

# Variational Inference for Probabilistic Poisson PCA

StatMathAppli2017 - September 2017

*Julien Chiquet<sup>1</sup>, Mahendra Mariadassou<sup>2</sup>, Stéphane Robin<sup>1</sup>*

<sup>1</sup> *MIA 518, AgroParitech/INRA, Université Paris-Saclay, 75005 Paris, France*

<sup>2</sup> *MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France*

Many application domains such as ecology or genomics have to deal with multivariate non Gaussian observations. A typical example is the joint observation of the respective abundances of a set of species in a series of sites, aimed at understanding the co-variations between these species. The Gaussian setting provides a canonical way to model such dependencies, but does not apply in general to such data.

We consider here the multivariate exponential family framework for which we introduce a generic hierarchical model with multivariate Gaussian latent variables. This model can be seen as an extension of probabilistic Principal Component Analysis (pPCA) to non Gaussian settings and enables us to account for covariates and offsets at little additional cost.

Unlike the purely Gaussian setting, the likelihood is generally not tractable in this framework. We resort instead to a variational approximation for parameter inference and solve the corresponding optimization problem using gradient descent. Formal expression of the gradient depends on the exponential family at hand and does not always have an analytical expression. However, coordinates of the gradient depend only on one dimensional integrals of the form  $\int_{\mathbb{R}} f(x)e^{-x^2} dx$  for smooth functions  $f$ , which can be evaluated efficiently using Gauss-Hermite quadrature.

We then focus on the case of the Poisson-lognormal model, for which both the variational approximation and its gradient have closed-form expressions, in the context of microbial ecology. We illustrate the importance of accounting for offsets and covariates on two datasets. Finally, we sketch some promising extensions of the framework, most notably to inference of co-occurrence networks.

Chiquet, Julien, Mahendra Mariadassou, and Stéphane Robin. 2017. "Variational Inference for Probabilistic Poisson Pca." *ArXiv E-Prints*, March. <https://arxiv.org/abs/1703.06633>.