Partially Ordered Sets in Socio-Economic Data Analysis

Subjects: Mathematics, Applied

Contributor: Marco Fattore , Lucio De Capitani

A *partially ordered set* (or a *poset*, for short) is a set endowed with a *partial order relation*, i.e., with a *reflexive*, *anti-symmetric*, and *transitive* binary relation. As mathematical objects, posets have been intensively studied in the last century, coming to play essential roles in pure mathematics, logic, and theoretical computer science. More recently, they have been increasingly employed in data analysis, multi-criteria decision-making, and social sciences, particularly for building synthetic indicators and extracting rankings from multidimensional systems of ordinal data. Posets naturally represent systems and phenomena where some elements can be compared and ordered, while others cannot be and are then *incomparable*. This makes them a powerful data structure to describe collections of units assessed against multidimensional variable systems, preserving the nuanced and multi-faceted nature of the underlying domains. Moreover, poset theory collects the proper mathematical tools to treat ordinal data, fully respecting their non-numerical nature, and to extract information out of order relations, providing the proper setting for the statistical analysis of multidimensional ordinal data. Currently, their use is expanding both to solve open methodological issues in ordinal data analysis and to address evaluation problems in socio-economic sciences, from multidimensional poverty, well-being, or quality-of-life assessment to the measurement of financial literacy, from the construction of knowledge spaces in mathematical psychology and education theory to the measurement of multidimensional ordinal inequality/polarization.

incomparability formal concept analysis Hasse diagram lattice multidimensional ordinal data multi-indicator system order relation ranking socio-economic assessment synthetic indicators

Partially ordered sets (*posets*) are relevant in socio-economic studies and data analysis for both conceptual and technical reasons. Phenomena like poverty, well-being, and sustainability are complex, nuanced, and multi-faceted, and trying to reduce them to a single point of view is at odds with their nature. For example, not only can individuals be poor with different intensity, but also in alternative and different ways that cannot be compared and that require different actions to be contrasted. Any statistical representation of this kind of trait must preserve such nuances and irreducible facets, and posets that allow for both *comparabilities* and *incomparabilities* between objects/units are the proper data structure to do that. Moreover, in social sciences, data are often of an ordinal kind and cannot be treated through the usual metric tools employed in the statistical analysis of numerical variables. *Order theory*, a branch of discrete mathematics designed to deal with order relations and order structures, provides the concepts and the tools for dealing with this kind of input and is the natural setting for the multidimensional analysis of ordinal data.

Historically, posets emerged in pure mathematics in the XIX century and appear in many mathematical areas as a ubiquitous structure. In particular, lattices (a special kind of partially ordered sets, whose systematic and general study solidified around 1930) play a key role in abstract algebra and in logic. In the 1960s, posets also became relevant in theoretical computer science, where they provide the mathematical framework to formalize the semantics of computer programs. Progressively, they were recognized not only as pure mathematical objects but also as a tool to describe, model, and represent problems and data in different scientific domains. They made their appearance in applied data analysis and social sciences in the 1980s (although some traces of them can be found even before), becoming a relevant tool in socio-economic statistics at the beginning of this century. This occurred mainly in connection with the problem of building synthetic indicators and rankings out of multidimensional variable systems, avoiding both the aggregative/compensative procedures, typical of composite indicators, and the inconsistency of treating ordinal data as numerical ones (a quite usual practice in data analysis). The increasing awareness of the complexity of socio-economic phenomena and the necessity of capturing and representing it also contributed to raising interest in posets. Advances in the study of their mathematical properties and improvements in computational efficiency led poset theory to intertwine with multi-criteria decision-making and the development of evaluation tools for multi-faceted traits. At the same time, order theory began to be acknowledged as the proper mathematical foundation of multidimensional ordinal data analysis, in perspective making it an asset of methodological statistics.