Algorithms for Approximating the Normalizing Constant of the G-Wishart Distribution

Bokgyeong Kang

Pennsylvania State University

bxk487@psu.edu

December 3, 2019

Gaussian Graphical Models

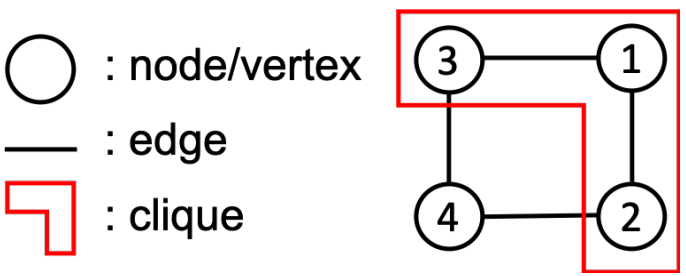
Useful: Understanding dependence relationship between variables.

- Let G=(V,E) be an undirected graph.
- The vertices V is associated with a p-dimensional vector $\mathbf{X}$,

$$\mathbf{X} = (X_1, \cdots, X_p)^\top \sim N_p(0, \Sigma), \quad K = \Sigma^{-1},$$

where $K_{ij} = 0$ if X_i and X_j are conditionally independent.

Example: A 4-dimensional cycle graph.



$$\mathbf{X} \sim N_5(0, \Sigma)$$

$$K = \begin{pmatrix} * & * & * & 0 \\ * & * & 0 & * \\ * & 0 & * & * \\ 0 & * & * & * \end{pmatrix}$$

The G-Wishart Distributions

$$\mathbf{X} = (X_1, \cdots, X_p)^\top \sim N_p(0, \Sigma), \quad K = \Sigma^{-1},$$

Interest: Bayesian inference on K.

The G-Wishart distribution is a conjugate prior for K.

$$\text{(prior)} \quad K|G \sim W_G(\delta, D) \tag{1}$$
$$\text{(posterior)} \quad K|(X, G) \sim W_G(\delta^*, D^*) \tag{2}$$

Computational challenge: If $K \sim W_G(\delta, D)$,

$$P(K|G) = \frac{1}{I_G(\delta, D)} exp\{h_{\delta,D}(K)\}. \quad \text{Intractible!}$$

Goal of this project: Compare two existing methods to approximate it.

Monte Carlo method *vs.* **Laplace method**

Monte Carlo Approximation (Atay-Kayis and Massam, 2005)

Goal: Approximate $I_G(\delta, D) = \int e^{h_{\delta,D}(K)} dK$

Theorem: Atay-Kayis and Massam (2005)

$$I_G(\delta, D) = C_{\delta,D} E\left[f_D(\psi_{\mathcal{V}})\right]$$
$$\psi_{ii} \sim \sqrt{\chi^2_{\delta,\nu_i}}, \quad i = 1, \cdots, p,$$
$$\psi_{ij} \sim N(0,1), \quad (i,j) \in \{(i,j) \in \mathcal{V} : i \neq j\},$$

where $\mathcal{V}$ is the set of indices of the free elements.

Idea: Use Monte Carlo method.

1. Generate $\psi_{ii}^{(1)}, \cdots, \psi_{ii}^{(N)}$ from $\sqrt{\chi^2_{\delta,\nu_i}}, \quad i = 1, \cdots, p$
2. Generate $\psi_{ij}^{(1)}, \cdots, \psi_{ij}^{(N)}$ from $N(0,1), \quad (i,j) \in \{(i,j) \in \mathcal{V} : i \neq j\}$

$$\hat{I}_G^{mc}(\delta, D) = C_{\delta,D} \frac{1}{N} \sum_{\ell=1}^{N} f_D(\psi_{\mathcal{V}}^{(\ell)})$$

Laplace Approximation (Lenkoski and Dobra, 2011)

Motivation: In MC, the number of iterations relies on the dimension.

Goal: Efficiently approximate $I_G(\delta, D) = \int e^{h_{\delta,D}(K)} \prod_{(i,j)\in\mathcal{V}} dK_{ij}$.

Idea: Use Taylor's expansion and Gaussian assumption to approximate integrals of the form $\int e^{h(x)} dx$.

Suppose $h(x)$ has an unique global maximum at $\hat{x}$.

$$h(x) \approx h(\hat{x}) - \frac{1}{2}|h''(\hat{x})|(x-\hat{x})^2 \quad \because \text{Taylor's theorem} \tag{3}$$

$$\int e^{h(x)} dx \approx e^{h(\hat{x})} \int e^{-\frac{1}{2}|h''(\hat{x})|(x-\hat{x})^2} dx \tag{4}$$

$$\approx (2\pi)^{1/2} |h''(\hat{x})|^{-1/2} e^{h(\hat{x})} \quad \because x \sim N(\hat{x}, |h''(\hat{x})|^{-1}) \tag{5}$$

Approximation to $I_G(\delta, D)$:

$$\hat{I}_G^{lap}(\delta, D) = (2\pi)^{|\mathcal{V}|/2} [\det|H_\delta(\hat{\mathbf{K}})|]^{-1/2} e^{h_{\delta,D}(\hat{\mathbf{K}})} = J_{\delta,D}(\hat{\mathbf{K}})$$

Laplace Approximation (Lenkoski and Dobra, 2011)

Approximation to $I_G(\delta, D)$:

$$\hat{I}_G^{lap}(\delta, D) = (2\pi)^{|\mathcal{V}|/2}[\det|H_\delta(\hat{\mathbf{K}})|]^{-1/2} e^{h_{\delta,D}(\hat{\mathbf{K}})} = J_{\delta,D}(\hat{\mathbf{K}}),$$

where $\hat{\mathbf{K}}$ is the mode of $h_{\delta,D}(K)$ (i.e., the mode of $W_G(\delta, D)$) and is obtained by an iterative algorithm called "iterative proportional algorithm".

Properties of the Laplace approximation:

- The accuracy of the approximation depends on the degree to which the density resembles a Gaussian distribution.

  - We assumed that $\mathbf{K} \sim N(\hat{\mathbf{K}}, |H_\delta(\hat{\mathbf{K}})|^{-1})$.
  - As δ increases, the variance for each K_{ij} decreases.

- The number of iterations does not depend on the dimension, p.

Simulation Study

Simulation settings: Approximation to $I_G(\delta, D)$

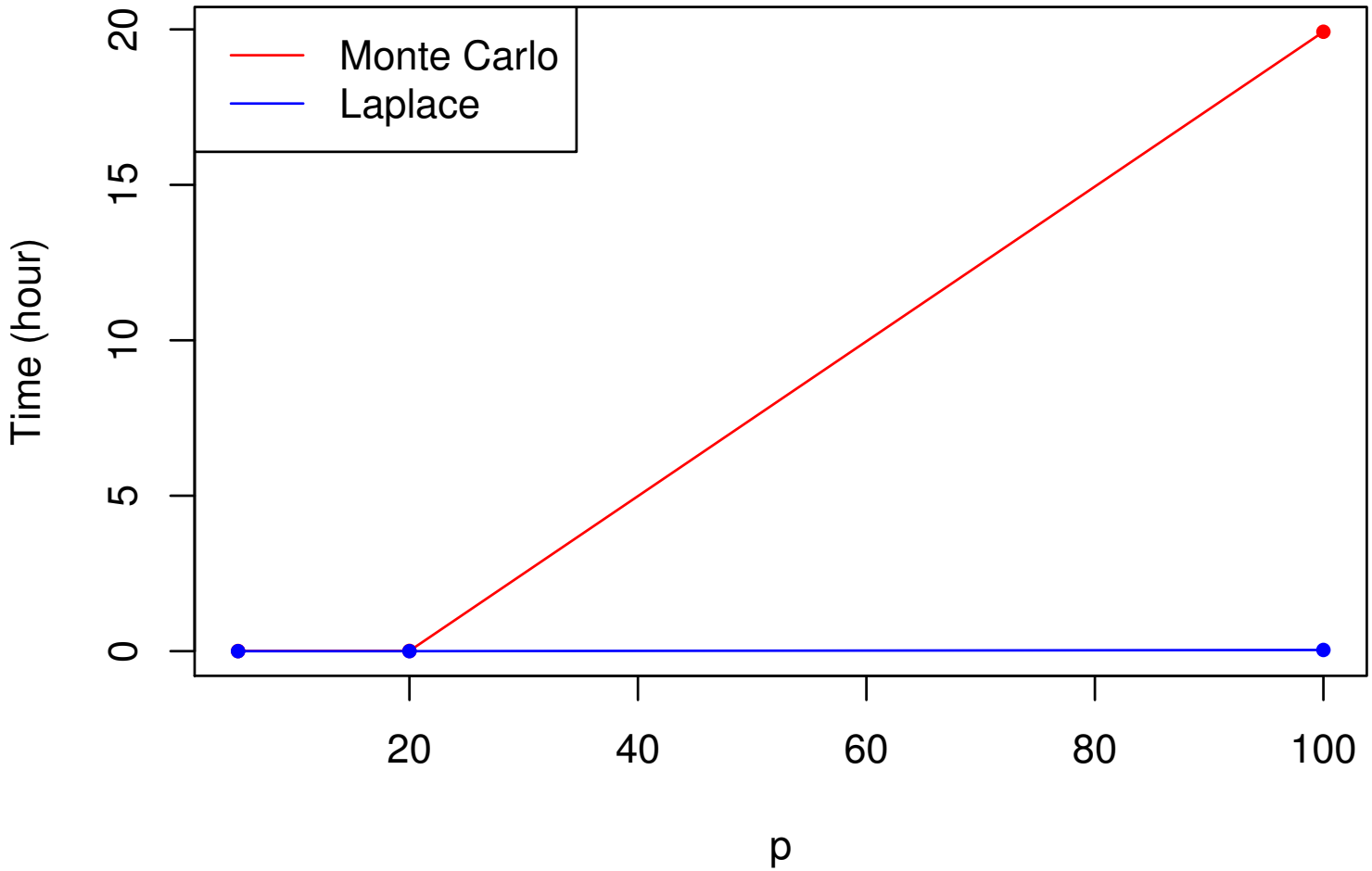
- $\delta \in \{3, 10, 25, 50, 100\}$
- $p \in \{5, 20, 100, 200, 500\}$
- G is a p-dimensional cycle graph
- D is an arbitrary matrix that allows the normalizing constant to be finite.
- N = 25,000 (MC sample size)

As δ increases: when $p = 5$

	δ					Time (sec)
	3	10	25	50	100	
$log\,\hat{I}_G^{mc}(\delta, D)$	14.201 (8e-04)	50.570 (3e-04)	166.236 (9e-06)	407.693 (3e-08)	975.966 (2e-14)	14.104
$log\,\hat{I}_G^{lap}(\delta, D)$	11.068	50.044	166.149	407.618	977.033	0.069

Simulation Study

As p increases:



Conclusion

Based on the simulation study:

- The Monte Carlo method is accurate but computationally demanding for a high-dimensional dataset.
- The Laplace method is computationally efficient even for the high dimension.
- The Laplace method approximates $I_G(\delta, D)$ better for larger values of δ.

Recommendation:

- For small δ, Monte Carlo method.
- For large δ or very high-dimensional data, Laplace method.