Supplementary material for “Clustering microbiome data using mixtures of logistic normal multinomial models”

Yuan Fang\textsuperscript{1} \quad Sanjeena Subedi\textsuperscript{2,*}

\textsuperscript{1}School of Pharmacy and Pharmaceutical Sciences, Binghamton University, State University of New York, 4400 Vestal Parkway East, Binghamton, NY 13902, USA.
\textsuperscript{2}School of Mathematics & Statistics, 4302 Herzberg Laboratories, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada.
\textsuperscript{*Correspondence to: sanjeena.dang@carleton.ca.}

1 Mathematical Detail

Consider the following transformed parameter $\eta$ from $Y$:

$$
\eta = BY,
$$

where $B = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ 0 & 0 & \ldots & 0 \end{pmatrix}$

is a $(K + 1) \times K$ matrix which takes the form as an identity matrix attached by a row of $K$ zeros. Given the assumption that the true distribution of $Y$ is $N(\mu, \Sigma)$, we have the true distribution of $\eta$ to be Gaussian too, with mean $\tilde{\mu}$ and covariance matrix $\tilde{\Sigma}$, where

$$
\tilde{\mu} = B\mu = (\mu, 0)^\top;
\tilde{\Sigma} = B\Sigma B^\top = \begin{pmatrix} \Sigma & 0_{K \times 1} \\ 0_{1 \times K} & 0 \end{pmatrix}.
$$

For computational convenience, we further assume that $V$ has a diagonal structure, with each diagonal element denoted as $v_k^2$ such that

$$
v_k^2 = \begin{cases} v_k^2, & k = 1, \ldots, K \\ 0, & k = K + 1. \end{cases}
$$

We also denote the $k$–th element of $m$ as $m_k$ such that

$$
m_k = \begin{cases} m_k, & k = 1, \ldots, K \\ 0, & k = K + 1. \end{cases}
$$

Recall that we have the following decomposition of the ELBO

$$
F(q(\eta), w) = F(m, V) = -\int q(\eta) \log q(\eta) d\eta + \int q(\eta) \log p(\eta) d\eta + \int q(\eta) \log p(w | \eta) d\eta;
$$
among which, the first integral by definition is the entropy of the variational Gaussian distribution $q(\eta|m, V)$:

$$- \int q(\eta) \log q(\eta) d\eta = -E_{q(\eta|m, V)}(q(\eta)) = \frac{1}{2} \sum_{k=1}^{K} \log(v_k^2) + \frac{K}{2} \log(2\pi) + \frac{K}{2}.$$  

The second integral can be evaluated explicitly as well, which turn into the expected value of the log density function of $p(\eta) = N(\eta|\mu, \Sigma)$ with respect to $q(\eta|m, V)$:

$$\int q(\eta) \log p(\eta) d\eta = E_{q(\eta|m, V)}(\log p(\eta))$$

$$= E_{q(\eta|m, V)} \left( -\frac{K}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\eta - \mu)^\top \Sigma^{-1} (\eta - \mu) \right)$$

$$= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (m - \mu)^\top \Sigma^{-1} (m - \mu) - \frac{1}{2} \text{Tr}(\Sigma^{-1} V).$$

Due to the special structure of $\tilde{\Sigma}$, we have $|\tilde{\Sigma}| = 0$ and $\tilde{\Sigma}^{-1}$ does not exist, which brings in a computational issue. Therefore, we substitute $|\tilde{\Sigma}|$ by $|\Sigma| = |B^\top \Sigma B|$ and $\tilde{\Sigma}^{-1}$ by the generalized inverse of $\Sigma$

$$\tilde{\Sigma}^* = \begin{pmatrix} \Sigma^{-1} & 0_K \\ 0_{K \times 1} & 0 \end{pmatrix}.$$  

Hence we have

$$\int q(\eta) \log p(\eta) d\eta = -\frac{K}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (m - \mu)^\top \tilde{\Sigma}^* (m - \mu) - \frac{1}{2} \text{Tr}(\tilde{\Sigma}^* V)$$

$$= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \log |B^\top \Sigma B| - \frac{1}{2} (m - \mu)^\top \tilde{\Sigma}^* (m - \mu) - \frac{1}{2} \text{Tr}(\tilde{\Sigma}^* V).$$

The third integral is intractable, because of the log-sum exponential term. We upper bound this term with a Taylor expansion similar to previous literature (Blei and Lafferty (2007)) resulting in the following

$$E_{q(\eta|m, V)} \left[ \log \left( \sum_{k=1}^{K+1} \exp \eta_k \right) \right] \leq \xi^{-1} \left\{ \sum_{k=1}^{K+1} E_{q(\eta|m, V)} [\exp(\eta_k)] \right\} - 1 + \log(\xi)$$

$$= \xi^{-1} \left\{ \sum_{k=1}^{K+1} \exp \left( m_k + \frac{v_k^2}{2} \right) \right\} - 1 + \log(\xi).$$

Therefore, the third integral is lower bounded by

$$\int q(\eta) \log p(w|\eta) d\eta = E_{q(\eta|m, V)} \left[ \log \left( \sum_{k=1}^{K+1} \exp \eta_k \right) \right]$$

$$= w^\top \hat{m} - \left( \sum_{k=1}^{K+1} w_k \right) E_{q(\eta|m, V)} \left[ \log \left( \sum_{k=1}^{K+1} \exp \eta_k \right) \right]$$

$$\geq w^\top \hat{m} - \left( \sum_{k=1}^{K+1} w_k \right) \left\{ \xi^{-1} \sum_{k=1}^{K+1} \exp \left( m_k + \frac{v_k^2}{2} \right) \right\} - 1 + \log(\xi).$$
Combining all three integrals, we obtain a concave variational Gaussian lower bound to the model evidence

\[
\tilde{F}(m, V, \tilde{\mu}, \tilde{\Sigma}, \xi) = w^\top m - \left( \sum_{k=1}^{K+1} w_k \right) \left\{ \xi^{-1} \left[ \sum_{k=1}^{K+1} \exp \left( m_k + \frac{v_k^2}{2} \right) \right] - 1 + \log(\xi) \right\} \\
- \frac{1}{2} \log |B^\top \tilde{\Sigma} B| - \frac{1}{2} (m - \tilde{\mu})^\top \tilde{\Sigma}^\ast (m - \tilde{\mu}) - \frac{1}{2} \text{Tr}(\tilde{\Sigma}^\ast V) + \frac{1}{2} \sum_{k=1}^{K} \log(v_k^2) + \frac{K}{2}.
\]

We maximize this lower bound with respect to the variational parameters \(\xi, m, V\).

First, we maximize the lower bound \(\tilde{F}\) with respect to \(\xi\). The derivative with respect to \(\xi\) is

\[
\frac{\partial \tilde{F}}{\partial \xi} = \left( \sum_{k=1}^{K+1} w_k \right) \left\{ -\xi^{-2} \left[ \sum_{k=1}^{K+1} \exp \left( m_k + \frac{v_k^2}{2} \right) \right] + \xi^{-1} \right\},
\]

which yields an optimizer at

\[
\hat{\xi} = \sum_{k=1}^{K+1} \exp \left( m_k + \frac{v_k^2}{2} \right).
\]

Second, we maximize with respect to \(m\), of which the derivative is given as

\[
\frac{\partial \tilde{F}}{\partial m} = w - \tilde{\Sigma}^\ast (m - \tilde{\mu}) - \left( \sum_{k=1}^{K+1} w_k \right) \xi^{-1} \exp \left( m + \frac{v^2}{2} \right),
\]

with \(v^2 = (v_1^2, \ldots, v_K^2, 0)\) denoting the diagonal element of \(V\) as a vector. There is no analytical solution to this derivative and so we use Newton’s method to approximate the root to this derivative, with a constrain that the \((K + 1)\)-th element is zero. The procedure requires the Hessian matrix with respect to \(m\):

\[
H_m = -\tilde{\Sigma}^\ast - \left( \sum_{k=1}^{K+1} w_k \right) \xi^{-1} \text{diag} \left\{ \exp \left( m + \frac{v^2}{2} \right) \right\}.
\]

Finally, we optimize with respect to \(v_k\), for \(k = 1, \ldots, K\) and always set \(v_{K+1}\) as zero. Again, there are no analytical solutions and Newton’s method is used for each coordinate. The first and second derivatives with respect to \(v_k\) for \(k = 1, \ldots, K\) are given as follows

\[
\frac{\partial \tilde{F}}{\partial v_k} = v_k^{-1} - v_k \tilde{\Sigma}^\ast_{k,k} - \left( \sum_{k=1}^{K+1} w_k \right) \xi^{-1} \exp \left( m_k + \frac{v_k^2}{2} \right) v_k;
\]

\[
\frac{\partial^2 \tilde{F}}{\partial v_k^2} = -v_k^{-2} - \left( \sum_{k=1}^{K+1} w_k \right) \xi^{-1} \exp \left( m_k + \frac{v_k^2}{2} \right) (v_k^2 + 1).
\]

At each iteration of the variational EM algorithm, when we maximize the variational Gaussian lower bound \(\tilde{F}(m, V, \tilde{\mu}, \tilde{\Sigma}, \xi)\) with respect to the variational parameter set \((\xi, m, V)\), we take one step of update based on the optimization discussed above.
## 2 Parameter recovery of Simulation Studies

Table 1: True and estimated parameters along with the standard deviations from the one hundred datasets from Simulation Study 1 using VGA approach; average ARI= 0.94 (0.02).

| Component 1 ($n = 600$) | Parameter | True | Average of the estimates (sd) |
|-------------------------|-----------|------|------------------------------|
| $\mu$                   | [5, 2, 1] | [5.00 (0.05), 2.01 (0.05), 1.01 (0.04)] |
| $\Sigma$                | $\begin{bmatrix} 1 & 0.4 & 0 \\ 0.4 & 1.2 & -0.5 \\ 0 & -0.5 & 1 \end{bmatrix}$ | $\begin{bmatrix} 1.05 (0.07) & 0.44 (0.06) & 0.02 (0.05) \\ 0.44 (0.06) & 1.24 (0.08) & -0.48 (0.06) \\ 0.02 (0.05) & -0.48 (0.06) & 1.00 (0.06) \end{bmatrix}$ |

| Component 2 ($n = 400$) | Parameter | True | Average of the estimates (sd) |
|-------------------------|-----------|------|------------------------------|
| $\mu$                   | [1, 3, 2] | [1.01 (0.07), 3.01 (0.05), 2.01 (0.05)] |
| $\Sigma$                | $\begin{bmatrix} 1.4 & 0.2 & -0.65 \\ 0.2 & 1 & 0 \\ -0.65 & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} 1.39 (0.11) & 0.21 (0.07) & -0.63 (0.08) \\ 0.21 (0.07) & 1.00 (0.08) & -0.01 (0.05) \\ -0.63 (0.08) & -0.01 (0.05) & 0.96 (0.08) \end{bmatrix}$ |

Table 2: True and estimated parameters along with the standard deviations from the one hundred datasets for the latent variable parameters in Simulation Study 1 using the hybrid approach.

| Component 1 ($n = 600$) | Parameter | True | Average of the estimates (sd) |
|-------------------------|-----------|------|------------------------------|
| $\mu$                   | [5, 2, 1] | [4.99(0.05), 2.00(0.05), 0.99(0.04)] |
| $\Sigma$                | $\begin{bmatrix} 1 & 0.4 & 0 \\ 0.4 & 1.2 & -0.5 \\ 0 & -0.5 & 1 \end{bmatrix}$ | $\begin{bmatrix} 0.98(0.07) & 0.41(0.05) & 0.00(0.05) \\ 0.41(0.05) & 1.21(0.08) & -0.50(0.06) \\ 0.00(0.05) & -0.50(0.06) & 0.99(0.06) \end{bmatrix}$ |

| Component 2 ($n = 400$) | Parameter | True | Average of the estimates (sd) |
|-------------------------|-----------|------|------------------------------|
| $\mu$                   | [1, 3, 2] | [0.98(0.07), 3.01(0.05), 2.02(0.05)] |
| $\Sigma$                | $\begin{bmatrix} 1.4 & 0.2 & -0.65 \\ 0.2 & 1 & 0 \\ -0.65 & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} 1.32(0.11) & 0.23(0.07) & -0.59(0.07) \\ 0.23(0.07) & 1.00(0.08) & -0.01(0.05) \\ -0.59(0.07) & -0.01(0.05) & 0.95(0.08) \end{bmatrix}$ |
Table 3: True and estimated values for the parameters from Simulation Study 2 using VGA approach; Average ARI= 0.93 (sd 0.02).

| Parameter | True | Estimated (sd) |
|-----------|------|----------------|
| **Component 1 (n = 300)** |      |                |
| $\mu$    | $[5, 2, 1, 2, 3]$ | $[5.02 (0.09), 2.02 (0.06), 1.01 (0.07), 2.01 (0.07), 3.02 (0.06)]$ |
|          | $[2, -0.2, 0.8, -1, 0]$ | $[2.06 (0.17), -0.14 (0.09), 0.85 (0.13), -0.96 (0.12), 0.06 (0.10)]$ |
| $\Sigma$ | $[0.8, -0.2, 1.4, 0.6, 0]$ | $[0.85 (0.13), -0.15 (0.07), 1.46 (0.13), 0.64 (0.10), 0.04 (0.09)]$ |
|          | $[-1, 0, 0.6, 1.6, 0.2]$ | $[-0.96 (0.12), 0.01 (0.08), 0.64 (0.10), 1.64 (0.14), 0.22 (0.08)]$ |
|          | $[0, -0.4, 0, 0.2, 1.2]$ | $[0.06 (0.10), -0.36 (0.07), 0.04 (0.09), 0.22 (0.08), 1.25 (0.11)]$ |
| **Component 2 (n = 400)** |      |                |
| $\mu$    | $[2, 3, 4, 1, 2]$ | $[2.04 (0.07), 3.04 (0.06), 4.04 (0.06), 1.02 (0.06), 2.02 (0.07)]$ |
|          | $[1.4, 0.65, 0.4, 0, 0]$ | $[1.34 (0.10), 0.61 (0.06), 0.36 (0.08), -0.02 (0.08), -0.02 (0.09)]$ |
| $\Sigma$ | $[0.65, 1, 0.2, 0, 0.4]$ | $[0.61 (0.06), 0.96 (0.08), 0.16 (0.06), -0.02 (0.06), 0.39 (0.08)]$ |
|          | $[0.4, 0.2, 1, 0.6, 0]$ | $[0.36 (0.08), 0.16 (0.06), 0.96 (0.08), 0.59 (0.06), -0.02 (0.08)]$ |
|          | $[0, 0, 0.6, 1.2, 0.8]$ | $[-0.02 (0.08), 0.39 (0.08), -0.02 (0.08), 0.80 (0.09)]$ |
|          | $[0, 0.4, 0, 0.8, 2]$ | $[-0.02 (0.09), 0.39 (0.08), -0.02 (0.08), 0.80 (0.09), 2.01 (0.15)]$ |
| **Component 3 (n = 200)** |      |                |
| $\mu$    | $[1, 1, 1, 1, 1]$ | $[0.99 (0.08), 1.03 (0.08), 1.10 (0.10), 0.96 (0.09), 1.03 (0.09)]$ |
|          | $[1, 0, 0, 0, 0]$ | $[0.99 (0.12), -0.00 (0.08), -0.03 (0.08), -0.01 (0.08), -0.03 (0.06)]$ |
| $\Sigma$ | $[0, 1, 0, 0, 0]$ | $[0.00 (0.08), 1.00 (0.11), 0.06 (0.10), -0.03 (0.08), 0.03 (0.08)]$ |
|          | $[0, 0, 1, 0, 0]$ | $[-0.03 (0.08), 0.06 (0.10), 1.11 (0.14), -0.05 (0.08), 0.03 (0.09)]$ |
|          | $[0, 0, 0, 1, 0]$ | $[-0.01 (0.08), -0.03 (0.08), -0.05 (0.08), 1.00 (0.11), 0.00 (0.07)]$ |
|          | $[0, 0, 0, 0, 1]$ | $[-0.03 (0.06), 0.03 (0.08), 0.03 (0.09), 0.00 (0.07), 1.00 (0.10)]$ |
Table 4: True and estimated parameters (mean (sd)) for Simulation Study 2 using the hybrid approach.

| Parameter | True | Estimated (sd) |
|-----------|------|----------------|
| **Component 1** \((n = 300)\) | | |
| \(\mu\) | \([5, 2, 1, 2, 3]\) | \([5.00(0.09), 2.00(0.06), 0.99(0.08), 2.00(0.08), 3.00(0.06)]\) |
| | | |
| | | |
| | | |
| **Component 2** \((n = 400)\) | | |
| \(\mu\) | \([2, 3, 4, 1, 2]\) | \([1.99(0.07), 2.99(0.06), 3.99(0.06), 1.00(0.06), 2.00(0.07)]\) |
| | | |
| | | |
| **Component 3** \((n = 200)\) | | |
| \(\mu\) | \([1, 1, 1, 1, 1]\) | \([0.99(0.08), 0.96(0.08), 0.97(0.09), 1.01(0.09), 1.01(0.09)]\) |
| | | |
| | | |
| | | |
| | | |
Table 5: Summary of the number of times various $G$ are selected for simulation studies described in the Additional Simulation Studies section.

| Simulation setting | Proposed algorithm | Dirichlet-multinomial mixture model | GMM on ALR transformed data |
|--------------------|--------------------|------------------------------------|-----------------------------|
|                    | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ |
| K=5, n=100         | 100   | 99    | 1     | 2     | 6     | 36    | 56    |
| K=5, n=200         | 100   | 94    | 5     | 1     | 11    | 89    |
| K=5, n=500         | 100   | 1     | 3     | 40    | 26    | 30    | 100   |
| K=10, n=100        | 100   | 26    | 64    | 10    | 7     | 31    | 62    |
| K=10, n=200        | 100   | 88    | 10    | 1     | 1     | 1     | 99    |
| K=10, n=500        | 100   | 0     | 17    | 83    | 4     | 96    |
| K=20, n=100        | 100   | 80    | 20    | 28    | 43    | 20    | 9     |
| K=20, n=200        | 100   | 99    | 1     | 20    | 79    |
| K=20, n=500        | 100   | 0     | 30    | 34    | 36    | 100   |
| DMM (k=5, n=200)   | 1     | 61    | 29    | 8     | 1     | 100   | 7     | 93    |
| High dimensional K=50, n=500 | 100 | 27 | 30 | 17 | 17 | 9 | 29 | 71* |

*Note that $G > 2$ encountered computational issues when fitting GMM with unrestricted spherical cluster model (“VII”) on ALR transformed data for high dimensional $K = 50$ scenario for all datasets. Thus, only $G = 1$ and $G = 2$ were fitted and $G = 1, \ldots, 5$ could only be fitted for the model with equal spherical covariance across components (“EII”).
Table 6: Average $L_1$ norm with standard error of the true parameters and estimated values for simulation studies described in the Additional Simulation Studies section.

| Simulation setting | Component 1 ($\pi_1 = 0.5$) | Component 2 ($\pi_2 = 0.5$) |
|--------------------|-------------------------------|-------------------------------|
|                    | Average (sd) of $\hat{\pi}_1$ | Average (sd) of $|\hat{\mu}_1 - \mu_1|_{L_1}$ | Average (sd) of $|\hat{\Sigma}_1 - \Sigma_1|_{L_1}$ | Average (sd) of $|\hat{\pi}_2$ | Average (sd) of $|\hat{\mu}_2 - \mu_2|_{L_1}$ | Average (sd) of $|\hat{\Sigma}_2 - \Sigma_2|_{L_1}$ |
| K=5, n=100         | 0.4996 (0.009)                | 0.3789 (0.1535)               | 1.281 (0.3175) | 0.5004 (0.009)                | 0.3094 (0.1344)               | 0.9111 (0.2978) |
| K=5, n=200         | 0.5005 (0.0038)               | 0.2425 (0.1021)               | 0.8694 (0.2536) | 0.4995 (0.0038)               | 0.203 (0.0812)                | 0.6462 (0.2292) |
| K=5, n=500         | 0.4999 (0.0023)               | 0.1489 (0.0592)               | 0.574 (0.1597) | 0.5001 (0.0023)               | 0.1293 (0.0604)               | 0.3882 (0.1385) |
| K=10, n=100        | 0.4896 (0.0129)               | 0.9975 (0.2892)               | 10.1231 (1.8573) | 0.496 (0.0067)               | 0.7483 (0.2515)               | 7.1662 (1.2323) |
| K=10, n=200        | 0.496 (0.0067)                | 0.588 (0.1657)                | 7.8635 (2.0106) | 0.5005 (0.0024)               | 0.588 (0.1657)                | 7.8635 (2.0106) |
| K=10, n=500        | 0.4998 (6e-04)                | 0.7108 (0.1562)               | 10.1909 (0.9694) | 0.5001 (0.0023)               | 0.1293 (0.0604)               | 0.3882 (0.1385) |
| K=20, n=100        | 0.4546 (0.0212)               | 1.7992 (0.3343)               | 26.4108 (4.9824) | 0.4991 (0.0047)               | 1.111 (0.266)                 | 15.718 (1.5242) |
| K=20, n=200        | 0.4991 (0.0047)               | 1.111 (0.266)                 | 15.718 (1.5242) | 0.4998 (6e-04)                | 0.7108 (0.1562)               | 10.1909 (0.9694) |
| K=20, n=500        | 0.4995 (6e-04)                | 0.7108 (0.1562)               | 10.1909 (0.9694) | 0.4995 (6e-04)                | 0.7108 (0.1562)               | 10.1909 (0.9694) |
| K=20, n=500        | 0.4995 (6e-04)                | 0.7108 (0.1562)               | 10.1909 (0.9694) | 0.4995 (6e-04)                | 0.7108 (0.1562)               | 10.1909 (0.9694) |