Learning from rank data

Elja Arjas

University of Oslo, Norway, elja.arjas@helsinki.fi

Comparing and ranking of items according to some selected measure of quality, strength or performance is a natural way of collecting and organizing information in many areas of practical interest. The Mallows rank model (1957), a member of the exponential family, would offer an attractive alternative for describing and analysing such rank data; however, apart from situations in which the number of items is very small, the computational complexity of the Mallows model has limited its use to a particular form based on Kendall distance. In this talk I consider computationally tractable methods for Bayesian inference in Mallows models, which apply for any right-invariant metric on the space of permutations. The proposed method performs inference on the consensus ranking of the items, also when based on data on only partial rankings such as top-k items or pairwise comparisons. When the population of assessors is heterogeneous, a mixture model for clustering the assessors into homogeneous subgroups is proposed, with cluster-specific consensus rankings. An approximate stochastic sampling algorithm is introduced, enabling a fully probabilistic analysis, including quantification of the uncertainties involved. In particular, individual preferences can be predicted in situations in which they were missing in the data, and assessors can be assigned to classes after they had ranked only a subset of the items. The talk is based on joint work with Valeria Vitelli, Øystein Sørensen and Arnoldo Frigessi.