

International Society for Bayesian Analysis, 9<sup>th</sup> World Meeting,  
Hamilton Island, Australia, 2008.

---

## CLUSTERING AND VARIABLE SELECTION USING FISHER DISTRIBUTIONS

Yanan Fan<sup>1\*</sup>

<sup>1</sup> University of New South Wales, Sydney, Australia

\*Y.Fan@unsw.edu.au

High dimensional data such as those found in data mining and microarray gene expression experiments are often inherently directional. We present a novel approach to model-based clustering of high dimensional data via the use of a mixture of Fisher distributions. The proposed method carries out simultaneous variable selection and clustering. The resulting clustering depends on the amount of correlation between observations given the selected variables. A Bayesian approach is adopted, where the determination of the number of clusters, cluster allocation and variable selection is carried out simultaneously via the use of trans-dimensional Markov chain Monte Carlo.