A statistical depth function for non-standard data based on formal concept analysis

Statistical depth functions measure outlyingness and depth of a data point with respect to a sampled data cloud or an underlying distribution. Based on the theory of formal concept analysis, we introduce a general notion of statistical depth functions for non-standard data. The definition relies to a large extent on conceptual scaling that is a suitable transformation of the sampled data to a set of binary attributes. We obtain a closure system on the power set of the sample which can be uniquely described by formal implications between subsets of data points. Thereby, a premise of an implication can be considered as a characterization of the shared attributes of the premise, the conclusion or the single observations itself.

We reduce this set, consisting of many redundant implications, by introducing what we call the *union-free generic family of implications*. This is still a sufficient description of the closure system. The family contains implications where the premise is minimal and the conclusion is maximal. The main idea is that we delete also those implications that can be divided into proper smaller implications. Thereby, the premise equals the union of the proper premises and the same is true for the conclusion. These deleted implications can be seen as generalized or redundant descriptions of the conclusion and the observations. Based on the union-free generic family of implications, the introduced statistical depth function is defined as the proportion of the conclusions containing the data point of interest.

Since this is a more general definition, concepts like *monotonicity relative to deepest point* and *vanishing at infinity* have to be revisited.

To illustrate the generality of this approach, we apply the statistical depth function on different data types. For nominal data the deepest point corresponds to the modus of the sampled data, and for ordinal data we obtain the median as the deepest point. A more complex example are data points in a finite dimensional Euclidean space. In this case, the statistical depth function is similar to the simplicial depth.