



Department of Decision Sciences  
Statistics Seminars

## Bayesian rank-based clustering of high-dimensional data with the Mallows rank model

**Valeria Vitelli**

University of Oslo

Thursday, 21<sup>st</sup> November 2024

12:00 pm **Room 3-E4-SR03** Via Roentgen 1 Milano

### Abstract

Rank-based models can be used to estimate individual behaviours and preferences in several areas, such as marketing and politics. The Mallows model is a popular model for rankings, as it flexibly adapts to different types of preference data, and the previously proposed Bayesian Mallows Model (BMM) offers a computationally efficient framework for Bayesian inference also allowing capturing the users' heterogeneity, via a finite mixture. However, the Mallows model does not scale, and also seems unrealistic, when the pool of items is large. Furthermore, BMM does not currently allow the use of additional user-related information (covariates), which can potentially lead to a better accuracy in individual predictions, by enhancing the understanding of the users' personal profiles.

In this talk, the author will introduce recent extensions of BMM in a number of directions: (i) embedding covariate information related to assessors, via a similarity function that a priori favours the aggregation of people into a cluster when their covariates are similar. (ii) Then, combining a lower-dimensional version of BMM (lowBMM), that scales to large datasets, with a Bayesian mixture of Mallows models, and showcasing its performance in the context of cancer genomics. (iii) Finally, proposing a variational inferential setting to help with scalability.

This is joint work with Luca Coraggio, Emilie Eliseussen, Arnaldo Frigessi, Haakon Mugerud, and Ida Scheel.