Copula variational inference

Dustin Tran[1,2], David M. Blei[2], and Edoardo M. Airoldi[1]

[1] Harvard University, [2] Columbia University

Goal

We aim to do **scalable** and **generic** Bayesian inference:

$$p(\mathbf{z} \mid \mathbf{x}) \approx q(\mathbf{z}; \boldsymbol{\lambda})$$

- Mean-field VI is fast but highly biased, underestimates the variance, and is sensitive to local optima
- Structured VI incorporates dependency but requires explicit knowledge of model and is difficult to construct

Our approach automatically **learns the dependency structure** within a **black box** framework, and **generalizes** both approaches.

Background

Variational inference minimizes $\mathrm{KL}(q||p)$ by maximizing the ELBO

$$\mathcal{L}(\boldsymbol{\lambda}) = \mathbb{E}_{q(\mathbf{z};\boldsymbol{\lambda})}[\log p(\mathbf{x}, \mathbf{z}) - \log q(\mathbf{z}; \boldsymbol{\lambda})].$$

Any random variable $\mathbf{z} = \{\mathbf{z}_1, \dots, \mathbf{z}_d\} \sim q$ can be factorized as

$$q(\mathbf{z}) = \left[\prod_{i=1}^{d} q(\mathbf{z}_i)\right] c(q(\mathbf{z}_1), \dots, q(\mathbf{z}_d)), \tag{1}$$

where c is a joint density known as the **copula**. For example, the bivariate Gaussian copula is

$$c(\mathbf{u}_1, \mathbf{u}_2; \rho) = \Phi_\rho(\Phi^{-1}(\mathbf{u}_1), \Phi^{-1}(\mathbf{u}_2)), \tag{2}$$

which corresponds to the Pearson correlation ρ between z_1 and z_2. One can factorize a multivariate copula, for example, as

$$c(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4) = \Big[c(\mathbf{u}_1, \mathbf{u}_3) c(\mathbf{u}_2, \mathbf{u}_3) c(\mathbf{u}_3, \mathbf{u}_4)\Big]\Big[c(\mathbf{u}_1, \mathbf{u}_2|\mathbf{u}_3) c(\mathbf{u}_1, \mathbf{u}_4|\mathbf{u}_3)\Big] \Big[c(\mathbf{u}_2, \mathbf{u}_4|\mathbf{u}_1, \mathbf{u}_3)\Big],$$

where each pair copula can be of a different family.
We learn a choice of this factorization and perform model selection to choose the parametric family for each pair copula. This provides us very **flexible** models of the dependency structure.

Method

Let $\boldsymbol{\lambda}$ be the original parameters (mean-field or structured) and $\boldsymbol{\eta}$ be the augmented parameters (copula). Consider the factorization of the variational distribution

$$q(\mathbf{z}; \boldsymbol{\lambda}, \boldsymbol{\eta}) = \underbrace{\left[\prod_{i=1}^{d} q(\mathbf{z}_i; \boldsymbol{\lambda})\right]}_{\text{mean-field}} \underbrace{c(q(\mathbf{z}_1; \boldsymbol{\lambda}), \dots, q(\mathbf{z}_d; \boldsymbol{\lambda}); \boldsymbol{\eta})}_{\text{copula}}.$$

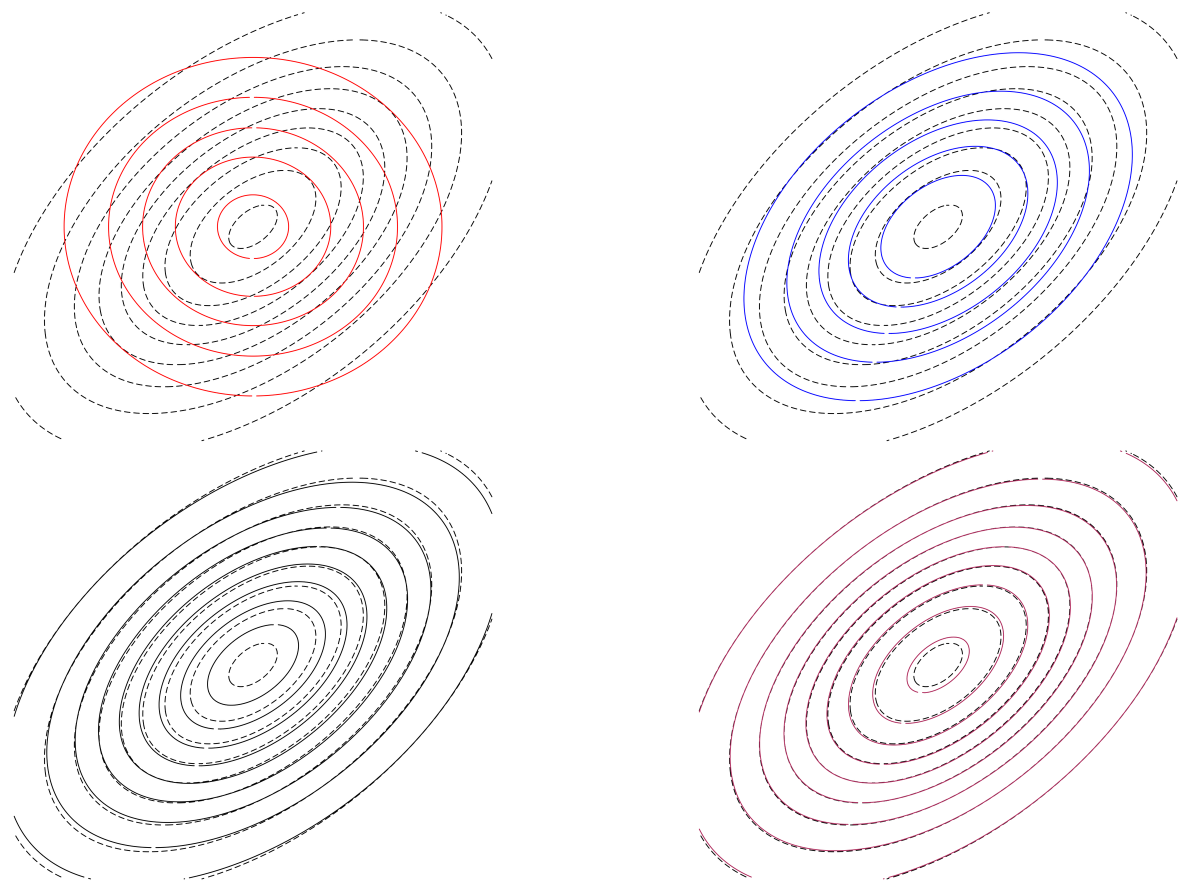

Figure 1: Approximations to an elliptical Gaussian. First step (red) runs mean-field; second step (blue) fits a copula; third (green) refits the mean-field; fourth (cyan) refits the copula.

Algorithm 1: Copula variational inference (COPULA VI)

Input: Data $\mathbf{x}$, Model $p(\mathbf{x}, \mathbf{z})$, Variational family q.
Initialize $\boldsymbol{\lambda}$ randomly, $\boldsymbol{\eta}$ so that c is uniform.
while *change in* ELBO *is above some threshold* **do**
  // Fix $\boldsymbol{\eta}$, maximize over $\boldsymbol{\lambda}$.
  Set iteration counter $t = 1$.
  while *not converged* **do**
    Draw sample $\mathbf{u} \sim \mathrm{Unif}([0,1]^d)$.
    Update $\boldsymbol{\lambda} = \boldsymbol{\lambda} + \rho_t \nabla_{\boldsymbol{\lambda}} \mathcal{L}$. (Eq.5, Eq.6)
    Increment t.
  end
  // Fix $\boldsymbol{\lambda}$, maximize over $\boldsymbol{\eta}$.
  Set iteration counter $t = 1$.
  while *not converged* **do**
    Draw sample $\mathbf{u} \sim \mathrm{Unif}([0,1]^d)$.
    Update $\boldsymbol{\eta} = \boldsymbol{\eta} + \rho_t \nabla_{\boldsymbol{\eta}} \mathcal{L}$. (Eq.7)
    Increment t.
  end
end
Output: Variational parameters $(\boldsymbol{\lambda}, \boldsymbol{\eta})$.

Experiments

Gaussian mixture model

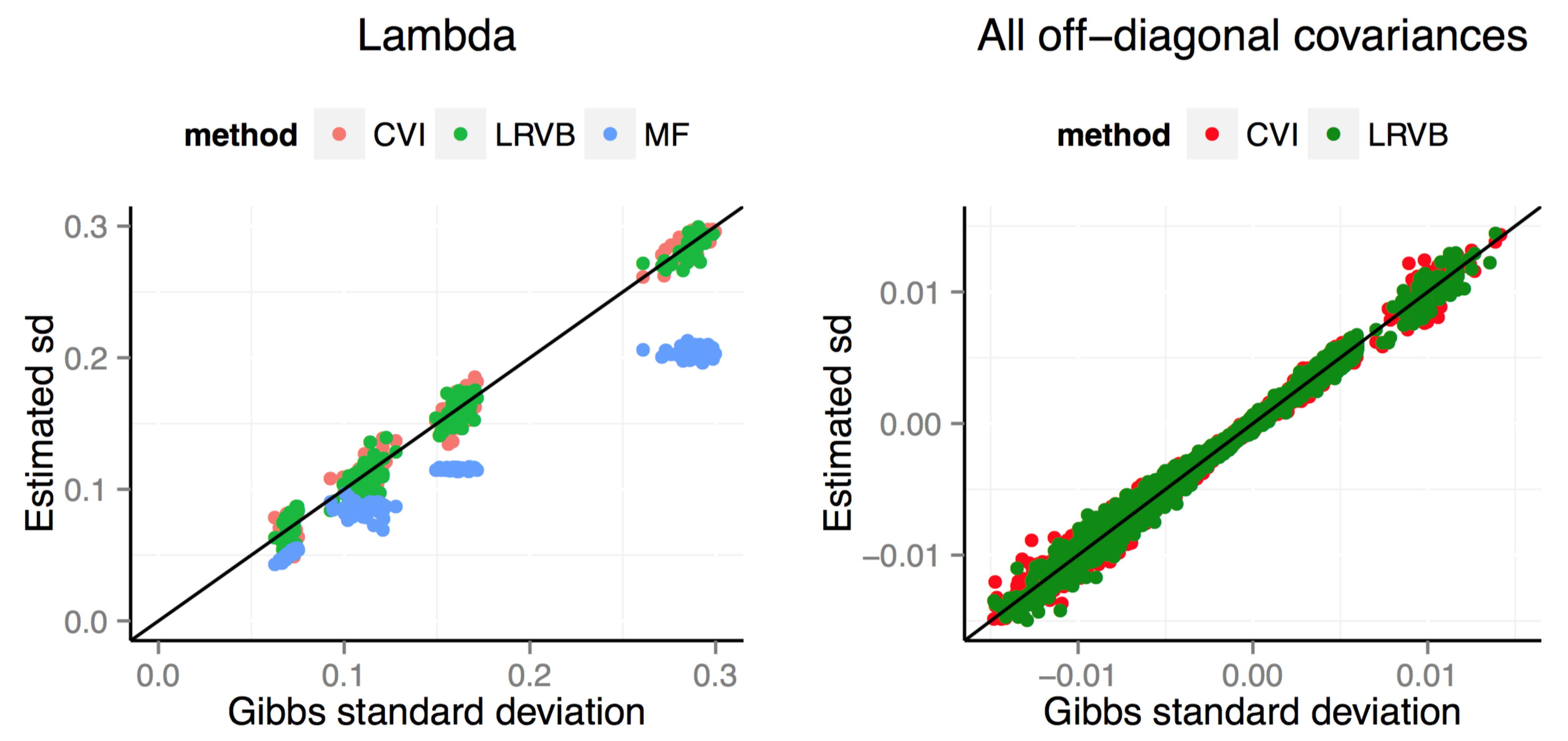

Figure 2: 10,000 samples, 2 mixture components, and 2 dimensional Gaussian distributions.

Latent space model

Variational inference methods	Predictive Likelihood	Runtime
Mean-field	-383.2	15 min.
LRVB	-330.5	38 min.
COPULA VI (2 steps)	-303.2	32 min.
COPULA VI (5 steps)	-80.2	1 hr. 17 min.
COPULA VI (converged)	-50.5	2 hr.

Table 1: 100,000 node network with with latent node attributes from a $K = 10$ dimensional normal distribution $\mathbf{z}_n \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$.

COPULA VI dominates mean-field and linear response variational Bayes (LRVB) in accuracy, is less sensitive to local optima and hyperparameters, and is more robust than both methods.

References

[1] Dustin Tran, David M. Blei, and Edoardo M. Airoldi.
Copula variational inference.
In *Neural Information Processing Systems*, 2015.

[2] Rajesh Ranganath, Sean Gerrish, and David M. Blei.
Black box variational inference.
In *Artificial Intelligence and Statistics*, 2014.

[3] Rajesh Ranganath, Dustin Tran, and David M. Blei.
Hierarchical variational models.
arXiv preprint arXiv:1511.02386, 2015.