

Fisher Information Matrix of Multivariate Gaussian random effect model

Dongyue Xie

1 Model

Assume

$$\mathbf{x}_i \sim N(\mathbf{0}, \mathbf{U} + \mathbf{D} + \mathbf{V}_i), \quad (1)$$

where $\mathbf{V}_i \in \mathbb{R}^{R \times R}$ is a known matrix, \mathbf{D} is a diagonal matrix and \mathbf{U} is an unstructured covariance matrix. The goal is to derive the Fisher information matrix of parameters in \mathbf{U}, \mathbf{D} .

Denote $\mathbf{S}_i = \mathbf{U} + \mathbf{D} + \mathbf{V}_i$, and $\boldsymbol{\theta} = (\mathbf{u}, \boldsymbol{\sigma}^2)^T$, where $\mathbf{u} = \text{vec}(\mathbf{U})$ and contains only unique elements so $\mathbf{u} \in \mathbb{R}^{R(R+1)/2}$, and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_R^2)$. Note that elements of \mathbf{S}_i are functions of $\boldsymbol{\theta}$ and for simplicity we write \mathbf{S}_i instead of $\mathbf{S}_i(\boldsymbol{\theta})$.

The log likelihood of \mathbf{x}_i is

$$l(\boldsymbol{\theta}; \mathbf{x}_i) = -\frac{1}{2} \log |\mathbf{S}_i| - \frac{1}{2} \mathbf{x}_i^T \mathbf{S}_i^{-1} \mathbf{x}_i. \quad (2)$$

The gradient of $l(\boldsymbol{\theta}; \mathbf{x}_i)$ with respect to the j th entry of $\boldsymbol{\theta}$ gives the score function

$$\frac{\partial l(\boldsymbol{\theta}; \mathbf{x}_i)}{\partial \theta_j} = -\frac{1}{2} \text{Tr}(\mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_j}) + \frac{1}{2} \mathbf{x}_i^T \mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_j} \mathbf{S}_i^{-1} \mathbf{x}_i. \quad (3)$$

The second order derivative of $l(\boldsymbol{\theta}; \mathbf{x}_i)$ with respect to $\boldsymbol{\theta}$ gives the Hessian,

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{x}_i)}{\partial \theta_j \partial \theta_k} = & -\frac{1}{2} \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial^2 \mathbf{S}_i}{\partial \theta_j \partial \theta_k} - \mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_j} \mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_k} \right) \\ & + \frac{1}{2} \mathbf{x}_i^T \mathbf{S}_i^{-1} \left(\frac{\partial^2 \mathbf{S}_i}{\partial \theta_j \partial \theta_k} - 2 \frac{\partial \mathbf{S}_i}{\partial \theta_j} \mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_k} \right) \mathbf{S}_i^{-1} \mathbf{x}_i. \end{aligned} \quad (4)$$

The Fisher information matrix is given by the expectation of the negative Hessian,

$$-\mathbb{E} \left(\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{x}_i)}{\partial \theta_j \partial \theta_k} \right) = \frac{1}{2} \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_j} \mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_k} \right). \quad (5)$$

We denote the Fisher information matrix as $\mathbf{I}(\boldsymbol{\theta})$, and it's j, k th element is

$$\mathbf{I}_{jk}(\boldsymbol{\theta}) = \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_j} \mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i}{\partial \theta_k} \right). \quad (6)$$

The asymptotic variance of $\hat{\boldsymbol{\theta}}^{mle}$ is given by $I(\boldsymbol{\theta})^{-1}$, and can be estimated by $I(\hat{\boldsymbol{\theta}})^{-1}$.

We can partition the information matrix into the following four blocks

$$\begin{aligned} \mathbf{I}(\boldsymbol{\theta}) &= \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mathbf{u} \partial \mathbf{u}^T} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \mathbf{u} \partial (\boldsymbol{\sigma}^2)^T} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}^{2^T} \partial \mathbf{u}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}^{2^T} \partial \boldsymbol{\sigma}^2} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix} \end{aligned} \quad (7)$$

We order the elements in \mathbf{u} as the unique elements in each row of matrix \mathbf{U} , i.e.

$$\mathbf{u} = (u_{11}, u_{12}, \dots, u_{1R}, u_{22}, \dots, u_{2R}, \dots, u_{RR}). \quad (8)$$

The matrix \mathbf{A} of dimension $\frac{R(R+1)}{2} \times \frac{R(R+1)}{2}$ is

$$\mathbf{A} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 u_{11}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial u_{12}} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial u_{RR}} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial u_{11}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 u_{12}} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial u_{RR}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial u_{11}} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial u_{12}} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 u_{RR}} \end{pmatrix}. \quad (9)$$

The matrix \mathbf{B} of dimension $\frac{R(R+1)}{2} \times R$ is

$$\mathbf{B} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{11} \partial \sigma_R^2} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{12} \partial \sigma_R^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial u_{RR} \partial \sigma_R^2} \end{pmatrix}. \quad (10)$$

The matrix \mathbf{C} of dimension $R \times R$ is

$$\mathbf{C} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_1^2 \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_1^2 \partial \sigma_R^2} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_2^2 \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_2^2 \partial \sigma_R^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_R^2 \partial \sigma_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma_R^2 \partial \sigma_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial^2 \sigma_R^2} \end{pmatrix}. \quad (11)$$

Let \mathbf{e}_j denote a vector that the j th entry is 1 and others are 0. The following derivatives hold

$$\begin{aligned}\frac{\partial \mathbf{S}_i}{\partial u_{rr}} &= \frac{\partial \mathbf{U}}{\partial u_{rr}} = \mathbf{e}_r \mathbf{e}_r^T, \\ \frac{\partial \mathbf{S}_i}{\partial u_{r_1 r_2}} &= \frac{\partial \mathbf{U}}{\partial u_{r_1 r_2}} = \mathbf{e}_{r_1} \mathbf{e}_{r_2}^T + \mathbf{e}_{r_2} \mathbf{e}_{r_1}^T, \\ \frac{\partial \mathbf{S}_i}{\partial \sigma_r^2} &= \frac{\partial \mathbf{D}}{\partial \sigma_r^2} = \mathbf{e}_r \mathbf{e}_r^T.\end{aligned}\tag{12}$$

1.1 Calculation of matrix \mathbf{A} :

Denote $\tilde{\mathbf{S}}_i = \mathbf{S}_i^{-1}$, with column vectors $\tilde{\mathbf{s}}_{i1}, \dots, \tilde{\mathbf{s}}_{iR}$.

Denote the (r_1, r_2) th entry of matrix $\tilde{\mathbf{S}}^{-1}$ as $\tilde{s}_{i, r_1 r_2}$.

We denote the entries of matrix \mathbf{A} as a_{rj_r, lk_l} , for $r, l \in [R]$, $j_r = r, r+1, \dots, R-r+1$, and $k_l = l, l+1, \dots, R-l+1$. Each entry is given by

$$a_{rj_r, lk_l} = \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i^{-1}}{\partial u_{rj_r}} \mathbf{S}_i^{-1} \frac{\partial \mathbf{S}_i^{-1}}{\partial u_{lk_l}} \right).\tag{13}$$

The matrix \mathbf{A} is symmetric since $a_{rj_r, lk_l} = a_{lk_l, rj_r}$. We consider the following cases:

- When $r = j_r, l = k_l$,

$$\begin{aligned}a_{rr, ll} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rr}} \mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{ll}} \right) \\ &= \frac{1}{2} \sum_i \text{Tr} (\mathbf{S}_i^{-1} \mathbf{e}_r \mathbf{e}_r^T \mathbf{S}_i^{-1} \mathbf{e}_l \mathbf{e}_l^T) \\ &= \frac{1}{2} \sum_i \mathbf{e}_l^T (\mathbf{S}_i^{-1} \mathbf{e}_r \mathbf{e}_r^T \mathbf{S}_i^{-1}) \mathbf{e}_l \\ &= \frac{1}{2} \sum_i (\tilde{s}_{i, rl})^2.\end{aligned}\tag{14}$$

- When $r = j_r, l \neq k_l$,

$$\begin{aligned}a_{rr, lk_l} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rr}} \mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{lk_l}} \right) \\ &= \frac{1}{2} \sum_i \text{Tr} (\mathbf{S}_i^{-1} \mathbf{e}_r \mathbf{e}_r^T \mathbf{S}_i^{-1} (\mathbf{e}_l \mathbf{e}_{k_l}^T + \mathbf{e}_{k_l} \mathbf{e}_l^T)) \\ &= \frac{1}{2} \sum_i (\mathbf{e}_{k_l}^T (\mathbf{S}_i^{-1} \mathbf{e}_r \mathbf{e}_r^T \mathbf{S}_i^{-1}) \mathbf{e}_l + \mathbf{e}_l^T (\mathbf{S}_i^{-1} \mathbf{e}_r \mathbf{e}_r^T \mathbf{S}_i^{-1}) \mathbf{e}_{k_l}) \\ &= \sum_i \tilde{s}_{i, rl} \tilde{s}_{i, rk_l}.\end{aligned}\tag{15}$$

- When $r \neq j_r, l \neq k_l$,

$$\begin{aligned}
a_{rj_r, lk_l} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rj_r}} \mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{lk_l}} \right) \\
&= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} (\mathbf{e}_r \mathbf{e}_{j_r}^T + \mathbf{e}_{j_r} \mathbf{e}_r^T) \mathbf{S}_i^{-1} (\mathbf{e}_l \mathbf{e}_{k_l}^T + \mathbf{e}_{k_l} \mathbf{e}_l^T) \right) \\
&= \sum_i (\tilde{s}_{i, lj_r} \tilde{s}_{i, rk_l} + \tilde{s}_{i, rl} \tilde{s}_{i, j_r k_l}).
\end{aligned} \tag{16}$$

1.2 Calculation of matrix \mathbf{B}

Denote the entries of matrix \mathbf{B} as $b_{rj_r, l}$, for $r, l \in [R]$, $j_r = r, r+1, \dots, R-r+1$. Each entry is given by

$$b_{rj_r, l} = \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rj_r}} \mathbf{S}_i^{-1} \frac{\partial \mathbf{D}}{\partial \sigma_l^2} \right). \tag{17}$$

We consider the following cases:

- When $r = j_r$,

$$\begin{aligned}
b_{rr, l} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rr}} \mathbf{S}_i^{-1} \frac{\partial \mathbf{D}}{\partial \sigma_l^2} \right) \\
&= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} (\mathbf{e}_r \mathbf{e}_r^T) \mathbf{S}_i^{-1} (\mathbf{e}_l \mathbf{e}_l^T) \right) \\
&= \frac{1}{2} \sum_i (\tilde{s}_{i, rl})^2.
\end{aligned} \tag{18}$$

- When $r \neq j_r$,

$$\begin{aligned}
b_{rj_r, l} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rj_r}} \mathbf{S}_i^{-1} \frac{\partial \mathbf{D}}{\partial \sigma_l^2} \right) \\
&= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} (\mathbf{e}_r \mathbf{e}_{j_r}^T + \mathbf{e}_{j_r} \mathbf{e}_r^T) \mathbf{S}_i^{-1} (\mathbf{e}_l \mathbf{e}_l^T) \right) \\
&= \sum_i (\tilde{s}_{i, rl} \tilde{s}_{i, j_r l}).
\end{aligned} \tag{19}$$

In the special case where $\mathbf{D} = \sigma^2 \mathbf{I}$, $\mathbf{B} = \mathbf{b}$ is a vector of length $R(R+1)/2$, and

- When $r = j_r$,

$$\begin{aligned}
b_{rr} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rr}} \mathbf{S}_i^{-1} \right) \\
&= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} (\mathbf{e}_r \mathbf{e}_r^T) \mathbf{S}_i^{-1} \right) \\
&= \frac{1}{2} \sum_i \sum_{l=1}^R (\tilde{s}_{i, rl})^2.
\end{aligned} \tag{20}$$

- When $r \neq j_r$,

$$\begin{aligned}
b_{rj_r} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{U}}{\partial u_{rj_r}} \mathbf{S}_i^{-1} \right) \\
&= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} (\mathbf{e}_r \mathbf{e}_{j_r}^T + \mathbf{e}_{j_r} \mathbf{e}_r^T) \mathbf{S}_i^{-1} \right) \\
&= \sum_i \sum_{l=1}^R (\tilde{s}_{i,rl} \tilde{s}_{i,j_rl}).
\end{aligned} \tag{21}$$

1.3 Calculation of matrix \mathbf{C}

Denote the entries of \mathbf{C} as c_{rl} , for $r, l \in [R]$. Each entry of \mathbf{C} is given by

$$\begin{aligned}
c_{rl} &= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} \frac{\partial \mathbf{D}}{\partial \sigma_r^2} \mathbf{S}_i^{-1} \frac{\partial \mathbf{D}}{\partial \sigma_l^2} \right) \\
&= \frac{1}{2} \sum_i \text{Tr} \left(\mathbf{S}_i^{-1} (\mathbf{e}_r \mathbf{e}_r^T) \mathbf{S}_i^{-1} (\mathbf{e}_l \mathbf{e}_l^T) \right) \\
&= \frac{1}{2} \sum_i \tilde{s}_{i,rl}^2
\end{aligned} \tag{22}$$

In the special case where $\mathbf{D} = \sigma^2 \mathbf{I}$, $\mathbf{C} = c$ is a scalar, and

$$\begin{aligned}
c &= \frac{1}{2} \sum_i \text{Tr} (\mathbf{S}_i^{-1} \mathbf{S}_i^{-1}) \\
&= \frac{1}{2} \sum_i \sum_r \sum_l \tilde{s}_{i,rl}^2.
\end{aligned} \tag{23}$$

2 Inverse Fisher Information Matrix

We are interested in the variance of σ^2 , which is given by

$$(\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B})^{-1}. \tag{24}$$

Let's consider the case $\mathbf{D} = \sigma^2 \mathbf{I}$ so $\text{var}(\hat{\sigma}^2) = \frac{1}{c - \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}}$. The difficulty of obtaining $\text{var}(\hat{\sigma}^2)$ is from formulating \mathbf{b} , \mathbf{A} and get the inverse of \mathbf{A} .

Since the Fisher information matrix is always positive semidefinite and we have assumed it's nonsingular, we have $\mathbf{b}^T \mathbf{A}^{-1} \mathbf{b} > 0$, and

$$\text{var}(\hat{\sigma}^2) = \frac{1}{c - \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}} > \frac{1}{c}. \tag{25}$$

We can use $1/c$ as a lower bound of $\text{var}(\hat{\sigma}^2)$ in practice (if not evaluating \mathbf{A}).

For a general \mathbf{D} , we are interested in the diagonal of $\text{cov}(\hat{\sigma}^2) = (\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B})^{-1}$.

Lemma 2.1. Let \mathbf{X} be a positive definite matrix, then $(\mathbf{X}^{-1})_{jj} \geq \frac{1}{\mathbf{x}_{jj}}$.

Proof. The eigen-decomposition of \mathbf{X} is $\mathbf{X} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$. Then

$$\mathbf{X}_{jj} = \mathbf{q}_j^T \mathbf{\Lambda} \mathbf{q}_j = \sum_k q_{jk}^2 \lambda_k,$$

and

$$(\mathbf{X})_{jj}^{-1} = \mathbf{q}_j^T \mathbf{\Lambda}^{-1} \mathbf{q}_j = \sum_k q_{jk}^2 / \lambda_k.$$

□

By Cauchy–Schwarz inequality,

$$\mathbf{X}_{jj}(\mathbf{X})_{jj}^{-1} \geq (\sum_k q_{jk})^2 = 1.$$

Since by assumption $\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B}$ is positive definite, by lemma (2.1), we have

$$(\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B})_{rr}^{-1} \geq 1/(\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B})_{rr} = \frac{1}{\mathbf{C}_{rr} - \mathbf{b}_r^T \mathbf{A}^{-1} \mathbf{b}_r} \geq \frac{1}{\mathbf{C}_{rr}}. \quad (26)$$

We summarize the results below.

1. When $\mathbf{V}_i = \mathbf{I}$,
 - if $\mathbf{D} = \sigma^2 \mathbf{I}$, $\text{var}(\hat{\sigma}^2) \geq \frac{2}{N \text{Tr}((\mathbf{U} + \mathbf{I})^{-1}(\mathbf{U} + \mathbf{I})^{-1})}$.
 - if $\mathbf{D} = \text{diag}(\boldsymbol{\sigma}^2)$, $\text{var}(\hat{\sigma}_r^2) \geq \frac{2}{N((\mathbf{U} + \mathbf{I})_{rr}^{-1})^2}$.
2. When $\mathbf{V}_i = \mathbf{V}$,
 - if $\mathbf{D} = \sigma^2 \mathbf{I}$, $\text{var}(\hat{\sigma}^2) \geq \frac{2}{N \text{Tr}((\mathbf{U} + \mathbf{V})^{-1}(\mathbf{U} + \mathbf{V})^{-1})}$.
 - if $\mathbf{D} = \text{diag}(\boldsymbol{\sigma}^2)$, $\text{var}(\hat{\sigma}_r^2) \geq \frac{2}{N((\mathbf{U} + \mathbf{V})_{rr}^{-1})^2}$.
3. When \mathbf{V}_i varies with samples,
 - if $\mathbf{D} = \sigma^2 \mathbf{I}$, $\text{var}(\hat{\sigma}^2) \geq \frac{2}{\sum_i \text{Tr}((\mathbf{U} + \mathbf{V}_i)^{-1}(\mathbf{U} + \mathbf{V}_i)^{-1})}$.
 - if $\mathbf{D} = \text{diag}(\boldsymbol{\sigma}^2)$, $\text{var}(\hat{\sigma}_r^2) \geq \frac{2}{\sum_i ((\mathbf{U} + \mathbf{V}_i)_{rr}^{-1})^2}$.

3 Fisher scoring for weighted likelihood

Consider estimating $\boldsymbol{\theta} = \{\mathbf{U}, \mathbf{D}\}$ in the model

$$\mathbf{x}_i \sim N(\mathbf{0}, \mathbf{U} + \mathbf{D} + \mathbf{V}_i). \quad (27)$$

Let w_i be the weight of sample i , and $\sum_i w_i = 1$, then the weighted log-likelihood is

$$l(\boldsymbol{\theta}) = \sum_i \gamma_i \left(-\frac{1}{2} \log |\mathbf{U} + \mathbf{D} + \mathbf{V}_i| - \frac{1}{2} \mathbf{x}_i^T (\mathbf{U} + \mathbf{D} + \mathbf{V}_i)^{-1} \mathbf{x}_i \right). \quad (28)$$

Denote the score and Fisher information as $\mathbf{s}(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$ respectively, then the Fisher scoring iteration is

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \mathbf{I}(\boldsymbol{\theta}^{(t)})^{-1} \mathbf{s}(\boldsymbol{\theta}^{(t)}). \quad (29)$$

This can be applied to the M-step in multivariate deconvolution problem.

4 Appendix

Matrix calculus: let \mathbf{X} be a nonsingular matrix.

$$\partial a \mathbf{X} = a \partial \mathbf{X}.$$

$$\partial \text{Tr}(\mathbf{X}) = \text{Tr}(\partial \mathbf{X}).$$

$$\partial \log |\mathbf{X}| = \text{Tr}(\mathbf{X}^{-1} \partial \mathbf{X}).$$

$$\partial \mathbf{X}^{-1} = -\mathbf{X}^{-1} \partial \mathbf{X} \mathbf{X}^{-1}.$$

$$\partial(\mathbf{X}\mathbf{Y}) = \partial(\mathbf{X})\mathbf{Y} + \mathbf{X}\partial(\mathbf{Y}).$$

Lemma 4.1. *Let \mathbf{X} be a positive definite matrix, and $\boldsymbol{\Omega}$ be a diagonal matrix, then the solution to*

$$\min_{\boldsymbol{\Omega}} \|\mathbf{X}\boldsymbol{\Omega} - \mathbf{I}\|_F^2$$

is given by

$$\boldsymbol{\Omega}_{jj} = \frac{\mathbf{X}_{jj}}{\|\mathbf{x}_j\|^2}.$$