Variational inference with copula augmentation

Dustin Tran*, David M. Blei†, and Edoardo M. Airoldi*

*Harvard University, †Columbia University

Goal

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

$$p(z \mid x) \approx q(z \mid \lambda)$$

- 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

$$\mathcal{L}(\lambda) \equiv \underbrace{\mathbb{E}_q[\log p(x,z)]}_{\text{energy}} \underbrace{-\mathbb{E}_q[\log q(z \mid \lambda)]}_{\text{entropy}}$$

Any random variable $z = \{z_1, \ldots, z_d\} \sim q$ can be factorized as

$$q(z) = \left[\prod_{i=1}^{d} q(z_i)\right] c(Q(z_1), \ldots, Q(z_d)) \tag{1}$$

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

$$c^{\text{Gauss}}(u_1, u_2; \rho) \equiv \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \tag{2}$$

which corresponds to the Pearson correlation ρ between z_1 and z_2. One can factorize a multivariate copula, e.g., as follows:

$$\begin{aligned} c(u_1, u_2, u_3, u_4, u_5) = {} & c(u_1, u_5)c(u_1, u_3)c(u_2, u_3)c(u_3, u_4) \cdot \\ & c(u_1, u_2|u_3)c(u_3, u_5|u_1)c(u_1, u_4|u_3) \cdot \\ & c(u_2, u_4|u_1, u_3)c(u_2, u_5|u_1, u_3) \cdot \\ & c(u_4, u_5|u_1, u_2, u_3) \end{aligned}$$

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

Methodology

Let λ be the original parameters (mean-field or structured) and η be the augmented parameters (copula). Consider the factorization of the variational distribution

$$q(z \mid \lambda, \eta) = \underbrace{\left[\prod_{i=1}^{d} q(z_i \mid \lambda)\right]}_{\text{mean-field}} \underbrace{c(Q(z_1 \mid \lambda), \ldots, Q(z_d \mid \lambda); \eta)}_{\text{copula}}$$

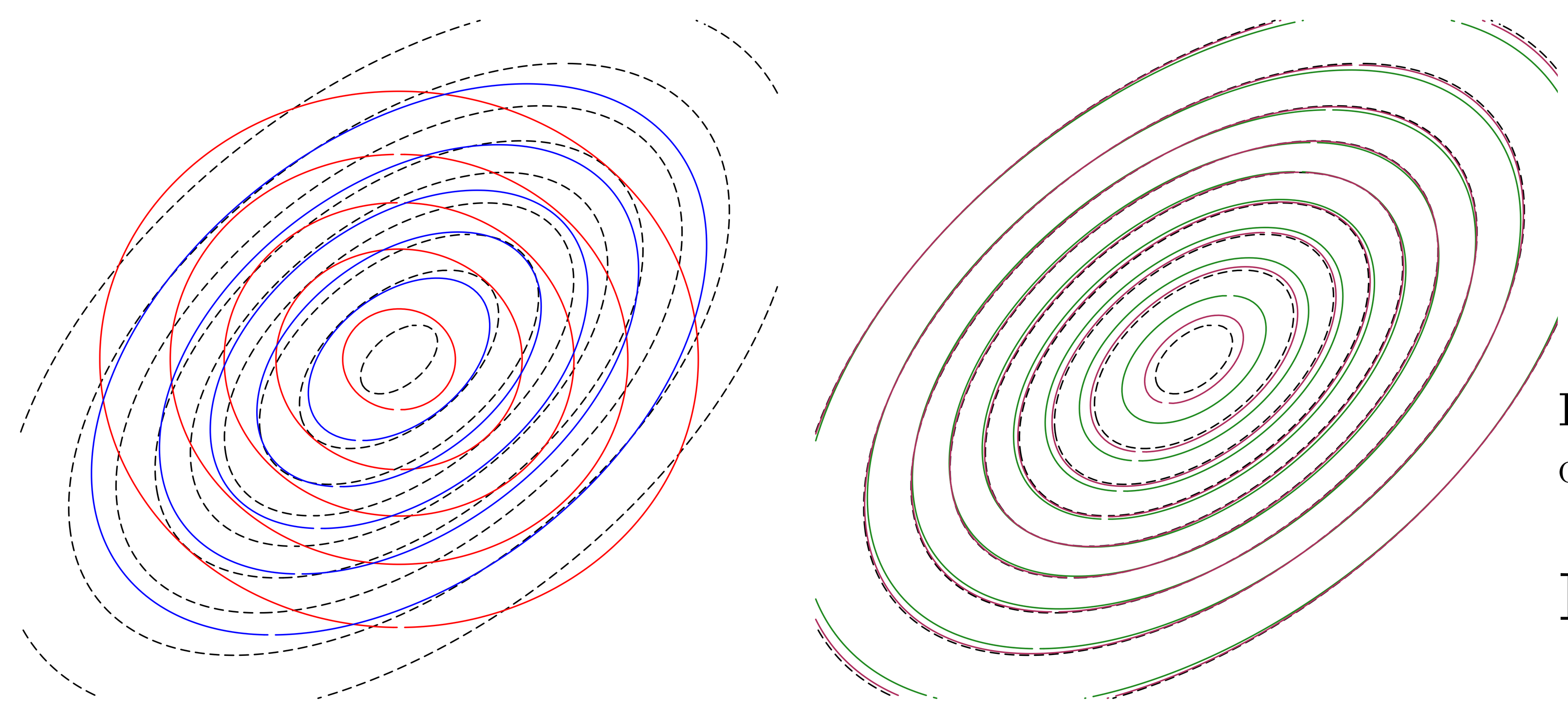

Figure: Variational approximations to an elliptical Gaussian posterior (black). First step (red on left) runs mean-field; second step (blue on left) fits a copula; third (green on right) refits the mean-field; fourth (purple on right) refits the copula.

Algorithm 1 Copula variational inference (CVI)

Input: data x, joint distribution p, variational family q
Initialize λ randomly, η so that c is uniform
repeat
  // Fix η, maximize over λ
  Initialize $t = 1$
  repeat
    $z \sim q(z \mid \lambda, \eta)$
    $\lambda = \lambda + \rho_t(\nabla_\lambda \log q(z \mid \lambda, \eta) \cdot (\log p(x, z) - \log q(z \mid \lambda, \eta)))$
    $t = t + 1$
  until change of λ is less than a small ϵ_1
  // Fix λ, maximize over η
  Initialize $t = 1$
  repeat
    $z \sim q(z \mid \lambda, \eta)$
    $\eta = \eta + \rho_t(\nabla_\eta \log q(z \mid \lambda, \eta) \cdot (\log p(x, z) - \log q(z \mid \lambda, \eta)))$
    $t = t + 1$
  until change of η is less than a small ϵ_2
until change of λ and η is less than a small ϵ_1 and ϵ_2 respectively

Experiments

Gaussian mixture model

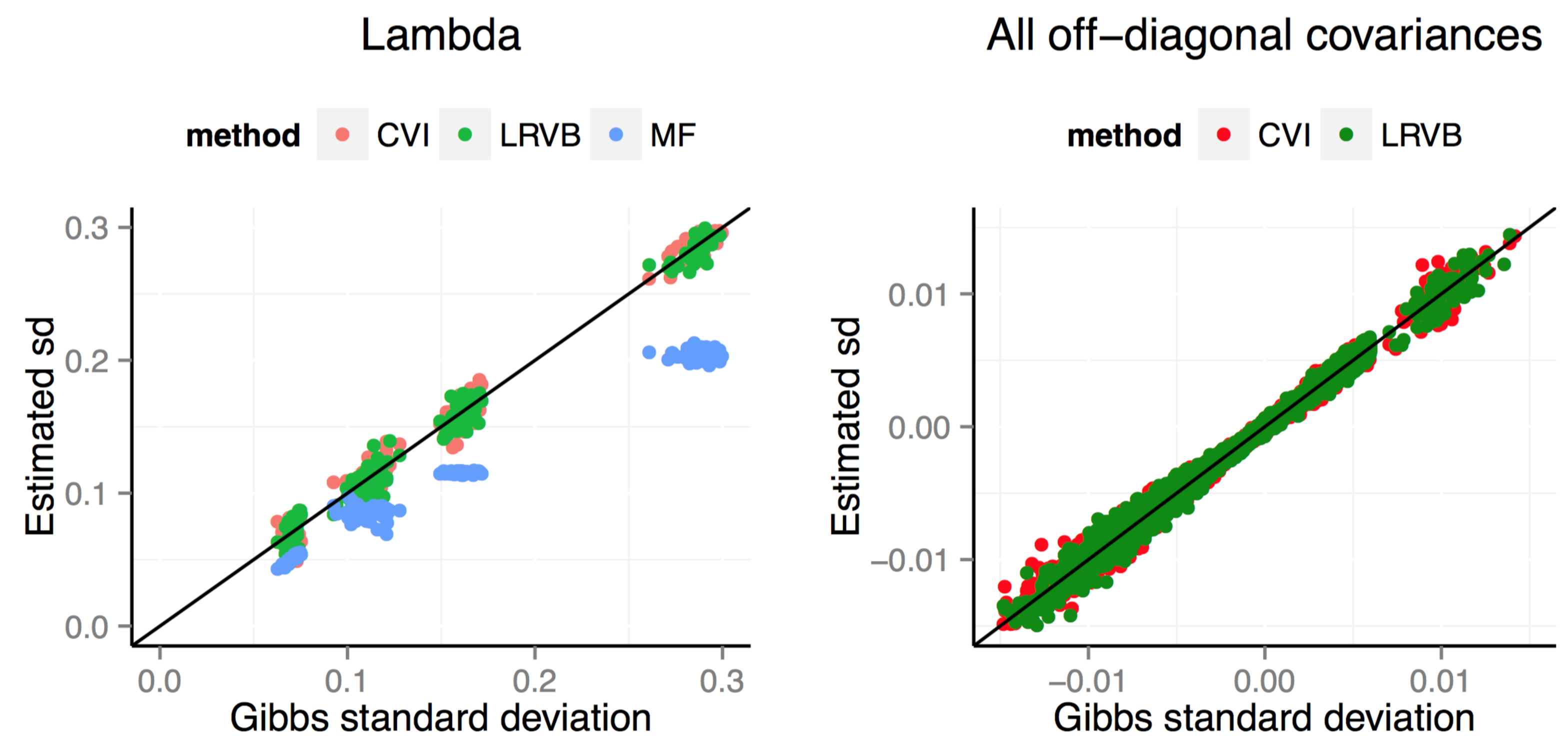

Figure: 10,000 samples, 2 mixture components, and 2 dimensional Gaussian distributions.

Latent space model

Variational inference methods	Predictive Likelihood
Mean-field	-383.2
LRVB	-330.5
CVI (2 iterations)	-303.2
CVI (5 iterations)	-80.2
CVI (converged)	-50.5

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

CVI 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] Alp Kucukelbir, Rajesh Ranganath, Andrew Gelman, and David M. Blei. Automatic variational inference in stan. *arXiv:1506.03431*, 2015.

[2] Dustin Tran, David M. Blei, and Edoardo M. Airoldi. Variational inference with copula augmentation. *arXiv:1506.03159*, 2015.

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