

# Overdispersed Black-Box Variational Inference: Supplement

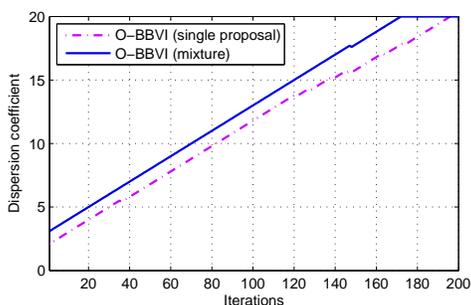
**Francisco J. R. Ruiz**  
Data Science Institute  
Dept. of Computer Science  
Columbia University

**Michalis K. Titsias**  
Dept. of Informatics  
Athens University of  
Economics and Business

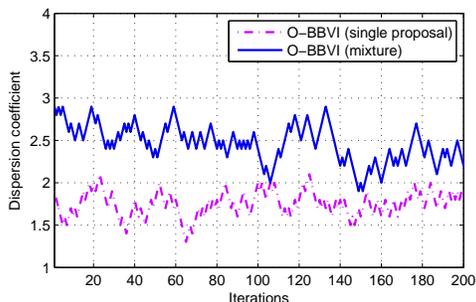
**David M. Blei**  
Data Science Institute  
Dept. of Computer Science and Statistics  
Columbia University

## 1 EMPIRICAL STUDY: DISPERSION COEFFICIENTS

As an example, we show here some plots of the evolution of the dispersion coefficients  $\tau_{nj}$ . Specifically, we plot  $\tau_{nj}$  for the two overdispersed black-box variational inference (O-BBVI) methods under study, namely, the method with a single proposal ( $J = 1$ ) and the mixture method that makes use of two proposals ( $J = 2$ ). For the latter, we only plot  $\tau_{n2}$ , since  $\tau_{n1}$  is fixed to one.



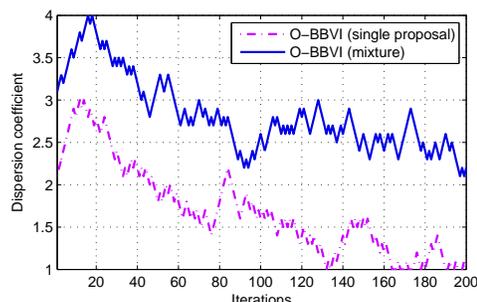
(a) Dispersion coefficient for the Gaussian distribution over  $w_{11}$ .



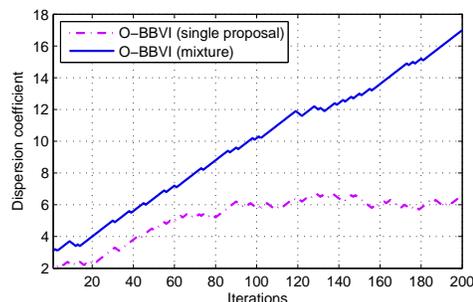
(b) Dispersion coefficient for the gamma distribution over  $z_{111}$ .

Figure 1: Evolution of the dispersion coefficients for the gamma-normal time series model (GN-TS) model. The single-proposal algorithm uses a lower dispersion coefficient.

We show<sup>1</sup> in Figures 1 and 2 the evolution of  $\tau_{nj}$  for the two considered models: the GN-TS and the Poisson deep exponential family (DEF), respectively. We only plot the first 200 iterations of the algorithm. For each model, we choose two specific latent variables (shown in the caption of the Figures), covering all the different variational distributions that we make use of (Gaussian, gamma, and Poisson). As expected, the dispersion coefficient for the mixture case is higher than in the single-proposal case.



(a) Dispersion coefficient for the gamma distribution over  $w_{11}^{(0)}$ .



(b) Dispersion coefficient for the Poisson distribution over  $z_{11}^{(1)}$ .

Figure 2: Evolution of the dispersion coefficients for the Poisson DEF model. The single-proposal algorithm uses a lower dispersion coefficient.

<sup>1</sup>We truncated  $\tau_{nj} \leq 20$  to avoid numerical instabilities.