

HARVARD UNIVERSITY
DEPARTMENT OF STATISTICS



Tel. 617-495-5496
Fax. 617-496-8057
statdept@stat.harvard.edu
<http://www.stat.harvard.edu/>

Science Center 7th Fl.
One Oxford Street
Cambridge, MA 02138

STATISTICS COLLOQUIUM

MONDAY, NOVEMBER 9, 2015
TALK: 4:15 PM — SCIENCE CENTER 300H
RECEPTION: 3:50 PM

“Fast Bayesian Factor Analysis via Automatic Rotations to Sparsity”

[Veronika Rockova](#)
Statistics Department
The Wharton School
University of Pennsylvania

ABSTRACT

Rotational post-hoc transformations have traditionally played a key role in enhancing the interpretability of factor analysis. Regularization methods also serve to achieve this goal by prioritizing sparse loading matrices. In this work, we bridge these two paradigms with a unifying Bayesian framework. Our approach deploys intermediate factor rotations throughout the learning process, greatly enhancing the effectiveness of sparsity inducing priors. These automatic rotations to sparsity are embedded within a PXL-EM algorithm, a Bayesian variant of parameter-expanded EM for posterior mode detection. By iterating between soft-thresholding of small factor loadings and transformations of the factor basis, we obtain (a) dramatic accelerations, (b) robustness against poor initializations and (c) better oriented sparse solutions. To avoid the pre-specification of the factor cardinality, we extend the loading matrix to have infinite number of columns with the Indian Buffet Process (IBP) prior. The factor dimensionality is learned from the posterior, which is shown to concentrate on sparse matrices. Our deployment of PXL-EM performs a dynamic posterior exploration, outputting a solution path indexed by a sequence of spike-and-slab priors. For accurate recovery of the factor loadings, we deploy the Spike-and-Slab LASSO prior, a two- component refinement of the Laplace prior (Rockova (2015)). A companion criterion, motivated as an integral lower bound, is provided to effectively select the best recovery. The potential of the proposed procedure is demonstrated on both simulated and real high-dimensional data, which would render posterior simulation impractical. This is joint work with Ed George.