Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 135 (2007) 196–213
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

On the learnability of majority rule

Yuval Salant

Stanford Graduate School of Business, 518 Memorial Way, Stanford, CA 94305-5015, USA

Received 9 August 2005; final version received 23 March 2006
Available online 30 June 2006

Abstract

We establish how large a sample of past decisions is required to predict future decisions of a committee with few members. The committee uses majority rule to choose between pairs of alternatives. Each member's vote is derived from a linear ordering over all the alternatives. We prove that there are cases in which an observer cannot predict *precisely* any decision of a committee based on its past decisions. Nonetheless, *approximate* prediction is possible after observing relatively few random past decisions.

JEL classification: D71; D83

Keywords: Social choice; Learning; Majority rule; Committees; Tournaments; Choice functions

1. Introduction

This paper establishes how large a sample of past decisions is required to forecast future decisions of a social institution that chooses between pairs of alternatives via simple majority rule. We first show that there are cases in which an observer cannot *exactly* forecast any future decision of an institution based on its past decisions. We then show that *approximate* forecasting is possible after observing relatively few decisions, provided the institution has few members. Rubinstein [12] and Kalai [6] establish the basic information requirements for learning rational choice. Our results extend their analysis to an important form of social choice.

The standard social choice model assumes that each member of a group has a rational preference relation (i.e., complete and transitive) over a finite set of alternatives. The model then applies an aggregation rule to formulate the choice rule of the group. One of the most popular aggregation rules is simple majority rule. The group chooses alternative a over alternative b if more than half of the group members prefer a to b. We refer to a group choosing between pairs of alternatives

E-mail address: salant@stanford.edu.

0022-0531/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jet.2006.03.012

via simple majority as a *committee*. Committees are widely observed in practice; Legislatures, courts, juries, boards of directors and many other institutions decide according to simple majority. In addition to explicit voting procedures, many groups such as households, cartels and computer networks use aggregation rules that resemble simple majority. Because committees are so popular, and their decisions often very influential, one may be interested in predicting future choices of committees based on their past choices.

Learning a committee's choices is more difficult than learning rational choice. As pointed out by Condorcet [2], a committee's choices may be intransitive even in the case of a three-member committee. Namely, unlike the case of rational choice, if one learns that the committee chooses a over b and b over c, one cannot infer that the committee necessarily chooses a over c. Arrow [1] shows that this discouraging property is inherent to every aggregation rule that satisfies a few desirable conditions. Maskin [9] and Dasgupta and Maskin [4] depart from Arrow's analysis and show that a majority rule results in transitive choices on a larger domain of individual preference profiles than any other rule satisfying slightly stronger conditions than Arrow's. This result, together with other results in the literature (e.g., [10]), provides a possible explanation for the popularity of majority rule in practice, and motivates further study of the properties of majority rule.

We study whether choice functions of small committees are learnable. [1] We first investigate this question in the context of three-member committees, and then extend the results to larger committees and decisive societies. While focused on learning, our results also establish the basic information requirements for econometric studies of a committee's decisions. As such, they help advance empirical and experimental analysis of committees in economics and other political and social sciences.

We analyze two-staged learning procedures. First, an observer sees examples of the committee's choices. An example is a pair of alternatives and the chosen element from this pair. Then, the observer formulates a hypothesis intended to predict future choices of the committee. We distinguish between two types of learning according to the desired quality of prediction. In *exact* learning, the learner's goal is to predict future choices of the committee with certainty. In *approximate* learning, the learner's goal is to predict future choices with high accuracy.

The first notion we examine is exact learning. Following Rubinstein [12], we consider a model in which the committee wishes to communicate its choice function to a student using the minimal possible number of examples. This model is appealing because it assumes that examples are communicated optimally to a learner. Hence, the number of examples needed for learning in this model serves as a lower bound to the number of examples needed for exact learning in other models, in which the examples are picked by the learner or generated by some random process.

In Section 2 we show that there exist cases in which a committee cannot communicate its choice function to a student without describing all of its choices. The intuition is straightforward. Suppose there are five elements, numbered 1–5, and that the committee's choice function is induced by the rational preference relation $1 \succ 2 \succ 3 \succ 4 \succ 5$. Assume also that the committee communicates to the student how it chooses between all pairs of elements except for the choice between 1 and 5, which appears to be the easiest to deduce. The learner cannot deduce with certainty that the committee chooses 1 over 5; a committee with members' preference relations $1 \succ 2 \succ 3 \succ 4 \succ 5$, $5 \succ 1 \succ 2 \succ 3 \succ 4$, and $2 \succ 3 \succ 4 \succ 5 \succ 1$ agrees with all the examples provided, yet chooses 5 over 1.

[1] A *choice function* assigns to every pair of elements a chosen element from the pair.

This negative result about exact learning motivates the study of the weaker concept of Probably Approximately Correct (PAC) Learning (see [7,15]). In the PAC model, examples of the committee's choices are revealed to an observer randomly and independently, according to some fixed probability measure over the choice pairs. After seeing the sample set, the observer has to formulate a hypothesis that will enable him to predict future choices of the committee with high probability (with respect to the same measure). Thus, in the PAC model the examples are drawn *at random* (instead of *optimally*), but the learner has to predict only *most* of the future choices (instead of *all* of them).

In Section 3 we show that the choices of three-member committees are PAC-learnable from a number of examples that is linear in the number of alternatives N, and that asymptotically fewer examples do not suffice. Kalai [6] shows that rational choice is PAC-learnable from $O(N)$ examples.[2] Our result implies that an asymptotically similar number of examples suffices for three-member committees.

We discuss larger committees with $r>3$ members in Section 4. We show that if the number of members r is relatively small, the choices of the committee are still PAC-learnable from $f(r, N)\cdot N$ examples, where $A_1 \cdot r \leqslant f(r, N) \leqslant A_2 \cdot \min\{r^2 \log_2 r, r \log_2 N\}$, and A_1 and A_2 are constants.

McGarvey [11] shows that any asymmetric binary relation can be induced by a majority vote of a large committee. Hence, PAC-learning the choices of large committees requires large samples. Our results support this claim by indicating that the number of examples needed for PAC-learning increases at least linearly in the number of members on the committee. Nonetheless, there are interesting cases in which the choices of large committees are still PAC-learnable, and we consider one such example in Section 4. A *society* is a committee with potentially many members. An α-*decisive society* is a society in which every choice is α-decisive, i.e., at least a fraction of $(\frac{1}{2}+\alpha)$ of the society's members (not necessarily the same members) agree with every decision. We show that the choices of α-decisive societies are PAC-learnable from at most $f(\alpha)\cdot N$ examples, where $f(\alpha) = O((\frac{1}{\alpha})^2 \log_2 \frac{1}{\alpha})$. Thus, if a society is decisive, it is much easier to PAC-learn its choices regardless of the number of members in the society.

2. Exact learning

Suppose that Alice wants to communicate a choice function to Bob. Bob knows the family to which the choice function belongs (e.g., it is induced by a three-member committee), but does not initially know the choice function. Knowledge is communicated via examples. An example is a pair of elements and the chosen element from this pair. Alice selects examples and communicates them to Bob. Bob's task is to deduce the entire choice function from the examples. Generating examples, communicating them, and deducing from them is costly. Thus, Alice and Bob want the number of examples to be as small as possible. Describability is the minimal number of examples needed to describe any choice function in the family.

The notion of describability was introduced by Rubinstein [12], who seeks "to explain the fact that certain properties of binary relations are frequently observed in natural language." One of the features Rubinstein investigates is the describability of a relation, i.e., the ease with which the relation can be described by means of examples. We find this notion appealing for two reasons. First, describability is an intuitive measure of supervised exact learning, in which an instructor guides a student through the learning process. Second, describability is a "first best" notion

[2] We use the following notation throughout the paper. Let $f, g : \mathbb{N} \longrightarrow \mathbb{R}_+$. We write $f(n) = O(g(n))$ if there is a constant $A > 0$ such that $f(n) \leqslant A \cdot g(n)$ for every n.

in the sense that it assumes that examples are communicated *optimally* to the learner. That is, describability serves as a lower bound on the number of examples needed for exact learning in other scenarios, in which the examples are picked by the learner or generated by some random process.

2.1. Definitions

Let $X = \{x_1, \ldots, x_N\}$ be a finite set of N elements. Let $Y = \{(x_i, x_j) : i < j\}$ be the collection of pairs of distinct elements of X. The set Y contains $\binom{N}{2} = \frac{N(N-1)}{2}$ pairs. A *choice function* $c : Y \to X$ assigns to every choice problem $(x_i, x_j) \in Y$ an element $c(x_i, x_j) \in \{x_i, x_j\}$. In other words, a choice function is a *tournament*, i.e., a complete asymmetric binary relation, on X. A choice function is *rational* if it satisfies transitivity. We identify a rational choice function with the linear ordering it induces on X.

We explore the learnability of committees' choice functions. A *committee* is a collection of r members, where $r \geqslant 3$ is an odd integer. Every member of the committee has a linear ordering on X. For every pair of elements x_i and x_j, the committee chooses x_i over x_j if more than half of the committee's members rank x_i higher than x_j.[3] We denote by *rMaj* the family of all choice functions of r-member committees.

Our benchmark measure of exact learning is describability. Let C be a family of choice functions. The *describability* of C is the minimal integer k such that every choice function in C is uniquely determined by k examples or less. Formally,

Definition 2.1. The describability of a family of choice functions C is $desc(C) = \max_{c \in C}\{d_C(c)\}$ where $d_C(c)$ denotes the minimal integer m such that there exist m pairs, $y_1, y_2, \ldots, y_m \in Y$, which obey the following:

$$\text{if } c' \in C \quad \text{and} \quad c'(y_i) = c(y_i) \quad \text{for all } i = 1, 2, \ldots, m \quad \text{then } c' = c.$$

For example, Rubinstein [12] shows that the describability of the family of all linear orderings is $N - 1$. Indeed, any linear ordering of the form $x_1 \succ x_2 \succ x_3 \succ \ldots \succ x_{N-1} \succ x_N$ can be described by the examples "x_i is chosen over x_{i+1}" for $1 \leqslant i \leqslant N - 1$. On the other hand, a linear ordering cannot be described by less than $N - 1$ examples, because one cannot deduce the order between two elements that are never chosen in the examples.

Definition 2.1 implies that for every two families of tournaments (i.e., choice functions), C_1 and C_2, if $C_1 \subseteq C_2$, then $desc(C_1) \leqslant desc(C_2)$. Thus, as a family of tournaments expands, it becomes weakly more difficult to describe the tournaments in the family. Moreover, the describability of the family of all tournaments on N alternatives is $\binom{N}{2}$, because if even one example is missing we can always find two tournaments that agree on all the examples given, but disagree on the missing example. These two observations imply that for any family of tournaments C, $desc(C) \leqslant \binom{N}{2}$.

2.2. Three-member committees

We now establish the describability of the family of three-member committees. The family $3Maj$ contains a relatively small number of tournaments (at most $(N!)^3$) in comparison to the

[3] Since the number of members is odd and they each have a linear ordering on X, the committee's choice function is well-defined.

total number of tournaments on X, which is $2^{\binom{N}{2}}$. One might expect the describability of $3Maj$ to be approximately similar to the describability of the family of linear orderings. However,

Proposition 2.2. *The describability of* $3Maj$ *over a set* X *of* N *elements is* $\binom{N}{2}$.

Proof. Consider the family $C = C_1 \cup C_2$, where C_1 is the family of all tournaments induced by linear orderings, and C_2 is the family of all tournaments that deviate from some linear ordering in exactly one pair. The describability of C is $\binom{N}{2}$. Indeed, any tournament $c_1 \in C_1$ cannot be described by less than $\binom{N}{2}$ examples, because for every set of $\binom{N}{2} - 1$ examples used to describe c_1, there is a tournament $c_2 \in C_2$ that agrees with this set of examples and still disagrees with c_1 on the missing example.

Moreover, $C \subset 3Maj$. Indeed, $C_1 \subset 3Maj$, because we can replicate any linear ordering three times and receive a tournament in $3Maj$. $C_2 \subset 3Maj$, because we can generate any tournament with one deviation from a linear ordering as a majority vote of three linear orderings. Without loss of generality, we illustrate this for a tournament which is consistent with the linear ordering $x_1 \succ \ldots \succ x_i \succ \ldots \succ x_j \succ \ldots \succ x_N$ except for one deviation $x_j \succ x_i$. This tournament can be obtained as a majority vote of the following three linear orderings:

$$\begin{aligned} &x_1 \succ \ldots \succ x_i \succ \ldots \succ x_j \succ \ldots \succ x_N,\\ &x_j \succ x_i \succ x_1 \succ \ldots \succ x_N,\\ &x_1 \succ \ldots \succ x_N \succ x_j \succ x_i. \end{aligned}$$

Consequently, we get that $\binom{N}{2} = desc(C) \leqslant desc(3Maj) \leqslant \binom{N}{2}$. That is, $desc(3Maj) = \binom{N}{2}$. □

Proposition 2.2 extends to larger committees. Since $3Maj \subseteq rMaj$ for any odd integer $r \geqslant 3$, we get that $desc(rMaj) = \binom{N}{2}$ as well.

Describability refers to *supervised* exact learning, in which a teacher provides both the questions and the answers. One can also think of scenarios of *independent* exact learning, in which the student has to figure out by himself the "right" questions to ask and the teacher only provides the answers.[4] For example, a graduate student who wants to learn which research questions are interesting, repeatedly presents various research topics to his advisors (or thesis committee) who point out the most interesting one among them. In our context, the independent exact learning problem is formulated as follows: How many questions does a student need to ask a teacher in order to learn a choice function in $3Maj$? Proposition 2.2 implies the following.

Corollary 2.3. *An independent learner who wants to discover a choice function of a three-member committee needs to ask* $\binom{N}{2}$ *questions in the worst case.*

Suppose that the learner has already learned the choice function of the committee over the N alternatives, and that a new alternative becomes available to the committee members. Assuming

[4] Independent exact learning of linear orderings is extensively discussed in the computer science literature under the title of comparison-based *sorting* algorithms. See [3,8] for details.

that the new alternative does not alter the relations between the N incumbent alternatives, the learner's task is to learn the relation between the new alternative and the N incumbent ones. Of course, N queries or examples will suffice to do so. Proposition 2.4 suggests that one cannot do any better.

Proposition 2.4. *A teacher who wants to communicate to a student how to add a new element z to a tournament in 3Maj must use N examples in the worst case.*

Proof. Consider a committee that has a linear ordering $x_1 \succ x_2 \ldots \succ x_N$ on X. Assume that the new element z is located somewhere within the ordering. Specifically, the committee's new linear ordering is $x_1 \succ \ldots \succ x_i \succ z \succ x_{i+1} \succ \ldots \succ x_N$. A teacher cannot communicate this fact to a student without describing all the relations between z and the elements of X.

Indeed, assume that the teacher communicates to the student the relations between z and all the elements of X except for one arbitrary element x_j. Without loss of generality, assume that the committee chooses x_j over z. The student cannot deduce the relation between x_j and z. A committee with members' preference relations

$$x_1 \succ \ldots \succ x_i \succ z \succ x_{i+1} \succ \ldots \succ x_N,$$
$$z \succ x_j \succ x_1 \succ \ldots \succ x_N,$$
$$x_1 \succ \ldots \succ x_N \succ z \succ x_j,$$

where x_j's location in the first ordering is the same as in the committee's true linear ordering, agrees on all the examples provided yet chooses z over x_j. □

Note the difference from a scenario in which one restricts attention to the family of linear orderings. In this case, a teacher has to communicate at most two examples to a student in order to describe how to add z to a known linear ordering on X. It suffices to communicate the relations between z and its immediate predecessor, and z and its immediate successor, when they exist.

2.3. Economic interpretations

The results about exact learning suggest that in learning "aggregated choice" there is a large gap between a situation in which learning is based only on observing the committee's choices and a situation in which learning is also based on observing the choices of individual committee members. Namely, if a teacher can communicate the choices of individual committee members to a student, then the student can learn the committee's choice function from at most $3(N-1)$ examples. However, if one has access only to the choices of the committee, then in the worst case one cannot learn the committee's choice function before seeing all of its choices.

Proposition 2.2 and Corollary 2.3 imply another result. Suppose that an observer does not care about learning the committee's choices, and only wishes to verify that the choices of the committee are transitive. As can be inferred from the proof of Proposition 2.2, he has no way of doing so without seeing all the committee's choices. Moreover, suppose that an observer knows that the committee's choices are transitive, and only wants to verify that they remain so after a new alternative z becomes available. The proof of Proposition 2.4 implies that he cannot do so before seeing the relation between z and all the other elements.

While our results are phrased in the context of social choice, they apply to learning individual choice as well. Instead of an r-member committee, one can think of a single decision maker (DM)

with r criteria according to which she ranks the alternatives. The DM chooses x_i over x_j, if x_i is ranked higher than x_j in more than half of the criteria. Proposition 2.2 and Corollary 2.3 then imply that it is much easier to exactly learn a DM's choices when one can identify the different criteria the DM uses and learn about each of them separately, as opposed to a case in which one observes only the DM's choices.

3. Probably approximately correct learning

The above results about exact learning motivate our study of a weaker concept of learning, called probably approximately correct learning (henceforth, PAC-learning). Kearns and Vazirani [7] and Vidyasagar [15] provide a detailed analysis of PAC-learning. In the PAC model, a sample set of the choices of a three-member committee is revealed to an observer randomly according to some probability measure on all the choice pairs. The observer's task is to predict approximately the choices of the committee. That is, given the sample set, the observer should predict future choices of the committee with high accuracy. Note the difference from the describability notion, where the sample set is chosen optimally (and not according to some probability measure), and where we demand exact prediction of future choices (and not prediction with high accuracy). [5]

3.1. Definitions

Let C be a family of Boolean functions from an instance space Z to $\{0, 1\}$. We assume that C is known, and we want to learn a specific target function $c \in C$. Note that choice functions are Boolean functions over the set Y of all choice pairs, if we interpret $c(x_i, x_j) = 0$ as implying that x_i is chosen over x_j. Let $\mathcal{P}$ be a probability measure over Z. The measure $\mathcal{P}$ provides a natural measure of error between any function $h \in C$ and c. Namely, we define $error_c(h) = Pr[x \in Z : c(x) \neq h(x)]$. Let $0 < \varepsilon, \delta < 1/2$.

Definition 3.1. A family of Boolean functions C is PAC-learnable from t examples with confidence $1 - \delta$ and accuracy $1 - \varepsilon$ with respect to $\mathcal{P}$ if:

For every $c \in C$, if $z_1, \ldots, z_t$ are drawn at random and independently according to $\mathcal{P}$, then with probability at least $1 - \delta$:

$$\text{if } h \in C \text{ satisfies } h(z_i) = c(z_i) \quad \text{for } i = 1, \ldots, t \text{ then } \text{error}_c(h) \leqslant \varepsilon.$$

If this holds for every measure $\mathcal{P}$, then we say that C is PAC-learnable from t examples with confidence $1 - \delta$ and accuracy $1 - \varepsilon$.

Fig. 1 provides graphical intuition. On the right side, the large oval represents a family of functions C. A particular function $c \in C$ is represented by a point. The probability measure $\mathcal{P}$ induces a distance function (or an error measure) on C. The small grey oval includes all the functions whose distance from c is $\leqslant \varepsilon$. The probability measure $\mathcal{P}$ also induces a probability measure over samples of t examples. On the left side, these samples are classified into "good" and "bad" samples. A sample is good if any $h \in C$ that agrees with the sample lies in the grey

[5] See Kalai [6] for a discussion on PAC-learnability and describability.

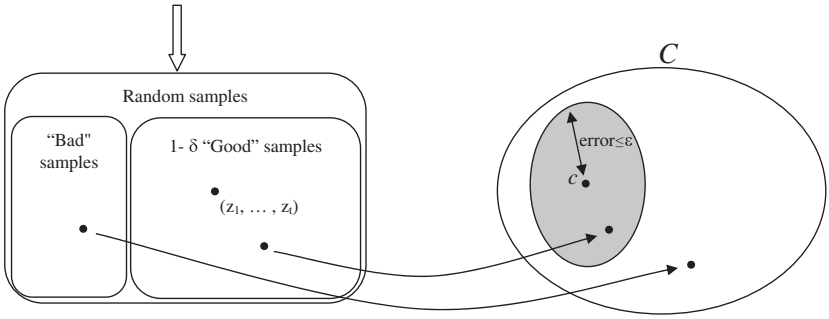

Fig. 1. PAC-learning.

oval around c. The family C is PAC-learnable from t examples if for every $c \in C$, the proportion (w.r.t. to the probability measure $\mathcal{P}$) of good samples is at least $1 - \delta$.

Thus, if a family of functions C is PAC-learnable from t examples, then with high probability, after seeing a random sample of t examples of some function $c \in C$, *any* function $h \in C$ that "agrees" with the examples will predict a large proportion of the values of c; Hence the name probably approximately correct learning.

Learning in the PAC model is susceptible to two kinds of failure. The confidence parameter δ is necessary since a random sample may be "unrepresentative" of the underlying function one wants to learn. For example, the sample might include repeated draws of the same example despite the fact that $\mathcal{P}$ is a uniform measure. The accuracy parameter ε is necessary since a small random sample may not distinguish between two functions that differ on only a few examples.

A fundamental aspect of PAC-learnability is the number of examples needed to learn a family of functions C. This number is closely connected to the notion of the Vapnik–Chervonenkis dimension. More specifically, let $S = \{s_1, s_2, \ldots, s_m\} \subseteq Z$, and denote by

$$\Pi_C(S) = \{(c(s_1), c(s_2), \ldots, c(s_m)) : c \in C\} \subseteq \{0, 1\}^m$$

the set of all the configurations of S that are realized by C. If $\Pi_C(S) = \{0, 1\}^m$ then we say that C *attains* S. Thus, C attains S if C realizes all the possible configurations of S.

Definition 3.2. The Vapnik–Chervonenkis dimension of C, denoted as $VCD(C)$, is the cardinality d of the largest set $S = \{s_1, s_2, \ldots, s_d\}$ attained by C. If C attains arbitrarily large finite sets then $VCD(C) = \infty$.

The definition implies three important things. First, it follows from the definition that $s_1, s_2, \ldots, s_d$ must be distinct. Second, in order to prove that $VCD(C)$ is at least d, one has to find *some* attained set of size d. Third, in order to prove that $VCD(C)$ is at most d, one has to show that no set of size $d + 1$ is attained by C.

For example, let $X = \{x_1, x_2, x_3, x_4\}$, and let c be the choice function induced by the linear ordering $x_1 \succ x_2 \succ x_3 \succ x_4$. Consider the family C of all choice functions that "agree" with c on all pairs $(x_i, x_j) \in Y$ except for at most two pairs. The VC-dimension of C is two. Indeed, C attains any two pairs in Y because we allow for two "deviations" from c. However, no three pairs are attained by C because this would imply that there is a function in the family that disagrees

with c on at least three pairs (the function that chooses the second element from every pair.) This example can be generalized as follows.

Example 3.3. Let c be a rational choice function. Let C_K be the family of all choice functions that agree with c on all $y \in Y$ except for at most K arbitrary pairs. Then, $VCD(C_K) = K$.

Example 3.3 suggests that as the number of allowed "deviations" from c increases, $VCD(C_K)$ increases. Intuitively, one might also argue that as the number of deviations increases, C_K becomes more "complex" and hence more difficult to learn. The tight connection between PAC-learning and the VC-dimension is established in the following theorem.

Theorem 3.4. *For fixed values of δ and ε, the number of examples needed to PAC-learn a family of Boolean functions with confidence $1-\delta$ and accuracy $1-\varepsilon$ is bounded above and below by linear functions of the VC-dimension.* [6]

Thus, in order to evaluate how many examples are needed to learn a family of functions C, it is enough to investigate the VC-dimension of the family. A simple observation is that if the VC-dimension of C is d, then C must contain at least 2^d functions (otherwise, it would be impossible to attain a set of size d).

Proposition 3.5. *Let C be a family of Boolean functions. Then, $VCD(C) \leqslant \log_2 |C|$.*

The following theorem, which was proved independently by Sauer [13], and Shelah and Perles [14], provides another connection between $VCD(C)$ and the number of functions in C.

Theorem 3.6. *Let C be a family of Boolean functions from a space of m elements to $\{0,1\}$. If $VCD(C) \leqslant d$, then the number of functions in C is at most $g_d(m) = \sum_{i=0}^{d} \binom{m}{i}$ where $\binom{m}{i} = \frac{m!}{i!(m-i)!}$. Hence, if the number of functions in C is at least $g_d(m)+1$, then $VCD(C) \geqslant d+1$.*

3.2. PAC-learnability of 3Maj

We return now to the family of three-member committees. We first prove that $VCD(3Maj)$ is linear in the number of alternatives N. We then use Theorem 3.4 to conclude that for fixed values of δ and ε the family of three-member committees is PAC-learnable from $O(N)$ examples. Note that by using the result of Proposition 3.5 along with the fact that $|3Maj| < (N!)^3$, we get that $VCD(3Maj) < 3N \log_2 N$. We obtain an asymptotic improvement on this upper bound in Proposition 3.9.

We start by obtaining a lower bound on the VC-dimension.

Proposition 3.7. *The VC-dimension of 3Maj is at least $3(N-2)$.*

Proof. Let $X = \{x_1, x_2, \ldots, x_N\}$. In order to prove that $VCD(3Maj) \geqslant 3(N-2)$, we introduce a set of $3(N-2)$ choice pairs that $3Maj$ attains. First, we introduce an attained set of $3(N-3)$

[6] For further details about the connection between the VC-dimension and PAC-learning, and between the number of examples and δ and ε, see Kearns and Vazirani [7], Chapter 3.

pairs, and then we add three more pairs. The set of $3(N-3)$ pairs is separated into three types:

$$T_1: \ \forall\, 4 \leqslant j \leqslant N \ \ (x_1, x_j),$$

$$T_2: \ \forall\, 4 \leqslant j \leqslant N \ \ (x_2, x_j),$$

$$T_3: \ \forall\, 4 \leqslant j \leqslant N \ \ (x_3, x_j).$$

Given a configuration of choices from these pairs (i.e., a vector in $\{0,1\}^{3(N-3)}$), we construct a function c in $3Maj$ that realizes this configuration by introducing three orderings O_1, O_2, and O_3 that induce c. The idea is that the ordering O_i "takes care" of the T_i-pairs in the sense that the remaining two orderings disagree on these pairs and O_i resolves this disagreement according to the configuration. More formally, in O_i we place x_i above all the elements x_j, $4 \leqslant j \leqslant N$, for which $c(x_i, x_j)$ should be 0, and below all the elements x_j for which $c(x_i, x_j)$ should be 1. In the other two orderings, we place x_i once below all the other elements and once above all the other elements. Therefore, O_i determines the realization of the T_i examples, and this realization is consistent with the given configuration.

We now add three additional pairs (x_1, x_2), (x_1, x_3), and (x_2, x_3). The construction of the orderings O_i still leaves a few "degrees of freedom" that allow to realize all the configurations of the additional pairs. We distinguish between two cases.

Case 1: $c(x_1, x_3) = 0$. Then, the orderings are

$$O_1: \ldots \succ x_1 \succ \ldots \succ x_2, x_3,$$

$$O_2: x_3 \succ \ldots \succ x_2 \succ \ldots \succ x_1,$$

$$O_3: x_1, x_2 \succ \ldots \succ x_3 \succ \ldots .$$

Changing the order between x_2 and x_3 in O_1 and between x_1 and x_2 in O_3 allows us to realize all the configurations in which $c(x_1, x_3) = 0$.

Case 2: $c(x_1, x_3) = 1$. Then, the orderings are

$$O_1: x_2, x_3 \succ \ldots \succ x_1 \succ \ldots,$$

$$O_2: x_1 \succ \ldots \succ x_2 \succ \ldots \succ x_3,$$

$$O_3: \ldots \succ x_3 \succ \ldots \succ x_1, x_2.$$

Changing the order between x_2 and x_3 in O_1 and between x_1 and x_2 in O_3 allows us to realize all the configurations in which $c(x_1, x_3) = 1$.

This gives us a set of $3(N-3)+3 = 3(N-2)$ pairs that $3Maj$ attains, as required. □

We now prove an upper bound on $VCD(3Maj)$. We use the following proposition in the proof.

Proposition 3.8 (*Kalai [6]*). *The VC-dimension of the family of linear orderings is* $N-1$.

Note that Proposition 3.8 and Theorem 3.4 imply that for fixed values of δ and ε, the number of examples needed to PAC-learn the family of rational choice functions is linear in the number of alternatives N.

Proposition 3.9. *The VC-dimension of* $3Maj$ *is less than* $99N$.

Proof. Assume the VC-dimension of 3*Maj* is M. Then there are M pairs of elements, $y_1, y_2, \ldots, y_M \in Y$, such that every configuration of choices from these pairs is realized by a tournament in 3*Maj*. Thus, given a configuration of choices from the M pairs (i.e., a vector in $\{0,1\}^M$), there exist 3 linear orderings such that for every coordinate (or choice) of the configuration at least 2 of the 3 orderings "agree" with it. Consequently, there is one ordering (or more) that agrees with at least $\frac{2M}{3}$ coordinates of the configuration. In that case, we say that the ordering "covers" the configuration. What is the minimal number of different orderings needed to cover all the possible configurations of the M pairs? A single ordering can agree with $\binom{M}{i}$ configurations on $M-i$ coordinates (we take the configuration induced by the ordering, and we have $\binom{M}{i}$ options to choose the i coordinates that disagree with the ordering). Consequently, a single ordering can cover at most $\binom{M}{0}+\binom{M}{1}+\ldots+\binom{M}{M-\frac{2M}{3}}$ configurations. Therefore, as the total number of configurations is 2^M, the number of different orderings needed is at least

$$U=\frac{2^M}{\binom{M}{0}+\binom{M}{1}+\ldots+\binom{M}{M-\frac{2M}{3}}}\geqslant 2^{M(1-H(\frac{1}{3}))},$$

where $H(\lambda)=-\lambda\log_2\lambda-(1-\lambda)\log_2(1-\lambda)$, $0<\lambda<1$, is the binary entropy function, and the inequality is derived from Conclusion A.3 in Appendix A.1.

Let us think of these U orderings as U different vectors in $\{0,1\}^M$, where we identify each ordering with the configuration it induces. According to Theorem 3.6, if the number of vectors exceeds $g_{N-1}(M)=\sum_{i=0}^{N-1}\binom{M}{i}$, then the VC-dimension of these linear orderings (and, consequently, the VC-dimension of all linear orderings) is at least N. This is impossible due to Proposition 3.8. Therefore, we get that

$$2^{M(1-H(\frac{1}{3}))}\leqslant U\leqslant\sum_{i=0}^{N-1}\binom{M}{i}\leqslant 2^{MH(\frac{N}{M})},$$

where the right inequality is derived from Conclusion A.3. Taking $\log_2$ of both sides and dividing by M, we get that

$$1-H\left(\frac{1}{3}\right)\leqslant H\left(\frac{N}{M}\right),$$

which implies that $M<99N$. □

Remark. The proof of Proposition 3.9 can be applied to societies of potentially many members—every member with a linear ordering on the alternatives and one vote—in which every choice is supported by two-thirds or more of the votes.

Propositions 3.7 and 3.9 establish that $VCD(3Maj)$ is linear in N. Consequently,

Theorem 3.10. *For fixed values of δ and ε, the number of examples needed to PAC-learn the family of choice functions of three-member committees with confidence $1-\delta$ and accuracy $1-\varepsilon$ is linear in the number of alternatives N.*

Thus, with high probability, after seeing $A \cdot N$ independent random examples of the choices of a three-member committee, any choice function in 3*Maj* that "agrees" with the examples will predict a large proportion of the committee's future choices.

4. PAC-learning of larger committees and decisive societies

The results of Section 3 can be generalized to larger committees. Consider the family *rMaj* for an odd integer $r \geqslant 3$. The number of functions in *rMaj* depends on both the number of alternatives N and the number of members r. If r is very large then any tournament on N alternatives can be realized by a committee of r members, and consequently $VCD(rMaj) = \binom{N}{2}$. In fact, Erdős and Moser [5] show that every tournament on N alternatives can be realized by a majority vote of $O(\frac{N}{\log_2 N})$ orderings. Therefore, we limit attention to committees of at most $r \ll \frac{N}{\log_2 N}$ members.

According to Proposition 3.5, $VCD(rMaj) < rN \log_2 N$, because the number of functions in *rMaj* is less than $(N!)^r$. The same line of argument used in the proofs of Propositions 3.7 and 3.9 can be used to obtain the following result (see Appendix A.2 for a detailed proof).

Theorem 4.1. *Let $r \geqslant 3$ be an odd integer. Then,*
1. *The VC-dimension of rMaj is at least $Nr - r^2$.*
2. *The VC-dimension of rMaj is at most $N \cdot f(r, N)$, where $f(r, N) = \min\{10r^2 \log_2 r, r \log_2 N\}$.*

Consequently, for fixed values of δ and ε, the number of examples needed to PAC-learn rMaj is at least $A_1(Nr - r^2)$ and at most $A_2 \cdot Nf(r, N)$, where A_1 and A_2 are constants.

Another interesting family of choice functions is the family of functions induced by α-*decisive societies*. A *society* is a committee with potentially many members. An α-*decisive society* is a society in which every choice is α-decisive, i.e., at least a fraction of $(\frac{1}{2} + \alpha)$ of the society members (not necessarily the same members) agree with every choice, where $0 < \alpha < 1/2$. In other words, the choices of an α-decisive society are not sensitive to a small fraction of people changing their minds. It is easy to verify that any r-member committee, where $r \geqslant 3$ is odd, is a decisive society for $\alpha = \frac{1}{2r}$. In the other direction (which is more difficult), if we randomly sample a committee of $\frac{\ln N}{\alpha^2}$ members from the society, then with probability $> \frac{1}{2}$, the choices of the committee will coincide with the choices of the society. Erdős and Moser [5] show that a large society can realize any tournament on N elements, and therefore PAC-learning the choices of large societies is difficult. If, however, we know that a society is decisive, it is much easier to PAC-learn its choices regardless of the size of the society.

Proposition 4.2. *For fixed values of δ and ε, the family of choice functions of α-decisive societies is PAC-learnable with confidence $1 - \delta$ and accuracy $1 - \varepsilon$ from at most $f(\alpha) \cdot N$ examples, where $f(\alpha) = O((\frac{1}{\alpha})^2 \log_2 \frac{1}{\alpha})$.*

The proof of Proposition 4.2 is similar to that of Proposition A.3 in Appendix A.2, and is left to the reader.

Proposition 4.2 is non-trivial when $(\frac{1}{\alpha})^2 \log_2 \frac{1}{\alpha} \ll N$. A sufficient condition for this is that $\alpha \gg \sqrt{\frac{\log_2 N}{N}}$. Thus, as the number of alternatives N grows, one can learn from a reasonably small number of examples (with respect to N) the choices of societies which are less and less "decisive". When $N \to \infty$ and we allow $\alpha \to 0$ at a slow enough rate, the choices of an α-decisive society are still learnable from a number of examples that is relatively small with respect to N.

5. Concluding remarks

This paper explores whether it is possible to learn the choices of a small committee from examples. The first part of the paper discusses exact learning. We show that in the worst case $\binom{N}{2}$ examples are needed to describe a choice function of a three-member committee. It is an open problem whether fewer examples suffice in the average case, when the linear orderings of the members are uniformly and independently distributed.

The second part of the paper discusses PAC-learning. The results we obtain are asymptotic in nature. Namely, we study situations in which the number of alternatives N is large. This follows the basic paradigm of theoretical computer science, which draws its main insights into the behavior of algorithms from their asymptotic behavior. For example, we prove that $VCD(3Maj) < 3N \log_2 N$ and that $VCD(3Maj) < 99N$. Of course, it might be the case that the constant 99 in the second inequality can be significantly improved, but as it stands the first inequality is stronger when $N < 2^{33}$, i.e., for all practical purposes. Nevertheless, the second inequality provides an insight that cannot be deduced from the first. The number of examples needed for PAC-learning a choice function of a three-member committee is asymptotically similar to the number of examples needed for PAC-learning a rational choice function; i.e., both are PAC-learnable from $O(N)$ examples.

The analysis in the PAC model raises a complementary algorithmic question. Given a sample set of choices of a three-member committee, what can be deduced about the other choices of the committee, and how? We argue that it is possible to deduce most of the committee choices after seeing a relatively small number of them. However, we do not present an efficient algorithm that does so; i.e., an algorithm that finds a committee that agrees with the examples after a number of steps which is polynomial in the number of alternatives N.

A basic assumption of the PAC model is that examples are drawn at random and independently. While this assumption is a reasonable approximation in some settings, it is less plausible in others. For example, a legislature often decides between a status quo option (which is the chosen option from the previous stage) and a new (possibly, random) option, and not between a pair of options drawn at random. Extending our results to such a scenario is a challenge for future work.

It may also be interesting to examine which additional properties of rational choice functions extend to three-member committees. In particular, are there simple regularities that characterize choice functions of committees? For example, we know that a choice function is rational if it is rational when it is restricted to every subset of three alternatives. Is there a similar characterization for choice functions of three-member committees (with three replaced by a larger constant)? A positive answer to this question would provide a positive answer to the following question. Given a set of examples of choices by a social institution, is there an efficient way to decide whether a three-member committee can generate them? We leave these questions as well as applying the PAC model to additional questions of economic interest for future research.

Acknowledgments

I am indebted to Gil Kalai for his devoted guidance, encouragement, and most important comments. I thank Bob Wilson for rewarding discussions and insightful comments, and Ron Siegel for many valuable suggestions. I also thank Elchanan Ben-Porath, Jeremy Bulow, Ariel Rubinstein, the associate editor of this journal, and an anonymous referee for most helpful comments. This research was supported in part by the ISF Bikura grant.

Appendix A.

A.1. Combinatorial approximations

The main combinatorial result we use throughout this section is Stirling's approximation:

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leqslant n! \leqslant \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

Proposition A.1. *Let* $0<\lambda<1$. *Then,* $\binom{n}{\lambda n} \leqslant \frac{2}{\sqrt{n}} 2^{nH(\lambda)}$ *where* $H(\lambda) = -\lambda \log_2 \lambda - (1-\lambda)\log_2(1-\lambda)$ *is the binary entropy function.*

Proof. Using Stirling's approximation, we have

$$\begin{aligned}
\binom{n}{\lambda n} &\leqslant \frac{2\sqrt{\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi\lambda n}\left(\frac{\lambda n}{e}\right)^{\lambda n}\sqrt{2\pi(1-\lambda)n}\left(\frac{(1-\lambda)n}{e}\right)^{(1-\lambda)n}} \\
&\leqslant \frac{\sqrt{n}}{\sqrt{\lambda n}\sqrt{(1-\lambda)n}} \cdot \frac{n^n}{(\lambda n)^{\lambda n}((1-\lambda)n)^{(1-\lambda)n}} \\
&= \frac{1}{\sqrt{\lambda(1-\lambda)n}} \cdot \left(\frac{1}{\lambda^{\lambda}(1-\lambda)^{(1-\lambda)}}\right)^n \\
&= \frac{1}{\sqrt{\lambda(1-\lambda)n}} 2^{nH(\lambda)} \\
&\leqslant \frac{2}{\sqrt{n}} 2^{nH(\lambda)} \qquad \square
\end{aligned}$$

Proposition A.2. *Let* $0 \leqslant k < \frac{1}{2}n$. *Then,* $\sum_{i=0}^{k}\binom{n}{i} \leqslant \binom{n}{k}\frac{n-k}{n-2k}$.

Proof. It is easy to verify that $\binom{n}{k-i} \leqslant \left(\frac{k}{n-k+1}\right)^i \binom{n}{k}$. Indeed, the inequality holds for $i = 0, 1$, and for $i > 1$ we get by induction on i that

$$\binom{n}{k-i} = \frac{k-i+1}{n-k+i}\binom{n}{k-(i-1)} \leqslant \frac{k}{n-k+1}\binom{n}{k-(i-1)} \leqslant \left(\frac{k}{n-k+1}\right)^i \binom{n}{k}.$$

Thus,

$$\begin{aligned}
\sum_{i=0}^{k}\binom{n}{i} &\leqslant \binom{n}{k}\sum_{i=0}^{k}\left(\frac{k}{n-k+1}\right)^i \leqslant \binom{n}{k}\sum_{i=0}^{k}\left(\frac{k}{n-k}\right)^i \leqslant \binom{n}{k}\frac{1}{1-\frac{k}{n-k}} \\
&= \binom{n}{k}\frac{n-k}{n-2k}. \qquad \square
\end{aligned}$$

Conclusion A.3. *Let* $0 < \lambda \leqslant \frac{\sqrt{n}-2}{2\sqrt{n}-2}$. *Then,* $\sum_{i=0}^{\lambda n}\binom{n}{i} \leqslant 2^{nH(\lambda)}$.

Proof. According to Proposition A.2,

$$\sum_{i=0}^{\lambda n}\binom{n}{i}\leqslant\frac{1-\lambda}{1-2\lambda}\binom{n}{\lambda n}.$$

Using Proposition A.1 and the inequality $\lambda\leqslant\frac{\sqrt{n}-2}{2\sqrt{n}-2}$, we obtain the conclusion. □

A.2. Proof of Theorem 4.1

The following two propositions imply the result of Theorem 4.1.

Proposition *A.2*. *Let r be a positive integer. Then*,

$$VDC((2r+1)Maj)\geqslant(2r+1)N-\binom{r+1}{2}-(r+1)(2r+1).$$

Proof. We introduce a set of $(2r+1)N-\binom{r+1}{2}-(r+1)(2r+1)$ choice pairs that the family $(2r+1)Maj$ attains. The pairs are of two types, T1 and T2, as follows:

$$\begin{aligned}
T1:\ &1. && \forall\, 2\leqslant j\leqslant N && (x_1,x_j),\\
&2. && \forall\, 3\leqslant j\leqslant N && (x_2,x_j),\\
&\vdots\\
&r. && \forall\, r+1\leqslant j\leqslant N && (x_r,x_j),\\
T2:\ &r+1. && \forall\, 2r+2\leqslant j\leqslant N && (x_{r+1},x_j),\\
&r+2. && \forall\, 2r+2\leqslant j\leqslant N && (x_{r+2},x_j),\\
&\vdots\\
&2r+1. && \forall\, 2r+2\leqslant j\leqslant N && (x_{2r+1},x_j).
\end{aligned}$$

There are $Nr-\binom{r+1}{2}$ pairs of type T1, and $(r+1)(N-(2r+1))$ pairs of type T2. Consequently, the total number of pairs is $(2r+1)N-\binom{r+1}{2}-(r+1)(2r+1)$.

Given a configuration of choices from these pairs, we construct the $2r+1$ orderings, denoted by $O_1,\ldots,O_{2r+1}$, as follows. Generally, every ordering O_i has four regions:

$$\underbrace{\text{A}}\succ\underbrace{\text{B}}\succ x_i\succ\underbrace{\text{C}}\succ\underbrace{\text{D}}.$$

Regions A and D are "balance" regions, which assure that every x_i, $1\leqslant i\leqslant 2r+1$, appears r times as a "small" element and r times as a "big" element in the $2r$ orderings except for O_i. Then,

the ordering O_i, by manipulating elements in regions B and C, determines the realization of the pairs in which x_i appears first. More specifically, the orderings are divided into two types and constructed as follows.

Type 1: There are r orderings of this type, denoted by $O_1, O_2, \ldots, O_r$. The ordering O_i is "responsible" for the pairs of type T1–i, that is, pairs in which x_i appears first. The ordering O_i has four regions as described above. Regions A and D of O_i are "balance" regions, which include the elements $x_1, \ldots, x_{i-1}$. Regions B and C include the elements $x_{i+1}, \ldots, x_N$. Region B includes all the elements out of $x_{i+1}, \ldots, x_N$ such that $c(x_i, x_j) = 1$. Region C includes all the elements out of $x_{i+1}, \ldots, x_N$ such that $c(x_i, x_j) = 0$.

The ordering of the elements within the regions obeys the following rule: The smaller the index of the element, the further it is located from x_i. For example, if both x_j and $x_{j'}$ appear in region A (or D), and $j < j'$, then $x_j \succ x_{j'}$ (or $x_{j'} \succ x_j$).

Type 2: There are $r + 1$ orderings of this type, denoted by $O_{r+1}, \ldots, O_{2r+1}$. The ordering O_i is "responsible" for the pairs of type T2–i, and has four regions, as described above. Regions A and D are the "balance" regions, which include the elements $x_1, \ldots, x_{2r+1}$ except x_i (note that more elements appear in these regions with respect to the orderings of type 1). Regions B and C include the elements $x_{2r+2}, \ldots, x_N$. Region B includes all the elements out of $x_{2r+2}, \ldots, x_N$ such that $c(x_i, x_j) = 1$. Region C includes all the elements out of $x_{2r+2}, \ldots, x_N$ such that $c(x_i, x_j) = 0$. The ordering of the elements within the regions obeys the same rule as in type 1 orderings.

It still remains to describe the balancing process, i.e., how to position the elements in regions A and D in the $2r + 1$ orderings. The construction of the orderings implies that x_i, $1 \leqslant i \leqslant 2r + 1$, may appear in regions B and C only in the orderings $O_1, \ldots, O_r$; that is, x_i appears in regions B and C at most r times. We have to balance these appearances with appearances in regions A and D. For example, if x_i appears 2 times in region B and 4 times in region C, then it is located in the remaining orderings (except for the ordering O_i) $r - 2$ times in region A and $r - 4$ times in region D. The ordering of the elements within regions combined with the fact that for $j > i$, x_j joins regions A and D not before x_i joins these regions, implies that $x_i \succ x_j$ in orderings (except for O_i) in which x_i appears in either region A or B, and that $x_j \succ x_i$ whenever x_i appears in regions C or D. Therefore, the balancing process assures that for every $j > i$, $x_i \succ x_j$ in r orderings (not including O_i), and $x_j \succ x_i$ in r orderings (not including O_i). Consequently, O_i alone realizes the examples in which x_i appears first, according to the location of the elements in regions B and C in this ordering. As every O_i agrees with the given configuration on the pairs in which x_i appears first, we conclude that the $2r + 1$ orderings realize the configuration. □

Proposition A.3. *Let r be a positive integer. Then*,

$$VDC((2r+1)Maj) \leqslant \min\{(2r+1)N \log_2 N, 10N(2r+1)^2 \log_2 (2r+1)\}.$$

Proof. We showed earlier that $VCD((2r + 1)Maj) < (2r + 1)N \log_2 N$. It is left to show that $VDC((2r + 1)Maj) \leqslant 10N(2r + 1)^2 \log_2 (2r + 1)$. Denote the VC-dimension by M. Using the same arguments as in the proof of Proposition 3.9 we get the inequality:

$$1 \leqslant H\left(\frac{r}{2r+1}\right) + H\left(\frac{N}{M}\right).$$

A series of algebraic manipulations will allow us to derive the result of the proposition.

Step 1: Approximate $H(\frac{r}{2r+1})$.

For a small x, we can approximate $H\left(\frac{1}{2}+x\right)$ using Taylor's formula around $x_0=\frac{1}{2}$:

$$H\left(\frac{1}{2}+x\right)=H\left(\frac{1}{2}\right)+H'\left(\frac{1}{2}\right)x+H''\left(\frac{1}{2}\right)\frac{x^2}{2}+H^{(3)}\left(\frac{1}{2}\right)\frac{x^3}{6}+R_4(x),$$

where $R_4(x)$ is the remainder in Taylor's formula. The term $R_4(x)$ is negative because every odd derivative of H at $x_0=\frac{1}{2}$ is zero and every even derivative is negative. Substituting for numbers in the above formula and using the fact that $R_4(x)$ is negative, we have:

$$H\left(\frac{1}{2}+x\right)\leqslant 1-2x^2\log_2 e \qquad \Longrightarrow_{\text{substitute } x \text{ with } -\frac{1}{4r+2}}$$

$$H\left(\frac{r}{2r+1}\right)\leqslant 1-2\log_2 e\frac{1}{(4r+2)^2} \Longrightarrow_{\text{use the fact that } 1\leqslant H\left(\frac{r}{2r+1}\right)+H\left(\frac{N}{M}\right)}$$

$$1\leqslant 1-2\log_2 e\frac{1}{(4r+2)^2}+H\left(\frac{N}{M}\right) \Longrightarrow$$

$$2\log_2 e\frac{1}{(4r+2)^2}\leqslant H\left(\frac{N}{M}\right) \qquad \Longrightarrow_{\log_2 e>1}$$

$$\frac{1}{2(2r+1)^2}\leqslant H\left(\frac{N}{M}\right)$$

Step 2: Approximate $H\left(\frac{N}{M}\right)$.

Let $0<t<1$. The sum of the geometric progression with multiplier t is

$$\frac{1}{1-t}=1+t+t^2+t^3+\ldots \qquad \Longrightarrow_{\text{use integration on both sides}}$$

$$-\log_2(1-t)=t+\tfrac{1}{2}t^2+\tfrac{1}{3}t^3+\ldots \qquad \Longrightarrow_{\text{multiply by } (1-t)}$$

$$-(1-t)\log_2(1-t)=t-\left(\tfrac{1}{2}t^2+\tfrac{1}{6}t^3+\ldots\right)\leqslant t \Longrightarrow_{H(t)=-t\log_2 t-(1-t)\log_2(1-t)}$$

$$H(t)\leqslant -t\log_2 t+t \qquad \Longrightarrow_{\frac{1}{2(2r+1)^2}\leqslant H\left(\frac{N}{M}\right)}$$

$$\frac{1}{2(2r+1)^2}\leqslant -\frac{N}{M}\log_2\frac{N}{M}+\frac{N}{M}$$

Denote $M=cN$. Then, the last inequality can be written as $\frac{1}{2(2r+1)^2}\leqslant\frac{1}{c}\log_2 c+\frac{1}{c}$. This inequality implies that $c=O((2r+1)^2\log_2(2r+1))$. It remains to determine the constant of

the $O(\cdot)$. Denote $c = 2d(2r+1)^2 \log_2(2r+1)$. Then,

$$\frac{1}{2(2r+1)^2} \leqslant \frac{1}{c}(\log_2 c + 1) \qquad \Longrightarrow_{c=2d(2r+1)^2 \log_2(2r+1)}$$

$$\frac{1}{2(2r+1)^2} \leqslant \frac{1}{2d(2r+1)^2 \log_2(2r+1)} \log_2(2d(2r+1)^2 \log_2(2r+1)) + 1 \qquad \Longrightarrow$$

$$1 \leqslant \frac{1}{d \log_2(2r+1)} \log_2 2d + 2\log_2(2r+1) + \log_2 \log_2(2r+1) + 1 \qquad \Longrightarrow_{\text{multiply by } d}$$

$$d \leqslant 2 + \frac{\log_2 2d + \log_2 \log_2(2r+1) + 1}{\log_2(2r+1)} \leqslant 2 + 3 = 5$$

Consequently, the VC-dimension is at most $10N(2r+1)^2 \log_2(2r+1)$. □

References

[1] K.J. Arrow, Social Choice and Individual Values, second ed., Wiley, New York, 1963.
[2] M. Condorcet, Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix, L'imprimerie Royale, Paris, 1785.
[3] T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms, MIT Press, Cambridge, MA, 1990.
[4] P. Dasgupta, E. Maskin, On the robustness of majority rule and unanimity rule, Economics working paper no. 36, Institute for Advanced Study, School of Social Science, Princeton, 2004.
[5] P. Erdős, L. Moser, On the representation of directed graphs as unions of orderings, Magyar Tud. Akad. Mat. Kutató Int Közl 9 (1964) 125–132.
[6] G. Kalai, Learnability and rationality of choice, J. Economic Theory 113 (1) (2003) 104–117.
[7] M.J. Kearns, U.V. Vazirani, An Introduction to Computational Learning Theory, MIT Press, Cambridge, MA, 1994.
[8] D.E. Knuth, The art of computer programming, vol. 3, Sorting and Searching, Addison-Wesley, Cambridge, MA, 1973.
[9] E. Maskin, Majority rule, social welfare functions, and game forms, in: K. Basu, P.K. Pattanaik, K. Suzumura (Eds.), Choice, Welfare and Development: A Festschrift in Honour of Amartya K. Sen, Clarendon Press, Oxford, 1995.
[10] K.O. May, A set of independent necessary and sufficient conditions for simple majority decision, Econometrica 20 (4) (1952) 680–684.
[11] D.C. McGarvey, A theorem on the construction of voting paradoxes, Econometrica 21 (4) (1953) 608–610.
[12] A. Rubinstein, Why are certain properties of binary relations relatively more common in natural language?, Econometrica 64 (2) (1996) 343–355.
[13] N. Sauer, On the density of families of sets, J. Combin. Theory, Series A 13 (1) (1972) 145–147.
[14] S. Shelah, A combinatorial problem; stability and order for models and theories in infinitary languages, Pacific J. Math. 41 (1) (1972) 247–261.
[15] M. Vidyasagar, A Theory of Learning and Generalization: With Applications to Neural Networks and Control Systems, Springer, London, 1997.