Mathematical Social Sciences 59 (2010) 88–101

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

A critique of distributional analysis in the spatial model

Craig A. Tovey *

School of Industrial and Systems Engineering, College of Computing, Georgia Institute of Technology, Atlanta, Ga 30332-0205, USA

ARTICLE INFO

Article history:
Received 11 October 2006
Received in revised form
23 July 2009
Accepted 16 September 2009
Available online 29 September 2009

JEL classification:
B23
C62
D71
D72

Keywords:
Probability
Euclidean preference
Voting
Minimax
Convergence
Yolk

ABSTRACT

Distributional analysis is widely used to study social choice in Euclidean models (Tullock, 1967a,b; Arrow, 1969; Davis et al., 1972; Grandmont, 1978; McKelvey et al., 1980; Demange, 1982; Caplin and Nalebuff, 1988, e.g). This method assumes a continuum of voters distributed according to a probability measure. Since infinite populations do not exist, the goal of distributional analysis is to give an insight into the behavior of large finite populations. However, the 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 (Arrow, 1969) 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 Caplin and Nalebuff (1988). Uniform convergence of empirical measures (Pollard, 1984) 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 (Tullock, 1967a,b; Arrow, 1969; Davis et al., 1972; Grandmont, 1978; McKelvey et al., 1980; Demange, 1982; Schofield, 1985; Caplin and Nalebuff, 1988, e.g.) (see also Davis et al. (1970) and Riker and Ordeshook (1973) 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

* Tel.: +1 404 894 3034; fax: +1 404 894 2301.
E-mail address: ctovey@isye.gatech.edu.

0165-4896/$ – see front matter
doi:10.1016/j.mathsocsci.2009.09.005

analysis must be to give an 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 limiting case is to give an 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, the 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 Caplin and Nalebuff (1988). 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 is as 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 Arrow (1969) 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