Accelerating Bayesian Structure Learning in Sparse Gaussian Graphical Models

Reza Mohammadi
Department of Business Analytics, University of Amsterdam
and
Hélène Massam *
Department of Mathematics and Statistics, York University
and
Gérard Letac
Laboratoire de Statistique et Probabilités, Université Paul Sabatier

October 18, 2021

The supplementary material contains technical proofs and additional simulation results and organized as follows:

- Section A provides a proposition which is needed in Section 3.1 of the manuscript.
- Section B includes the prove of Lemma 1 for reformation of the ratio of normalizing constants.
- Section C includes the prove of Lemma 2 for the ratio of normalizing constants in the case of the disjoint paths.
- Section D contains the prove of Theorem 1 as the first of our two main theoretical results regarding the error made of our approximation.
- Section E contains the prove of Theorem 2 as the second main theoretical result.
- Section F provides a pseudo-code for computing the ratio I_1/I_2 in Theorem 2.
- Section G includes additional results for the simulations in Sections 5.1 and 6 of the manuscript.

*This work is dedicated to the memory of Hélène Massam who passed away on August 22d, 2020.

A Proposition

The following proposition is used to compute $\mathbb{E}\left(e^{-\frac{A^2}{2}}\right)$.

Proposition A.1. *Let $U_1, \ldots, U_k$, $V_1 \ldots, V_k$, and Q be independent random variables such that U_i and V_i are $N(0,1)$ and $Q \sim \chi^2_\delta$. Then*

$$E\left(e^{-\frac{1}{2Q}\left(\sum_{i=1}^k U_iV_i\right)^2}\right) = \frac{\Gamma\left(\frac{\delta+k}{2}\right)\Gamma\left(\frac{\delta+1}{2}\right)}{\Gamma\left(\frac{\delta}{2}\right)\Gamma\left(\frac{\delta+k+1}{2}\right)}.$$

Proof. We have

$$\sum_{i=1}^k U_iV_i \sim \left(U_1^2 + \cdots + U_k^2\right)^{\frac{1}{2}} V_1 = X^{\frac{1}{2}} V_1, \tag{1}$$

where $X \sim \Gamma(\frac{k}{2}, \frac{1}{2})$ and is independent of V_1. To prove Equation 1 is enough to compute the Laplace transforms of both sides. From Equation 1, we now get

$$E\left(e^{-\frac{1}{2Q}\left(\sum_{i=1}^k U_iV_i\right)^2}\right) = E\left(e^{-\frac{XV_1^2}{2Q}}\right) = E\left(\left(1+\frac{X}{Q}\right)^{\frac{-1}{2}}\right),$$

where the last equality is due to integrating with regard to variable V_1. Since $X \sim \Gamma(\frac{k}{2}, \frac{1}{2})$ and $Q \sim \Gamma(\frac{d}{2}, \frac{1}{2})$, we have $U = \frac{X}{Q} \sim B_2\left(\frac{k}{2}, \frac{d}{2}\right)$ which is a Beta distribution of second kind. Thus

$$E\left(\left(1+\frac{X}{Q}\right)^{\frac{-1}{2}}\right) = \frac{\Gamma\left(\frac{\delta+k}{2}\right)\Gamma\left(\frac{\delta+1}{2}\right)}{\Gamma\left(\frac{\delta}{2}\right)\Gamma\left(\frac{\delta+k+1}{2}\right)}.$$

B Proof of Lemma 1

By considering $\Psi_\cap^- = \{\psi_{ij} : (i,j) \in E \setminus (E_q \cap E_p)\}$, we have

$$\mathbb{E}\left(e^{-\frac{D}{2}-\frac{(A+b)^2}{2}}\right) = \mathbb{E}\left(e^{-\frac{D}{2}}\mathbb{E}\left(e^{-\frac{(A+b)^2}{2}}\middle|\Psi_\cap^-\right)\right).$$

Note that D and b_1 are $\Psi_\cap^-$-measurable, that is, are functions of the elements of $\Psi_\cap^-$ only. Due to Equation 1, we have

$$\mathbb{E}\left(e^{\frac{-(A_1+b_1)^2}{2\psi_{qq}}}\middle|\Psi_\cap^-\right) = \mathbb{E}\left(e^{-\frac{1}{2}\frac{(UV+b_1)^2}{\psi_{qq}}}\middle|\Psi_\cap^-\right)$$

where U, V, and ψ_{qq} are independent random variables such that $U^2 \sim \chi^2_d$, $V \sim N(0,1)$, and $\psi_{qq} \sim \chi^2_\delta$. Integrating with respect to $V \sim N(0,1)$, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{-\frac{1}{2}\frac{(UV+b_1)^2}{\psi_{qq}}}\middle|\Psi_\cap^-\right) &= \mathbb{E}\left(e^{-\frac{1}{2}\frac{b_1^2}{U^2+\psi_{qq}}}\left(\frac{\psi_{qq}}{U^2+\psi_{qq}}\right)^{\frac{1}{2}}\middle|\Psi_\cap^-\right) \\ &= \mathbb{E}\left(\sqrt{B}\right)\mathbb{E}\left(e^{\frac{-b_1^2}{2Y}}\middle|\Psi_\cap^-\right),\end{aligned}$$

where $B = \frac{\psi_{qq}}{U^2+\psi_{qq}} \sim Beta(\frac{\delta}{2}, \frac{d}{2})$ and is independent of $Y = U^2 + \psi_{qq} \sim \chi^2_{\delta+d}$. Thus

$$\mathbb{E}\left(\sqrt{B}\right) = \frac{\Gamma(\frac{\delta+1}{2})\Gamma(\frac{\delta+d}{2})}{\Gamma(\frac{\delta+d+1}{2})\Gamma(\frac{\delta}{2})},$$

which is equal to $\mathbb{E}\left(e^{-\frac{A^2}{2}}\right)$. Since $Y \sim \chi^2_{\delta+d}$, we have

$$\mathbb{E}\left(e^{\frac{-b_1^2}{2Y}} \middle| \Psi^-_{\cap}\right) = \mathbb{E}\left(h(b_1,\delta^*) \mid \Psi^-_{\cap}\right).$$

Regarding that $\Psi^-_{\cup} \subset \Psi^-_{\cap}$ we have

$$\mathbb{E}\left(h(b_1,\delta^*) \mid \Psi^-_{\cap}\right) = \mathbb{E}\left(h(b_1,\delta^*) \mid \Psi^-_{\cup}\right).$$

Note that, while b_1 is $\Psi^-_{\cap}$-measurable, it is not $\Psi^-_{\cup}$-measurable.

C Proof of Lemma 2

We note three important facts. First, the elements of the first row of the matrix ψ are all zero except for those corresponding to the edges of the path λ_1, i.e.

$$\psi_{1v} = 0, \quad \text{for} \;\; v \in \cup_{\lambda\in\Lambda} V_\lambda \;\; \text{and} \;\; v \neq \{1,2,q\}.$$

Second, based on the above fact and Equation 10 of the manuscript, the remaining non-free entries in all the columns of ψ except for the columns q and p, are equal to zero. Third, due to the first entry ψ_{1q} of column q being free, none of the entries of column q are necessarily zero. However, for each $\lambda \in \Lambda$, using iteratively Equation 10 of the manuscript, we see the non-free entries of column p are zero except for the last one ψ_{qp}. Considering these facts and applying Equation 10 of the manuscript yields

$$\psi_{qp} = \frac{-1}{\psi_{qq}} \sum_{\lambda\in\Lambda} \sum_{i_{\lambda\in V_\Lambda}} \psi_{i_\lambda q}\psi_{i_\lambda p} = \frac{-1}{\psi_{qq}} \sum_{\lambda\in\Lambda} \psi_{\ell_\lambda q}\psi_{\ell_\lambda p}. \tag{2}$$

The entries $\psi_{\ell_\lambda p}$, $\lambda \in \Lambda$ are free. The entries $\psi_{\ell_\lambda q}$ are obtained by successively applying Equation 10 of the manuscript and the fact that $\psi_{(j-1)_\lambda j_\lambda}$, $j = \{1,\ldots,(l-1)\}$ are free and the non-free entries of Ψ are equal to zero expect for the columns q and p. That is

$$\begin{aligned}
\psi_{\ell_\lambda q} &= -\frac{\psi_{(\ell-1)_\lambda \ell_\lambda}\psi_{(\ell-1)_\lambda q}}{\psi_{(\ell-1)_\lambda(\ell-1)_\lambda}} = +\frac{\psi_{(\ell-1)_\lambda \ell_\lambda}\psi_{(\ell-2)_\lambda \ell_\lambda}\psi_{(\ell-2)_\lambda q}}{\psi_{(l-1)_\lambda(l-1)_\lambda}\psi_{(l-2)_\lambda(l-2)_\lambda}} \\
&= \ldots \\
&= (-1)^{\ell_\lambda - 1}\frac{\psi_{1_\lambda q}\prod_{j=1}^{\ell-1}\psi_{j_\lambda(j+1)_\lambda}}{\prod_{j=2}^{\ell}\psi_{j_\lambda j_\lambda}} = (-1)^{\ell_\lambda-1}\frac{\psi_{1_\lambda q}\prod_{j=1}^{\ell-1}\psi_{j_\lambda(j+1)_\lambda}}{\prod_{j=1}^{\ell-1}\psi_{(j+1)_\lambda(j+1)_\lambda}}, \\
&= (-1)^{\ell_\lambda-1}\psi_{1_\lambda q}\prod_{j=1}^{l-1}\frac{\psi_{j_\lambda(j+1)_\lambda}}{\psi_{(j+1)_\lambda(j+1)_\lambda}}.
\end{aligned}$$

Above equality and Equation 2 together yield

$$\psi_{qp} = \frac{1}{\psi_{qq}}\sum_{\lambda\in\Lambda}(-1)^{\ell_\lambda}\frac{\psi_{1_\lambda,q}\psi_{\ell_\lambda,p}\prod_{j=1}^{\ell-1}\psi_{j_\lambda,(j+1)_\lambda}}{\prod_{j=2}^{l}\psi_{j_\lambda,j_\lambda}} = \frac{1}{\psi_{qq}}\sum_{\lambda\in\Lambda}(-1)^{\ell_\lambda}\psi_{\ell_\lambda,p}\psi_{1_\lambda,q}\prod_{j=1}^{\ell-1}\frac{\psi_{j_\lambda,(j+1)_\lambda}}{\psi_{(j+1)_\lambda,(j+1)_\lambda}},$$

which is identical to Equation 17 of the manuscript.

D Proof of Theorem 1

For the case where the paths between q and p are disjoint, we can rewrite A_1 and b_1 as

$$A_1 = \sum_{\lambda\in\Lambda, l_{\lambda=1}} \psi_{\ell_\lambda,q}\psi_{\ell_\lambda,p}, \quad b_1 = \sum_{\lambda\in\Lambda,\ell_\lambda\geq 2} b_{1\lambda},$$

where

$$b_{1\lambda} = (-1)^{\ell_\lambda}\psi_{1_\lambda,q}\psi_{\ell_\lambda,p}\prod_{j_\lambda=1}^{\ell_\lambda-1}\frac{\psi_{j_\lambda,(j+1)_\lambda}}{\psi_{(j+1)_\lambda,(j+1)_\lambda}}$$

with the convention that $b_{1\lambda}=0$ if $\ell_\lambda = 1$, and

$$D = \sum_{\lambda\in\Lambda} D_\lambda \text{ where } D_\lambda = \sum_{k=2}^{\ell_\lambda}\left((-1)^{k-1}\psi_{1_\lambda,q}\prod_{j_\lambda=1}^{k-1}\frac{\psi_{j_\lambda,(j+1)_\lambda}}{\psi_{(j+1)_\lambda,(j+1)_\lambda}}\right)^2.$$

All the entries appearing in the expression for A_1 and b_1 are free variables independent of each other and those appearing in $b_{1\lambda}, \lambda\in\Lambda, \ell_\lambda\geq 2$ are different from those appearing in A_1. Thus, A_1 and $\sum_{\lambda\in\Lambda,\ell_\lambda\geq 2} b_{1\lambda}$ are stochastically independent. Moreover, according to Proposition 1 of the Manuscript, all $\psi_{ij}, i\neq j$ are $N(0,1)$ random variables while ψ_{ii}^2 follow a $\chi^2_{\delta+\nu_i}$ distribution. In particular $\psi_{qq}^2\sim\chi^2_\delta$.

To find a lower bound for the I_1/I_2, we use a Gaussian equality as follows: if $Z\sim N(0,\sigma^2)$, then

$$\mathbb{E}\left(e^{itZ}\right) = \int_{-\infty}^{+\infty} e^{itz}e^{-\frac{z^2}{2\sigma^2}}\frac{dz}{\sigma\sqrt{2\pi}} = e^{-\frac{\sigma^2t^2}{2}}.$$

Applying above equality with $t=b_1$ and $\sigma^2=\frac{1}{y}$, we have

$$\begin{aligned}
h(b_1,\delta^*) &= \frac{2^{-\delta^*}}{\Gamma(\delta^*)}\int_0^{+\infty} y^{\delta^*-1}e^{\frac{-y}{2}}e^{\frac{-b_1^2}{2y}}dy\\
&= \frac{2^{-\delta^*}}{\Gamma(\delta^*)}\int_0^{+\infty} y^{\delta^*-1}e^{\frac{-y}{2}}\left(\int_{-\infty}^{+\infty} e^{ib_1z}e^{-\frac{yz^2}{2}}\sqrt{y}\frac{dz}{\sqrt{2\pi}}\right)dy\\
&= \frac{2^{-\delta^*}}{\Gamma(\delta^*)\sqrt{2\pi}}\int_{+\infty}^{+\infty} e^{ib_1z}\left(\int_0^{+\infty} y^{\delta^*-\frac{1}{2}}e^{\frac{-(1+z^2)y}{2}}dy\right)dz\\
&= \frac{2^{-\delta^*}}{\Gamma(\delta^*)\sqrt{2\pi}}\int_{+\infty}^{+\infty} e^{ib_1z}\left(\frac{\Gamma(\delta^*+\frac{1}{2})}{\left(\frac{1+z^2}{2}\right)^{\delta^*+\frac{1}{2}}}\right)dz\\
&= \int_{+\infty}^{+\infty} e^{ib_1z}f(z)dz,
\end{aligned}$$

where

$$f(z) = \frac{\Gamma(\delta^*+\frac{1}{2})}{\sqrt{\pi}\Gamma(\delta^*)}\left(1+z^2\right)^{-\delta^*+\frac{1}{2}}.$$

Thus

$$\begin{aligned}
\mathbb{E}\left(e^{-\frac{D}{2}}\mathbb{E}\left(h(b_1,\delta^*)\mid\Psi_\cup^-\right)\right) &= \int_{+\infty}^{+\infty}\mathbb{E}\left(e^{-\frac{D}{2}+ib_1z}\right)f(z)dz\\
&= \prod_{\lambda\in\Lambda}\int_{+\infty}^{+\infty}\mathbb{E}\left(e^{-\frac{D_\lambda}{2}+ib_{1\lambda}z}\right)f(z)dz
\end{aligned}.$$

Similarly, we have

$$\mathbb{E}\left(e^{-\frac{D}{2}}\right) = \prod_{\lambda\in\Lambda} \mathbb{E}\left(e^{-\frac{D_\lambda}{2}}\right).$$

Consider independent identically distributed random variables $X_1, \dots, X_n, \dots$ such that $X_1 \sim Z/\sqrt{Q}$ with $Z \sim N(0,1)$ independent of $Q \sim \chi^2_{\delta+1}$. For $\ell = \ell_\lambda, \lambda \in \Lambda$, we define

$$S_\ell = X_1^2 + X_1^2 X_2^2 + \cdots + (X_1 \dots X_{\ell-1})^2 \text{ and } B_\ell = X_1 X_2 \dots X_{\ell-1}. \tag{3}$$

We see that for $\lambda \in \Lambda$, we have

$$\begin{aligned} D_\lambda &\sim N_{1_\lambda q} S_{\ell_\lambda}, \\ b_{1\lambda} &\sim N_{1_\lambda q} N_{\ell_\lambda p} B_{\ell_\lambda}, \end{aligned}$$

where $N_{1_\lambda q}$ and $N_{\ell_\lambda p}$ are independent $N(0,1)$ random variables, independent of $X_1, \dots, X_\ell, \dots$. We note that, from the independence of the entries of ψ_E, we have

$$(b_{1\lambda}, D_\lambda, N_{1_\lambda q}, N_{\ell_\lambda p}), \text{ for } \lambda \in \Lambda,$$

are mutually independent.

Omitting the index λ on ℓ_λ, and simplifying $N_{1_\lambda q}$ to N_q and $N_{\ell_\lambda p}$ to N_p, we define

$$g_\ell(x) = \mathbb{E}\left(e^{-\frac{N_q^2 S_\ell}{2} + iN_p N_q B_\ell x}\right).$$

Then

$$I_1 = \prod_{\lambda\in\Lambda} \int_{-\infty}^{\infty} g_{\ell_\lambda}(x) f(x) dx \quad \text{and} \quad I_2 = \prod_{\lambda\in\Lambda} g_{\ell_\lambda}(0).$$

Therefore we can write

$$\begin{aligned} \frac{I_2 - I_1}{I_2} &= \int_{-\infty}^{\infty} \frac{\prod_{\lambda\in\Lambda} g_{\ell_\lambda}(0) - \prod_{\lambda\in\Lambda} g_{\ell_\lambda}(x)}{\prod_{\lambda\in\Lambda} g_{\ell_\lambda}(0)} f(x) dx \\ &\leq \int_{-\infty}^{\infty} \sum_{\lambda\in\Lambda} \frac{g_{\ell_\lambda}(0) - g_{\ell_\lambda}(x)}{g_{\ell_\lambda}(0)} f(x) dx, \end{aligned} \tag{4}$$

where the last inequality is based on Lemma D.2 applied to $a_\lambda = g_{\ell_\lambda}(0)$ and $b_\lambda = g_{\ell_\lambda}(x)$. Writing ℓ for ℓ_λ, we have

$$\begin{aligned} g_\ell(0) - g_\ell(x) &= \mathbb{E}\left(e^{-\frac{N_q^2 S_\ell}{2}} \left(1 - e^{iN_p N_q B_\ell x}\right)\right) \\ &\leq \mathbb{E}\left(e^{-\frac{N_q^2 S_\ell}{2}} |N_p N_q X_1 \dots X_{\ell-1}|\right) |x| \\ &\leq \mathbb{E}\left(e^{-\frac{N_q^2 X_1^2}{2}} |N_q X_1|\right) \mathbb{E}\left(|N_p X_2 \dots X_{\ell-1}|\right) |x| \\ &= \frac{2}{\pi} \frac{\delta}{\delta+2} r(\delta)^\ell \, |x|, \end{aligned}$$

where the first inequality is due to the fact that $|1 - e^{iN_pN_qB_\ell x}| \leq |N_pN_qB_\ell|\ |x|$, the second inequality is due to the fact that $X_1^2 \leq S_\ell$ and the independence of (N_q, X_1) and $(N_p, X_2, \ldots, X_{\ell-1})$, and the last equality is obtained using Equations 5 and 6. Moreover, by using Equation 7, we have

$$\frac{g_\ell(0) - g_\ell(x)}{g_\ell(0)} \quad \leq \quad \frac{\delta^2}{\pi(\delta+2)} \left[\frac{\Gamma(\frac{\delta}{2})}{\Gamma(\frac{\delta+1}{2})}\right]^2 r(\delta)^\ell\ |x|.$$

Finally, Equation 4 yields

$$\begin{aligned} 0 \leq \frac{I_2 - I_1}{I_2} &\leq \frac{\delta^2}{\pi(\delta+2)} \left(\frac{\Gamma(\frac{\delta}{2})}{\Gamma(\frac{\delta+1}{2})}\right)^2 \left(\sum_{\lambda\in\Lambda} r(\delta)^{\ell_\lambda}\right) \int_{-\infty}^{\infty} |x| f(x) dx \\ &\leq \frac{\delta^2}{\pi(\delta+2)} \left(\frac{\Gamma(\frac{\delta}{2})}{\Gamma(\frac{\delta+1}{2})}\right)^2 \left(\sum_{\lambda\in\Lambda} r(\delta)^{\ell_\lambda}\right) \frac{\Gamma(\delta^* - \frac{1}{2})}{\sqrt{\pi}\Gamma(\delta^*)} \\ &= \frac{\delta^2}{\pi(\delta+2)} \left(\frac{\Gamma(\frac{\delta}{2})}{\Gamma(\frac{\delta+1}{2})}\right)^2 \left(\sum_{\lambda\in\Lambda} r(\delta)^{\ell_\lambda}\right) r(\delta + d - 1) \end{aligned}$$

which leads to Equation 18 of the manuscript.

The following lemmas are used in the proof of Theorem 1.

Lemma D.1. *Let $X_1, \ldots, X_{\ell-1}$ be independent identically distributed random variables such that $X_1 \sim Z/\sqrt{Q}$ with $Z \sim N(0,1)$ independent of $Q \sim \chi^2_{\delta+1}$ where $\delta \geq 3$. Let N_p and N_q also be standard normal $N(0,1)$ random variables, mutually independent and independent of $X_1, \ldots, X_{\ell-1}$. We then have*

$$\mathbb{E}\left(e^{-\frac{N_q^2X_1^2}{2}}\ |N_qX_1|\right) = \sqrt{\frac{2}{\pi}} \frac{\delta}{\delta+2} r(\delta) \tag{5}$$

and

$$\mathbb{E}\left(|N_pX_1 \ldots X_{\ell-1}|\right) = \sqrt{\frac{2}{\pi}} r(\delta)^{\ell-1}, \tag{6}$$

where $r(\delta) = \frac{\Gamma(\frac{\delta}{2})}{\sqrt{\pi}\Gamma(\frac{\delta+1}{2})}$. Let S_ℓ as defined in Equation 3 then

$$\mathbb{E}\left(e^{-\frac{N_q^2S_\ell}{2}}\right) > \frac{2}{\delta}\left(\frac{\Gamma\left(\frac{\delta+1}{2}\right)}{\Gamma\left(\frac{\delta}{2}\right)}\right)^2. \tag{7}$$

Proof. For Equation 5, the variable $Y = X_1^2 \sim B_2\left(\frac{1}{2}, \frac{\delta+1}{2}\right)$ which is a Beta distribution of the second kind. Thus

$$\mathbb{E}\left(e^{-\frac{N_q^2X_1^2}{2}} |N_qX_1|\right) \quad = \quad \mathbb{E}\left(\mathbb{E}\left(e^{-\frac{N_q^2X_1^2}{2}} |N_qX_1| \,\Big|\, X_1\right)\right)$$

where

$$\mathbb{E}\left(e^{-\frac{N_q^2X_1^2}{2}} |N_qX_1| \,\Big|\, X_1\right) \quad = \quad \frac{|X_1|}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |u| e^{-\frac{(1+X_1^2)}{2}u^2} du = \sqrt{\frac{2}{\pi}} \frac{|X_1|}{1+X_1^2}.$$

Thus

$$\mathbb{E}\left(e^{-\frac{N_q^2X_1^2}{2}}|N_qX_1|\right) = \sqrt{\frac{2}{\pi}}\mathbb{E}\left(\frac{|X_1|}{1+X_1^2}\right) = \sqrt{\frac{2}{\pi}}\mathbb{E}\left(\frac{\sqrt{Y}}{1+Y}\right) = \sqrt{\frac{2}{\pi}}\frac{\delta}{\delta+2}r(\delta).$$

For Equation 6, by using the mutual independence of N_p and $X_1, \ldots, X_{\ell-1}$ and the fact that $X_i^2 \sim B_2\left(\frac{1}{2}, \frac{\delta+1}{2}\right)$ for $i = \{1, \ldots, \ell-1\}$, we have

$$\mathbb{E}\left(|N_pX_1 \ldots X_{\ell-1}|\right) = \mathbb{E}\left(|N_p|\right)\mathbb{E}\left(|X_1|\right)^{\ell-1} = \sqrt{\frac{2}{\pi}}\left(\frac{\Gamma(\frac{\delta}{2})}{\sqrt{\pi}\Gamma(\frac{\delta+1}{2})}\right)^{\ell-1} = \sqrt{\frac{2}{\pi}}r(\delta)^{\ell-1}.$$

For Equation 7, from Chamayou and Letac (1991, Example 9), if $U \sim B_2(a,b)$ for $b > a$ and V are independent, then $U(1+V) \sim V$ if and only if $V \sim B_2(a, b-a)$. One applies this to $U = X_1^2$, $V = S$, $a = \frac{1}{2}$, and $b = \frac{\delta+1}{2}$ since $S' = \sum_{i=2}^{\infty} X_2^2 \ldots X_i^2 \sim S$ and $X_1^2(1+S') = S$. Thus $S \sim B_2\left(\frac{1}{2}, \frac{\delta}{2}\right)$. Since $S_\ell < S$, we have

$$\mathbb{E}\left(e^{-\frac{N_q^2S_\ell}{2}}\right) > \mathbb{E}\left(e^{-\frac{N_q^2S}{2}}\right) = \frac{2}{\delta}\left(\frac{\Gamma\left(\frac{\delta+1}{2}\right)}{\Gamma\left(\frac{\delta}{2}\right)}\right)^2.$$

□

Lemma D.2. *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be complex numbers such that $|b_i| \leq |a_i|$, for $i = 1, \ldots, n$. Then*

$$\left|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i\right| \leq \prod_{i=1}^n |a_i| \sum_{j=1}^n \frac{|a_j - b_j|}{|a_j|}.$$

Proof.

$$\begin{aligned} |a_1 \ldots a_n - b_1 \ldots b_n| &= |(a_1 \ldots a_n - b_1a_2 \ldots a_n) + (b_1a_2 \ldots a_n - b_1b_2a_3 \ldots a_n) \\ &\qquad + \ldots\ldots + (b_1 \ldots b_{n-1}a_n - b_1b_2 \ldots b_n)| \\ &\leq \sum_{j=1}^n \left|\prod_{i=1}^{j-1} b_i \prod_{i=j}^n a_i - \prod_{i=1}^j b_i \prod_{i=j+1}^n a_i\right| = \sum_{j=1}^n \left|\prod_{i=1}^{j-1} b_i \prod_{i=j+1}^n a_i\right| |a_j - b_j| \\ &\leq \sum_{j=1}^n \left|\prod_{i=1}^{j-1} a_i \prod_{i=j+1}^n a_i\right| |a_j - b_j| = \prod_{i=1}^n |a_i| \sum_{j=1}^n \left|1 - \frac{b_j}{a_j}\right|. \end{aligned}$$

□

E Proof of Theorem 2

The proof of Theorem 2 is long and is done in the next three subsections. In Subsection E.1, we show in Proposition E.1 that, conditional on the set of random variables Ψ_{U}^-, b_1 can be expressed as a bilinear form. In Proposition E.2 of Subsection E.2, using its expression as a bilinear form, we show that b_1 is distributed like the continuous scale mixture of centered Gaussian variables. This allows us to deduce that there exists a unique v_D such that the normal $N(0, v_D)$ distribution best approximates the distribution of b_1. Finally, in

Subsection E.3, we prove Theorem 2: under the assumption that v_D is small, I_1/I_2 can accurately be approximated by 1 or equivalently the ratio of normalizing constants can accurately be approximated by Equation 8 of the manuscript, which is what we want to prove.

E.1 Expression of b_1 as a bilinear form

Here we want to show that the distribution of b_1 is a scale mixture of normal distributions which can be approximated by another $N(0, v_D)$ distribution where the variance v_D depends on $\Psi_{\cup}^{-}$. To do so, we must first express b_1 as a bilinear form in two standard normal random vectors.

Regarding $E_q = \{(i,j) : (i,q) \in E\}$ and $E_p = \{(i,j) : (i,p) \in E\}$, we define

$$\Psi_{E_q^-} = \{\psi_{ij} : (i,j) \in E_q \setminus E_p\} \text{ and } \Psi_{E_p^-} = \{\psi_{ij} : (i,j) \in E_p \setminus E_q\}.$$

The set $\Psi_{E_q^-}$ represents the free elements of matrix Ψ regarding to just the neighbor of q and the same for p. In the following, we express D and b_1 as polynomials in $\Psi_{E_q^-}$ and $\Psi_{E_p^-}$.

Proposition E.1. *Let N_q^- denote the set of nodes that are neighbours of q but not of p and N_p^- denote the set of nodes that are neighbours of p but not of q. There exist vectors $M_i^q \in \mathbb{R}^{N_q^-}$ and $M_i^p \in \mathbb{R}^{N_p^-}$, $i = \{1, \dots, q-d-1\}$, functions of $\Psi_{\cup}^-$, such that if C is the $|N_q^-| \times |N_p^-|$-dimensional matrix*

$$C = \sum_{i=1}^{q-d-1} M_i^q (M_i^p)^t$$

then we have

$$b_1 = tr\left(\Psi_{E_q^-} C \Psi_{E_p^-}\right). \tag{8}$$

Furthermore, $\{M_i^q, M_i^p\}_{i=1}^{q-d-1}$, $\Psi_{E_q^-}$, and $\Psi_{E_p^-}$ are independent.

Proof. From the expression of b_1 of the manuscript, we have

$$b_1 = \sum_{i=1}^{q-d-1} \psi_{iq}\psi_{ip},$$

which is based on assume the nodes which are neighbours to both q and p, are numbered $q-d, q-d+1, \dots, p$. By Equation 10 of the manuscript, each $\psi_{iq}, (i,q) \in \overline{E}$ is equal to the sum of products

$$\psi_{iq} = \frac{-1}{\psi_{ii}} \sum_{l=1}^{i-1} \psi_{li}\psi_{lq}, \tag{9}$$

where each of these ψ_{li} or ψ_{lq}, $l = \{1, \dots, q-d-1\}$ may be free or not free. If ψ_{lq} is free, l necessarily belongs to N_q^- because it is a neighbour of q and, since $i \leq q-d$ and $l \leq i$, it cannot be a neighbour of both q and p. If ψ_{lq} is not free, then, we write the expression of ψ_{lq} according to Equation 10 of the manuscript and we repeat this process until ψ_{lq} has been expressed in terms of a ratio of a product of $\psi_{uv}, (u,v) \in E, u \leq l, v \leq q$, one of which is necessarily (since the sum in Equation 9 is finite) equal to $\psi_{u_l q}$ for some $u_l \leq l, u_l \in N_q^-$, and a product of $\psi_{vv}, v \leq l$. Similarly ψ_{li} is free or not free. If not free, it will be expressed as a ratio of products of elements of ψ_E, none of which, in the numerator, can be equal to $\psi_{u_l q}$ since it is the product of entries ψ_{uv} of ψ with $u \leq l$ and $v \leq i < q$. Thus, from Equation 9, we can write, for each $i = \{1, \ldots, q-d-1\}$

$$\psi_{iq} = \frac{-1}{\psi_{ii}} \sum_{l=1}^{i-1} \psi_{li}\psi_{lq} = \sum_{l \in N_q^-} (M_i^q)_l \; \psi_{u_l,q} = \text{tr}\left(M_i^q \Psi_{E_q^-}\right), \tag{10}$$

where $(M_i^q)_l, l \in N_q^-$ are the components of M_i^q, some of which can be equal to 0, if the Cholesky equations (in Equation 10 of the manuscript) do not lead to that particular $l \in N_q^-$, or 1, if $l \in N_q^-$. Similarly, we have

$$\psi_{ip} = \frac{-1}{\psi_{ii}} \sum_{l=1}^{i-1} \psi_{li}\psi_{lp} = \sum_{l \in N_p^-} (M_i^p)_l \; \psi_{v_l,p} = \text{tr}\left(M_i^p \Psi_{E_p^-}\right), \tag{11}$$

for some $v_l < l, v_l \in N_p^-$ and Equation 8 follows from Equations 10 and 11. □

Example E.1. *Consider the graph in Figure 3 (right) of the manuscript where* $q = 6$, $p = 7$, $N_q = \{1,2,5\}$, $N_p = \{3,4,5\}$, $N_q^- = \{1,2\}$, $N_p^- = \{3,4\}$, $d = 1$, *and* $\psi_E = \{\psi_{14}, \psi_{16}, \psi_{23}, \psi_{24}, \psi_{26}, \; \psi_{37}, \psi_{47}, \psi_{56}, \psi_{57}\}$. *Thus* $\Psi_{E_q^-} = (\psi_{16}, \psi_{26})$ *and* $\Psi_{E_p^-} = (\psi_{37}, \psi_{47})$. *Using the notation* $X_{ij} = \psi_{ij}/\psi_{jj}$ *for convenience, the non-free entries are*

$$\begin{aligned}
\psi_{34} &= -\psi_{24}X_{23}, \\
\psi_{36} &= -\psi_{26}X_{23} = -(\psi_{16}, \psi_{26})(0, X_{23})^t = tr(\Psi_{E_q^-} M_3^q), \\
M_3^q &= (0, X_{23}), \\
\psi_{46} &= -\psi_{26}X_{23}^2X_{24} - \psi_{26}X_{24} - \psi_{16}X_{14} = -(\psi_{16}, \psi_{26})\left(X_{14}, X_{24} + X_{23}^2X_{24}\right)^t = tr\left(\Psi_{E_q^-} M_4^q\right), \\
M_4^q &= -\left(X_{14}, X_{24} + X_{23}^2X_{24}\right)^t, \\
\psi_{67} &= \frac{1}{\psi_{66}}\left(-\psi_{57}\psi_{56} + \psi_{26}X_{23}^2X_{24}\psi_{47} + \psi_{26}X_{24}\psi_{47} + \psi_{16}X_{14}\psi_{47} + \psi_{26}X_{23}\psi_{37}\right) \\
&= \frac{1}{\psi_{66}}(A_1 + b_1),
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= -\psi_{57}\psi_{56}, \\
b_1 &= \psi_{26}X_{23}^2X_{24}\psi_{47} + \psi_{26}X_{24}\psi_{47} + \psi_{16}X_{14}\psi_{47} + \psi_{26}X_{23}\psi_{37}.
\end{aligned}$$

It leads $b_1 = \Psi_{E_q^-}^t C \Psi_{E_p^-}$ *where*

$$C = \begin{bmatrix} 0 & X_{14} \\ X_{23} & X_{24} + X_{23}^2X_{24} \end{bmatrix}.$$

It also follows from the definition of $\Psi_{E_q^-}$ *that* $M_1^q = (1,0)^t$ *and* $M_2^q = (0,1)^t$. *From the definition of* $\Psi_{E_p^-}$, *we have* $M_1^p = M_2^p = (0,0)^t$ *and* $M_3^p = (1,0)^t$ *and* $M_4^p = (0,1)^t$. *We can then verify*

$$tr\left(\Psi_{E_q^-}^t C \Psi_{E_p^-}\right) = \sum_{i=1}^{4} tr\left(\Psi_{E_q^-} M_i^q\right) tr\left(\Psi_{E_p^-} M_i^p\right).$$

E.2 A normal approximation to the distribution of b_1

We see in Equation 8, that if we condition on $\Psi^-_{\cup}$, then b_1 can be expressed as the bilinear form of $\Psi_{E_q^-}$ and $\Psi_{E_p^-}$ as

$$b_1 = \Psi^t_{E_q^-} C \Psi_{E_p^-},$$

where $C = \sum_{i=1}^{q-d-1} (M_i^q)^t M_i^p$ is a matrix of rank $m \le \min\left(|N_q^-|, |N_p^-|\right)$. Once $\Psi^-_{\cup}$ is known, C is fixed. We are now going to show that, conditional on $\Psi^-_{\cup}$, the distribution of b_1 has the following property.

Proposition E.2. *When conditioned by $\Psi^-_{\cup}$, the distribution of b_1 is a continuous scale mixture of centered normal distributions. More precisely, b_1 follows the same distribution as $X\sqrt{Y/2}$ where X and Y are independent with $X \sim N(0,1)$ and*

$$Y = \sum_{i=1}^{m} \frac{Y_i}{\lambda_i} \quad \textit{where} \quad Y_i \overset{iid}{\sim} \chi^2_2$$

and where $1/\lambda_1, \ldots, 1/\lambda_m$ are the non-zero eigenvalues of $C^T C$.

Proof. We first show that b_1 follows the same distribution as $X\sqrt{Y/2}$. To do so, it suffices to show the two Laplace transforms $\mathbb{E}\left(e^{sb_1}\right)$ and $\mathbb{E}\left(e^{sX\sqrt{Y/2}}\right)$ coincide. Integrating this last expected value first with respect to X, holding Y fixed, and then with respect to Y, we obtain

$$\mathbb{E}\left(e^{sX\sqrt{Y/2}}\right) = \mathbb{E}\left(e^{s^2 Y/2}\right) = \prod_{j=1}^{m} \left(1 - \frac{\lambda_j s^2}{2}\right)^{\frac{-1}{2}}.$$

Next, from Equation 8 and then integrating with respect to $\Psi_{E_q^-}$, we have

$$\mathbb{E}\left(e^{sb_1}\right) = \mathbb{E}\left(e^{s\mathrm{tr}\left(\Psi_{E_q^-} C \Psi_{E_p^-}\right)}\right) = \mathbb{E}\left(e^{\frac{s^2}{2}\mathrm{tr}\left(\Psi_{E_p^-} C^t C \Psi_{E_p^-}\right)}\right).$$

Now we have

$$\mathrm{tr}\left(\Psi_{E_p^-} C^t C \Psi_{E_p^-}\right) \sim Z^t \mathrm{diag}\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_m}\right) Z,$$

where $Z = (Z_1, \ldots, Z_m)$ are independent $N(0,1)$ random variables. Thus b_1 and $X\sqrt{Y/2}$ have the same Laplace transform. $\square$

To show the distribution of b_1 is a scale mixture of centered normals, we note that if $X \sim N(0,1)$ and $V = Y/2$ is any positive random variable with distribution $\mu(dv)$, then if $U = X\sqrt{Y/2}$, the density of U is

$$f_U(u) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi v}} e^{-\frac{u^2}{2v}} \mu(dv).$$

So, the distribution of U, that is the distribution of b_1, is a mixture of normal $N(0, v)$ distributions. Following Letac and Massam (2020, Theorem 3.1.), with the distribution of b_1 for f, we deduce that there exists a unique v_D such that the normal $N(0, v_D)$ distribution best approximates the distribution of b_1.

E.3 Expression I_1/I_2 regarding $b_1 \sim N(0, v_D)$

We will now derive an expression for I_1/I_2 when we approximate the distribution of b_1 by the $N(0, v_D)$ distribution. We start with the following lemma.

Lemma E.1. *Under the approximation $b_1 \sim N(0, v_D)$, we have*

$$\mathbb{E}\left(h(b_1, \delta^*)\middle|\Psi_{\cup}^-\right) = \frac{v_D^{\delta^*}}{2^{\delta^*}\Gamma(\delta^*)}\int_0^{\infty} t^{\delta^*-\frac{1}{2}}(1+t)^{\frac{-1}{2}}e^{-\frac{v_D t}{2}}dt.$$

Moreover, when v_D is small, we have

$$\mathbb{E}\left(h(b_1, \delta^*)\middle|\Psi_{\cup}^-\right) = 1 - \frac{\Gamma(\delta^* + \frac{1}{2})}{\Gamma(\delta^*)}\left(\frac{v_D}{2}\right)^{\delta^*}\mathcal{O}\left(\left|\frac{v_D}{2}\right|^{\delta^*-1}\right). \tag{12}$$

Proof. For $b_1 \sim N(0, v_D)$, we have

$$\begin{aligned}
\mathbb{E}\left(h(b_1, \delta^*)\middle|\Psi_{\cup}^-\right) &= \frac{2^{-\delta^*}}{\Gamma(\delta^*)}\int_{-\infty}^{+\infty}\left(\int_0^{+\infty} y^{\delta^*-1}e^{\frac{-1}{2}\left(y+\frac{b_1^2}{y}\right)}dy\right)\frac{e^{-\frac{b_1^2}{2v_D}}}{\sqrt{2\pi v_D}}db_1 \\
&= \frac{2^{-\delta^*}}{\Gamma(\delta^*)}\int_0^{+\infty}\left(\int_{-\infty}^{+\infty}\frac{e^{-\frac{y+v_D}{2v_D y}b_1^2}}{\sqrt{2\pi v_D}}db_1\right)y^{\delta^*-1}e^{-\frac{y}{2}}dy \\
&= \frac{2^{-\delta^*}}{\Gamma(\delta^*)}\int_0^{+\infty}\left(\frac{y}{y+v_D}\right)^{\frac{1}{2}}e^{-\frac{y}{2}}y^{\delta^*-1}dy \\
&= \frac{2^{-\delta^*}v_D^{\delta^*}}{\Gamma(\delta^*)}\int_0^{\infty} t^{\delta^*-\frac{1}{2}}(1+t)^{\frac{-1}{2}}e^{-\frac{v_D t}{2}}dt.
\end{aligned}$$

We note that the above integral is a confluent hypergeometric function of the form

$$\Gamma(a)U(a, b, z) = \int_0^{+\infty} e^{-zt}t^{a-1}(1+t)^{b-a-1}dt,$$

with $z = \frac{v_D}{2}$, $a = \delta^* + \frac{1}{2}$, and $b = \delta^* + 1$ (see Abramovitz and Stegun (1972), p.505, formula 13.2.5) and from p.508, formula 13.5.6 of the same, we know when $|z| \to 0$ and $b > 2$, then

$$U(a, b, z) = \frac{\Gamma(b-1)}{\Gamma(a)}z^{1-b} + \mathcal{O}(|z|^{b-2}).$$

This yields Equation 12. □

F Pseudo-code for I_1/I_2 in Theorem 2

Since we cannot evaluate the approximation in Theorem 2 directly, we write instead

$$\frac{I_1}{I_2} = \frac{\mathbb{E}\left(e^{-\frac{D}{2}}g\left(\delta^*, v_D\right)\right)}{\mathbb{E}\left(e^{-\frac{D}{2}}\right)} = \frac{\int g\left(\delta^*, v_D\right)e^{-\frac{D}{2}}\pi(D)dD}{\int e^{-\frac{D}{2}}\pi(D)dD} = \int g\left(\delta^*, v_D\right)\pi_1(D)dD$$

where $\pi(D)$ is the unknown density of D, and

$$\pi_1(D) = \frac{e^{-\frac{D}{2}}\pi(D)}{\int e^{-\frac{D}{2}}\pi(D)dD}.$$

We then approximate I_1/I_2 by the following sequence of steps:

1. Generate $D_i, i = 1, \ldots, N$ the usual way. Divide the range of D into appropriate small intervals $int^{(q)}, q = 1, \ldots, Q$, and for each interval $int^{(q)}$, compute the relative frequency $f^{(q)}$.

2. Compute $D^{(q)} = \frac{1}{Nf^{(q)}}\sum_{D_i \in int^{(q)}} D_i$ and $r^{(q)} = \frac{e^{-\frac{D^{(q)}}{2}}f^{(q)}}{\sum_{j=1}^{Q} e^{-\frac{D^{(j)}}{2}}f^{(j)}}$, $q = 1, \ldots, Q$.

3. Sample M values of $D^{(m)}, m = 1, \ldots, M$ with probabilities given by the empirical distribution of the $r^{(q)}, q = 1, \ldots, Q$.

4. For each $D^{(m)}$, generate $b_1^{(m,k)}, k = 1, \ldots, K$ the usual way. Compute $v_{D^{(m)}} = \frac{1}{K}\sum_{k=1}^{K}(b_1^{(m,k)} - \overline{b_1}^{(m)})^2$ where $\overline{b_1}^{(m)} = \frac{1}{K}b_1^{(lmk)}$.

5. Compute

$$I_3(v_{D^{(m)}}) = \mathbb{E}\left(\sqrt{\frac{t}{t+1}}\right)$$

by simulating from the $\Gamma(\delta^*, \frac{v_{D^{(m)}}}{2})$ distribution for t.

6. Take the average $\frac{1}{M}\sum_{m=1}^{M} I_3(v_{D^{(m)}})$ as the estimate of I_1/I_2.

G Additional simulation results

Figures G.1, G.2, G.3, and G.4 are the results for the simulation in Section 5.1 of the manuscript. Figures G.5 and G.6 are the ROC plots for Section 6 of the manuscript.

References

Abramovitz, M. and I. Stegun (1972). *Handbook of Mathematical Functions*, Volume 55. National Bureau of Standards, Applied Mathematics.

Atay-Kayis, A. and H. Massam (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika 92*(2), 317–335.

Chamayou, J.-F. and G. Letac (1991). Explicit stationary distributions for compositions of random functions and products of random matrices. *Journal of Theoretical Probability 4*(1), 3–36.

Letac, G. and H. Massam (2020). Gaussian approximation of gaussian scale mixture. *kybersetika 56*(6), 1063–1080.

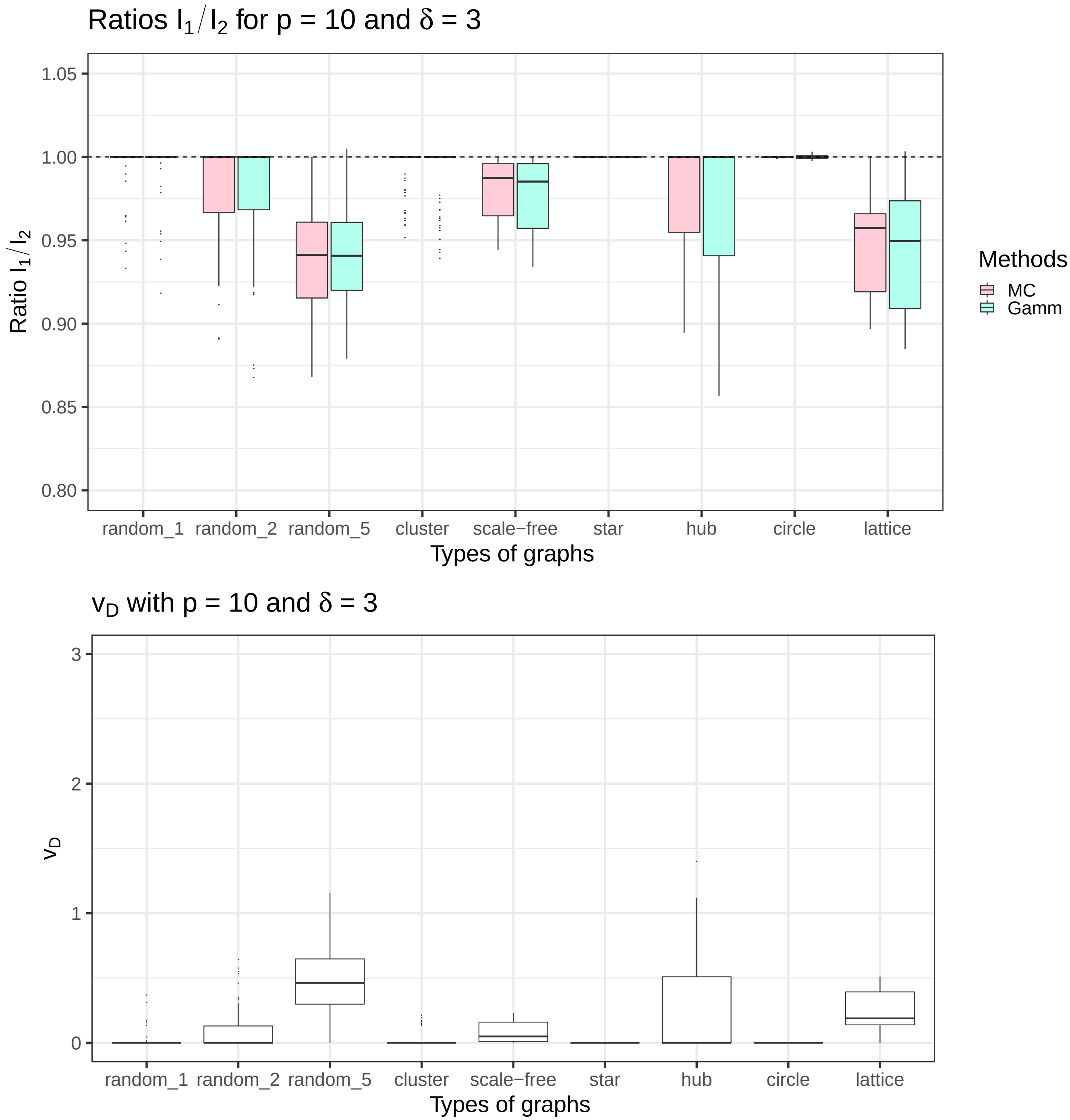
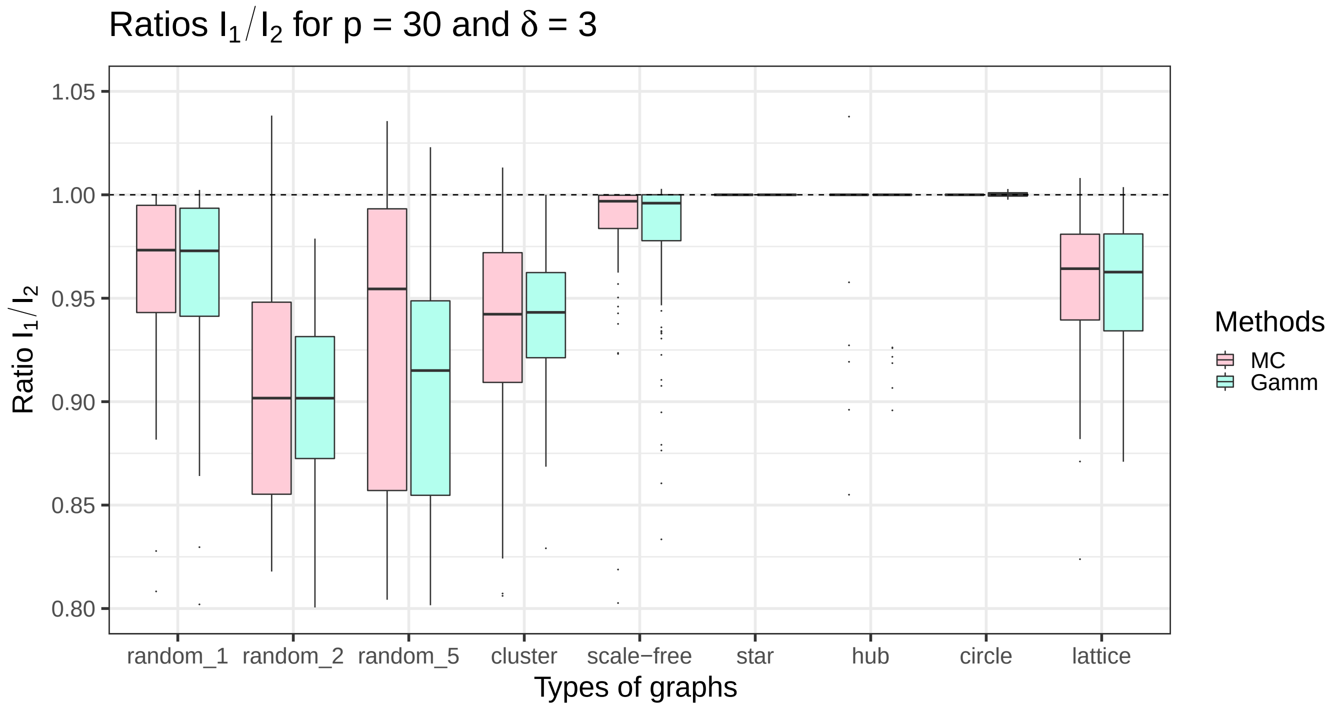


Figure G.1: (Top) The boxplot for the ratio I_1/I_2 computed by the MC approach of Atay-Kayis and Massam (2005) (in red), with 500 samples, and our approximation in Theorem 1 (in green). (Bottom) The boxplot of the variance v_D of variable b_1 for the corresponding graphs. These computations are done over 100 replications for nine different graphs with $p = 10$ nodes and $\delta = 3$.



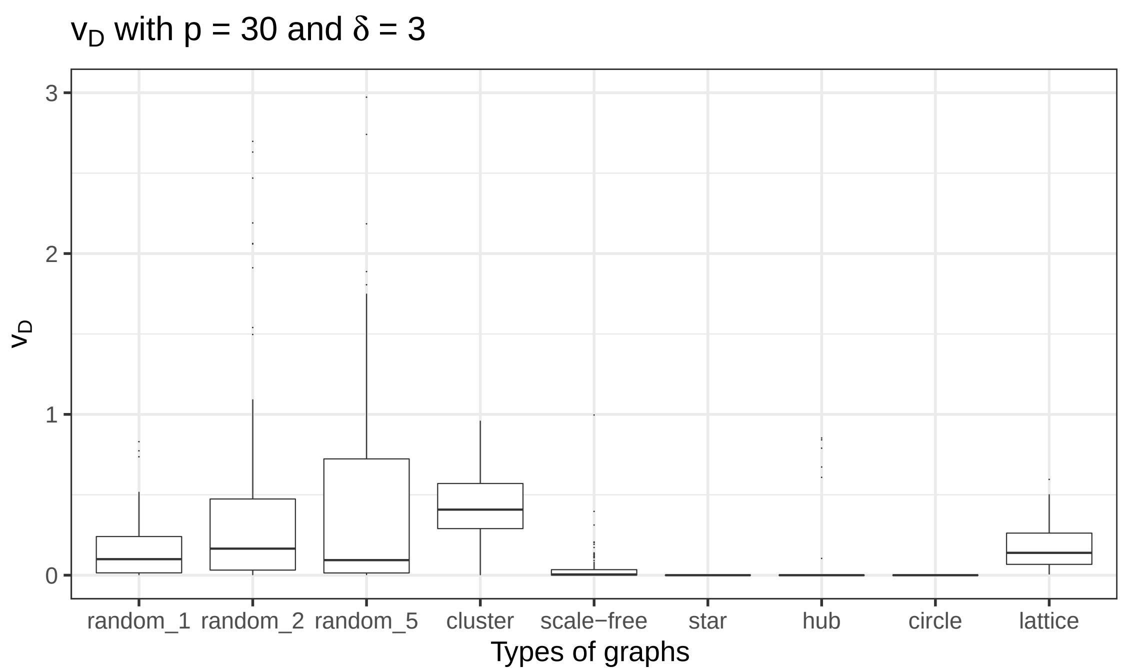
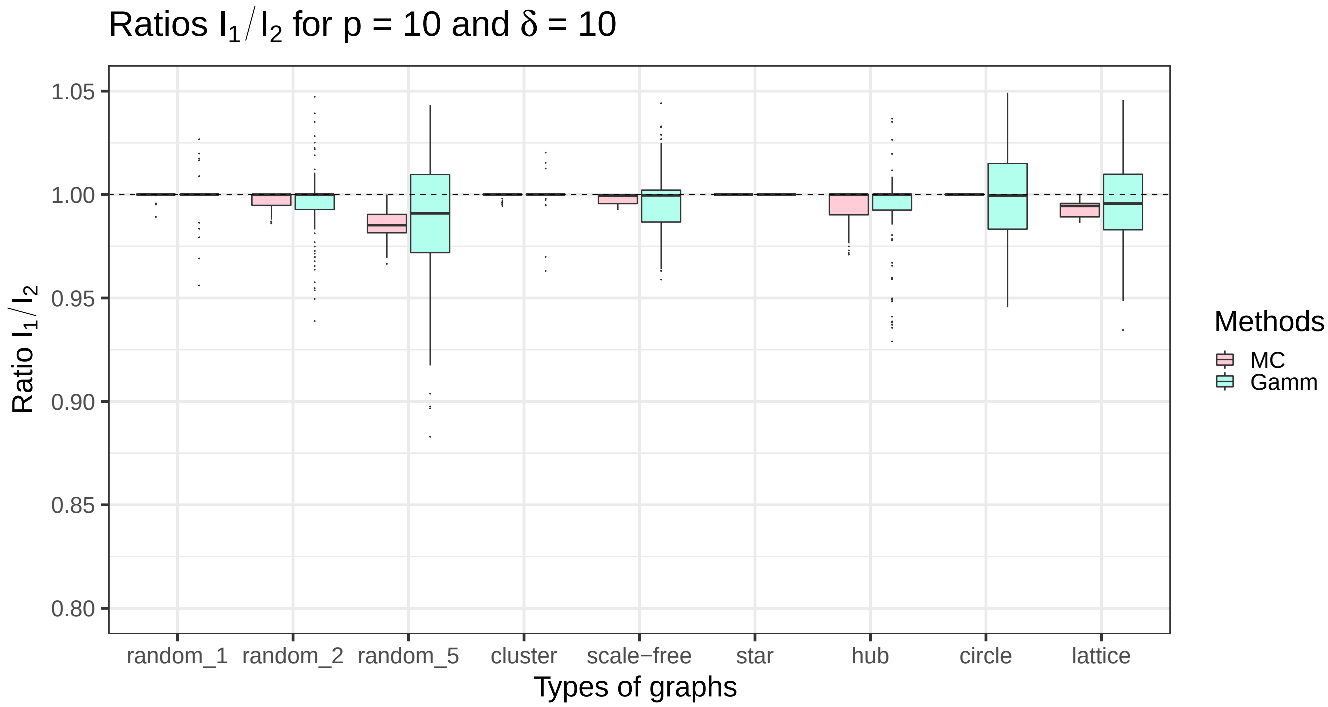


Figure G.2: (Top) The boxplot for the ratio I_1/I_2 computed by the MC approach of Atay-Kayis and Massam (2005) (in red), with 500 samples, and our approximation in Theorem 1 (in green). (Bottom) The boxplot of the variance v_D of variable b_1 for the corresponding graphs. These computations are done over 100 replications for nine different graphs with $p = 30$ nodes and $\delta = 3$.



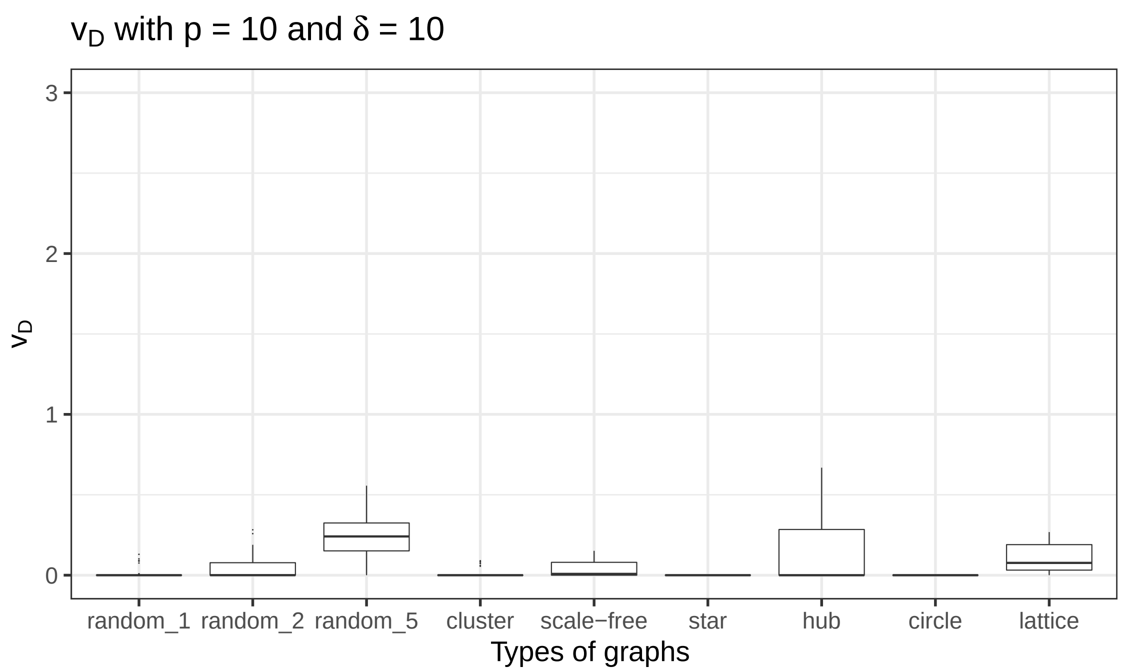


Figure G.3: (Top) The boxplot for the ratio I_1/I_2 computed by the MC approach of Atay-Kayis and Massam (2005) (in red), with 500 samples, and our approximation in Theorem 1 (in green). (Bottom) The boxplot of the variance v_D of variable b_1 for the corresponding graphs. These computations are done over 100 replications for nine different graphs with $p = 10$ nodes and $\delta = 10$.

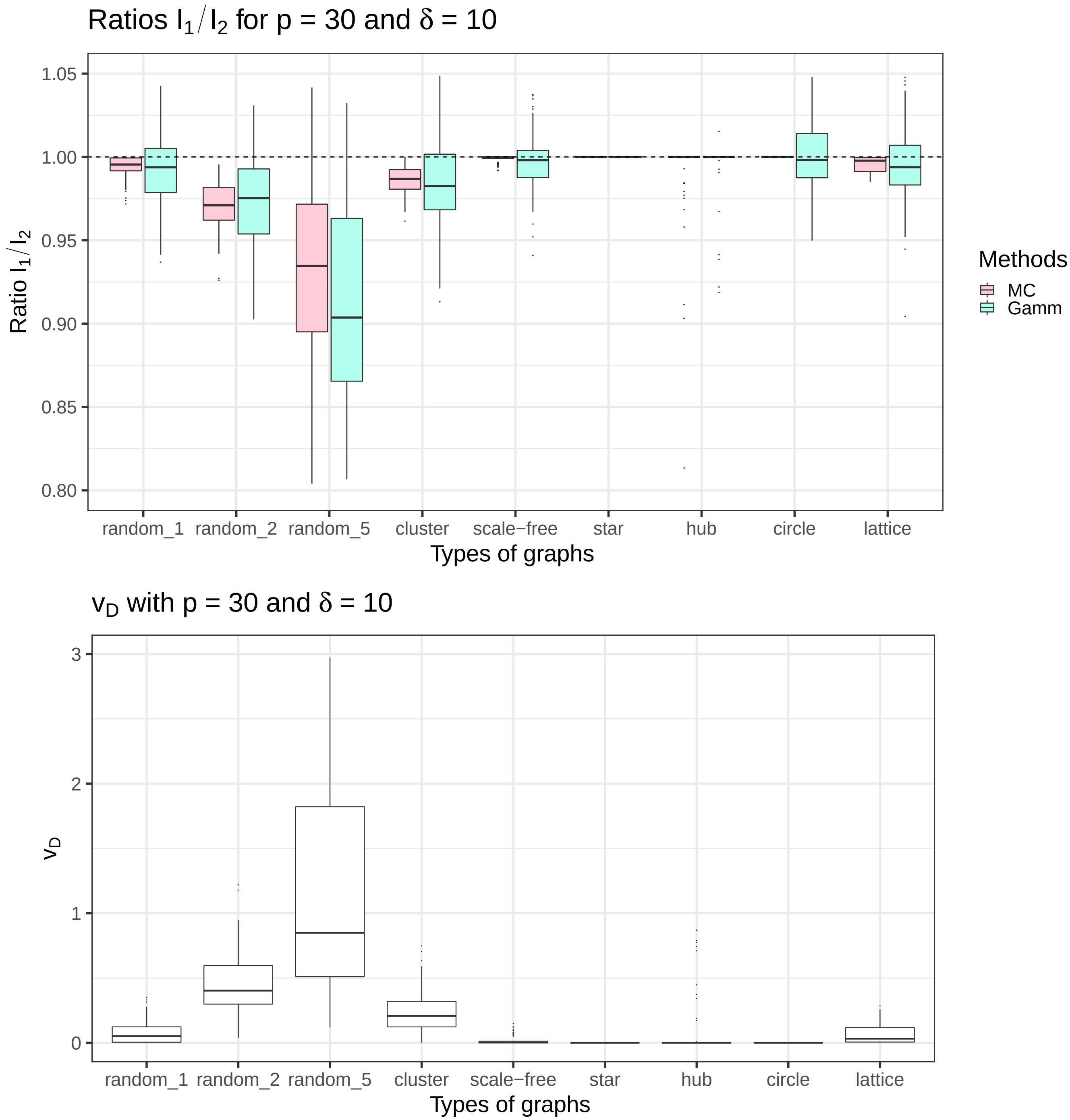


Figure G.4: (Top) The boxplot for the ratio I_1/I_2 computed by the MC approach of Atay-Kayis and Massam (2005) (in red), with 500 samples, and our approximation in Theorem 1 (in green). (Bottom) The boxplot of the variance v_D of variable b_1 for the corresponding graphs. These computations are done over 100 replications for nine different graphs with $p = 30$ nodes and $\delta = 10$.

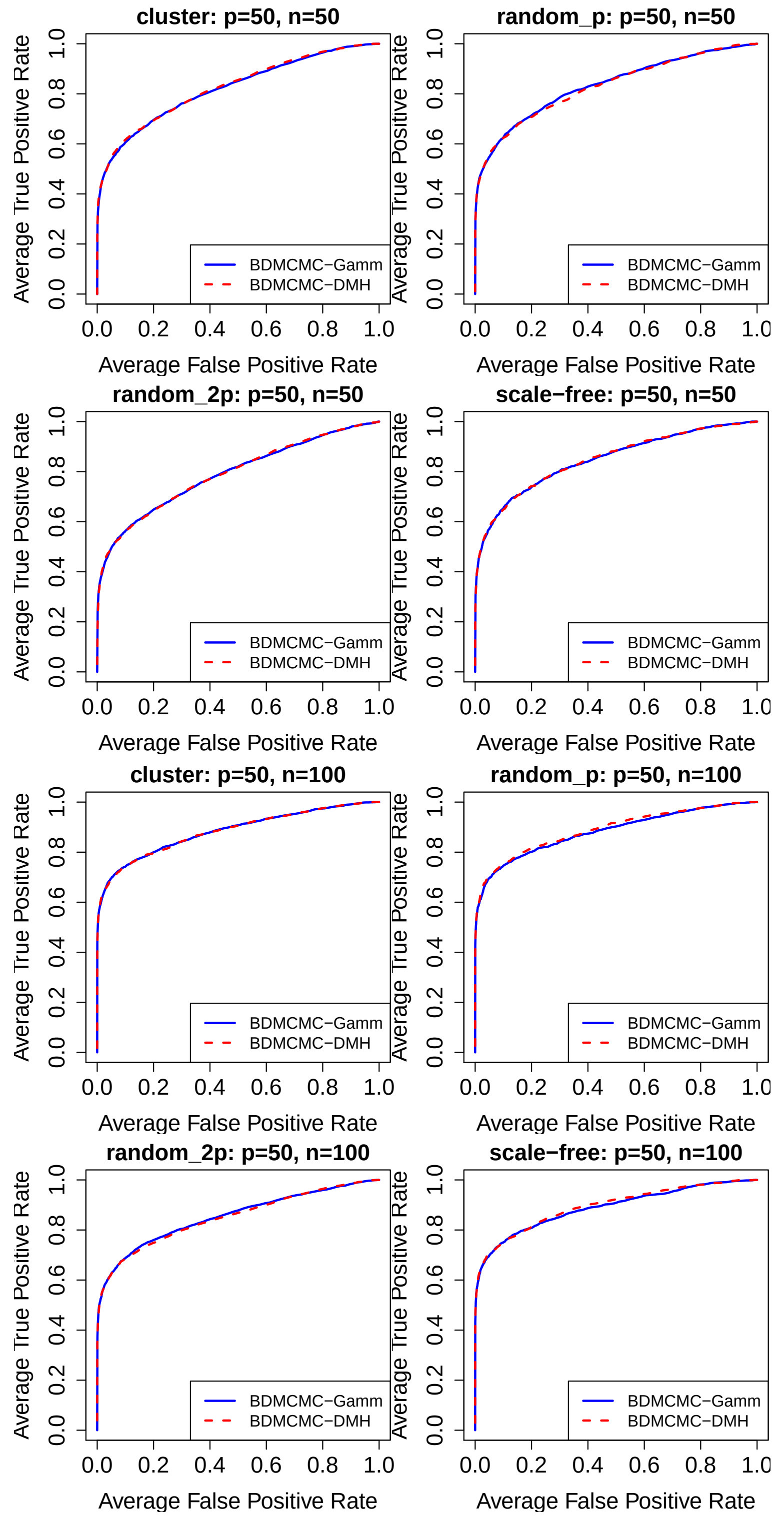


Figure G.5: ROC curves for the BDMCMC algorithm with our approximation in Equation 8 (BDMCMC-Gamm) and BDMCMC algorithm with exchange algorithm (BDMCMC-DMH), over 50 replications. Here, $p = 50$, $n \in \{50, 100\}$, and 4 different graph structures.

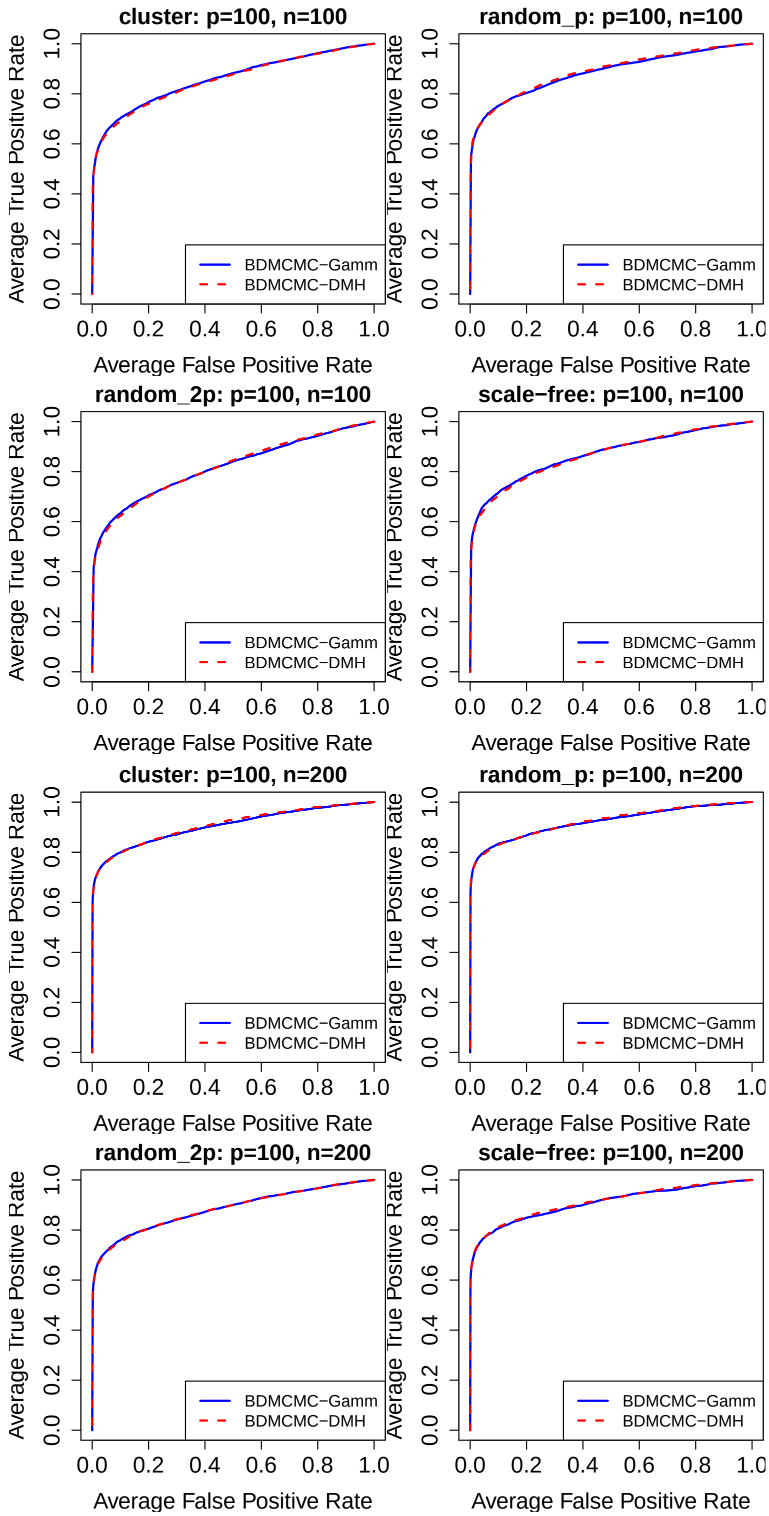


Figure G.6: ROC curves for the BDMCMC algorithm with our approximation in Equation 8 (BDMCMC-Gamm) and BDMCMC algorithm with exchange algorithm (BDMCMC-DMH), over 50 replications. Here, $p = 100$, $n \in \{100, 200\}$, and 4 different graph structures.