Statistica Sinica

Supplementary Material for "VALISE: A Robust Vertex Hunting Algorithm"

Dieyi Chen[*], Zheng Tracy Ke[*] and Shuyi Zhang[†]

[*]*Department of Statistics, Harvard University, Cambridge, MA 02138, USA*

KLATASDS-MOE, School of Statistics,

Academy of Statistics and Interdisciplinary Sciences,

East China Normal University, Shanghai 200062, China

This supplementary material provides Lemmas S3.1 – S4.3, and includes detailed proofs of Proposition 2.1, Theorems 3.1 – 3.4 and Corollary 3.1.

Contents

S1. Notation

Table S1.1: Notations

Symbol	Description
X_i	i-th observation
X_i^*	i-th denoised observation
Z_i	i-th noise vector
$\pi_i(k)$	k-th barycentric coordinat of X_i^*
V_k^*	k-th true vertex
$\hat{V}_k$	estimate of k-th true vertex under correctly specified models
$\tilde{V}_k$	estimate of k-th true vertex under misspecified models
$\mathcal{S}^*$	simplex spanned by $V_1^*, V_2^*, \ldots, V_K^*$
$\tilde{S}$	simplex spanned by $\tilde{V}_1, \tilde{V}_2, \ldots, \tilde{V}_K$
$\sigma^{2,*}$	true variance of noise
$\hat{\sigma}^2$	estimate of $\sigma^{2,*}$ under correctly specified models
$\tilde{\sigma}^2$	estimate of $\sigma^{2,*}$ under misspecified models
α	parameter of the Dirichlet distribution
$f(\pi; \alpha)$	density of Dirichlet distribution with parameter α
$\phi_r(x; \mu, \Sigma)$	density of $\mathcal{N}_r(\mu, \Sigma)$
$\ell(V, \sigma^2, \alpha)$	minus pseudo log-likelihood
$\mathcal{N}_k$	set of pure nodes in community k
$\mathcal{N}_k(\eta)$	set of “nearly” pure nodes in community k
$\mathcal{M}$	set of mixed nodes
$\Vert \cdot \Vert$	Euclidean/L_2 norm of a vector

S2. Proof of Proposition 2.1

We prove Proposition 2.1, which provides an equivalent form of the minus pseudo log-likelihood function. In the proof, we give a more general form of the minus pseudo log-likelihood function for $r \geq K-1$.

Proof. Notice that if we let $\tilde{\pi} = (\pi(1), \ldots, \pi(K-1))'$ for any $\pi \in \mathbb{R}^K$, we can obtain another expression for the K-th order standard simplex $\mathcal{S}_{K-1}$ as follows

$$\mathcal{S}_{K-1} = \{\pi \in \mathbb{R}_+^K : \ \|\pi\|_1 = 1\} \overset{1-1}{\leftrightarrow} \{\tilde{\pi} \in \mathbb{R}_+^{K-1} : \ 0 < \|\tilde{\pi}\|_1 < 1\} =: \Delta_{K-1}.$$

In the above, $\overset{1-1}{\leftrightarrow}$ represents a one-to-one transformation from the set on the left to that on the right. When $\alpha = (1, 1, \ldots, 1)'$, the minus log-likelihood function can be re-written as

$$\ell(V, \sigma^2, \alpha) = -\sum_{i=1}^{n} \log \left(\int \exp\left\{ -\frac{1}{2\sigma^2} \|X_i - V\pi_i\|^2 \right\} \mathbb{I}\{\tilde{\pi}_i \in \Delta_{K-1}\} d\tilde{\pi}_i \right) + \frac{nr}{2}\log(\sigma^2) - n\sum_{j=1}^{K-1} \log j + \frac{nr}{2}\log(2\pi), \tag{S2.1}$$

where $\pi_i = (\tilde{\pi}_i', 1 - \|\tilde{\pi}_i\|_1)'$. To understand the impact of V on the log-likelihood function, we will focus on the first part of the right hand side of the above equation. Let

$$g(\tilde{\pi}_i) = \exp\left\{ -\frac{1}{2\sigma^2} \|X_i - V\pi_i\|^2 \right\} \mathbb{I}\{\tilde{\pi}_i \in \Delta_{K-1}\} \quad \text{and}$$

$$G(V, \sigma^2) = \int g(\tilde{\pi}_i) d\tilde{\pi}_i.$$

The key is to apply the proper change of variables to the integral. Let

$$\tilde{V} = (V_1 - V_K, \ldots, V_{K-1} - V_K)', \ \tilde{\pi}_i = (\pi_i(1), \ldots, \pi_i(K-1))',$$

$$\tilde{y}_i = \tilde{V}\tilde{\pi}_i \text{ and } y_i = V\pi_i.$$

Then $y_i = \tilde{y}_i + V_K$ and

$$g(\tilde{\pi}_i) = \exp\left\{-\frac{1}{2\sigma^2}\left\|X_i - \tilde{V}\tilde{\pi}_i - V_K\right\|^2\right\}\mathbb{I}\{\tilde{\pi}_i \in \Delta_{K-1}\}.$$

Since $V_1, \ldots, V_K$ are affinely independent, $\text{rank}(\tilde{V}) = K-1$. Define $\bar{V}_K, \bar{V}_{K+1}, \ldots, \bar{V}_r$ as a group of unit orthogonal basis of the orthogonal complimen-t of the colume space of $\tilde{V}$ in $\mathbb{R}^r$, and let $V^* = (\bar{V}_K, \bar{V}_{K+1}, \ldots, \bar{V}_r) \in \mathbb{R}^{r\times(r-K+1)}$. Define $\pi_i^* = (\bar{\pi}_i(K), \bar{\pi}_i(K+1), \ldots, \bar{\pi}_i(r))'$ for any $\bar{\pi}_i(K), \bar{\pi}_i(K+1), \ldots, \bar{\pi}_i(r) \in \mathbb{R}_+$ and $y_i^* = V^*\pi_i^*$, and let

$$\bar{V} = (\tilde{V}, V^*), \ \bar{\pi}_i = (\tilde{\pi}_i', (\pi_i^*)')' \text{ and } \bar{y}_i = \bar{V}\bar{\pi}_i.$$

Then it can be derived that

$$\bar{y}_i = \tilde{V}\tilde{\pi}_i + V^*\pi_i^* = \tilde{y}_i + y_i^*.$$

Notice that $G(V, \sigma^2)$ can be re-written as

$$G(V, \sigma^2) = \int g(\tilde{\pi}_i)\mathbb{I}\{\pi_i^* \in (0,1)^{r-K+1}\}d\bar{\pi}_i.$$

Since $\bar{y}_i = \bar{V}\bar{\pi}_i$ and $\bar{V}$ is invertable, $\bar{\pi}_i = \bar{V}^{-1}\bar{y}_i$ and $d\bar{\pi}_i = |\det(\bar{V}^{-1})|d\bar{y}_i = |\det(\bar{V})|^{-1}d\bar{y}_i$. By the change of variables from $\bar{\pi}_i$ to $\bar{y}_i$ ($\bar{y}_i = \bar{V}\bar{\pi}_i$),

$$G(V, \sigma^2) = \int g((\bar{V}^{-1}\bar{y}_i)_{1,\ldots,K-1})\mathbb{I}\{(\bar{V}^{-1}\bar{y}_i)_{K,\ldots,r} \in (0,1)^{r-K+1}\}|\det(\bar{V})|^{-1}d\bar{y}_i, \tag{S2.2}$$

where $(\bar{V}^{-1}\bar{y}_i)_{1,\ldots,K-1}$ and $(\bar{V}^{-1}\bar{y}_i)_{K,\ldots,r}$ represent the first $K-1$ and the last $r-K+1$ components of the vector $\bar{V}^{-1}\bar{y}_i$, respectively.

To move forward, we define $\bar{V}_1, \bar{V}_2, \ldots, \bar{V}_{K-1}$ as a group of unit orthogonal basis of the colume space of $\tilde{V}$ and let $V_{01} = (\bar{V}_1, \bar{V}_2, \ldots, \bar{V}_{K-1}) \in \mathbb{R}^{r\times(K-1)}$. Then there exists an invertable matrix $A \in \mathbb{R}^{(K-1)\times(K-1)}$ such that $\tilde{V} = V_{01}A$. For notation consistency, we denote $V_{02} = V^*$ homogeneously and define $V_0 = (V_{01}, V_{02}) = (\bar{V}_1, \bar{V}_2, \ldots, \bar{V}_r) \in \mathbb{R}^{r\times r}$ which is an orthogonal matrix such that $V_0^{-1} = V_0'$, $V_{01}'V_{02} = \mathbf{0}$ and $V_{02}'V_{01} = \mathbf{0}$. Let

$$A_0 = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & I_{r-K+1} \end{pmatrix}.$$

Then $\bar{V} = V_0 A_0$. Furthermore, we have

$$\bar{y}_i = \bar{V}\bar{\pi}_i = V_0 A_0 \bar{\pi}_i = V_0 \begin{pmatrix} A\tilde{\pi}_i \\ \pi_i^* \end{pmatrix}.$$

Then

$$\tilde{\pi}_i = (A^{-1}\ \mathbf{0})V_0^{-1}\bar{y}_i = (A^{-1}\ \mathbf{0})V_0'\bar{y}_i = (A^{-1}\ \mathbf{0}) \begin{pmatrix} V_{01}' \\ V_{02}' \end{pmatrix} \bar{y}_i = A^{-1}V_{01}'\bar{y}_i \quad \text{and}$$

$$\pi_i^* = (\mathbf{0}\ I_{r-K+1})V_0^{-1}\bar{y}_i = (\mathbf{0}\ I_{r-K+1})V_0'\bar{y}_i = (\mathbf{0}\ I_{r-K+1})\begin{pmatrix} V_{01}' \\ \\ V_{02}' \end{pmatrix}\bar{y}_i = V_{02}'\bar{y}_i.$$

In other words, $(\bar{V}^{-1}\bar{y}_i)_{1,\ldots,K-1} = A^{-1}V_{01}'\bar{y}_i$ and $(\bar{V}^{-1}\bar{y}_i)_{K,\ldots,r-K+1} = V_{02}'\bar{y}_i$. Plugging in the expression in (S2.2), we can obtain

$$G(V,\sigma^2) = |\det(\bar{V})|^{-1}\int g(A^{-1}V_{01}'\bar{y}_i)\mathbb{I}\{V_{02}'\bar{y}_i \in (0,1)^{r-K+1}\}d\bar{y}_i. \quad \text{(S2.3)}$$

Next, we use the change of variables the second time by applying the above orthogonal representation (S2.3). Let

$$y_i^0 = V_0'\bar{y}_i = \begin{pmatrix} V_{01}'\bar{y}_i \\ \\ V_{02}'\bar{y}_i \end{pmatrix}.$$

Since V_0 is orthogonal matrix, $\det(V_0') = 1$ and hence $dy_i^0 = |\det(V_0')|d\bar{y}_i = d\bar{y}_i$. By the change of variables from $\bar{y}_i$ to y_i^0 ($y_i^0 = V_0'\bar{y}_i$),

$$\begin{aligned} G(V,\sigma^2) &= |\det(\bar{V})|^{-1}\int g(A^{-1}(y_i^0)_{1,\ldots,K-1})\mathbb{I}\{(y_i^0)_{K,\ldots,r} \in (0,1)^{r-K+1}\}dy_i^0 \\ &= |\det(\bar{V})|^{-1}\int g(A^{-1}(y_i^0)_{1,\ldots,K-1})d(y_i^0)_{1,\ldots,K-1}. \end{aligned}$$

Let $y_i^+ = (y_i^0)_{1,\ldots,K-1} \in \mathbb{R}^{K-1}$. Then

$$\begin{aligned} &G(V,\sigma^2) = |\det(\bar{V})|^{-1}\int g(A^{-1}y_i^+)dy_i^+ \\ &= |\det(\bar{V})|^{-1}\int \exp\left\{-\frac{1}{2\sigma^2}\left\|X_i - \tilde{V}A^{-1}y_i^+ - V_K\right\|^2\right\}\mathbb{I}\{A^{-1}y_i^+ \in \Delta_{K-1}\}dy_i^+ \\ &= |\det(\bar{V})|^{-1}\int \exp\left\{-\frac{1}{2\sigma^2}\left\|X_i - V_{01}y_i^+ - V_K\right\|^2\right\}\mathbb{I}\{A^{-1}y_i^+ \in \Delta_{K-1}\}dy_i^+, \end{aligned}$$

Notice that

$$A^{-1}y_i^+ \in \Delta_{K-1} \iff (V_1, V_2, \ldots, V_{K-1})A^{-1}y_i^+ + V_K(1 - \left\|A^{-1}y_i^+\right\|_1) \in \mathcal{S}$$

$$\iff \tilde{V}A^{-1}y_i^+ + V_K \in \mathcal{S} \iff V_{01}y_i^+ + V_K \in \mathcal{S},$$

where $\mathcal{S}$ denote the simplex spanned by V. Thus

$$G(V, \sigma^2) = |\det(\bar{V})|^{-1} \cdot \int \exp\left\{-\frac{1}{2\sigma^2}\left\|X_i - (V_{01}y_i^+ + V_K)\right\|^2\right\} \mathbb{I}\{V_{01}y_i^+ + V_K \in \mathcal{S}\}dy_i^+.$$

One important thing we should realize is the relationship between the volume of the simplex $\mathcal{S}$ and the determinant of matrix $\bar{V}$. On one hand,

$$\det(\bar{V}) = \det(V_0 A_0) = \det(V_0)\det(A_0) = \det(A_0)$$
$$= \det(A)\det(I_{r-K+1}) = \det(A).$$

On the other hand, according to the results on the Cayley-Menger determinant of simplex ((Sommerville, 1958; Gritzmann and Klee, 1994)),

$$|\det(A)| = (K-1)!\text{Vol}(\mathcal{S}),$$

where $\text{Vol}(\mathcal{S})$ represents the volume of the simplex $\mathcal{S}$. The last equation in the above uses the fact that the orthogonal transformation is isometric and hence keeps the volume of the geometry unchanged. Thus,

$$|\det(\bar{V})| = (K-1)!\text{Vol}(\mathcal{S}).$$

Hence,

$$G(V, \sigma^2) = [(K-1)!\text{Vol}(\mathcal{S})]^{-1}$$

$$\cdot \int \exp\left\{-\frac{1}{2\sigma^2}\left\|X_i - (V_{01}y_i^+ + V_K)\right\|^2\right\}\mathbb{I}\{V_{01}y_i^+ + V_K \in \mathcal{S}\}dy_i^+.$$

Plugging the above expression of $G(V,\sigma^2)$ into (S2.1), we have

$$\begin{aligned}
&\ell(V,\sigma^2,\alpha)\\
&= -\sum_{i=1}^n \log\left(\int \exp\left\{-\frac{\left\|X_i - (V_{01}y_i^+ + V_K)\right\|^2}{2\sigma^2}\right\}\mathbb{I}\{V_{01}y_i^+ + V_K \in \mathcal{S}\}dy_i^+\right)\\
&\quad + n\log \mathrm{Vol}(\mathcal{S}) + \frac{nr}{2}\log(\sigma^2) + \frac{nr}{2}\log(2\pi). \qquad \text{(S2.4)}
\end{aligned}$$

Specially, when $r = K-1$, we have $A = A_0$, $V_{01} = V_0$ and $\bar{V} = \tilde{V}$. Furthermore, we can choose $V_{01} = V_0 = I_r$ such that $A = A_0 = \tilde{V} = \bar{V}$. In this case,

$$\begin{aligned}
\ell(V,\sigma^2,\alpha) = &-\sum_{i=1}^n \log\left(\int_{x\in\mathcal{S}} \exp\left\{-\frac{1}{2\sigma^2}\left\|X_i - x\right\|^2\right\}dx\right) + n\log\mathrm{Vol}(\mathcal{S})\\
&+ \frac{nr}{2}\log(\sigma^2) + C.
\end{aligned}$$

where $C = nr\log(2\pi)/2$. The claim has been proved. □

S3. Proof of Theorem 3.1

We provide some useful lemmas and their proofs, and prove Theorem 3.1 from Section 3.

Lemma S3.1. Suppose Assumptions 3.1 and 3.2 hold. If $\theta^{(0)} = (\mathrm{vec}(V^{(0)})', \sigma^{2,(0)}, \alpha^{(0)})'$ satisfies $(\mathbf{1}_K', (V^{(0)})')'$ is invertible, for all $x \in \mathcal{X} \subset \mathbb{R}^r$, $\theta \to \log f(x;\theta)$ is continuous at $\theta^{(0)}$ with respect to the distance $d(\cdot,\cdot)$.

Proof. Let $\theta^{(n)} = (\mathrm{vec}(V^{(n)})', \sigma^{2,(n)}, \alpha^{(n)})'$ be a sequence of parameter vectors such that $\theta^{(n)} \to \theta^{(0)}$ as $n \to \infty$ with respect to the distance $d(\cdot,\cdot)$. Let $W^{(n)} = ((V^{(n)})', (\alpha^{(n)})')'$ and $W = (V', \alpha')'$. In this case, we can reparametrize the probability density function by $f(x;\theta) = f(x; W, \sigma^2)$. According to the definition of $d(\cdot,\cdot)$, we have $\min_\tau \max_{1\le k\le K} \|W^{(n)}_{\tau(k)} - W^{(0)}_k\| \to 0$ and $\sigma^{2,(n)} \to \sigma^{2,(0)}$ as $n \to \infty$. Under Assumption 3.1, $\phi_r(x; V^{(n)}\pi, \sigma^{2,(n)} I_r) \le (2\pi\epsilon_0^\sigma)^{-r/2}$. Then by the dominated convergence theorem, $f(x;\theta^{(n)}) \to f(x;\theta)|_{W=W^{(n)},\sigma^2=\sigma^{2,(0)}}$ as $\sigma^{2,(n)} \to \sigma^{2,(0)}$. In the following, we will show $f(x;\theta)|_{W=W^{(n)},\sigma^2=\sigma^{2,(0)}} \to f(x;\theta^{(0)})$ as $\min_\tau \max_{1\le k\le K} \|W^{(n)}_{\tau(k)} - W^{(0)}_k\| \to 0$.

Since $\phi_r(x;\mu,\sigma^{2,(0)} I_r)$ is continuous with respect to μ, for any $\epsilon > 0$, there exists $\delta_1 > 0$ such that for any $\|\mu_1 - \mu_2\| \le \delta_1$, $|\phi_r(x;\mu_1,\sigma^{2,(0)} I_r) - \phi_r(x;\mu_2,\sigma^{2,(0)} I_r)| < \epsilon/2$. Let $\tilde V^{(0)} = (V^{(0)}_1 - V^{(0)}_K, V^{(0)}_2 - V^{(0)}_K, \ldots, V^{(0)}_{K-1} - V^{(0)}_K)$. Then $\tilde V^{(0)}$ is invertible if and only if $(\mathbf{1}'_K, (V^{(0)})')'$ is invertible and $\det(\tilde V^{(0)}) = \det((\mathbf{1}'_K, (V^{(0)})')')$. For the above $\epsilon > 0$, there exists $\delta_2 > 0$ such that for any $\|\alpha - \tilde\alpha\| < \delta_2$, $|f(\pi;\alpha) - f(\pi;\tilde\alpha)| < |\det(\tilde V^{(0)})|\epsilon/2$. Let $\delta_0 = \min(\delta_1,\delta_2)/\sqrt{K}$. Since $\min_\tau \max_{1\le k\le K} \|W^{(n)}_{\tau(k)} - W^{(0)}_k\| \to 0$, there exist $N_\epsilon \in \mathbb{N}_+$ such that for any $n \ge N_\epsilon$, there exists a permutation τ_{n,δ_0} such that $\max_{1\le k\le K} \|W^{(n)}_{\tau_{n,\delta_0}(k)} - W^{(0)}_k\| \le \delta_0$.

In fact, τ_{n,δ_0} is free of δ_0 when δ_0 is sufficiently small. For any $\delta, \tilde\delta \in (0, \min_{k_1\ne k_2} \|W^{(0)}_{k_1} - W^{(0)}_{k_2}\|/2)$ such that $\tilde\delta < \delta$, we denote $\tau_{n,\delta}$ and $\tau_{n,\tilde\delta}$ as defined in the preceding paragraph. For any permutation $\tau \neq \tau_{n,\delta}$, there exists a $k = 1, \ldots, K$ such that $\tau(k) \neq \tau_{n,\delta}(k)$. Since $\tau_{n,\delta}$ is a surjection on $\{1, \ldots, K\}$, there exists $k' \neq k$ such that $\tau_{n,\delta}(k') = \tau(k)$. Then

$$\|W_k^{(0)} - W_{\tau(k)}^{(n)}\| = \|W_k^{(0)} - W_{\tau_{n,\delta}(k')}^{(n)}\| \geq \|W_k^{(0)} - W_{k'}^{(0)}\| - \|W_{k'}^{(0)} - W_{\tau_{n,\delta}(k')}^{(n)}\|$$
$$> \|W_k^{(0)} - W_{k'}^{(0)}\| - \delta > \|W_k^{(0)} - W_{k'}^{(0)}\|/2 > \tilde{\delta}.$$

Thus $\tau \neq \tau_{n,\tilde{\delta}}$. Then we have $\tau_{n,\tilde{\delta}} = \tau_{n,\delta}$. Then we can omit δ_0 in τ_{n,δ_0} and write τ_n for short for $\delta_0 \in (0, \min_{k_1 \neq k_2} \|W_{k_1} - W_{k_2}\|/2)$.

It can be easily derived from $\max_{1 \leq k \leq K} \|W_{\tau_n(k)}^{(n)} - W_k^{(0)}\| \leq \delta_0$ that $\max_{1 \leq k \leq K} \|V_{\tau_n(k)}^{(n)} - V_k^{(0)}\| \leq \delta_0 < \delta_1$ and $\|\alpha_{\tau_n}^{(n)} - \alpha\| < \sqrt{K}\delta_0 < \delta_2$. Then for any $\pi \in \mathcal{S}_{K-1}$,

$$\|V_{\tau_n}^{(n)}\pi - V^{(0)}\pi\| \leq \sum_{k=1}^{K} \|V_{\tau_n(k)}^{(n)} - V_{0,k}\|\pi(k) \leq \delta_1.$$

Let $\tilde{\pi} = (\pi_1, \pi_2, \ldots, \pi_{K-1})'$. Then $V^{(0)}\pi = V_K + \tilde{V}^{(0)}\tilde{\pi}$. Notice that the probability density function $f(x;\theta)$ is invariant under permutations of the columns of W. It follows that

$$\begin{aligned}
&|f(x; W^{(n)}, \sigma^{2,(0)}) - f(x; W^{(0)}, \sigma^{2,(0)})| = |f(x; W_{\tau_n}^{(n)}, \sigma^{2,(0)}) - f(x; W^{(0)}, \sigma^{2,(0)})| \\
&\leq \int |\phi_r(x; V_{\tau_n}^{(n)}\pi, \sigma^{2,(0)}I_r) - \phi_r(x; V^{(0)}\pi, \sigma^{2,(0)}I_r)| f(\pi, \alpha^{(n)}) d\pi \\
&\quad + \int \phi_r(x; V^{(0)}\pi, \sigma^{2,(0)}I_r) |f(\pi, \alpha^{(n)}) - f(\pi, \alpha^{(0)})| d\pi \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon |\det(\tilde{V}^{(0)})|}{2} \int \phi_r(x; V^{(0)}\pi, \sigma^{2,(0)}I_r) d\pi
\end{aligned}$$

$$\leq \frac{\epsilon}{2} + \frac{\epsilon|\det(\tilde{V}^{(0)})|}{2} \int \phi_r(x - V_K; \tilde{V}^{(0)}\tilde{\pi}, \sigma^{2,(0)} I_r) d\tilde{\pi}$$
$$= \frac{\epsilon}{2} + \frac{\epsilon|\det(\tilde{V}^{(0)})|}{2} |\det(\tilde{V}^{(0)})|^{-1} \int \phi_r(x - V_K; u, \sigma^{2,(0)} I_r) du \leq \epsilon/2 + \epsilon/2 = \epsilon.$$

The claim follows immediately. □

Lemma S3.2. Under model identifiability, $\mathbb{E}_{X\sim f(x;\theta^*)}[\log f(X;\theta)]$ attains its maximum uniquely at the equivalence class with representative θ^*.

Proof. Since $\log x \leq 2(\sqrt{x} - 1)$ for $x \geq 0$, we have

$$\mathbb{E}_{X\sim f(x;\theta^*)}[\log f(X;\theta)] - \mathbb{E}_{X\sim f(x;\theta^*)}[\log f(X;\theta^*)]$$
$$= \mathbb{E}_{X\sim f(x;\theta^*)}\left[\log \frac{f(X;\theta)}{f(X;\theta^*)}\right] \leq 2\mathbb{E}_{X\sim f(x;\theta^*)}\left[\sqrt{\frac{f(X;\theta)}{f(X;\theta^*)}} - 1\right]$$
$$= 2\int \sqrt{f(x;\theta) f(x;\theta^*)} dx - 2 = -2\int [\sqrt{f(x;\theta)} - \sqrt{f(x;\theta^*)}]^2 dx \leq 0.$$

It can be seen that $\mathbb{E}_{X\sim f(x;\theta^*)}[\log f(X;\theta)] = \mathbb{E}_{X\sim f(x;\theta^*)}[\log f(X;\theta^*)]$ if and only if $f(x;\theta) = f(x;\theta^*)$ for all x. Under the model identifiability, this implies θ and θ^* are equivalent, which gives the claim. □

In the following, we will prove Theorem 3.1.

Proof. Under Assumptions 3.1 and 3.2, for $\theta^* = (\mathrm{vec}(V^*)', \sigma^{2,*}, \alpha^*)'$ with $(\mathbf{1}_K', (V^*)')'$ invertible, $\log f(\cdot;\theta)$ is continuous at θ^* with respect to the distance $d(\cdot,\cdot)$ by Lemma S3.1. On the other hand, since $\phi_r(x; V\pi, \sigma^2 I_r) \leq$

$(2\pi\epsilon_0^\sigma)^{-r/2}$, $f(x;\theta) \le (2\pi\epsilon_0^\sigma)^{-r/2}$. Then for any sufficiently small ball U,

$$\mathbb{E}[\sup_{\theta\in U} \log f(x;\theta)] \le -r/2 \log(2\pi\epsilon_0^\sigma) < \infty.$$

According to Theorem 5.14 in Van der Vaart (2000), the MLE $\hat\theta_n$ satisfies for any $\epsilon > 0$, $\mathbb{P}(\{d(\hat\theta_n, \theta^*) > \epsilon\} \cap \{\hat\theta_n \in \mathcal{H}\}) \to 0$ as $n \to \infty$. This gives the first claim in Theorem 3.1.

Next, we consider the second claim. By the first claim, for any $\epsilon > 0$, $\mathbb{P}(d(\hat\theta_n, \theta^*) > \epsilon) \to 0$ as $n \to \infty$. Since $d(\hat\theta_n, \theta^*) = \min_\tau \max_{1\le k\le K} \|\hat W_{n,\tau(k)} - W_k^*\| + |\hat\sigma_n^2 - \sigma^{2,*}|$, we have

$$\mathbb{P}(\min_\tau \max_{1\le k\le K} \|\hat W_{n,\tau(k)} - W_k^*\| > \epsilon) \to 0 \quad\text{and}\quad \mathbb{P}(|\hat\sigma_n^2 - \sigma^{2,*}| > \epsilon) \to 0.$$

The first limit in the above gives

$$\mathbb{P}(\exists \tau_{n,\epsilon} \text{ such that } \max_{1\le k\le K} \|\hat W_{n,\tau_{n,\epsilon}(k)} - W_k^*\| \le \epsilon) \to 1. \tag{S3.5}$$

In the following we will show $\tau_{n,\epsilon}$ is free of ϵ when ϵ is sufficiently small. For $\epsilon_1, \epsilon_2 \in (0, \min_{k_1\ne k_2} \|W_{k_1} - W_{k_2}\|/2)$ such that $\epsilon_2 < \epsilon_1$, let $\tau_{n,\epsilon_1}, \tau_{n,\epsilon_2}$ be the two sequences of permutations as defined in the above. For any permutation $\tau \ne \tau_{n,\epsilon_1}$, there exists a $k = 1, \dots, K$ such that $\tau(k) \ne \tau_{n,\epsilon_1}(k)$. Since τ_{n,ϵ_1} is a surjection on $\{1, \dots, K\}$, there exists $k' \ne k$ such that $\tau_{n,\epsilon_1}(k') = \tau(k)$. Then

$$\begin{aligned}\|W_k^* - \hat W_{\tau(k)}\| = \|W_k^* - \hat W_{\tau_{n,\epsilon_1}(k')}\| &\ge \|W_k^* - W_{k'}^*\| - \|W_{k'}^* - \hat W_{\tau_{n,\epsilon_1}(k')}\| \\ &> \|W_k^* - W_{k'}^*\| - \epsilon_1 > \|W_k^* - W_{k'}^*\|/2 > \epsilon_2.\end{aligned}$$

It gives $\tau_{n,\epsilon_2} = \tau_{n,\epsilon_1}$. Thus we can omit ϵ in $\tau_{n,\epsilon}$ in (S3.5). Let τ_n be the sequence of permutations defined in (S3.5). Let $\hat{\theta}_{n,\tau_n} = (\mathrm{vec}(\hat{V}_{n,\tau_n})', \hat{\sigma}_n^2, \hat{\alpha}_{n,\tau_n})'$. Then

$$\begin{aligned}
\|\hat{\theta}_{n,\tau_n} - \theta^*\| &= \{\|\mathrm{vec}(\hat{V}_{n,\tau_n})' - \mathrm{vec}(V^*)'\|^2 + (\hat{\sigma}_n^2 - \hat{\sigma}^2)^2 + \|\hat{\alpha}_{n,\tau_n} - \alpha^*\|^2\}^{1/2} \\
&= \left\{ \sum_{k=1}^{K} \|\hat{W}_{n,\tau_n(k)} - W_k^*\|^2 + (\hat{\sigma}_n^2 - \hat{\sigma}^2)^2 \right\}^{1/2} \\
&\leq \{K \max_{1 \leq k \leq K} \|\hat{W}_{n,\tau_n(k)} - W_k^*\|^2 + (\hat{\sigma}_n^2 - \hat{\sigma}^2)^2\}^{1/2} \\
&\leq \sqrt{2}\{\sqrt{K} \max_{1 \leq k \leq K} \|\hat{W}_{n,\tau_n(k)} - W_k^*\| + |\hat{\sigma}_n^2 - \hat{\sigma}^2|\} \\
&\leq \sqrt{2K} d(\hat{\theta}_{n,\tau_n}, \theta^*).
\end{aligned}$$

In the above inequalities, we use the fact $x^2 + y^2 \leq 2(x+y)^2$ for $x, y \geq 0$. Thus we have for any $\epsilon > 0$, $\mathbb{P}(\|\hat{\theta}_{n,\tau_n} - \theta^*\| > \epsilon) \leq \mathbb{P}(d(\hat{\theta}_n, \theta^*) > \epsilon/\sqrt{2K})$, which proves the second claim in Theorem 3.1. □

S4. Proof of Theorem 3.2

We first provide two lemmas from Lemma 7.6 and Theorem 5.39 in Van der Vaart (2000), and prove Theorem 3.2 based on these two lemmas.

Lemma S4.1. For every θ in an open subset of $\mathbb{R}^k$, let p_θ be a μ-probability density. Assume that the map $\theta \mapsto s_\theta(x) = \sqrt{p_\theta(x)}$ is continuously differentiable for every x. If the elements of the matrix $I_\theta = \int (\dot{p}_\theta / p_\theta)(\dot{p}_\theta' / p_\theta) p_\theta d\mu$

are well defined and continuous in θ, then the map $\theta \mapsto \sqrt{p_\theta}$ is differentiable in quadratic mean such that

$$\int [\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - h'\dot{l}_\theta \sqrt{p_\theta}/2]^2 d\mu = o(\|h\|^2),\ h \to 0,$$

with $\dot{l}_\theta$ given by $\dot{p}_\theta / p_\theta$.

Lemma S4.2. Under Assumptions 3.1 and 3.2, the map $\theta \mapsto \sqrt{f(x;\theta)}$ is differentiable in quadratic mean at θ^*.

Proof. We will first prove $f(x;\theta)$ is continuously differentiable for every x at θ^* with respect to the L_2 distance on $\mathbb{R}^{rK+K+1}$. It can be derived that

$$\nabla_V \phi_r(x; V\pi, \sigma^2 I_r) = (\sigma^2)^{-1}(x - V\pi)\pi' \phi_r(x; V\pi, \sigma^2 I_r).$$

For a matrix $A = (a_{j_1 j_2})$, let $\|A\|_{\infty,\infty} = \max_{j_1,j_2} |a_{j_1 j_2}|$ and $\|A\|_\infty = \max_{j_1} \sum_{j_2} |a_{j_1 j_2}|$. For a vector $a = (a_1, a_2, \ldots, a_l)'$, let $\|a\|_\infty = \max_i |a_i|$ and $\|a\|_1 = \sum_i |a_i|$. Since the parameter space $\mathcal{H}$ is compact, $\|V\|_\infty \le C_0$ for a positive constant C_0. Let $V = (v_1, v_2, \ldots, v_r)'$. Then for $\pi \in \mathcal{S}_{K-1}$,

$$\|V\pi\|_\infty = \|(v_1'\pi, v_2'\pi, \ldots, v_r'\pi)'\|_\infty = \max_i |v_i'\pi| \le \max_i \|v_i\|_1 = \|V\|_\infty \le C_0.$$

Thus we have, for $\pi \in \mathcal{S}_{K-1}$,

$$\begin{aligned} &(\sigma^2)^{-1}\|(x - V\pi)\pi'\|_{\infty,\infty} \phi_r(x; V\pi, \sigma^2 I_r) \\ &\le (\sigma^2)^{-1}\|x - V\pi\|_\infty \phi_r(x; V\pi, \sigma^2 I_r) \le (\epsilon_0^\sigma)^{-1}(2\pi\epsilon_0^\sigma)^{-r/2}(\|x\|_\infty + \|V\pi\|_\infty) \\ &\le (\epsilon_0^\sigma)^{-1}(2\pi\epsilon_0^\sigma)^{-r/2}(\|x\|_\infty + C_0), \end{aligned} \tag{S4.6}$$

In a neighborhood of V with respect to the L_2 distance, it can be shown from (S4.6) that $\sup \|f(\pi;\alpha)\nabla_V \phi_r(x;V\pi,\sigma^2 I_r)\|_{\infty,\infty} \le (\epsilon_0^\sigma)^{-1}(2\pi\epsilon_0^\sigma)^{-r/2}(\|x\|_\infty + C_0)f(\pi;\alpha)$ which is integrable. By the dominated convergence theorem, $f(x;\theta)$ is differentiable with respect to V and we can take the differentiation under the integral. Then

$$\begin{aligned}\nabla_V f(x;\theta) &= \int \nabla_V \phi_r(x;V\pi,\sigma^2 I_r)f(\pi;\alpha)d\pi \\ &= (\sigma^2)^{-1}\int (x-V\pi)\pi'\phi_r(x;V\pi,\sigma^2 I_r)f(\pi;\alpha)d\pi. \qquad \text{(S4.7)}\end{aligned}$$

According to (S4.6), by the dominated convergence theorem, it can also be seen that $\nabla_V f(x;\theta)$ is continuous about V and σ^2. On the other hand, for any α and $\tilde{\alpha}$ such that $|f(\pi;\alpha)-f(\pi;\tilde{\alpha})|<\epsilon$, we have

$$\begin{aligned}&\|\nabla_V f(x;V,\sigma^2,\alpha) - \nabla_V f(x;V,\sigma^2,\tilde{\alpha})|\|_{\infty,\infty} \\ \le\ & (\epsilon_0^\sigma)^{-1}(\|x\|_\infty + C_0)\epsilon \int \phi_r(x;V\pi,\sigma^2 I_r)d\pi \\ \le\ & (\epsilon_0^\sigma)^{-1}(\|x\|_\infty + C_0)|\det(\tilde{V})|^{-1}\epsilon, \qquad \text{(S4.8)}\end{aligned}$$

where the last inequality holds if $\tilde{V} = (V_1 - V_k, V_2 - V_K, \ldots, V_{K-1} - V_K)$ is invertible. Under Assumption 3.2, $\nabla_V f(x;\theta)$ is continuous at θ^*.

Analogously, by the dominated convergence theorem, we can obtain the derivatives of $f(x;\theta)$ with respect to σ^2 and α. Let $G_1(\alpha) = (\psi(\alpha_1), \ldots, \psi(\alpha_K))'$ and $G_2(\pi) = (\log \pi_1, \ldots, \log \pi_K)'$ where $\psi(\cdot)$ is the digamma function. Then it can be derived that

$$\nabla_{\sigma^2} f(x;\theta) = \int \left(\frac{\|x - V\pi\|^2}{2(\sigma^2)^2} - \frac{r}{2\sigma^2} \right) \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi, \quad \text{(S4.9)}$$

$$\nabla_{\alpha} f(x;\theta) = \int [\psi(\|\alpha\|_1) - G_1(\alpha) + G_2(\pi)] \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi, \quad \text{(S4.10)}$$

where $\|\alpha\|_1 = \sum_{k=1}^{K} |\alpha_k|$. It can be shown similarly as $\nabla_V f(x;\theta)$ that $\nabla_{\sigma^2} f(x;\theta)$ and $\nabla_\alpha f(x;\theta)$ is continuous at θ^*. Hence $f(x;\theta)$ is continuously differentiable at θ^* for every x.

In the next, we will show elements of the Fisher information matrix $I_\theta = \int \nabla_\theta f(x;\theta) \nabla'_\theta f(x;\theta)/f(x;\theta) dx$ are well defined and continuous in θ. From (S4.7), we have

$$\nabla_{\text{vec}(V)} f(x;\theta) = (\sigma^2)^{-1} \int \pi \otimes (x - V\pi) \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi, \quad \text{(S4.11)}$$

where $\otimes$ represents the Kronecker product. Under Assumption 3.1,

$$\begin{aligned} \|\nabla_{\text{vec}(V)} f(x; V, \sigma^2, \alpha)\|_\infty &\le (\epsilon_0^\sigma)^{-1} (\|x\|_\infty + C_0) f(x;\theta) \\ &=: b_1(x) f(x;\theta). \end{aligned} \quad \text{(S4.12)}$$

Recall that $V = (v_1, \ldots, v_r)' = (V_1, V_2, \ldots, V_K)$. Notice that

$$\begin{aligned} |x'V\pi| &\le \sum_{i,k} |x_i V_k(i) \pi_k| = \sum_i |x_i| \sum_k |V_k(i)\pi_k| \\ &\le \sum_i |x_i| \sum_k |V_k(i)| \le \|x\|_1 \|V\|_\infty \quad \text{and} \end{aligned}$$

$$\|V\pi\|^2 = \|(v_1'\pi, \ldots, v_r'\pi)'\|^2 = \sum_i (v_i'\pi)^2$$

$$\leq \sum_i \|v_i\|_1^2 \leq r \max_i \|v_i\|_1^2 = r\|V\|_\infty^2. \tag{S4.13}$$

Thus we have

$$\|x - V\pi\|^2 = \|x\|^2 - 2x'V\pi + \|V\pi\|^2 \leq \|x\|^2 + 2\|x\|_1\|V\|_\infty + r\|V\|_\infty^2$$
$$\leq \|x\|^2 + 2C_0\|x\|_1 + rC_0^2.$$

From (S4.9), it can be derived that

$$\|\nabla_{\sigma^2} f(x;\theta)\|_\infty \leq [(\|x\|^2 + 2C_0\|x\|_1 + rC_0^2)/\epsilon_0^\sigma + r]/(2\epsilon_0^\sigma)f(x;\theta)$$
$$=: b_2(x)f(x;\theta). \tag{S4.14}$$

Then we deal with $\nabla_\alpha f(x;\theta)$. Since $\psi(\cdot)$ is a continuous function in $\mathbb{R}_+$, under Assumption 3.1, there exists a positive constant $M^\alpha > 0$ such that $|\psi(\|\alpha\|_1)| \leq M^\alpha$ and $\max_k |\psi(\alpha_k)| \leq M^\alpha$. Thus $\|\psi(\|\alpha\|_1) - G_1(\alpha)\|_\infty \leq 2M^\alpha$. On the other hand, we have

$$\int G_2(\pi)\phi_r(x;V\pi,\sigma^2 I_r)f(\pi;\alpha)d\pi = \frac{e^{-\frac{x'x}{2\sigma^2}}}{(2\pi\sigma^2)^{r/2}}\int e^{\frac{x'V\pi}{\sigma^2}-\frac{\|V\pi\|^2}{2\sigma^2}} f(\pi;\alpha)G_2(\pi)d\pi,$$
$$f(x;\theta) = \frac{1}{(2\pi\sigma^2)^{r/2}}e^{-\frac{x'x}{2\sigma^2}}\int e^{\frac{x'V\pi}{\sigma^2}-\frac{\|V\pi\|^2}{2\sigma^2}} f(\pi;\alpha)d\pi.$$

Thus

$$f(x;\theta)^{-1}\int G_2(\pi)\phi_r(x;V\pi,\sigma^2 I_r)f(\pi;\alpha)d\pi = \frac{\int e^{\frac{x'V\pi}{\sigma^2}-\frac{\|V\pi\|^2}{2\sigma^2}} f(\pi;\alpha)G_2(\pi)d\pi}{\int e^{\frac{x'V\pi}{\sigma^2}-\frac{\|V\pi\|^2}{2\sigma^2}} f(\pi;\alpha)d\pi}.$$

Under Assumption 3.1, according to (S4.13),

$$\exp\left\{\frac{-C_0\|x\|_1}{\epsilon_0^\sigma} - \frac{rC_0^2}{2\epsilon_0^\sigma}\right\} \leq \exp\left\{\frac{x'V\pi}{\sigma^2} - \frac{\|V\pi\|^2}{2\sigma^2}\right\} \leq \exp\left\{\frac{C_0\|x\|_1}{\epsilon_0^\sigma}\right\}.$$

Then we have

$$
\begin{aligned}
0 &\leq -f(x;\theta)^{-1} \int G_2(\pi)\phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi \\
&\leq \exp\left\{ \frac{C_0\|x\|_1}{\epsilon_0^{\sigma}} + \frac{C_0\|x\|_1}{\epsilon_0^{\sigma}} + \frac{rC_0^2}{2\epsilon_0^{\sigma}} \right\} \int -f(\pi;\alpha) G_2(\pi) d\pi \\
&= \exp\left\{ \frac{2C_0\|x\|_1}{\epsilon_0^{\sigma}} + \frac{rC_0^2}{2\epsilon_0^{\sigma}} \right\} [\psi(\|\alpha\|_1) - G_1(\alpha)],
\end{aligned}
$$

where the last equality holds by using the exponential family differential identities. Let $C_1 = 2C_0/\epsilon_0^{\sigma}$ and $C_2 = rC_0^2/(2\epsilon_0^{\sigma})$. Then

$$
0 \leq -f(x;\theta)^{-1} \int G_2(\pi)\phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi \leq 2M^{\alpha} \exp\{C_1\|x\|_1 + C_2\}.
$$

By considering the form of $\nabla_{\alpha} f(x;\theta)$ in (S4.10), we are able to obtain its bound as follows,

$$
\|\nabla_{\alpha} f(x;\theta)\|_{\infty} \leq 2M^{\alpha}[\exp\{C_1\|x\|_1 + C_2\} + 1] f(x;\theta) =: b_3(x) f(x;\theta). \tag{S4.15}
$$

According to (S4.12), (S4.14) and (S4.15), if letting $b(x) = \max\{b_1(x), b_2(x), b_3(x)\}$, we can obtain the upper bound for $\nabla_{\theta} f(x;\theta)$ by $\|\nabla_{\theta} f(x;\theta)\|_{\infty} \leq b(x) f(x;\theta)$. Then

$$
\|\nabla_{\theta} f(x;\theta) \nabla_{\theta}' f(x;\theta)\|_{\infty} / f(x;\theta) \leq b^2(x) f(x;\theta).
$$

By the definition of $b^2(x)$, it can be seen that $\int b^2(x) f(x;\theta) < \infty$. Thus elements of the matrix $I_{\theta} = \int \nabla_{\theta} f(x;\theta) \nabla_{\theta}' f(x;\theta) / f(x;\theta) dx$ are well defined. Moreover, notice that $b^2(x)\phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) \leq (2\pi\epsilon_0^{\sigma})^{-r/2} b^2(x) f(\pi;\alpha)$ which is integral with respect to π and x. By the dominated convergence theorem, I_θ is continuous in V and σ^2. The continuity of I_θ at α^* can be obtained similarly as in (S4.8). From Lemma S4.1, $\theta \mapsto \sqrt{f(x;\theta)}$ is differentiable in quadratic mean at θ^*. □

Lemma S4.3. Under Assumptions 3.1 and 3.2, there exists a measurable function $\dot{\ell}(x)$ with $\mathbb{E}_{X\sim f(x;\theta^*)}[\dot{\ell}^2(X)] < \infty$ such that for every θ_1 and θ_2 in a neighborhood of θ^*,

$$|\log f(x;\theta_1) - \log f(x;\theta_2)| \leq \dot{\ell}(x)\|\theta_1 - \theta_2\|.$$

Proof. By the mean value theorem, for any $\theta_1, \theta_2 \in \mathcal{H}$,

$$\begin{aligned}\log f(x;\theta_1) - \log f(x;\theta_2) &= \nabla'_\theta \log f(x;\theta)|_{\theta=\xi}(\theta_1 - \theta_2) \\ &= f(x;\xi)^{-1}\nabla'_\theta f(x;\theta)|_{\theta=\xi}(\theta_1 - \theta_2),\end{aligned}$$

where ξ is some point between θ_1 and θ_2. In the proof of Lemma S4.2, we have obtained that $\|\nabla_\theta f(x;\theta)\|_\infty \leq b(x)f(x;\theta)$ where $\mathbb{E}_{X\sim f(x;\theta^*)}[b^2(X)] < \infty$. Thus

$$\begin{aligned}|\log f(x;\theta_1) - \log f(x;\theta_2)| &\leq f(x;\xi)^{-1}\|\nabla_\theta f(x;\theta)|_{\theta=\xi}\|_\infty\|\theta_1 - \theta_2\|_1 \\ \leq b(x)\|\theta_1 - \theta_2\|_1 &\leq b(x)\sqrt{rK+K+1}\|\theta_1 - \theta_2\| =: \dot{\ell}(x)\|\theta_1 - \theta_2\|,\end{aligned}$$

where $\mathbb{E}_{X\sim f(x;\theta^*)}[\dot{\ell}^2(X)] \leq (rK+K+1)\mathbb{E}_{X\sim f(x;\theta^*)}[b^2(X)] < \infty$. Now the claim has been proved. □

In the following, we provide the proof of Theorem 3.2.

Proof. By the second claim in Theorem 3.1, we obtained a sequence of MLE $\{\hat{\theta}_{n,\tau_n}\}$ such that $\hat{\theta}_{n,\tau_n}$ is a consistent estimator for θ^* with respect to the L_2 distance. By Lemma S4.2, under Assumptions 3.1 and 3.2, the model $\{f(\cdot,\theta),\theta\in\mathcal{H}\}$ is differentiable in quadratic mean at θ^*. By Lemma S4.3, under Assumptions 3.1 and 3.2, there exists a measurable function $\dot{\ell}(x)$ with $\mathbb{E}_{X\sim f(x;\theta^*)}[\dot{\ell}^2(X)]<\infty$ such that for every θ_1 and θ_2 in a neighborhood of θ^*,

$$|\log f(x;\theta_1)-\log f(x;\theta_2)|\le\dot{\ell}(x)\|\theta_1-\theta_2\|.$$

Furthermore, if θ^* is an inner point of $\mathcal{H}$ and the Fisher information matrix I_{θ^*} is nonsingular, by Theorem 5.39 in Van der Vaart (2000),

$$\sqrt{n}(\hat{\theta}_{n,\tau_n}-\theta^*)=I_{\theta^*}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_\theta\log f(X_i;\theta^*)+o_P(1),$$

where the explicit form of $\nabla_\theta\log f(x;\theta)=(\nabla'_{\mathrm{vec}(V)}\log f(x;\theta),\nabla_{\sigma^2}\log f(x;\theta),$ $\nabla'_\alpha\log f(x;\theta))'$ can be derived by (S4.9), (S4.10) and (S4.11) such that

$$\nabla_{\mathrm{vec}(V)}\log f(x;\theta)=\frac{1}{\sigma^2 f(x;\theta)}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[\pi\otimes(x-V\pi)\phi_r(x;V\pi,\sigma^2I_r)],$$

$$\nabla_{\sigma^2}\log f(x;\theta)=\frac{1}{2(\sigma^2)^2f(x;\theta)}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[\|x-V\pi\|^2\phi_r(x;V\pi,\sigma^2I_r)]-\frac{r}{2\sigma^2},$$

$$\nabla_\alpha\log f(x;\theta)=\psi(\|\alpha\|_1)-G_1(\alpha)+\frac{1}{f(x;\theta)}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[G_2(\pi)\phi_r(x;V\pi,\sigma^2I_r)].$$

In particular, $\sqrt{n}(\hat{\theta}_{n,\tau}-\theta^*)$ is asymptotically normal with mean zero and covariance matrix $I_{\theta^*}^{-1}$.

Below, we derive the explicit form of the Fisher information matrix. By definition, the Fisher information matrix I_θ can be formulated as

$$I_\theta = -\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\theta\theta} \log f(X;\theta)] = -\int f(x;\theta)\nabla_{\theta\theta} \log f(x;\theta)dx.$$

Thus we need to derive the Hessian matrix of $\log f(x;\theta)$. According to (S4.9), (S4.10) and (S4.11) in Lemma S4.2, by some tedious algebra, we can obtain that

$$\begin{aligned}
&\nabla_{\mathrm{vec}(V),\mathrm{vec}(V)} \log f(x;\theta) \\
=& \frac{1}{f(x;\theta)(\sigma^2)^2}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[(\pi\pi') \otimes \{(x-V\pi)(x-V\pi)'\}\phi_r(x;V\pi,\sigma^2 I_r)] \\
&- \frac{1}{f(x;\theta)\sigma^2}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[(\pi\pi') \otimes I_r\phi_r(x;V\pi,\sigma^2 I_r)] \\
&- \frac{1}{f(x;\theta)^2}M_1(x;\theta)M_1'(x;\theta), \\
&\nabla_{\mathrm{vec}(V),\sigma^2} \log f(x;\theta) \\
=& \frac{1}{2f(x;\theta)(\sigma^2)^3}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[\|x-V\pi\|^2\pi \otimes (x-V\pi)\phi_r(x;V\pi,\sigma^2 I_r)] \\
&- \frac{1}{f(x;\theta)\sigma^2}M_1(x;\theta) - \frac{1}{f(x;\theta)^2}M_1(x;\theta)M_2(x;\theta), \\
&\nabla_{\mathrm{vec}(V),\alpha} \log f(x;\theta) \\
=& \frac{1}{f(x;\theta)\sigma^2}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[\pi \otimes (x-V\pi)\phi_r(x;V\pi,\sigma^2 I_r)G_2'(\pi)] \\
&- \frac{1}{f(x;\theta)^2}M_1(x;\theta)M_3'(x;\theta), \\
&\nabla_{\sigma^2,\sigma^2} \log f(x;\theta) \\
=& \frac{r}{2(\sigma^2)^2} + \frac{1}{4f(x;\theta)(\sigma^2)^4}\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[\|x-V\pi\|^4\phi_r(x;V\pi,\sigma^2 I_r)]
\end{aligned}$$

$$- \frac{2}{f(x;\theta)\sigma^2} M_2(x;\theta) - \frac{1}{f(x;\theta)^2} M_2^2(x;\theta),$$

$$\nabla_{\sigma^2,\alpha} \log f(x;\theta)$$

$$= \frac{1}{2f(x;\theta)(\sigma^2)^2} \mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[\|x - V\pi\|^2 \phi_r(x; V\pi, \sigma^2 I_r) G_2'(\pi)]$$

$$- \frac{1}{f(x;\theta)^2} M_2(x;\theta) M_3'(x;\theta) \quad \text{and}$$

$$\nabla_{\alpha,\alpha} \log f(x;\theta)$$

$$= \psi^{(1)}(\|\alpha\|_1) - G_3(\alpha) + \frac{1}{f(x;\theta)} \mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}[\phi_r(x; V\pi, \sigma^2 I_r) G_2(\pi) G_2'(\pi)]$$

$$- \frac{1}{f(x;\theta)^2} M_3(x;\theta) M_3'(x;\theta).$$

It can be derived that

$$\iint \|x - V\pi\|^2 \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi dx$$

$$= \iint \|u\|^2 \phi_r(u; \mathbf{0}, \sigma^2 I_r) f(\pi;\alpha) d\pi du = r\sigma^2 \text{ and}$$

$$\iint \pi \otimes (x - V\pi) \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi dx$$

$$= \iint (\pi \otimes u) \phi_r(u; \mathbf{0}, \sigma^2 I_r) f(\pi;\alpha) d\pi du = \mathbf{0}.$$

Analogously, we have

$$\iint \|x - V\pi\|^4 \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi dx = (r^2 + 2r)(\sigma^2)^2,$$

$$\iint \|x - V\pi\|^2 \pi \otimes (x - V\pi) \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi dx = \mathbf{0} \quad \text{and}$$

$$\iint \pi \otimes (x - V\pi) \phi_r(x; V\pi, \sigma^2 I_r) G_2'(\pi) f(\pi;\alpha) d\pi dx = \mathbf{0}.$$

Moreover, we can obtain that

$$\iint (\pi\pi') \otimes I_r \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi dx = \mathrm{E}_{\pi\sim\mathrm{Dir}(\alpha)}[(\pi\pi')\otimes I_r] \text{ and}$$

$$\iint (\pi\pi') \otimes \{(x - V\pi)(x - V\pi)'\} \phi_r(x; V\pi, \sigma^2 I_r) f(\pi;\alpha) d\pi dx$$

$$= \sigma^2 \mathrm{E}_{\pi\sim\mathrm{Dir}(\alpha)}[(\pi\pi')\otimes I_r].$$

By using the exponential family differential identities, $\mathbb{E}_{\pi\sim\mathrm{Dir}(\alpha)}(\log \pi_k) = \psi(\alpha_k) - \psi(\|\alpha\|_1)$ and $\mathrm{Cov}(\log \pi_{k_1}, \log \pi_{k_2}) = \psi^{(1)}(\alpha_{k_1})\delta_{k_1k_2} - \psi^{(1)}(\|\alpha\|_1)$ where $\psi^{(1)}(\cdot)$ is the trigamma function and $\delta_{k_1k_2} = \mathbb{I}\{k_1 = k_2\}$. Then it can be derived that

$$\iint \|x - V\pi\|^2 \phi_r(x; V\pi, \sigma^2 I_r) G_2(\pi) f(\pi;\alpha) d\pi dx = r\sigma^2[G_1(\alpha) - \psi(\|\alpha\|_1)]$$

$$\text{and} \quad \iint \phi_r(x; V\pi, \sigma^2 I_r) G_2(\pi) G_2'(\pi) f(\pi;\alpha) d\pi dx = G_4(\alpha).$$

According to the above results, we have

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\mathrm{vec}(V),\mathrm{vec}(V)} \log f(x;\theta)] = \int \frac{1}{f(x;\theta)} M_1(x;\theta) M_1'(x;\theta) dx,$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\mathrm{vec}(V),\sigma^2} \log f(x;\theta)] = \int \frac{1}{f(x;\theta)} M_1(x;\theta) M_2(x;\theta) dx,$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\mathrm{vec}(V),\alpha} \log f(x;\theta)] = \int \frac{1}{f(x;\theta)} M_1(x;\theta) M_3'(x;\theta) dx,$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\sigma^2,\sigma^2} \log f(x;\theta)] = -\frac{r^2}{4(\sigma^2)^2} + \int \frac{1}{f(x;\theta)} M_2^2(x;\theta) dx,$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\sigma^2,\alpha} \log f(x;\theta)] = \frac{r}{2\sigma^2}[\psi(\|\alpha\|_1) - G_1'(\alpha)] + \int \frac{1}{f(x;\theta)} M_2(x;\theta) M_3'(x;\theta) dx \text{ and}$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\alpha,\alpha} \log f(x;\theta)] = G_3(\alpha) - \psi^{(1)}(\|\alpha\|_1) - G_4(\alpha)$$

$$+\int \frac{1}{f(x;\theta)} M_3(x;\theta) M_3'(x;\theta) dx.$$

Since $I_\theta = -\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\theta\theta} \log f(X;\theta)]$, the claims in Theorem 3.2 follow immediately. □

S5. Proof of Corollary 3.1

Proof. Let us consider the case when $\alpha_1 = \alpha_2 = \cdots = \alpha_K = \|\alpha\|_1 / K$ for Theorem 3.2. In this case, if $\|\alpha\|_1 \to \infty$, $f(\pi;\alpha)$ will shrink to a point mass function concentrating at $\mathbf{1}_K/K$. Then $f(x;\theta) \to \phi_r(x; V\mathbf{1}_K/K, \sigma^2 I_r)$. It can be derived that

$$M_1(x;\theta) \to (\sigma^2)^{-1} K^{-1} \mathbf{1}_K \otimes (x - V\mathbf{1}_K/K)\phi_r(x; V\mathbf{1}_K/K, \sigma^2 I_r),$$

$$M_2(x;\theta) \to 2^{-1}(\sigma^2)^{-2} \|x - V\mathbf{1}_K/K\|^2 \phi_r(x; V\mathbf{1}_K/K, \sigma^2 I_r) \quad \text{and}$$

$$M_3(x;\theta) \to -\log K \mathbf{1}_K \phi_r(x; V\mathbf{1}_K/K, \sigma^2 I_r).$$

By the dominated convergence theorem, we have

$$\int f(x;\theta)^{-1} M_1(x;\theta) M_1'(x;\theta) \to (\sigma^2)^{-1} K^{-2} \mathbf{1}_K \mathbf{1}_K' \otimes I_r,$$

$$\int f(x;\theta)^{-1} M_1(x;\theta) M_2(x;\theta) \to \mathbf{0},$$

$$\int f(x;\theta)^{-1} M_1(x;\theta) M_3'(x;\theta) \to \mathbf{0},$$

$$\int f(x;\theta)^{-1} M_2^2(x;\theta) \to 4^{-1}(\sigma^2)^{-2}(r^2 + 2r),$$

$$\int f(x;\theta)^{-1} M_2(x;\theta) M_3'(x;\theta) \to -2^{-1}(\sigma^2)^{-1} r \log K \mathbf{1}_K \quad \text{and}$$

$$\int f(x;\theta)^{-1} M_3(x;\theta) M_3'(x;\theta) \to (\log K)^2 \mathbf{1}_K \mathbf{1}_K'.$$

Plugging the above limits into the formula of the Fisher information matrix in Theorem 3.2, we can obtain

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\mathrm{vec}(V),\mathrm{vec}(V)} \log f(x;\theta)] \to (\sigma^2)^{-1} K^{-2} \mathbf{1}_K \mathbf{1}_K' \otimes I_r,$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\mathrm{vec}(V),\sigma^2} \log f(x;\theta)] \to \mathbf{0},$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\mathrm{vec}(V),\alpha} \log f(x;\theta)] \to \mathbf{0} \quad \text{and}$$

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\sigma^2,\sigma^2} \log f(x;\theta)] \to 2^{-1} r (\sigma^2)^{-2}.$$

It can be shown that

$$G_1(\alpha) - \psi(\|\alpha\|_1) = \mathbb{E}_{\pi\sim \mathrm{Dir}(\alpha)}[G_2(\pi)] \to -\log K \mathbf{1}_K \quad \text{and}$$

$$G_4(\alpha) = \mathbb{E}_{\pi\sim \mathrm{Dir}(\alpha)}[G_2(\pi) G_2'(\pi)] \to (\log K)^2 \mathbf{1}_K \mathbf{1}_K'.$$

Moreover, by the definition of the trigamma function, $\psi^{(1)}(u) \to 0$ as $u \to \infty$. Thus $G_3(\alpha) - \psi^{(1)}(\|\alpha\|_1) \to 0$ as $\|\alpha\|_1 \to \infty$. Hence

$$-\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\sigma^2,\alpha} \log f(x;\theta)] \to \mathbf{0} \quad \text{and} \quad -\mathbb{E}_{X\sim f(x;\theta)}[\nabla_{\alpha,\alpha} \log f(x;\theta)] \to \mathbf{0}.$$

The claim in Corollary 3.1 has been proved. □

S6. Proof of Theorem 3.3

Proof. Let $\hat{h}_i = \|X_i - X_i^*\|$, $\hat{h} = \max_i \hat{h}_i$ and $\tilde{h} = \left\{\sum_{i=1}^n \hat{h}_i^2/n\right\}^{1/2}$. Then $\hat{h} \geq \tilde{h}$. Note that $Z_i = X_i - X_i^* \overset{iid}{\sim} \mathcal{N}_r(0, \sigma^2 I_r)$. Thus we have $\|Z_i\|^2/\sigma^2 \overset{iid}{\sim}$

$\chi^2(r)$. By Assumption 3.4, for any $\epsilon > 0$, there exists a positive integer n such that $\mathbb{P}(|\tilde{\sigma}^2/\sigma^2 - c_2| > c_2/2) \leq \epsilon/2$. For such n, there exist $M, \tilde{M} > 0$ such that $\mathbb{P}(3c_2M/2 \leq \tilde{h}^2/\sigma^2 \leq c_2\tilde{M}/2) \geq 1 - \epsilon/2$. Since

$$\begin{aligned}
&\mathbb{P}(3c_2M/2 \leq \tilde{h}^2/\sigma^2 \leq c_2\tilde{M}/2) \\
=& \mathbb{P}(3c_2M/2 \leq \tilde{h}^2/\tilde{\sigma}^2 \cdot \tilde{\sigma}^2/\sigma^2 \leq c_2\tilde{M}/2) \\
\leq& \mathbb{P}(M \leq \tilde{h}^2/\tilde{\sigma}^2 \leq \tilde{M}) + \mathbb{P}(|\tilde{\sigma}^2/\sigma^2 - c_2| > c_2/2) \\
\leq& \mathbb{P}(M \leq \tilde{h}^2/\tilde{\sigma}^2 \leq \tilde{M}) + \epsilon/2,
\end{aligned}$$

we have $\mathbb{P}(M \leq \tilde{h}^2/\tilde{\sigma}^2 \leq \tilde{M}) \geq 1 - \epsilon$. Below, we prove $\max_k \|\tilde{V}_k - V_k^*\| \leq C_0\tilde{h}$ for a constant C_0 to be decided if $M \leq \tilde{h}^2/\tilde{\sigma}^2 \leq \tilde{M}$ holds by contradiction. Define

$$\begin{aligned}
\tilde{\Delta}_i &= \int_{x\in\tilde{\mathcal{S}}} \exp\left\{-\frac{1}{2\tilde{\sigma}^2}\|X_i - x\|^2\right\} dx \quad \text{and} \\
\Delta_i &= \int_{x\in\mathcal{S}^*} \exp\left\{-\frac{1}{2\tilde{\sigma}^2}\|X_i - x\|^2\right\} dx.
\end{aligned}$$

Then by Proposition 2.1,

$$\begin{aligned}
\ell(\tilde{V}, \tilde{\sigma}^2, \mathbf{1}_K) &= -\sum_{i=1}^n \log(\tilde{\Delta}_i) + n\log\mathrm{Vol}(\tilde{\mathcal{S}}) + \frac{nr}{2}\log(\tilde{\sigma}^2) + C \quad \text{and} \\
\ell(V^*, \tilde{\sigma}^2, \mathbf{1}_K) &= -\sum_{i=1}^n \log(\Delta_i) + n\log\mathrm{Vol}(\mathcal{S}^*) + \frac{nr}{2}\log(\tilde{\sigma}^2) + C.
\end{aligned}$$

Thus we have

$$\frac{1}{n}\ell(\tilde{V}, \tilde{\sigma}^2, \mathbf{1}_K) - \frac{1}{n}\ell(V^*, \tilde{\sigma}^2, \mathbf{1}_K) = -\frac{1}{n}\sum_{i=1}^n \log\left\{\frac{\tilde{\Delta}_i/\mathrm{Vol}(\tilde{\mathcal{S}})}{\Delta_i/\mathrm{Vol}(\mathcal{S}^*)}\right\}.$$

If $\max_k \|\tilde{V}_k - V_k^*\| > C_0\tilde{h}$, we will prove $\ell(\tilde{V}, \tilde{\sigma}^2, \mathbf{1}_K) > \ell(V^*, \tilde{\sigma}^2, \mathbf{1}_K)$ based on the above equation in the following.

For any $i \in \mathcal{N}_k(\eta_n)$, we have $\|X_i - V_k^*\| \leq \hat{h}_i + \eta_n$. Then for any $x \in \mathbb{R}^r$,

$$\frac{1}{2}\|V_k^* - x\|^2 - \hat{h}_i^2 - \eta_n^2 \leq \|X_i - x\|^2 \leq 2\|V_k^* - x\|^2 + 2\hat{h}_i^2 + 2\eta_n^2.$$

Thus we have

$$\tilde{\Delta}_i \leq \exp\left\{\frac{\hat{h}_i^2 + \eta_n^2}{2\tilde{\sigma}^2}\right\} \int_{x\in\tilde{\mathcal{S}}} \exp\left\{-\frac{1}{4\tilde{\sigma}^2}\|V_k^* - x\|^2\right\} dx =: \exp\left\{\frac{\hat{h}_i^2 + \eta_n^2}{2\tilde{\sigma}^2}\right\} \tilde{U}_k,$$
$$\Delta_i \geq \exp\left\{-\frac{\hat{h}_i^2 + \eta_n^2}{\tilde{\sigma}^2}\right\} \int_{x\in\mathcal{S}^*} \exp\left\{-\frac{1}{\tilde{\sigma}^2}\|V_k^* - x\|^2\right\} dx =: \exp\left\{-\frac{\hat{h}_i^2 + \eta_n^2}{\tilde{\sigma}^2}\right\} U_k.$$

Then

$$\log\left\{\frac{\tilde{\Delta}_i/\mathrm{Vol}(\tilde{\mathcal{S}})}{\Delta_i/\mathrm{Vol}(\mathcal{S}^*)}\right\} \leq \frac{3(\hat{h}_i^2 + \eta_n^2)}{2\tilde{\sigma}^2} + \log\left\{\frac{\tilde{U}_k/\mathrm{Vol}(\tilde{S})}{U_k/\mathrm{Vol}(\mathcal{S}^*)}\right\}.$$

First, we study the lower bound of $U_k/\mathrm{Vol}(\mathcal{S}^*)$. Let Uniform$(\mathcal{S}^*)$ be the uniform distribution over the simplex $\mathcal{S}^*$ and $\tilde{C} = \max_k \mathbb{E}_{X\sim\mathrm{Uniform}(\mathcal{S}^*)}\|V_k^* - X\|^2$. By the Jessen's inequality, for any $k = 1, 2, \ldots, K$,

$$\begin{aligned}\frac{U_k}{\mathrm{Vol}(\mathcal{S}^*)} &= \mathbb{E}_{X\sim\mathrm{Uniform}(\mathcal{S}^*)} \exp\left\{-\frac{1}{\tilde{\sigma}^2}\|V_k^* - X\|^2\right\} \\ &\geq \exp\left\{-\frac{1}{\tilde{\sigma}^2}\mathbb{E}_{X\sim\mathrm{Uniform}(\mathcal{S}^*)}\|V_k^* - X\|^2\right\} \geq \exp\left\{-\frac{\tilde{C}}{\tilde{\sigma}^2}\right\}. \quad \text{(S6.16)}\end{aligned}$$

Second, we derive the upper bound of $\tilde{U}_k/\mathrm{Vol}(\tilde{\mathcal{S}})$ and hence the upper bound of $\{\tilde{U}_k/\mathrm{Vol}(\tilde{\mathcal{S}})\}/\{U_k/\mathrm{Vol}(\mathcal{S}^*)\}$. We start with the case $\mathcal{S}^* \subset \tilde{\mathcal{S}}$. Then if $\max_k \|\tilde{V}_k - V_k^*\| > C_0\tilde{h}$, there exists $k_0 \in \{1, 2, \ldots, K\}$ such that $\|\tilde{V}_{k_0} - V_{k_0}^*\| > C_0\tilde{h}$. For such k_0, there exists a positive constant $C_1 = O(C_0)$

such that $\mathcal{B}(V_{k_0}^*, C_1\tilde{h}) = \{x \in \mathbb{R}^r : \|x - V_{k_0}^*\| \le C_1\tilde{h}\} \subset \tilde{S}$. Then

$$\frac{\tilde{U}_{k_0}}{\mathrm{Vol}(\tilde{\mathcal{S}})} = \frac{A_1 + A_2}{B_1 + B_2} \le \max\left(\frac{A_1}{B_1}, \frac{A_2}{B_2}\right), \tag{S6.17}$$

where

$$A_1 = \int_{x \in \mathcal{B}(V_{k_0}^*, C_1\tilde{h})} \exp\left\{-\frac{1}{4\tilde{\sigma}^2}\left\|V_{k_0}^* - x\right\|^2\right\} dx,\ B_1 = \mathrm{Vol}\{\mathcal{B}(V_{k_0}^*, C_1\tilde{h})\},$$

$$A_2 = \int_{x \in \tilde{\mathcal{S}}/\mathcal{B}(V_{k_0}^*, C_1\tilde{h})} \exp\left\{-\frac{1}{4\tilde{\sigma}^2}\left\|V_{k_0}^* - x\right\|^2\right\} dx, B_2 = \mathrm{Vol}\{\tilde{\mathcal{S}}/\mathcal{B}(V_{k_0}^*, C_1\tilde{h})\}.$$

It can be derived that

$$\frac{A_1}{B_1} \le 2^{r/2}\Gamma(r/2+1)\left\{\frac{1 - \exp\{-C_1^2\tilde{h}^2/\tilde{\sigma}^2\}}{C_1^2\tilde{h}^2/\tilde{\sigma}^2}\right\}^{r/2} \quad \text{and}$$

$$\frac{A_2}{B_2} \le \exp\left\{-C_1^2\tilde{h}^2/(2\tilde{\sigma}^2)\right\},$$

where $\Gamma(\cdot)$ is the Gamma function. If $\tilde{h}^2/\tilde{\sigma}^2 \ge M$, since $(1 - e^{-x})/x$ is decreasing when $x \in (0, +\infty)$, we have

$$\frac{A_1}{B_1} \le 2^{r/2}\Gamma(r/2+1)\left\{\frac{1 - \exp\{-C_1^2 M\}}{C_1^2 M}\right\}^{r/2} \text{ and } \frac{A_2}{B_2} \le \exp\left\{-\frac{C_1^2 M}{2}\right\}.$$

Thus

$$\lim_{C_0 \to +\infty} A_1/B_1 = \lim_{C_0 \to +\infty} A_2/B_2 = 0. \tag{S6.18}$$

Combining (S6.16), (S6.23) and (S6.18), we have as $C_0 \to +\infty$,

$$\log\left\{\frac{\tilde{U}_{k_0}/\mathrm{Vol}(\tilde{\mathcal{S}})}{U_{k_0}/\mathrm{Vol}(\mathcal{S}^*)}\right\} \le \left\{\frac{\tilde{C}}{\tilde{\sigma}^2} + \log\max\left(\frac{A_1}{B_1}, \frac{A_2}{B_2}\right)\right\} \to -\infty. \tag{S6.19}$$

If $\|\tilde{V}_k - V_k^*\| \le C_0\tilde{h}$, note that

$$\tilde{U}_k = \int_{x \in \tilde{\mathcal{S}}} \exp\left\{-\frac{1}{4\tilde{\sigma}^2}\|V_k^* - x\|^2\right\} dx \le \int_{x \in \tilde{\mathcal{S}}} dx = \mathrm{Vol}(\tilde{\mathcal{S}}). \tag{S6.20}$$

Thus we have

$$\log\left\{\frac{\tilde{U}_k/\mathrm{Vol}(\tilde{\mathcal{S}})}{U_k/\mathrm{Vol}(\mathcal{S}^*)}\right\} \le \frac{\tilde{C}}{\tilde{\sigma}^2}. \tag{S6.21}$$

Then we derive the upper bound of $\{\tilde{U}_k/\mathrm{Vol}(\tilde{\mathcal{S}})\}/\{U_k/\mathrm{Vol}(\mathcal{S}^*)\}$ under the case $\mathcal{S}^* \cap \tilde{\mathcal{S}}^c \neq \emptyset$. In this case, there exists $k_1 \in \{1, 2, \ldots, K\}$ such that $V^*_{k_1} \notin \tilde{\mathcal{S}}$. If $\|\tilde{V}_{k_1} - V^*_{k_1}\| \le C_0\tilde{h}$, we have

$$\tilde{U}_{k_1} \le \mathrm{Vol}(\tilde{S}) \quad \text{and} \quad \log\left\{\frac{\tilde{U}_{k_1}/\mathrm{Vol}(\tilde{\mathcal{S}})}{U_{k_1}/\mathrm{Vol}(\mathcal{S}^*)}\right\} \le \frac{\tilde{C}}{\tilde{\sigma}^2}. \tag{S6.22}$$

according to (S6.20). If $\|\tilde{V}_{k_1} - V^*_{k_1}\| > C_0\tilde{h}$, there exists a positive constant $C_2 = o(C_0)$ such that $\mathcal{B}(\tilde{V}_{k_1}, C_2\tilde{h}) = \{x \in \mathbb{R}^r : \|x - \tilde{V}_{k_1}\| \le C_2\tilde{h}\} \subset \mathcal{S}^*$. Then

$$\frac{\tilde{U}_{k_1}}{\mathrm{Vol}(\tilde{\mathcal{S}})} = \frac{\tilde{A}_1 + \tilde{A}_2}{\tilde{B}_1 + \tilde{B}_2} \le \max\left(\frac{\tilde{A}_1}{\tilde{B}_1}, \frac{\tilde{A}_2}{\tilde{B}_2}\right), \tag{S6.23}$$

where

$$\tilde{A}_1 = \int_{x \in \mathcal{B}(\tilde{V}_{k_1}, C_2\tilde{h})} \exp\left\{-\frac{1}{4\tilde{\sigma}^2}\left\|V^*_{k_1} - x\right\|^2\right\} dx,\ \ \tilde{B}_1 = \mathrm{Vol}\{\mathcal{B}(\tilde{V}_{k_1}, C_2\tilde{h}) \cap \tilde{\mathcal{S}}\},$$

$$\tilde{A}_2 = \int_{x \in \tilde{\mathcal{S}}/\mathcal{B}(\tilde{V}_{k_1}, C_2\tilde{h})} \exp\left\{-\frac{1}{4\tilde{\sigma}^2}\left\|V^*_{k_1} - x\right\|^2\right\} dx, \tilde{B}_2 = \mathrm{Vol}\{\tilde{\mathcal{S}}/\mathcal{B}(\tilde{V}_{k_1}, C_2\tilde{h})\}.$$

It can be derived that

$$\frac{\tilde{A}_1}{\tilde{B}_1} \le \exp\{-(C_0 - C_2)^2\tilde{h}^2/4\tilde{\sigma}^2\} \quad \text{and} \quad \frac{\tilde{A}_2}{\tilde{B}_2} \quad \le \exp\left\{-C_2^2\tilde{h}^2/(2\tilde{\sigma}^2)\right\}.$$

If $\tilde{h}^2/\tilde{\sigma}^2 \ge M$, we have

$$\max\left(\frac{\tilde{A}_1}{\tilde{B}_1}, \frac{\tilde{A}_2}{\tilde{B}_2}\right) \le \exp\{-C_0^2 M\} = o(1),$$

where the last limit hold when $C_0 \to \infty$. Hence

$$\log\left\{\frac{\tilde{U}_{k_1}/\mathrm{Vol}(\tilde{\mathcal{S}})}{U_{k_1}/\mathrm{Vol}(\mathcal{S}^*)}\right\} \le \left\{\frac{\tilde{C}}{\tilde{\sigma}^2} + \log\max\left(\frac{\tilde{A}_1}{\tilde{B}_1}, \frac{\tilde{A}_2}{\tilde{B}_2}\right)\right\} \to -\infty. \tag{S6.24}$$

For $i \in \mathcal{M}$, it can be derived that

$$\tilde{\Delta}_i = \int_{x\in\tilde{\mathcal{S}}} \exp\left\{-\frac{1}{2\tilde{\sigma}^2}\|X_i - x\|^2\right\} dx \le \int_{x\in\tilde{\mathcal{S}}} dx = \mathrm{Vol}(\tilde{S}).$$

Moreover, since

$$\|X_i - x\|^2 \le 2\|X_i^* - x\|^2 + 2\hat{h}_i^2 \le 4\|X_i^* - V_k^*\|^2 + 4\|V_k^* - x\|^2 + 2\hat{h}_i^2,$$

by the Jessen's inequality, we have

$$\begin{aligned}
\frac{\Delta_i}{\mathrm{Vol}(\mathcal{S}^*)} &= \mathbb{E}_{X\sim\mathrm{Uniform}(\mathcal{S}^*)} \exp\left\{-\frac{1}{2\tilde{\sigma}^2}\|X_i - X\|^2\right\} \\
&\ge \exp\left\{-\frac{1}{2\tilde{\sigma}^2}\mathbb{E}_{X\sim\mathrm{Uniform}(\mathcal{S}^*)}\|X_i - X\|^2\right\} \\
&\ge \exp\left\{-\frac{1}{2\tilde{\sigma}^2}(4\|X_i^* - V_k^*\|^2 + 2\hat{h}_i^2 + 4\mathbb{E}_{X\sim\mathrm{Uniform}(\mathcal{S}^*)}\|V_k^* - X\|^2)\right\} \\
&\ge \exp\left\{-\frac{2\check{C} + \hat{h}_i^2 + 2\tilde{C}}{\tilde{\sigma}^2}\right\},
\end{aligned}$$

where $\check{C} = \max_{1\le k\ne l\le K}\|V_k^* - V_l^*\|^2$. Then

$$\log\left\{\frac{\tilde{\Delta}_i/\mathrm{Vol}(\tilde{\mathcal{S}})}{\Delta_i/\mathrm{Vol}(\mathcal{S}^*)}\right\} \le \frac{2\check{C} + \hat{h}_i^2 + 2\tilde{C}}{\tilde{\sigma}^2}. \tag{S6.25}$$

Let $\mathcal{K}_0 = \{k = 1, 2, \ldots, K : \|\tilde{V}_k - V_k^*\| > C_0\hat{h}\}$ and $\mathcal{K}_1 = \{k = 1, 2, \ldots, K : \|\tilde{V}_k - V_k^*\| \le C_0\hat{h}\}$. From (S6.19) – (S6.24) and (S6.25),

$$\begin{aligned}
&\frac{1}{n}\ell(\tilde{V}, \tilde{\sigma}^2, \mathbf{1}_K) - \frac{1}{n}\ell(V^*, \tilde{\sigma}^2, \mathbf{1}_K) \\
\ge& -\frac{1}{n}\sum_{k\in\mathcal{K}_0}|\mathcal{N}_k(\eta_n)|\left\{\frac{\tilde{C}}{\tilde{\sigma}^2} + \log\max\left(\frac{A_1}{B_1}, \frac{A_2}{B_2}\right)\right\} - \frac{1}{n}\sum_{k\in\mathcal{K}_1}|\mathcal{N}_k(\eta_n)|\left(\frac{\tilde{C}}{\tilde{\sigma}^2}\right)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{|\mathcal{M}|}{n}\cdot\frac{2\check{C}+2\tilde{C}}{\tilde{\sigma}^2}-\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{N}_k}\frac{3(\hat{h}_i^2+\eta_n^2)}{2\tilde{\sigma}^2}-\frac{1}{n}\sum_{i\in\mathcal{M}}\frac{\hat{h}_i^2}{\tilde{\sigma}^2}\\
=&-\frac{2|\mathcal{M}|\check{C}}{n\tilde{\sigma}^2}-\frac{3(\tilde{h}^2+\eta_n^2)}{2\tilde{\sigma}^2}-\frac{(n+|\mathcal{M}|)\tilde{C}}{n\tilde{\sigma}^2}-\frac{1}{n}\sum_{k\in\mathcal{K}_0}|\mathcal{N}_k(\eta_n)|\log\max\left(\frac{A_1}{B_1},\frac{A_2}{B_2}\right)\\
\geq&-\frac{2|\mathcal{M}|\check{C}}{n\tilde{\sigma}^2}-\frac{3(\tilde{h}^2+\eta_n^2)}{2\tilde{\sigma}^2}-\frac{(n+|\mathcal{M}|)\tilde{C}}{n\tilde{\sigma}^2}-\frac{\tilde{C}}{\tilde{\sigma}^2}-\frac{c_1 r_n}{n}|\mathcal{K}_0|\log\max\left(\frac{A_1}{B_1},\frac{A_2}{B_2}\right).
\end{aligned}
$$

Note that A_1, A_2, B_1 and B_2 in the above should be $\tilde{A}_1$, $\tilde{A}_2$, $\tilde{B}_1$ and $\tilde{B}_2$ as defined in (S6.24) and (S6.24) if $V_{k_1}\notin\mathcal{S}$. Here we unify the notation for simplicity. When r is fixed, we have $\tilde{C}/\tilde{\sigma}^2=O(1)$ and $\check{C}/\tilde{\sigma}^2=O(1)$. Under the condition of $\tilde{h}^2/\tilde{\sigma}^2\leq\tilde{M}$ and $\log^{-1}C_0=o(r_n/n)$, if $|\mathcal{K}_0|\geq 1$, we have

$$
\frac{1}{n}\ell(\tilde{V},\tilde{\sigma}^2,\mathbf{1}_K)-\frac{1}{n}\ell(V^*,\tilde{\sigma}^2,\mathbf{1}_K)\to+\infty.
$$

This yields a contradiction to the definition of $\tilde{V}$ and $\tilde{\sigma}^2$. Hence $\max_k\|\tilde{V}_k-V_k^*\|\leq C_0\tilde{h}\leq C_0\hat{h}$. Then

$$
\begin{aligned}
&\mathbb{P}(\max_k\|\tilde{V}_k-V_k^*\|\leq C_0\hat{h})\geq\mathbb{P}(\max_k\|\tilde{V}_k-V_k^*\|\leq C_0\tilde{h})\\
\geq&\ \mathbb{P}(M\leq\tilde{h}^2/\tilde{\sigma}^2\leq\tilde{M})\geq 1-\epsilon.
\end{aligned}
$$

This proves the claim. □

S7. Proof of Theorem 3.4

Proof. The proof of Theorem 3.4 is similar with that of Theorem 3.3 if we notice that Assumption 3.5 implies Assumption 3.3. □

References

Gritzmann, P. and V. Klee (1994). On the complexity of some basic problems in computational convexity. In *Polytopes: Abstract, Convex and Computational*, pp. 373–466.

Sommerville, D. M. Y. (1958). *An Introduction to the Geometry of N Dimensions.* New York: Dover.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, Volume 3. London: Cambridge University Press.