A Critique of Distributional Analysis In the Spatial Model

Craig A. Tovey*
School of Industrial and Systems Engineering and
College of Computing
Georgia Institute of Technology
Atlanta, Ga 30332-0205

July 23, 2009

Abstract

Distributional analysis is widely used to study social choice in Euclidean models [35, 36, 1, 5, 11, 19, 8, 2, e.g]. This method assumes a continuum of voters distributed according a probability measure. Since infinite populations do not exist, the goal of distributional analysis is to give insight into the behavior of large finite populations. However, properties of finite populations do not necessarily converge to the properties of infinite populations. Thus the method of distributional analysis is flawed. In some cases [1] it will predict that a point is in the core with probability 1, while the true probability converges to 0. In other cases it can be combined with probabilistic analysis to make accurate predictions about the asymptotic behavior of large populations, as in [2]. Uniform convergence of empirical measures [23] is employed here to yield a simpler, more general proof of α-majority convergence, a short proof of yolk shrinkage, and suggests a rule of thumb to determine the accuracy of distribution-based predictions. The results also help clarify the mathematical underpinnings of statistical analysis of empirical voting data.

1 Introduction

Distributional analysis has been a widely used technique in the study of social choice in Euclidean models [35, 36, 1, 5, 11, 19, 8, 30, 2, e.g.] (see also [7] and [25, Chaps. 11-12]). In distributional analysis, a continuum or infinite population of voters is analyzed, where the population follows some probability distribution μ. Infinite populations do not exist. Our concern is with finite populations. Therefore, the principal purpose of distributional analysis must be to give insight into the behavior of large but finite populations. In this paper it is shown that distributional analysis is flawed when applied to this end. The problem is essentially one of convergence: if the

*supported by a National Science Foundation Presidential Young Investigator Award ECS-8451032. A portion of this work was done while the author held a National Research Council Research Associateship at the Naval Postgraduate School

limiting case is to give insight into the large finite case, behavior of the latter should converge to behavior of the former as the population grows. Unfortunately, it turns out that properties of finite populations do not necessarily converge to the properties of infinite populations. In some cases a distributional analysis will predict that a point is in the core with probability 1, while the true probability converges to 0. Thus analysis of infinite populations may fail to yield any information about finite populations, however large.

An alternative to distributional analysis is termed here the *finite sample* method. In this method, n points are independently generated according to the distribution μ. This random finite sample from μ forms a *configuration* of n points whose properties are analyzed. A typical question would be: "what is the probability, as a function of n, that the configuration generated has nonempty core?" Typical answers to these questions are bounds or asymptotically close estimates for the desired probability.

It is sometimes possible to combine distributional analysis with finite sample analysis to make correct predictions about the asymptotic behavior of large populations. An example of this is found in [2]. We expose some key properties which enable the convergence in this case, yielding a simpler and more general proof of the convergence of (Simpson-Kramer) α-majority rule, and a simpler though less general proof of yolk shrinkage. The analysis suggests a rule of thumb as to when one might expect distributional analysis to give accurate or inaccurate predictions about the behavior of finite populations.

Another motivation for analyzing the distributional method is to help develop a rigorous foundation for statistical empirical study of group choice. One would like to poll the members of a committee, assembly, or population (or in some other way extract data on their preferences), and based on that data and some solution concept, make a prediction with some confidence regarding what the outcome will be. How can a solution concept be tested experimentally? When the data are sampled from a large population, there are issues of statistical accuracy. Even if preference data are extracted for each individual, issues remain concerning the robustness of the solution concept with respect to individual perturbations. In other words, a person's views on issues are not perfectly constant, and can even change in the voting booth. How can we know that a prediction based on polls taken one day will be close to the actual results the next day?

We may think of the preference data as a random sample from a probability distribution, and the population's actual vote as another random sample from this distribution. The problem is to establish rigorously the stability of a solution concept under this model. In statistical terms, the finite sample from μ is an empirical measure μ_n. A solution concept is a statistic, a function f operating on probability measures. If f is a consistent statistic, then the limiting behavior of $f(\mu_n)$ will (almost surely) be like $f(\mu)$, and the solution concept is stable. This issue has received a great deal of attention for the classical core or Nash equilibrium under the term "structural stability". The convergence theorems discussed in section 6 should aid in determining the stability of other more widely applicable solution concepts.

The outline of the paper follows: the remainder of this section reviews essential definitions of the spatial model. Section 2 introduces the two methods by way of a small example. Section 3 analyzes the distributional method. The difficulty with the method is shown to arise from the identification of two different definitions of majority rule preference. Section 4 demonstrates in greater detail a case from [1] where the distributional method gives a misleading result. Section 5 considers a case where the method may be used to achieve results valid for large finite populations, and demonstrates how to estimate how large the population must be. Section 6 introduces the use of uniform convergence of empirical measures, following a suggestion due to R. Foley, R. McKelvey, and G. Weiss, applies it to α-majority and yolk, and discusses in general when we may expect the distributional method to give accurate predictions.

1.1 Definition of the spatial model

In the Euclidean spatial model, a group of individuals must make a social choice from the set of alternatives $\Re^m$. Each individual i has a most preferred point $v_i \in \Re^m$. This point will be referred to as a voter point, or simply a voter. A voter at v_i strictly prefers y to z if $||v_i - y|| < ||v_i - z||$ and is indifferent between equidistant alternatives, where distances are with respect to the Euclidean norm. This model is more general than it appears: Davis *et al.* [5] show it is equivalent to any linearly transformed spatial model which maintains the properties of an inner product; Grandmont [11] observes that the essential property of the Euclidean model is often the "division-by-hyperplane" property (in the Euclidean case, the perpendicular bisector of two points separates those who prefer one point to the other), and so results in the Euclidean model usually apply to the more general class of "intermediate preferences", including constant elasticity of substitution (C.E.S.) utility functions (these extend the class of Davis *et al.* by allowing a change to an L^p norm from the L^2 norm).

A *finite configuration* is a list of (not necessarily distinct) voter points $V = \{v_1, \ldots, v_n\}$. Given a finite configuration V, we say that alternative y *dominates* alternative z under α-majority voting if $|\{i : ||v_i - y|| < ||v_i - z||\}| > \alpha n$, that is, more than a fraction α of the voters prefer y to z. The value $\alpha = \frac{1}{2}$ corresponds to simple majority voting. In geometric terms, y dominates z if the open halfspace containing y that is defined by the hyperplane bisecting and normal to the line segment $\overline{yz}$ contains more than αn voter points. The hyperplane is specified algebraically as $\{x : (y-z) \cdot x > \frac{(y-z)\cdot(y+z)}{2}\}$.

The *core* is the (possibly empty) set of alternatives that are not dominated by any alternative, where voting is assumed to be simple majority unless specified otherwise. The smallest α for which the α-majority core is nonempty is the *Simpson-Kramer*, or *minimax*, value; the corresponding core is the *minimax set*. A hyperplane is *median* if the two closed halfspaces it defines each contain at least half the population. The *yolk* is the smallest ball intersecting all median hyperplanes [10, 18, 14]. As is true of most solution concepts, both the minimax set and the yolk coincide with the simple majority rule core point if the latter exists. For recent work on the yolk and related concepts, see [4, 3].

For any set $S \subseteq \Re^m$ let $-S$ denote $\{-s | s \in S\}$. We say that μ is *sign-invariant* if $\mu(S) = \mu(-S)$ for all measurable subsets $S \subset \Re^m$. A probability measure μ on $\Re^m$ is *weakly centered* at 0 if for all nonzero vectors $c \in \Re^m$ the halfspace $h(c) \equiv \{x | c \cdot x \geq 0\}$ has measure $\mu(h(c)) \geq \frac{1}{2}$. In terms of distributional analysis, μ is weakly centered at 0 iff 0 in the core of a population μ. We say that μ is strictly centered at 0 if for all scalars $\epsilon > 0$ the halfspace $h(c, \epsilon) \equiv \{x | c \cdot x \geq \epsilon\}$ has measure $\mu(h(c, \epsilon)) < \frac{1}{2}$. Probability measure μ is weakly (resp. strictly) centered at p iff $\mu(S - p)$ is weakly (resp. strictly) centered at 0.

2 An example of the two methods

Let us begin with a simple two-dimensional model based on an example in [30]. Let μ be the probability distribution that is uniform on a circle (the boundary of a disk). Place a single voter v_1 at the center of the circle, which for convenience we locate at the origin. Randomly generate $n-1$ additional voter points $v_2, \ldots, v_n$, where n is even, according to μ.

We illustrate the method of finite sample analysis on the model just stated. A particular realization of this random process is a finite configuration. The question we pose is: what is the probability that v_1 is a core point of the configuration? A result of Schofield's [30] implies that the probability is positive; the exact probability turns out to be $1/2^{n-2}$, as follows. The proof is sketched here since it will be needed in Sections 4 and 5.

Theorem 1 Place v_1 at the origin in $\Re^2$ and generate $v_2, \ldots, v_n$ independently according to any nondegenerate sign-invariant distribution μ on $\Re^2$. Then for all even n, the probability v_1 is undominated with respect to simple majority voting is $1/2^{n-2}$.

Proof Sketch:([31])

Associate for each θ, $0 \leq \theta < \pi$, a line passing through the origin and an associated orientation. See Figure 1. Denote this line by $L(\theta)$. The open half space the line is oriented towards is the "front" and the other open half space is the "back" of the line $L(\theta)$.

Since the points are drawn from a nondegenerate distribution, the probability is 0 that any pair of the points $v_2, \ldots, v_n$ are collinear with the origin. Henceforth we assume this event does not occur.

For any $0 \leq \theta < \pi$ define the gap function $g(\theta)$ to equal the number of voter points in the front half plane of $L(\theta)$ minus the number of voter points in the back of $L(\theta)$. If $g(\theta) = -1$ or 1, the line $L(\theta)$ divides the $n-1$ points $v_2, \ldots, v_n$ as equally as possible given that n is even. If however the gap function $g(\theta)$ ever attains $|g(\theta)| \geq 3$ then one side of the line will contain at least $1 + n/2$ points and v_1 will not be a core point.

Starting at $\theta = 0$, increase θ continuously to π. Because no two points are collinear with v_1, $g(\theta)$ will change by either +2 or -2 as the line $L(\theta)$ crosses over a point v_i. Let $\theta_1, \ldots, \theta_{n-1}$ denote the values of θ at which $L(\theta)$ crosses over a voter and let $X_i = +2$ or -2 accordingly as the ith crossover increases or decreases $g(\theta)$. The key observation is that the gap function executes a random walk as θ goes from 0 to π.

Lemma 1. $X_1, \ldots, X_{n-1}$ are independent identically distributed variables taking values 2 with probability 1/2 and -2 with probability 1/2.

Proof of Lemma: the proof follows easily from the sign-invariance of μ. See Figure 2: the regions I and II are equally likely to contain the next point as $L(\theta)$ sweeps around. Details are given in [31].

There are 2^{n-1} possible paths for the random walk of the X_i to take. Of these, only two will keep the gap function at $|g(\theta) \leq 1|$. These are the alternating paths $X_i = 2(-1)^i$ and $X_i = -2(-1)^i$. By Lemma 1, each of the 2^{n-1} possible paths occurs with equal probability $1/2^{n-1}$. Therefore the probability that $\max_\theta |g(\theta)| \leq 1$ is $2/2^{n-1} = 1/2^{n-2}$ as desired.$\diamond$

The proof of Theorem 1 assumes only that μ is sign-invariant. Hence it applies to the uniform rectangle distribution in [36, 1] and elsewhere. For the model under discussion, Theorem 1 gives a stronger outcome, for obviously (w.p.1) no other point in $\Re^2$ can be undominated. Thus there is nonempty core with exact probability $1/2^{n-2}$. A less precise but more general result is given in [29]: under the conditions of Theorem 1, for an $\alpha(n) = \frac{n}{2} + k(n)$ supermajority voting rule, the probability of a core at the origin converges to 0 or 1 respectively as $\lim_{n\to\infty} \frac{k(n)}{\sqrt{n}} = 0$ or 1.

Now let us illustrate the distributional method on the same 2-dimensional model. (The following closely follows analyses in [36, 30, 5, 11]. Assume a continuum of voters uniformly distributed on the circle. Each halfspace h defined by a line through 0 has $\mu(h) \geq 1/2$. Thus 0 is in the core. In our terms it is weakly (and also strictly) centered. In fact by [5, theorem 1] or [19, theorem 2] it is the unique such point. (The reader who is concerned about the "extra" point at 0 may observe that this only improves the position of 0 with respect to equilibrium.)

The contrast between the two methods is evident. The finite sample method shows that the probability of 0 being undominated, indeed of a nonempty core, rapidly converges to 0. The distributional method says that for an infinite population, the probability of 0 being undominated is 1.

The example of this section reveals that there is a flaw in the distributional method. It would be desirable for the outcome of the distributional method to coincide with the limiting behavior of finite samples, since the goal must be insight into the behavior of finite populations. Yet there could hardly be less consonance than in the example just given. In the next section we analyze the distributional method to explain how this problem arises.

3 An analysis of the distributional method

We have observed that the outcomes of the two methods can differ. Let us point up an important distinction in how they operate. The distributional method works directly with μ, and quantities such as $\mu(h)$ are considered. On the other hand, in the finite sample method a configuration V is drawn from μ, and quantities such as $|V \cap h|$ are considered. Informally, the distributional method counts up voters by looking at the distribution function μ directly, while the finite sample method counts up voters by looking at configurations drawn from μ.

A brief history of distributional analysis

In the literature, the term *distribution* is used to mean both "configuration" and "distribution function" as defined here. As a consequence, distributional analyses are intertwined with analyses giving necessary and/or sufficient conditions for domination, local equilibrium, and/or global equilibrium [22, 19, 5, 1, 30, 20]. For instance, Plott's classic paper [22] is titled "A notion of equilibrium *and its possibility* [emphasis added] under majority rule." Plott performs no probabilistic analysis but observes (quite accurately) [IBID, page 792] that "it would only be an accident (and a highly improbable one) if an equilibrium exists at all." Later papers such as [19, 5, 20] generalize Plott's results to infinite populations and/or more general preference functions (also global rather than local equilibrium). For instance, Davis *et al.*[5, page 148] contrast their work with Plott's since the latter "allows only a finite number of individuals to be considered." Presumably Davis *et al.* view this limitation of Plott's analysis as undesirable because more insight is needed as to the behavior of large finite populations.

In 1981 however Tullock remarks [37, page 190] that his analysis was "not regarded as very reliable any more because McKelvey proved that majority voting can reach any part of the issue space." The analysis Tullock refers to ultimately showed (see [28, 30, 26, e.g.]) that the set of configurations for which equilibrium exists is measure 0, for $d \geq 3$ and also for $d = 2$ and odd n, confirming Plott's observation. These powerful results seem implicitly to invalidate the distributional analyses. Yet, this consequence does not even now appear to be fully assimilated in the literature. The only unresolved case was $d = 2, n$ even, (which was the author's original motive for undertaking this line of research.)

Analysis of distributional analysis

Distributional analysis has implications at odds with the instability theorems of McKelvey, Schofield, Rubenstein, and others [16, 17, 27, 26]. What is the source of this inconsistency? Consider part of Arrow's summary [1] of Tullock's analysis Tu67a,Tu67b:

> He [Tullock] assumes
> (1) that the number of voters is large, so large that we may consider them to constitute a continuum [1, page 108].

This assumption seems innocuous enough. In mathematics, passing to the limiting continuous case is a frequently used technique. The problem is that majority rule requires us to evaluate $n/2$ where $n =$ the number of voters, but the value $\infty/2$ is not well-defined. More precisely, if 0 is undominated and only one voter is located at 0 then placing two additional voters together at any location $x \neq 0$ must make 0 dominated (by the point ϵx for sufficiently small $\epsilon > 0$). But if n is treated as infinite no shifting of any finite number of voters changes the analysis, since $\infty/2 + 1 = \infty/2$.

What happens is that *a new definition is needed* when passing from the finite to the infinite case. Let us examine a specific definition from the literature. In an article by Davis, DeGroot, and Hinich [5], necessary and sufficient conditions are derived for the existence of a dominant point. As stated earlier, this analysis, unlike Plott's, is intended to apply to infinite populations. The critical definition of a non-dominance relation R is quoted below [5, page 149].

> Let P^* denote the distribution of most preferred points of the individuals. Let X be the most preferred point of an individual chosen at random from the population. [*note P^* is referred to as an infinite configuration in the previous sentence and as a probability function in the next sentence*] Given a (Borel) set $S \subset E_n$, Pr(S) will denote the probability that $X \in S$ under the distribution P^*.
>
> **Definition 1:** For any points $y \in E_n$ and $z \in E_n$, it is said that yRz if $\Pr(||y - X|| \leq ||z - X||) \geq \frac{1}{2}$.

The definition of the relation R just given is mathematically unambiguous. But there is a difficulty with the interpretation of the mathematical results. In [5] the passage just cited continues with the following interpretation:

> In other words, yRz if and only if at least half the population either prefers y to z or is indifferent between y and z.

What does the word "population" mean in the sentence just quoted? If we take it to mean the probability measure, then it would be accurate to say that

> yRz if and only the measure (mass) of the subset of the population, that either prefers y to z or is indifferent between y and z, is at least $1/2$.

But if the word "population" refers to a finite sample drawn from the distribution P^*, then the meaning of yRz is given by the following theorem.

Theorem 2. Suppose a finite number of points are drawn at random according to the distribution P^*. Then

> yRz if and only if the probability is at least $1/2$ that at least half the population either prefers y to z or is indifferent between y and z.

Proof: Suppose yRz. If we were to take a finite sample under the distribution P^*, each sample point would with probability at least $1/2$ be at least as close to y as to z. Then the number of points in the sample at least as close to y as to z follows a binomial distribution with "success" parameter $p \geq 1/2$. From elementary properties of the binomial distribution $p \geq 1/2$ implies the probability is at least $1/2$ that at least half the outcomes are "successes". Conversely, if the probability is at least $1/2$ that at least half of the Bernoulli trials end in success, it must be that the parameter $p \geq 1/2$, whence yRz.

The heart of the problem

We have arrived at the heart of the problem. When going from finite to infinite populations, a new definition of the nondominance relation R was needed. Succinctly, let $\mathcal{A}$ denote "at least half the population either prefers y to z or is indifferent between y and z." Then for any finite sample population, yRz means that $\mathcal{A}$ occurs with probability $1/2$. But the interpretation for infinite populations is, yRz means that $\mathcal{A}$ occurs.

If the purpose of the mathematical analysis of infinite populations is to gain insight into the behavior of large finite populations, then there should be a closer correspondence between the meanings of yRz for finite samples and for infinite populations.

Note: the gap between the finite sample (Theorem 2) and the distributional (Definition 1) methods just discussed is between $1/2$ and 1. In the earlier example of section 2 involving Theorem 1, the gap was (asymptotically) between 0 and 1. The larger gap in that example was due to the intersection of many events each with probability $1/2$.

4 An unsuccessful case: The Sonnenschein-Arrow Theorem

Let us now examine a specific case of analysis from the literature where the predictions of distributional analysis are misleading. In his article, Arrow continues by stating a theorem (he attributes to Sonnenschein) that generalizes Tullock's example [1, pages 108–109]:

> For any pair of alternatives x, y, let $N(x, y)$ be the number of individuals who prefer x to y. Then let xMy be the statement $N(x, y) \geq N(y, x)$ and $x\bar{M}y$ the statement that $N(x, y) > N(y, x)$....

Theorem. *Suppose that, for each alternative x^0, the set of alternatives x for which xMx^0 is closed, and [suppose] the set of alternatives [x] for which $x\bar{M}x^0$ is convex. Then for any compact (closed and bounded) convex set of alternatives S, there is (at least) one alternative x in S such that xMy for all y in S.*

Arrow later points out that "the hypotheses of the theorem are obviously fulfilled in Tullock's example." [IBID, page 110]. This is of course correct, but *only* subject to assumption (1) [IBID, page 108] quoted in section 3 (page 7). For if we employ the finite sample method of this paper, we find the *probability converges to 0 that the hypotheses of the Sonnenschein-Arrow theorem are fulfilled in Tullock's example.* The following theorem states and proves this statement precisely.

Theorem 3. Let n points be sampled from the uniform distribution on a rectangle with center $\mathbf{0}$ in $\Re^2$ (the hypothesis of Tullock's example), or more generally from any sign-invariant probability measure μ on $\Re^2$ such that $\mu(L) = 0$ for all lines L that are incident on $= \mathbf{0}$. Then the probability that the set $\left[x : x\bar{M}\mathbf{0}\right]$ is convex converges to 0 as $n \to \infty$. $\diamond$
Proof: Discard the zero-probability event that two sample points are incident on a line that passes through the origin, or that a sample point is incident on the horizontal axis. Let $v_\theta = (\sin\theta, \cos\theta)$. As in the proof of Theorem 1, define the gap function $g(\theta) : 0 \le \theta \le \pi$ to be the difference between the number of points in the halfplane pointed to from $\mathbf{0}$ by v_θ and in the complementary halfplane. A direct calculation shows that if there exist $\theta_1 < \theta_2 < \theta_3$ such that $g(\theta_1 > 0)$, $g(\theta_2 < 0)$, and $g(\theta_3 > 0)$, then the set $\left[x : x\bar{M}\mathbf{0}\right]$ is not convex. The set would fail to be convex because for all $\epsilon > 0$ the points $\epsilon v_{\theta_1, \epsilon v_{\theta_3}}$ (respectively, ϵv_{theta_2}) dominate (respectively, do not dominate) $\mathbf{0}$. Obviously such values of θ exist if $g(\theta)$ crosses 0, that is, strictly changes its sign, more than twice. The rest of the proof consists of showing that $g(\theta)$ crosses 0 at least three times with probability converging to 1 as $n \to \infty$.

Ignore the n isolated values of θ at which $g()$ is not continuous. The sign-invariance of μ implies that the changing values of $g(\theta)$ execute a random walk with step size 2 as θ increases from 0 to π. To be precise, each of the n times when the value $g(\theta)$ changes, it increases by 2 with probability $\frac{1}{2}$, and otherwise decreases by 2, independent of all other changes in value. From the point of view of the random walk, the value $g(0)$ is not determined at the start, but will satisfy $0 = \frac{1}{2}(g(0) + g(\pi))$. This follows from $g(0) = -g(\pi)$. If the random walk crosses z only once, z is called a *point of increase*. The probability is known to be less than $Clogn$ that there exists a point of increase anywhere in the walk, where C is a constant independent of n. (For an "elementary" proof see [21]; this result is very closely related to the classic theorem of Dvoretsky, Erdös and Kakutani [6] on the absence of points of increase of Brownian motion.) Therefore the probability that $g(\theta)$ crosses 0 at least twice converges to 1. Discard the event that the walk does not end where it starts, $g(0) \ne g(\pi)$, because for n even the probability of exactly $n/2$ successes out of n trials is $O(1/\sqrt{n}) \to 0$ as $n \to \infty$. Therefore $g(\theta)$ crosses 0 an odd number of times; if it crosses at least twice it crosses at least thrice. Hence as $n \to \infty$ the probability converges to 1 that the set $\left[x : x\bar{M}\mathbf{0}\right]$ is not convex. $\diamond$.

It has previously been observed that the Sonnenschein-Arrow Theorem can fail to be applicable. Greenberg [12], in a paper on d-majority equilibrium gives a deterministic example with $n = 4$ voters in which the set $[x : x\bar{M}0]$ is not convex. At the time it must have seemed that examples such Greenberg's would become less likely as n increased. For example Kramer [15, page 313] remarks,

> Several authors, ... have argued that this instability is a "small-sample" problem, and that majority equilibria will be more likely when the number of voters is large; examples and results supporting this thesis have been exhibited by ....

Theorem 3 proves that Greenberg's example is the rule, not the exception, as the number of voters gets large.

5 Estimating a minimum population to assure accurate predictions

Although in some cases the distributional method can mislead, it can in other cases make correct predictions for large enough populations. In these cases we want to know how large is enough. Intuitively, we need a margin of safety between the knife-edge requirement of a voting rule for finite populations, and the distributional properties. As a rule of thumb, if n is the population size, we will want the safety margin to be at least $\sqrt{n}$ because the most typical standard deviations involved are $\approx \sqrt{n}/2$.

To be concrete, we estimate the population size needed to assure accuracy of asymptotic predictions in a paper by Caplin and Nalebuff [2]. Employing the distributional method, they show (as was proved earlier by Grunbaum [13]) that if the distribution function μ is concave, then the minimax value can equal but not exceed $1 - (m/(m+1))^m$ ([2]). They continue and prove ([2, Theorem 3] that if a finite sample of size n is drawn at random from the concave distribution μ, then the minimax value of the sample converges to that of μ *a.e.* Demange [8, 5,2.4(iii), pp. 151-153; 5.3, p. 164] had previously proved this convergence under a continuity assumption.

At what population size can we be confident that these bounds apply? We first argue that a margin of safety is necessary. Sample n points independently from the uniform population density on an equilateral triangle (see Figure 4). The triangle center is the unique distributional minimax point, with $\alpha = \frac{5}{9}$. The mass of μ in the shaded region is $\frac{5}{9}$. Therefore the number of points falling in the shaded region is binomially distributed as $B(\frac{5}{9}, n)$. By the central limit theorem this converges to a normal distribution with mean $\frac{5}{9}$. Since the normal distribution is symmetric, the probability converges to $\frac{1}{2}$ that more than 5/9 of a random sample will fall in the shaded region, and the center will not be a 5/9-majority core point.

If, as a margin of safety, a $(5/9+\epsilon)$-majority rule were employed ($\epsilon > 0$), the probability would converge to 1 that the triangle center is a majority point [2]. However, the standard deviation of $B(\frac{5}{9}, n)$ is $\sqrt{20n}/9 \approx .50\sqrt{n}$. We would require a margin of at least two standard deviations. Hence it is necessary that $\epsilon \geq 2(.50)\sqrt{n}/n = 1/\sqrt{n}$ to enable confident predictions. This accords well with our rule-of-thumb stated at the beginning of this section. Our rule of thumb also works well in the case of a sign-invariant distribution in two dimensions. Theorem 4 below states we can expect an error of order $1/\sqrt{n}$.

Theorem 4. Place v_1 at the origin in $\Re^2$ and generate $v_2, \ldots, v_n$ independently according to any nondegenerate sign-invariant distribution μ on $\Re^2$. For n even, the largest majority that can be mustered against the origin has expected value $\geq \frac{n+\sqrt{n}}{2}$.

Proof: From the proof of Theorem 1, the gap function executes a random walk around 0. The expected absolute distance from 0 at the end of a random walk is on the order of $\sqrt{n}$ [9]. Dividing by the population size n gives the result. $\diamond$

With a committee size of 100 (e.g. U.S. Senate), $1/\sqrt{n}$ is a fairly substantial 10%. Therefore, concavity alone is not a sufficient explanation for the stability of $\frac{2}{3}$-majority rule in a group of this size. Concavity together with a limitation to 2 key issues (dimensions) may be adequate. If the population size is 10,000 or more, drawn from a concave density, the probability of stability under $\frac{2}{3}$-majority rule appears to be fairly close to 1. From Theorem 4 we heuristically may expect that the maximum gap will usually not exceed a few multiples of the expected value $\sqrt{n}/2$, say $3(\sqrt{n}/2)$. At $n = 10,000$ this gap as a fraction of population is 1.5%, or roughly half of the "cushion" between $2/3$ and the $1 - 1/e$ value assured by concavity in [2]. On the other hand, equilibrium appears unlikely to be guaranteed when the population size is $n \leq 250$. This is because Theorem 4 suggests a gap of at least $\sqrt{n}/2$ will occur quite often. At $n = 250$ this gap is approximately 3%, so the cushion is not large enough to make equilibrium likely without additional restrictions on voter preferences.

6 Uniform Convergence

Why do the distributional results discussed in section 5 apply to large finite populations, while some others do not? Part of the answer has to do with the distinction between non-dominance and strict dominance. Recall from section 4 that the finite sample meaning of the non-dominance relation R does not converge to the meaning in the distributional case. In contrast, the strict dominance relation $P : yPz$ iff yRz and not yRz does converge. That is, if yPz in the distributional sense, and a random sample of n points is taken, then yPz with respect to that finite sample with probability converging to 1 as $n \to \infty$. (This follows immediately from the weak law of large numbers and Davis *et al.*'s observation that "yPz if and only if $Pr(||y - X|| < ||z - X||) > 1/2$.")

This distinction is not enough to account for the difference. For example, suppose distribution μ is uniform in a square centered at y. Then for all $z \neq y$, yPz in the distributional sense. But if a finite sample of size $2n$ is taken, then by Theorem 1 with probability converging to 1 there will exist $z \neq y$ such that zPy. However, suppose y strictly dominated all z in some compact set Z. We might then argue, if μ were continuous, that the strict domination occurred with a minimum gap of some $\delta > 0$. If we could then find a way to reduce consideration of Z to a finite, relatively small (e.g. polynomial in n) number of points, we could establish the desired behavior of the finite sample. These ideas are found in the proof of Theorem 3 in [2], where compactness and Lemma 1 (page 807) provide the reduction to a finite number $(n + 1)$ of points. Similar ideas are found in [31], where the fundamental theorem of linear programming provides the reduction to a finite number.

The preceding suggests that the mathematical tools for the convergence of empirical measures may be appropriate to these questions[1]. This turns out to be the case. The interested reader should consult for example chapter 2, "Uniform Convergence of Empirical Measures" of Pollard's book[23]. A couple of the most pertinent results are cited below (specialized to our case and adapted to our terminology):

[1] I am indebted to Bob Foley, Richard McKelvey, and Gideon Weiss for suggesting this line of attack

Definition. Let n points be drawn at random according to a probability measure μ on $\Re^m$. The *empirical measure* μ_n is that which places mass $1/n$ at each of the n points (obviously they need not be distinct.)

Let $\mathcal{C}$ denote a class of sets in $\Re^m$. For any $c \in \mathcal{C}$, it follows that $\mu_n(c)$ simply equals the fraction of the points which fell in c (while $\mu(c)$ equals the expected fraction). The class $\mathcal{C}$ of most interest to us is the set of all closed and open halfspaces. Accordingly, let

$$\mathcal{C} \equiv \left\{c : c = \left[x : p \cdot x \leq p^0\right]; p \in \Re^m, p^0 \in \Re\right\}. \tag{1}$$

Also let $\mathcal{C}^+ \equiv [\bar{c}]$, the set of open halfspaces, and let $\mathcal{D} \equiv \mathcal{C} \cup \mathcal{C}^+$. The considerable generalization of the Glivenko-Cantelli uniform convergence theorem by Pollard ([23]) implies that the empirical measure converges to μ over these classes.

Theorem 5. Let μ be a probability measure on $\Re^m$. Then

$$\sup_{d \in \mathcal{D}} |\mu_n(d) - \mu(d)| \to 0 \text{ almost surely} \tag{2}$$

Proof: this follows from Theorem 14 (page 18), Lemma 15(i,ii)(page 18), and Lemma 18 (pages 20–21) of [23]. The key definition for our purposes is that a class of subsets $\mathcal{S}$ of $\Re^m$ has *polynomial discrimination* if the number of distinct subsets of n points $X \in \Re^m$ that can be represented as $X \bigcap S : S \in \mathcal{S}$ is polynomially bounded in n. For example, let $\mathcal{S}$ be the set of all balls centered at the origin. There are infinitely many such balls, but obviously for any set X of n points there are at most $n+1$ distinct subsets expressible as the intersection of a ball and X. Theorem 14 states that almost sure uniform convergence occurs on classes that have polynomial discrimination. The two lemmata cited prove that a huge variety of classes, including $\mathcal{D}$, have polynomial discrimination. It may be of additional interest that the class of sets of form $x \in \Re^m : x^T Ax + Bx \leq c$ for all $m \times m$ matrices A and B has polynomial discrimination as well. $\diamond$

This means that even if we simultaneously consider all half-spaces h, the largest gap between the fraction of points falling in the half-space, and the expected fraction ($\mu(h)$), converges to 0, with probability 1.

Recall that the minimax value $\alpha(\mu)$ is the smallest supermajority level at which the core is nonempty, and the *minimax set* is the corresponding core. To demonstrate the usefulness of Theorem 5, we invoke it to prove the convergence of the minimax value and set. Caplin and Nalebuff [2] previously proved that if μ has continuous density with compact support, then $\alpha(\mu_n) \to \alpha(\mu)$ *a.s.*. We generalize the result to arbitrary μ. Demange [8][2.4(iii) and Proposition 10] argues that for single-peaked continuous densities μ the minimax sets of the μ_n converge to a single point which is the minimax set of μ. We generalizes this result as well. Yet the proof of Theorem 6 is much shorter and simpler. This confirms the appropriateness of this line of attack.

Theorem 6. Let μ be a probability measure on $\Re^m$. Let n points be randomly independently sampled from μ. Then the min-max majority value of the sample, $\alpha(\mu_n)$ converges to the distributional min-max majority $\alpha(\mu)$ almost surely. If in addition μ is continuous and possesses unique min-max winner point z, then the min-max winner of the sample converges *a.s.* to z.

Proof: If z is an α-majority point with respect to μ then by Theorem 5 it will be an $\alpha + \epsilon$-majority point for μ_n eventually, for any positive ϵ. Thus $\limsup\{\alpha(\mu_n)\} \leq \alpha(\mu)$. Conversely, for any $\beta < \alpha(\mu)$, set $\delta = \alpha(\mu) - \beta$. For all $x \in \Re^m$, there exists a hyperplane h_x through x such that a halfspace h_x^+ defined by h_x has mass $\mu(h_x^+) \le \beta + \delta$. Again by Theorem 5, the supremum of the fractional discrepancies over all these halfspaces converges to 0 $a.s.$ Thus,

$$\inf_x |\mu_n(n_x^+)| > \beta + \delta/2 \tag{3}$$

eventually, with probability 1 (a fraction of at least $\beta + \delta/2$ can be mustered against every point). Hence $\liminf_n\{\alpha(\mu_n)\} \ge \alpha(\mu)$. This proves the first part of Theorem 6.

The proof of the first part has moreover established that z has limiting minimal winning supermajority fraction α. It remains to show that no point other than z can also be winning with fraction α. Accordingly let $\epsilon > 0$ be arbitrary. Let $S \subset \Re^m$ be an enormous ball containing z and with $\mu(S) > \alpha$, so that eventually with probability 1 no point outside S can be an α-majority winner. Let T denote S with the small ball of radius ϵ around z removed, $T = S \setminus B(z, \epsilon)$. By the compactness of T and continuity of μ, there exists β such that the min-max majority over all $x \in T$ equals β. By the uniqueness of z, we have $\beta > \alpha$.

Then by the same argument as led to inequality (3), every point in S is at least a fraction $\delta = \beta - \alpha$ from being an α-majority winner. As n increases, eventually with probability 1 we have:

$$\inf_{x \in T} \mu_n(h_x^+) > \beta - \delta/2 > \alpha.$$

Hence eventually no point in T will be an α-majority winner ($a.s.$). This completes the proof. $\diamond$

Theorem 6 ensures convergence of $\alpha(\mu_n)$ holds for *any* distribution. This is of particular importance for empirical applications, because spatial voting data are often discrete. For example, the Senate data in [14] and other studies [24] are taken from roll call votes. Similarly, most public opinion polls ask yes/no questions or limit answers to integers in a small range (e.g. 1–5). In all these cases the original data will be discrete. Even if kernel smoothing ([23, pp. 35,42]) were employed the resulting distributions might not be continuous. Also notice the following: if *two* groups of samples were taken from μ, Theorem 5 would ensure the convergence of the two empirical measures to each other, and Theorem 6 would ensure the convergence of the min-max values to each other. This matches the scenario described in section 1, where information from polls or past voting records is used to predict an outcome.

In general, we consider a function(al) f whose domain is the set of probability measures and whose range is the reals. For example, f might be an indicator function for the event "0 is undominated", or f_i might be the ith coordinate of the center of mass of the distribution. When f is continuous, the uniform convergence of the empirical measure will ensure the convergence of $f(\mu_n)$ to $f(\mu)$.

Consider the indicator function just defined. It is not continuous, in the following sense: there exists $\epsilon > 0$ such that for all $\lambda > 0$, there exist empirical distributions μ_n and $\tilde{\mu}_n$ satifying

$$\sup_{d \in \mathcal{D}} |\mu_n(d) - \tilde{\mu}_n(d)| \le \lambda$$

but $|f(\mu_n) - f(\tilde{\mu}_n)| > \epsilon$. (Just take $\epsilon = .9$). That is, there are empirical distributions arbitrarily close in the sup norm sense, which have function values more than .9 apart. Moreover the discontinuity occurs just at the distributions of interest, where the fraction on one side of a hyperplane is 1/2. From a more general point of view, this explains the failure of finite behavior to converge to distributional behavior as discussed in sections 3 and 4. The mathematical guideline for convergence is the continuity of the functional. Let us attempt to formulate a less technical rule of thumb to give a general sense of how to make accurate predictions for finite populations based on distributional results: if the event or quantity of interest depends on the precise way voters are split among regions, then a convergence problem is apt to arise; if it relies instead on having a certain fraction or more in a region, then the result is apt to apply to the large finite case, possibly with the fraction perturbed slightly.

To illustrate, consider simple majority voting in arbitrary dimension $m > 1$. From a distributional point of view, weak centeredness is a barely sufficient condition to make the origin a core point. Think of a simple majority as a supermajority $\alpha(n) = \alpha n$ with $\alpha = \frac{1}{2}$. Our rule of thumb suggests that the fraction α should be perturbed slightly to ensure the existence of a core. This is confirmed by a result in [29], that for any $\alpha > \frac{1}{2}$ the probability of a core converges to 1 for α-majority voting on a sample of n points from any weakly centered distribution, as $n \to \infty$.

Now let us apply these observations to the question of whether the yolk radius converges to zero. Elsewhere the author has shown that if μ is continuous with a compact region of support, then the radius (respectively center) of the sample yolk converges to the radius (respectively center) of the distributional yolk [33] (respectively [34]). Under exactly which circumstances can we expect the yolk radius to be small? From a distributional point of view[2], a yolk radius of 0 corresponds to a nonempty core. Necessary and sufficient conditions for a nonempty core, in the distributional sense, are (see [5, 17]) that μ be *weakly centered*: every hyperplane through 0 is a median hyperplane. Therefore a distributional analysis predicts that weak centeredness would be necessary and sufficient for the yolk radius of random samples to converge to 0.

Our rule of thumb suggests that there may be a problem with the exact 50 : 50 split of the weak centeredness condition, but that a $(50 + \epsilon) : (50 - \epsilon)$ splitting condition would be apt to work. A companion paper by the author [32] finds that the true necessary and sufficient condition is that μ be *strictly centered*: for every hyperplane *not* passing through 0, the halfspace it defines not containing the origin must contain strictly less than half the population. This outcome seems well in accord with the guidelines proposed above.

We can invoke Theorem 5 to prove the sufficiency of strict centeredness[3], though under an additional assumption of continuity of the distribution μ. Despite the lessened generality of Theorem 7, the ease and brevity of its proof are noteworthy.

Theorem 7. Let n points be sampled independently from μ, a strictly centered continuous distribution on $\Re^m$. Then the radius of the yolk of the sample converges to 0 *a.s.* as $n \to \infty$.

Proof: Following the proof in [32], we show that the largest distance from 0 to any median hyperplane converges to 0. Since this distance is an upper bound on the yolk radius, the result will follow.

For any $x \neq 0$, let h_z^+ denote the halfspace not containing the origin defined by the hyperplane normal at x. By strict centeredness $\mu(h_z^+) < 1/2$. By continuity $\mu(h_z^+)$ is continuous in x.

Let $\epsilon > 0$ be arbitrary. Clearly the largest vote attained against 0 by points ϵ or more away

[2]this distributional analysis is due to Richard McKelvey

[3]the essentials of this proof were suggested to me independently by Robert Foley, Richard McKelvey, and Gideon Weiss.

from 0 is attained by points ϵ away, or more precisely

$$\sup_{\|x\|\geq\epsilon} \mu(h_z^+) = \sup_{\|x\|=\epsilon} \mu(h_z^+).$$

By compactness of the set over which the latter supremum is taken, and continuity, the supremum is attained. Thus there exists $\beta < 1/2$ such that for all $\|x\| \geq \epsilon$, we have $\mu(h_z^+) \leq \beta$.

The halfspaces h_z^+ are contained in the class $\mathcal{C}$. Let the n points be sampled from μ. Apply Theorem 5 to find that with probability 1, as n increases,

$$\mu_n(h_z^+) \leq \frac{\beta + 1/2}{2} < 1/2 \quad \forall \|x\| \geq \epsilon.$$

This implies that there is no median hyperplane at distance ϵ or more from 0, whence the result follows. $\diamond$

Interestingly, the early paper of Tullock [35] actually discusses finite configurations, and only appeals to the infinite configurations as an intuitive aid. For example, in describing a distributional model, Tullock writes [35, page 259]: "Cycles are, therefore, possible, but they would become less and less important as the number of choosing individuals increases." Here the claim is not that cycles won't usually exist in large populations, but rather that they *won't matter.* Tullock gives a variety of practical and mathematically intuitive justifications of this claim, including a reluctance to "split hairs"[p.261], and a tendency for new proposals to lead towards the center[pp.261–262], a loose foreshadowing of the yolk. Later research work such as the establishment of the yolk and its properties [18, 10], together with finite sample analysis as in [32, 33], has enabled precise statements and proofs of some of these insights.

7 Acknowledgments

The author thanks George Dantzig, Mahmoud El-Gamal, Robert Foley, Bob Parks, Richard McKelvey, Loren Platzman, Norman Schofield, Richard Stone, Gideon Weiss, Lyn Whitaker, and Kevin Wood, for invaluable discussions and correspondence, and the anonymous reviewers for helpful comments. The author also owes particular thanks to Bernie Grofman and Don Saari for their inspiration and encouragement.

References

[1] Arrow, Kenneth. "Tullock and an Existence Theorem," *Public Choice* **6** (1969) pages 105–112.

[2] Caplin, Andrew, and Barry Nalebuff. "On 64% majority rule." *Econometrica 56* (1988).

[3] Banks, Jeffrey S., John Duggan and Michel Le Breton. "Social choice and electoral competition in the general spatial model" *Journal of Economic Theory* **126**, (2006) pages 194–234.

[4] Banks, Jeffrey S., John Duggan and Michel Le Breton. "Bounds for Mixed Strategy Equilibria and the Spatial Model of Elections" *Journal of Economic Theory* **103**, Issue 1, (2002), pages 88–105

[5] Davis, Otto A., Morris H. DeGroot, and Melvin J. Hinich. Social Preference Orderings and Majority Rule. *Econometrica* **40** (1972) pages 147–157.

[6] Dvoretzky, A., P. Erdös,and S. Kakutani. "Nonincrease everywhere of the Brownian Motion process," *4th Berkeley Symposium Mathematics Statistics Probability* **2** (1961) pages 103–116.

[7] Davis, Otto, Melvin J. Hinich, and Peter Ordeshook. "An expository development of a mathematical model of the electoral process" *American Polit. Sci. Review* **64** (1970) pages 426–448.

[8] Demange, Gabrielle. "A limit theorem on the minmax set." *Journal of Mathematical Economics* **9** (1982), pages 145–164.

[9] Feller, William. *An Introduction to Probability Theory and its Applications* (1965), John Wiley, New York.

[10] Ferejohn, John A., Richard McKelvey, and Edward Packel. "Limiting Distributions for Continuous State Markov Models", Social Choice and Welfare **1** (1984) pages 45–67.

[11] Grandmont, Jean-Michel. "Intermediate Preferences and the majority rule", *Econometrica* **46** (1978), pages 317–330.

[12] Greenberg, Joseph. "Consistent majority rules over compact sets of alternatives." *Econometrica* **47** (1979), pages 627–636.

[13] Grunbaum, B. "Partitions of mass-distributions and convex bodies by hyperplanes." *Pacific Journal Math.* **10**, (1960) pages 1257–1261.

[14] Koehler, D.H. The Size of the Yolk: Computations for Odd and Even-Numbered Committees. Social Choice and Welfare **7** (1990) 231–245.

[15] Kramer, Gerald H. "A dynamical model of political equilibrium". *Journal of Economic Theory* **16** (1977) pages 310–334.

[16] McKelvey, Richard D. "Intransitivities in multidimensional voting models and some implications for agenda control" *Journal of Economic Theory* **12** (1976), pages 472–482.

[17] McKelvey, Richard D. (1979). General conditions for global intransitivies in formal voting models. *Econometrica* **47** (1979) pages 1085–1112.

[18] McKelvey, Richard D. Covering, dominance, and institution free properties of social choice. American Journal Political Science **30** (1986) pages 283–314.

[19] McKelvey, Richard D., Peter C. Ordeshook, and Peter Ungar. "Conditions for voting equilibria in continuous voter distributions,"*SIAM Journal Appl. Math.* **39** (1980), pages 161–168. 1980.

[20] McKelvey, Richard D., and Richard E. Wendell. "Voting equilibria in multidimensional choice spaces," *Mathematics of Operations Research* **1** (1976), pages 144–158.

[21] Peres, Yuval. “Points of increase for random walks,” *Israel Journal of Mathematics* **95** (1996), pages 341-347.

[22] Plott, Charles. “A notion of equilibrium and its possibility under majority rule,” *American Economic Review* **57** (1967), pages 787–806.

[23] Pollard, David. Convergence of Stochastic Processes. Springer-Verlag, New York 1984.

[24] Poole, Keith, and Howard Rosenthal. “Patterns of Congressional Voting”, *American Journal of Political Science* **35** (1991).

[25] Riker, W. and Peter Ordeshook. *An Introduction to Positive Political Theory.* (1973) Prentice-Hall, Englewood Cliffs, New Jersey.

[26] Rubenstein, Ariel. “A note about the ‘nowhere denseness’ of societies having an equilibrium under majority rule”, *Econometrica* **47** (1979) pages 511-514.

[27] Schofield, Norman J. “Instability of simple dynamic games,” *Review of Economic Studies* **45** (1978), pages 575–594.

[28] Schofield, Norman J. “Generic instability of Majority Rule”, *Review of Economic Studies* **50** (1983), pages 696–705.

[29] Schofield, Norman J., and Craig A. Tovey. “Probability and Convergence for supra-majority rule with Euclidean preferences”. *Mathematical and Computer Modelling* **16** pages 41–58 (1992).

[30] Schofield, Norman J. *Social Choice and Democracy.* Springer-Verlag, Berlin (1985).

[31] Tovey, Craig A. “The probability of an undominated central voter in 2-dimensional spatial majority voting”. *Social Choice and Welfare* **9**, No. 1 (1992), pages 43–48.

[32] Tovey, Craig A. “The almost surely shrinking yolk.” *Mathematical Social Sciences*, this issue (2009).

[33] Tovey, Craig A. “The instability of instability,” Naval Postgraduate School Technical Report NPSOR-91-15, Monterey, California, (1991), in review with *Mathematical Social Sciences.*

[34] Tovey, Craig A. “Some foundations for empirical study in the Euclidean spatial model of social choice”. In W. Barnett, M. Hinich and N. Schofield, editors, *Political Economy: Institutions Information and Competition* Cambridge University Press, (1993).

[35] Tullock, Gordon. The General Irrelevance of the General Impossibility Theorem, *The Quarterly Journal of Economics* **81** (1967), pages 256–270.

[36] Tullock, Gordon. *Toward a Mathematics of Politics.* University of Michigan Press, Ann Arbor (1967).

[37] Tullock, Gordon. “Why so much stability?”, *Public Choice* **37** (1981) pages 189–202.