work (i) targets general joint distributions over voters' approval regions rather than the structured RIV mixture considered here; (ii) attains exact EJR+ rather than the 2-EJR+ (for RIV, Theorem 12) and 2-PJR+ (for general one-dimensional distributions, Theorem 13) guarantees obtained here; and (iii) does not require the distribution to be known exactly. In the one-dimensional case, their construction specializes to a rule that closely mirrors Algorithm 2: both place committee members near equally spaced quantile positions, but where Algorithm 2 selects the candidates closest to the marked points under the combined mixture CDF G, their rule instead quantizes the left-endpoint marginal CDF alone — a change that is enough to carry essentially the same construction from PJR+ to exact EJR+.