Variational Gaussian Copula Inference
Supplementary Material

Shaobo Han       Xuejun Liao       David B. Dunson1       Lawrence Carin
Department of ECE, Department of Statistical Science1, Duke University, Durham, NC 27708, USA

A: KL Additive Decomposition

Letting the variational proposal in Sklar’s representation be $q_{VC}(x) = c(u) \prod_{j=1}^{p} f_j(x_j)$, and the true posterior be $p(x | y) = c'(v) \prod_{j} f_j^p(x_j)$, where $u = F(x) = [F_1(x_1), \ldots, F_p(x_p)], v = F^*(x) = [F_1^*(x_1), \ldots, F_p^*(x_p)]$. The KL divergence decomposes into additive terms,

$$
\begin{align*}
\text{KL}(q(x) | p(x | y)) &= \int q(x) \left( \log \frac{q(x)}{p(x | y)} \right) dx \\
&= \int c[F(x)] \prod_j f_j(x_j) \left( \log \frac{c[F(x)]}{c[F^*(x)]} \right) dx \\
&= \int c[F(x)] \left( \log \frac{c(u)}{c[F^*(u)]} \right) du \\
&= \text{KL}(c(u) || \text{c'}[F^*(F^{-1}(u))]).
\end{align*}
$$

The first term in [1]

$$
\int c[F(x)] \left( \log \frac{c[F(x)]}{c[F^*(x)]} \right) \prod_j dF_j(x_j) \\
= \int c(u) \left( \log \frac{c(u)}{c[F^*(F^{-1}(u))]} \right) du \\
= \text{KL}(c(u) || \text{c'}[F^*(F^{-1}(u))]),
$$

The second term in [1]

$$
\begin{align*}
\int c[F(x)] \prod_j f_j(x_j) \left( \log \prod_j f_j(x_j) \right) \prod_j dx_j \\
= \sum_j \int c[F(x)] \prod_j f_j(x_j) \left( \log f_j(x_j) / f_j^p(x_j) \right) \prod_j dx_j \\
= \sum_j \int f_j(x_j) \left( \log f_j(x_j) / f_j^p(x_j) \right) dx_j \text{ (Marginal Closed Property)} \\
= \sum_j \text{KL}(f_j(x_j) || f_j^p(x_j)).
\end{align*}
$$

Therefore

$$
\begin{align*}
\text{KL}(q(x) || p(x | y)) &= \text{KL}(c(u) || c'[F^*(F^{-1}(u))]) \\
&+ \sum_j \text{KL}(f_j(x_j) || f_j^p(x_j)) \quad (2)
\end{align*}
$$

B: Model-Specific Derivations

B1: Skew Normal Distribution

1. $\ln p(x) \propto \ln \phi(x) + \ln \Phi(\alpha x)$ and $\partial \ln p(x) / \partial x = -x + \alpha \phi(\alpha x) / \Phi(\alpha x)$, $\alpha$ is the shape parameter

2. $\Psi(x)$ is predefined as CDF of $N(0,1)$

B2: Student’s t Distribution

1. $\ln p(x) \propto -(\nu + 1)/2 \ln (1 + x^2 / \nu)$ and $\partial \ln p(x) / \partial x = -(\nu + 1)x / (\nu + x^2)$, $\nu > 0$ is the degrees of freedom

2. $\Psi(x)$ is predefined as CDF of $N(0,1)$

B3: Gamma Distribution

1. $\ln p(x) \propto (\alpha - 1) \ln x - \beta x$ and $\partial \ln p(x) / \partial x = (\alpha - 1)x - \beta$, $\alpha$ is the shape parameter, $\beta$ is the rate parameter

2. $\Psi(x)$ is predefined as CDF of Exp(1)

B4: Beta Distribution

1. $\ln p(x) \propto (a - 1) \ln x + (b - 1) \ln (1 - x)$ and $\partial \ln p(x) / \partial x = (a - 1)x - (b - 1)/(1 - x)$, both $a, b > 0$

2. $\Psi(x)$ is predefined as CDF of Beta(2,2)
### B5: Bivariate Log-Normal

1. \( \ln p(x_1, x_2) \propto -\ln x_1 - \ln x_2 - \zeta/2 \) and

\[
\begin{align*}
\frac{\partial \ln f(x_1, x_2)}{\partial x_1} &= -\frac{1}{x_1} - \frac{\alpha_1(x_1) - \rho \alpha_2(x_2)}{(1 - \rho^2)x_1 \sigma_1} \\
\frac{\partial \ln f(x_1, x_2)}{\partial x_2} &= -\frac{1}{x_2} - \frac{\alpha_2(x_2) - \rho \alpha_1(x_1)}{(1 - \rho^2)x_2 \sigma_2}
\end{align*}
\]

2. \( \Psi(x) \) is pre-defined as CDF of \( \text{Exp}(1) \)

### C. Derivations in the Horseshoe Shrinkage Model

The equivalent hierarchical model is

\[
y|\tau \sim \mathcal{N}(0, \tau), \quad \tau|\gamma \sim \text{InvGa}(0.5, \gamma), \quad \gamma \sim \text{Ga}(0.5, 1)
\]

### C1: Gibbs Sampler

The full conditional posterior distributions are

\[
p(\tau|y, \gamma) = \text{InvGa}(1, y^2/2 + \gamma), \quad p(\gamma|\tau) = \text{Ga}(1, \tau^{-1} + 1)
\]

### C2: Mean-field Variational Bayes

The ELBO under MFVB is

\[
\mathcal{L}_{\text{MFVB}}[\nu_{\text{VB}}(\tau, \gamma)] = \mathbb{E}_{q(\tau)q(\gamma)}[\ln p(y, \tau, \gamma)] + H_1[q(\tau; \alpha_1, \beta_1)] + H_2[q(\gamma; \alpha_2, \beta_2)]
\]

where

\[
\begin{align*}
\mathbb{E}_{q(\tau)q(\gamma)}[\ln p(y, \tau, \gamma)] &= -0.5 \ln (2\pi) - 2 \ln (\Gamma(0.5)) - 2 \ln \tau - y^2 \left(\tau^{-1}\right) / 2 - \gamma \left(\tau^{-1}\right) - \gamma \\
H_1[q(\tau; \alpha_1, \beta_1)] &= \alpha_1 + \ln \beta_1 + \ln [\Gamma(\alpha_1)] - (1 + \alpha_1) \psi(\alpha_1) \\
H_2[q(\gamma; \alpha_2, \beta_2)] &= \alpha_2 - \ln \beta_2 + \ln [\Gamma(\alpha_2)] - (1 - \alpha_2) \psi(\beta_2)
\end{align*}
\]

The variational distribution

\[
q(\tau) = I\mathcal{G}(\tau; \alpha_1, \beta_1) = I\mathcal{G}(\tau; 1, y^2/2 + \gamma), \\
q(\gamma) = \mathcal{G}(\gamma; \alpha_2, \beta_2) = \mathcal{G}(\gamma, 1, (\tau^{-1}) + 1)
\]

where

\[
(\ln \tau) = \ln \beta_1 - \psi(\alpha_1) = \ln \left(y^2/2 + \gamma\right) - \psi(1), \\
\langle \tau^{-1} \rangle = \frac{\alpha_1}{\beta_1}, \quad \langle \gamma \rangle = \frac{\alpha_2}{\beta_2} = (\langle \tau^{-1} \rangle + 1)
\]

### C3: Deterministic VGC-LN

Denoting \( x = (x_1, x_2) = (\tau, \gamma) \), we construct a variational Gaussian copula proposal with (1) a bivariate Gaussian copula, and (2) fixed-form margin for both \( x_1 = \tau \in (0, \infty) \) and \( x_2 = \gamma \in (0, \infty) \); we employ

\[
\begin{align*}
 f_j(x_j; \mu_j, \sigma_j^2) &= \mathcal{L}\mathcal{N}(x_j; \mu_j, \sigma_j^2), \\
j = 1, 2, \quad \gamma_j = \exp(\tilde{z_j}) = \exp(\sigma_j \tilde{z_j} + \mu_j), \\
\end{align*}
\]

The ELBO of VGC-LN is

\[
\begin{align*}
\mathcal{L}_{\text{VGC}}(\mu, c) &= c_1 - \mu_1 + \mu_2 - \frac{y^2 \exp \left(- \mu_1 + c_1^2 \right)}{2} \\
- \ell_0 - \exp \left( \mu_2 + \frac{C_{21}^2 + C_{22}^2}{2} \right) + \ln |c| \\
\ell_0 &= \exp \left( \left(\mu_2 - \mu_1\right) + \frac{C_{11}^2 - 2 C_{12} C_{21} + C_{21}^2 + C_{22}^2}{2}\right)
\end{align*}
\]

where \( c_0 = -0.5 \ln (2\pi) - 2 \ln \Gamma(0.5) \), \( c_1 = c_0 + \ln (2\pi \epsilon) \). The gradients are

\[
\begin{align*}
\frac{\partial \mathcal{L}_{\text{VGC}}(\mu, C)}{\partial C_1} &= -1 + \frac{y^2}{2} \exp \left( \frac{C_{11}^2}{2} - \mu_1 \right) + \ell_0 \\
\frac{\partial \mathcal{L}_{\text{VGC}}(\mu, C)}{\partial C_2} &= 1 - \ell_0 - \exp \left( \mu_2 + \frac{C_{21}^2 + C_{22}^2}{2} \right)
\end{align*}
\]

### C4: Stochastic VGC-LN

The stochastic part of the ELBO is,

\[
\ell_s(\tilde{z}) = c_0 + \tilde{z}_2 - \tilde{z}_1 - \frac{y^2 \exp \left(- \tilde{z}_1 \right)}{2} - \exp(\tilde{z}_2 - \tilde{z}_1) - \exp(\tilde{z}_1)
\]

and

\[
\begin{align*}
\nabla_{\tilde{z}_1} \ell_s(\tilde{z}) &= -1 + \frac{y^2 \exp \left(- \tilde{z}_1 \right)}{2} + \exp(\tilde{z}_2 - \tilde{z}_1) \\
\nabla_{\tilde{z}_2} \ell_s(\tilde{z}) &= 1 - \exp(\tilde{z}_2 - \tilde{z}_1) - \exp(\tilde{z}_1)
\end{align*}
\]

### C5: Stochastic VGC-BP

1. \( \ln p(y, x_1, x_2) = c_0 - 2 \ln x_1 - y^2/(2x_1) - x_2/x_1 - x_2, \)

\[
\frac{\partial \ln p(y, x_1, x_2)}{\partial x_1} = -2/x_1 + y^2/(2x_1^2) + x_2/x_1^2, \\
\frac{\partial \ln p(y, x_1, x_2)}{\partial x_2} = -1/x_1 - 1
\]

2. \( \Psi(x) \) is pre-defined as CDF of \( \text{Exp}(0.01) \).
D. Derivations in Poisson Log Linear Regression

For $i = 1, \ldots, n$, the hierarchical model is

$$y_i \sim \text{Poisson}(\mu_i), \quad \log(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2,$$

$\beta_0 \sim N(0, \tau), \quad \beta_1 \sim N(0, \tau), \quad \beta_2 \sim N(0, \tau), \quad \tau \sim \text{Ga}(1, 1)$

The log likelihood and prior,

$$\ln p(y, \beta, \tau) = \sum_{i=1}^{n} \ln p(y_i | \beta) + \ln N(\beta_0; 0, \tau) + \ln N(\beta_1; 0, \tau)$$

$$+ \ln N(\beta_2; 0, \tau) + \ln \text{Ga}(\tau; 1, 1)$$

where $\ln p(y_i | \beta) = y_i \ln \mu_i - \mu_i - \ln y_i!$, and $\mu_i = \exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2)$.

The derivatives are

$$\frac{\partial \ln p(y, \beta, \tau)}{\partial \beta_0} = \sum_{i=1}^{n} (y_i - \mu_i) - \tau^{-1} \beta_0$$

$$\frac{\partial \ln p(y, \beta, \tau)}{\partial \beta_1} = \sum_{i=1}^{n} x_i (y_i - \mu_i) - \tau^{-1} \beta_1$$

$$\frac{\partial \ln p(y, \beta, \tau)}{\partial \beta_2} = \sum_{i=1}^{n} x_i^2 (y_i - \mu_i) - \tau^{-1} \beta_2$$

$$\frac{\partial \ln p(y, \beta, \tau)}{\partial \tau} = -\frac{3}{2\tau} + \frac{\beta_0^2 + \beta_1^2 + \beta_2^2}{2\tau^2} + \frac{a_0 - 1}{\tau} - b_0$$