Aggregated statistical rankings are arbitrary

Deanna B. Haunsperger
Department of Mathematics and Computer Science, Carleton College, Northfield, MN 55057, USA (e-mail: dhaunspe@mathcs.carleton.edu)

Abstract. In many areas of mathematics, statistics, and the social sciences, the intriguing, and somewhat unsettling, paradox occurs where the "parts" may give rise to a common decision, but the aggregate of those parts, the "whole", gives rise to a different decision. The Kruskal-Wallis nonparametric statistical test on \( n \) samples which can be used to rank-order a list of alternatives is subject to such a Simpson-like paradox of aggregation. That is, two or more data sets each may individually support a certain ordering of the samples under Kruskal-Wallis, yet their union, or aggregate, yields a different outcome. An analysis of this phenomenon yields a computable criterion which characterizes which matrices of ranked data, when aggregated, can give rise to such a paradox.

1 Introduction

Studies of the interesting and surprising outcomes that can occur when trying to draw a conclusion about the society by aggregating a number of individuals' preferences (rank orders, expert opinions, data sets, or the like) make for a rich and interesting literature in statistics, mathematical economics, and the social and decision sciences. An example of this behavior is often called an "aggregation paradox" because the "decision" of the aggregate seems contradictory to the decisions of the individual parts.

Social choice theory, which dates back at least to the mathematical theory of elections by Borda [5] and Condorcet [7], concerns itself with aggregating individuals' preferences to yield a collective choice (see [8], [25], [26], [27], or for a combinatorial and matrix-theoretic viewpoint, see [14]). A resurgence of interest and discussion in the area occurred following Arrow's "General
Possibility Theorem'' (see [1], [16]) in which it was shown that given certain reasonable, seemingly-innocuous conditions, there can be no ideally rational method for aggregating individual preferences.

Another interesting problem arises when aggregating a number of expert opinions which have been expressed in some numerical form (perhaps as a subjective probability distribution) to reflect individual uncertainty. Examples include members of a jury or partners in a firm trying to reach a collective decision. For information on Condorcet's Jury Theorem (which gives a condition under which the aggregation of the opinions of individuals with identical preferences is superior to dictatorship), see, for example, [2]. For a survey of the theoretical developments and an annotated bibliography of aggregation, see [9]. For a discussion of how one might aggregate a number of preferences or rank orders which occur probabilistically, see [17] and its cited references.

Recent developments in the study of aggregation paradoxes include several mathematical theories developed to understand these paradoxes from a geometric viewpoint [21] and a dynamical systems viewpoint using symbolic dynamics [23].

In statistics, a paradox that continues to draw attention is Simpson's (or Yule's) paradox, where the data for each of the "parts" support a common decision, but the data for the "whole" imply a different conclusion. (See, for example, [3], [4], [6], [10], [13], [18], [19], [21], [28], [29].) Certain statistical tests, such as the Kruskal-Wallis [15] nonparametric test, endow a set of data with an ordering of the candidates (samples, subjects, treatments). The Kruskal-Wallis test was created to determine the ordering of items whose values are observed random vectors, sampled independently from the same distribution. These random vectors may be of different lengths. Here, we consider the special case where the random vectors are also of the same length. In this case, the Kruskal-Wallis test can be interpreted as a projection of the data onto the order statistic of a standard uniform distribution. The Kruskal-Wallis test is subject to two distinct types of paradoxes: projection and the Simpson-like paradox of aggregation.

If a given data set is restricted to some subset of the candidates and the ordering induced by the test on this subset differs from the restriction of the ordering induced by the tests on the whole set, a projection paradox has occurred. In [11], a relationship between the Kruskal-Wallis test, $KW$, and the Borda method of voting using preferences is exploited to discover that $KW$ minimizes, over a large class of nonparametric tests, the number and kind of projection paradoxes which can occur. Also in [11], using results from [20] and [22], these paradoxes for $KW$, the Bhapkar V-test, the Deshpande class L of tests (which has, among its members, a statistical decision procedure which corresponds to the plurality method of voting and hence has all of the projection paradoxes that are related to plurality voting) and others are completely characterized. Analogous results for contingency tables, complete block designs, and the Borda method of voting are also exploited in [11].