Simultaneous estimation of normal means with side information

Sihai Dave Zhao

Department of Statistics, University of Illinois at Urbana-Champaign

Supplementary Material

These supplementary materials contain simulation results when the primary and auxiliary data are correlated, as well as proofs of Proposition 1 and Theorems 1–3 from the main text.

S1 Simulations with correlated X_{i1} and X_{i2}

When X_{i1} and X_{i2} are correlated, $\mathbf{X}_{\cdot 2}$ can provide useful information for estimating $\boldsymbol{\theta}_{\cdot 1}$ even if $\boldsymbol{\theta}_{\cdot 2}$ and $\boldsymbol{\theta}_{\cdot 1}$ are completely unrelated. This section illustrates this phenomenon in simulations that compare the oracle non-integrative rule

$$\delta_i^\star(\mathbf{X}_{\cdot 1}) = \frac{\sum_{j=1}^n \theta_{j1}\phi\{(X_{i1} - \theta_{j1})/\sigma_1\}/\sigma_1}{\sum_{j=1}^n \phi\{(X_{i1} - \theta_{j1})/\sigma_1\}/\sigma_1},$$

where $\phi(x)$ is the standard normal density, to the oracle integrative rule when (X_{i1}, X_{i2}) is bivariate normal with correlation r:

$$\delta_i^\star(\mathbf{X}_{\cdot 1}, \mathbf{X}_{\cdot 2}) = \frac{\sum_{j=1}^n \theta_{j1} \exp\{-(\mathbf{X}_{i\cdot} - \boldsymbol{\theta}_{i\cdot})^\top \Sigma^{-1} (\mathbf{X}_{i\cdot} - \boldsymbol{\theta}_{i\cdot})/2\}}{\sum_{j=1}^n \exp\{-(\mathbf{X}_{i\cdot} - \boldsymbol{\theta}_{i\cdot})^\top \Sigma^{-1} (\mathbf{X}_{i\cdot} - \boldsymbol{\theta}_{i\cdot})/2\}}, \tag{S1.1}$$

with $\mathbf{X}_{i\cdot} = (X_{i1}, X_{i2})^\top$, $\boldsymbol{\theta}_{i\cdot} = (\theta_{i1}, \theta_{i2})^\top$, and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

The θ_{i1} and non-informative θ_{i2} were generated as in Section 5, and the $X_{id} = \theta_{id} + \epsilon_{id}$ for $d = 1, 2$, where $(\epsilon_{i1}, \epsilon_{i2})$ are bivariate normal with mean zero and variance Σ. Figure 1 reports the average losses of these two oracle estimators over 200 simulations for $r = -0.9, 0, 0.9$ and shows that the integrative estimator significantly outperformed the non-integrative one when $r \neq 0$.

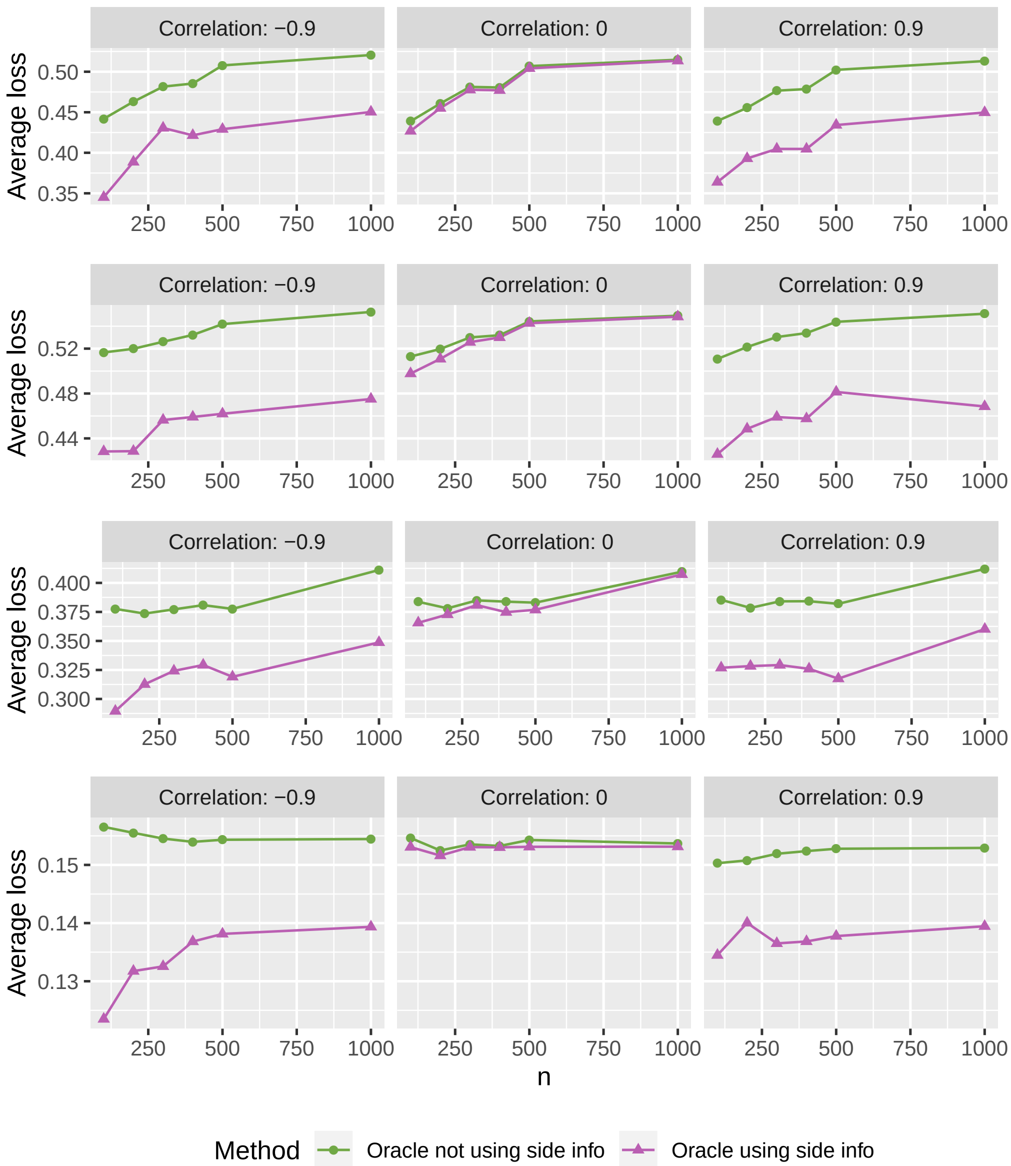

Figure 1: Average losses for four different configurations of $\boldsymbol{\theta}_{\cdot 1}$, non-informative $\boldsymbol{\theta}_{\cdot 2}$, and three levels of correlation between X_{i1} and X_{i2}.

S2 Proof of Proposition 1

For any separable rule $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n) \in \mathcal{S}$ (3.2) with $\delta_i(\mathbf{X}_{\cdot 1}, \mathbf{X}_{\cdot 2}) = f(X_{i1}, X_{i2})$,

$$\begin{aligned}
R_n(\boldsymbol{\theta}, \boldsymbol{\delta}) = &\frac{1}{n}\sum_{i=1}^n \int \{\theta_{i1} - f^\star(x_1, x_2)\}^2 p_i^0(x_1, x_2) dx_1 dx_2 + \\
&\frac{2}{n}\sum_{i=1}^n \int \{\theta_{i1} - f_i^\star(x_1, x_2)\}\{f^\star(x_1, x_2) - f(x_1, x_2)\} p_i^0(x_1, x_2) dx_1 dx_2 + \\
&\frac{1}{n}\sum_{i=1}^n \int \{f^\star(x_1, x_2) - f(x_1, x_2)\}^2 p_i^0(x_1, x_2) dx_1 dx_2,
\end{aligned}$$

where $p_i^0(x_1, x_2)$ equals the density of (X_{i1}, X_{i2}) as defined in (3.3). For the $f^\star$ given in (3.4), the middle term above equals zero, so

$$\begin{aligned}
R_n(\boldsymbol{\theta}, \boldsymbol{\delta}) = &\frac{1}{n}\sum_{i=1}^n \mathrm{E}_\theta[\{\theta_{i1} - f^\star(X_{i1}, X_{i2})\}^2] + \frac{1}{n}\sum_{i=1}^n \mathrm{E}_\theta[\{f^\star(X_{i1}, X_{i2}) - f(X_{i1}, X_{i2})\}^2] \\
\geq &R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star),
\end{aligned}$$

for any $\boldsymbol{\delta} \in \mathcal{S}$. Therefore $\boldsymbol{\delta}^\star$ achieves the minimum risk within the class of separable rules.

S3 Proof of Theorem 1

Denoting the density of (X_{i1}, X_{i2}) by $p_i^0(x_1, x_2)$ as in (3.3), the risk of the regularized oracle rule $\boldsymbol{\delta}_\rho^\star$ (3.6) can be written as

$$\begin{aligned}
R_n(\boldsymbol{\theta}, \boldsymbol{\delta}_\rho^\star) = &\frac{1}{n}\sum_{i=1}^n \mathrm{E}(\theta_{i1} - X_{i1})^2 - \\
&\frac{2}{n}\sum_{i=1}^n \int \frac{\sum_j (\theta_{j1} - x_1) p_j^0(x_1, x_2)}{\rho + \sum_j p_j^0(x_1, x_2)} (\theta_{i1} - x_1) p_i^0(x_1, x_2) dx_1 dx_2 + \\
&\frac{1}{n}\sum_{i=1}^n \int \left\{ \frac{\sum_j (\theta_{j1} - x_1) p_j^0(x_1, x_2)}{\rho + \sum_j p_j^0(x_1, x_2)} \right\}^2 p_i^0(x_1, x_2) dx_1 dx_2.
\end{aligned}$$

The second term above obeys

$$\begin{aligned}
&\frac{2}{n}\sum_{i=1}^n \int \frac{\sum_j (\theta_{j1} - x_1) p_j^0(x_1, x_2)}{\rho + \sum_j p_j^0(x_1, x_2)} (\theta_{i1} - x_1) p_i^0(x_1, x_2) dx_1 dx_2 \\
= &\frac{2}{n} \int \frac{\{\sum_j (\theta_{j1} - x_1) p_j^0(x_1, x_2)\}^2}{\rho + \sum_j p_j^0(x_1, x_2)} dx_1 dx_2 \\
\geq &\frac{2}{n} \int \left\{ \frac{\sum_j (\theta_{j1} - x_1) p_j^0(x_1, x_2)\}^2}{\rho + \sum_j p_j^0(x_1, x_2)} \right\}^2 \sum_j p_j^0(x_1, x_2) dx_1 dx_2,
\end{aligned}$$

since $\rho > 0$. Therefore

$$R_n(\boldsymbol{\theta}, \boldsymbol{\delta}_\rho^\star) \leq \sigma_1^2 - \frac{1}{n}\sum_{i=1}^n \int \left\{ \frac{\sum_j (\theta_{j1} - x_1) p_j^0(x_1, x_2)}{\rho + \sum_j p_j^0(x_1, x_2)} \right\}^2 p_i^0(x_1, x_2) dx_1 dx_2.$$

Similarly, the risk of the unregularized oracle rule $\boldsymbol{\delta}^\star$ (3.4) is

$$R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star) = \sigma_1^2 - \frac{1}{n}\sum_{i=1}^n \int \left\{ \frac{\sum_j (\theta_{j1} - x_1) p_j^0(x_1, x_2)}{\sum_j p_j^0(x_1, x_2)} \right\}^2 p_i^0(x_1, x_2) dx_1 dx_2.$$

Since

$$\begin{aligned}
&\left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^2 - \left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\rho+\sum_j p_j^0(x_1,x_2)}\right\}^2 \\
&= \left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^2 \left[1 - \left\{\frac{\sum_j p_j^0(x_1,x_2)}{\rho+\sum_j p_j^0(x_1,x_2)}\right\}^2\right] \\
&= \left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^2 \left\{1 - \frac{\sum_j p_j^0(x_1,x_2)}{\rho+\sum_j p_j^0(x_1,x_2)}\right\} \left\{1 + \frac{\sum_j p_j^0(x_1,x_2)}{\rho+\sum_j p_j^0(x_1,x_2)}\right\} \\
&\leq \left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^2 \frac{2\rho}{\rho+\sum_j p_j^0(x_1,x_2)\}},
\end{aligned}$$

it follows that $R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star_\rho) - R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star)$ is upper bounded by

$$\frac{2}{n}\int \left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^2 \frac{\rho}{\rho+\sum_j p_j^0(x_1,x_2)} \sum_{i=1}^n p_i^0(x_1,x_2)dx_1dx_2.$$

To simplify notation, for any set $\mathcal{C} \subseteq \mathbb{R}^2$, define

$$\Delta_n(\mathcal{C}) = \frac{2}{n}\int_{\mathcal{C}} \left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^2 \frac{\rho}{\rho+\sum_j p_j^0(x_1,x_2)} \sum_{i=1}^n p_i^0(x_1,x_2)dx_1dx_2, \tag{S3.1}$$

so that $R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star_\rho) - R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star) \leq \Delta_n(\mathbb{R}^2)$. It will be shown below that $\Delta_n(\mathbb{R}^2)$ tends to zero asymptotically. Since the regularized oracle rule $\boldsymbol{\delta}^\star_\rho$ is separable (3.2), $0 \leq R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star_\rho) - R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^\star)$ as well, by Proposition 1. This will conclude the proof.

To upper-bound $\Delta_n(\mathbb{R}^2)$, define the set

$$\mathcal{L}_n = \left\{(x_1,x_2) : \log n \leq \sum_i p_i^0(x_1,x_2)\right\},$$

so that, roughly, $\sum_i p_i^0(x_1,x_2)$ is large on $\mathcal{L}_n$. Then $\Delta_n(\mathbb{R}^2) = \Delta_n(\mathcal{L}_n) + \Delta_n(\mathcal{L}_n^c)$, and

$$\begin{aligned}\Delta_n(\mathcal{L}_n) &\le \frac{2\rho}{\rho+\log n}\frac{1}{n}\int_{\mathcal{L}_n}\left\{\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^2\sum_j p_j^0(x_1,x_2)dx_1dx_2\\ &\le \frac{2\rho}{\rho+\log n}\frac{1}{n}\int_{\mathcal{L}_n}\sum_j(\theta_{j1}-x_1)^2p_j^0(x_1,x_2)dx_1dx_2 \le \frac{2\rho}{\rho+\log n}\sigma_1^2 \to 0,\end{aligned}$$

where the second inequality follows from Jensen's inequality. It remains to show that $\Delta(\mathcal{L}_n^c)$ goes to zero. To this end, further define the sets

$$\begin{aligned}\mathcal{A}_1 &= \{(x_1,x_2) : x_1 \le -Cn^{1/4-\eta}\},\\ \mathcal{A}_2 &= \{(x_1,x_2) : -Cn^{1/4-\eta} < x_1 \le -Cn^{1/4-\eta}\},\\ \mathcal{A}_3 &= \{(x_1,x_2) : Cn^{1/4+\eta} < x_1\}.\end{aligned}$$

It will be shown that $\Delta_n(\mathcal{L}_n^c\cap\mathcal{A}_1)$, $\Delta_n(\mathcal{L}_n^c\cap\mathcal{A}_2)$, and $\Delta_n(\mathcal{L}_n^c\cap\mathcal{A}_3)$ converge to zero.

First,

$$\Delta_n(\mathcal{L}_n^c\cap\mathcal{A}_1) \le \frac{2}{n}\int_{\mathcal{L}_n^c\cap\mathcal{A}_1}\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\sum_{i=1}^n(\theta_{i1}-x_1)p_i^0(x_1,x_2)dx_1dx_2.$$

Defining the function $g(x) = \exp\{x/(2\sigma_1^2)\}$, Jensen's inequality gives

$$\begin{aligned}g\left\{\frac{\sum_j(\theta_{j1}-x_1)^2p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\} &\le \frac{\sum_j g\{(\theta_{j1}-x_1)^2\}p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\\ &= \frac{\sum_j(2\pi\sigma_1\sigma_2)^{-1}\exp\{-(x_2-\theta_{j2})/\sigma_2^2\}}{\sum_j p_j^0(x_1,x_2)} \le \frac{(2\pi\sigma_1\sigma_2)^{-1}n}{\sum_j p_j^0(x_1,x_2)},\end{aligned}$$

which is always at least 1. Combining this with another application of

Jensen's inequality implies

$$\left|\frac{\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right| \leq \left\{\frac{\sum_j(\theta_{j1}-x_1)^2 p_j^0(x_1,x_2)}{\sum_j p_j^0(x_1,x_2)}\right\}^{1/2} \leq \left\{2\sigma_1^2\log\frac{(2\pi\sigma_1\sigma_2)^{-1}n}{\sum_j p_j^0(x_1,x_2)}\right\}^{1/2}. \tag{S3.2}$$

Since $|\theta_{j1}| \leq Cn^{1/4-\eta}$ by Assumption 1, $\theta_{i1}-x_1 \geq 0$ on $\mathcal{A}_1$, so

$$\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_1) \leq \frac{2}{n}\int_{\mathcal{L}_n^c\cap\mathcal{A}_1}\left\{2\sigma_1^2\log\frac{(2\pi\sigma_1\sigma_2)^{-1}n}{\sum_j p_j^0(x_1,x_2)}\right\}^{1/2}\sum_{i=1}^n(\theta_{i1}-x_1)p_i^0(x_1,x_2)dx_1dx_2.$$

Now for each value of x_2, define the function

$$u_{x_2}(x_1) = n^{-1}2\pi\sigma_1\sigma_2\sum_j p_j^0(x_1,x_2)$$

so that

$$\frac{du_{x_2}}{dx_1} = n^{-1}2\pi\sigma_1\sigma_2\sum_j(\theta_{j1}-x_1)p_j^0(x_1,x_2)dx_1.$$

This implies that

$$\begin{aligned}
&\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_1)\\
&\leq \frac{1}{\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\int_{u_{x_2}(-\infty)}^{u_{x_2}(-Cn^{1/4-\eta})} I\{n^{-1}2\pi\sigma_1\sigma_2\log n \geq u_{x_2}(x_1)\}\left(2\sigma_1^2\log\frac{1}{u_{x_2}}\right)^{1/2}du_{x_2}dx_2\\
&= \frac{1}{\pi\sigma_2}\int_{-\infty}^{\infty}\int_{u_{x_2}(-\infty)}^{u_{x_2}(-Cn^{1/4-\eta})} I\{n^{-1}2\pi\sigma_1\sigma_2\log n \geq u_{x_2}(x_1)\}\left(\log\frac{1}{u_{x_2}}\right)^{1/2}du_{x_2}dx_2
\end{aligned}$$

Since $\theta_{j1}-x_1 > 0$ for all j on $\mathcal{A}_1$. Therefore, $u_{x_2}(x_1)$ is increasing on $\mathcal{A}_1$ and $u_{x_2}(-Cn^{1/4-\eta}) > 0$. Then

$$\begin{aligned}
\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_1) &\leq \frac{1}{\pi\sigma_2}\int_{-\infty}^{\infty}\int_0^{n^{-1}2\pi\sigma_1\sigma_2\log n \wedge u_{x_2}(-Cn^{1/4-\eta})}\left(\log\frac{1}{u_{x_2}^2}\right)^{1/2}du_{x_2}dx_2\\
&\leq \frac{1}{\pi\sigma_2}\int_{-\infty}^{\infty}\left\{\int_0^{u_{x_2}(-Cn^{1/4-\eta})}1du_{x_2}\right\}^{1/2}\left\{\int_0^{n^{-1}2\pi\sigma_1\sigma_2\log n}\log\frac{1}{u_{x_2}^2}du_{x_2}\right\}^{1/2}dx_2,
\end{aligned}$$

where the last line follows from the Cauchy-Schwarz inequality. Integrating by parts gives

$$\int_0^{n^{-1}2\pi\sigma_1\sigma_2\log n} \log\frac{1}{u_{x_2}^2} du_{x_2} = u_{x_2}\log\frac{1}{u_{x_2}^2}\bigg|_0^{n^{-1}2\pi\sigma_1\sigma_2\log n} + \int_0^{n^{-1}2\pi\sigma_1\sigma_2\log n} 2du_{x_2}dx_2$$
$$= \frac{2\pi\sigma_1\sigma_2\log n}{n}\left(2\log\frac{n}{2\pi\sigma_1\sigma_2\log n}+2\right),$$

so

$$\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_1) \leq \frac{2\sigma_1\log^{1/2} n}{n}\left(2\log\frac{n}{2\pi\sigma_1\sigma_2\log n}+2\right)^{1/2}\int_{-\infty}^{\infty}\left\{\sum_j p_j^0(-Cn^{1/4-\eta}, x_2)\right\}^{1/2} dx_2.$$

Finally, first integrate the remaining integral over $x_2 \leq -Cn^{1/4-\eta}$. Since $|\theta_{j2}| \leq Cn^{1/4-\eta}$ by Assumption 1, it follows that $\theta_{j2}-x_2 \geq -Cn^{1/4-\eta}-x_2 \geq 0$ for all $j = 1,\ldots,n$ and thus that $(\theta_{j2}-x_2)^2 \geq (-Cn^{1/4-\eta}-x_2)^2$. Therefore

$$\sum_j p_j^0(-Cn^{1/4-\eta}, x_2) = \frac{1}{2\pi\sigma_1\sigma_2}\sum_j \exp\left\{-\frac{1}{2\sigma_1^2}(-Cn^{1/4-\eta}-\theta_{j1})^2\right\}\exp\left\{-\frac{1}{2\sigma_2^2}(x_2-\theta_{j2})^2\right\}$$
$$\leq \frac{n}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{1}{2\sigma_2^2}(x_2+Cn^{1/4-\eta})^2\right\},$$

which implies that

$$\int_{-\infty}^{-Cn^{1/4-\eta}}\left\{\sum_j p_j^0(-Cn^{1/4-\eta}, x_2)\right\}^{1/2} dx_2$$
$$\leq \int_{-\infty}^{-Cn^{1/4-\eta}}\frac{n^{1/2}}{(2\pi\sigma_1\sigma_2)^{1/2}}\exp\left\{-\frac{1}{4\sigma_2^2}(x_2+Cn^{1/4-\eta})^2\right\}dx_2$$
$$\leq \frac{(2\sigma_2 n)^{1/2}}{\sigma_1^{1/2}}\int_{-\infty}^{-Cn^{1/4-\eta}}\frac{1}{(2\pi 2\sigma_2^2)^{1/2}}\exp\left\{-\frac{1}{4\sigma_2^2}(x_2+Cn^{1/4-\eta})^2\right\}dx_2 \leq \frac{(2\sigma_2 n)^{1/2}}{\sigma_1^{1/2}}.$$

A similar calculation shows that on $Cn^{1/4-\eta} < x_2$,

$$\int_{Cn^{1/4-\eta}}^{\infty} \left\{ \sum_j p_j^0(-Cn^{1/4-\eta}, x_2) \right\}^{1/2} dx_2 \leq \frac{(2\sigma_2 n)^{1/2}}{\sigma_1^{1/2}}$$

as well. Finally, on $-Cn^{1/4-\eta} < x_2 \leq Cn^{1/4-\eta}$,

$$\int_{-Cn^{1/4-\eta}}^{Cn^{1/4-\eta}} \left\{ \sum_j p_j^0(-Cn^{1/4-\eta}, x_2) \right\}^{1/2} dx_2 \leq \frac{n^{1/2}}{(2\pi\sigma_1\sigma_2)^{1/2}} 2Cn^{1/4-\eta}.$$

Putting these results together shows that $\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_1)$ is at most

$$\frac{2\sigma_1 \log^{1/2} n}{n} \left(2 \log \frac{n}{2\pi\sigma_1\sigma_2 \log n} + 2 \right)^{1/2} \left\{ 2 \frac{(2\sigma_2 n)^{1/2}}{\sigma_1^{1/2}} + \frac{n^{1/2}}{(2\pi\sigma_1\sigma_2)^{1/2}} 2Cn^{1/4-\eta} \right\} \to 0.$$

A very similar calculation can be used to show that $\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_3) \to 0$ as well.

It remains to bound $\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_2)$. Denote $u(x_1, x_2) = \sum_j p_j^0(x_1, x_2)$ and use (S3.2) see that $\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_2)$ is at most

$$\frac{4\sigma_1^2}{n} \int_{-\infty}^{\infty} \int_{-Cn^{1/4-\eta}}^{Cn^{1/4-\eta}} I\{\log n \geq u(x_1, x_2)\} \log \frac{(2\pi\sigma_1\sigma_2)^{-1} n}{u(x_1, x_2)} u(x_1, x_2) dx_1 dx_2.$$

Now for any constant K,

$$\frac{d}{du} \left(u^{1/2} \log \frac{K}{u} \right) = \frac{1}{u^{1/2}} \left(\frac{1}{2} \log \frac{K}{u} - 1 \right).$$

This implies that the function $u^{1/2} \log(K/u)$ increases for $u \leq Ke^{-2}$ and decreases thereafter. Now consider the function $2u^{1/2}$, which equals $u^{1/2} \log(K/u)$ at $u = Ke^{-2}$ and increases thereafter. Therefore $u^{1/2} \log(K/u)$ is upper-bounded by the nondecreasing function $u^{1/2}\{2 \vee \log(K/u)\}$. Therefore

$\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_2)$ is at most

$$\begin{aligned}
&\frac{4\sigma_1^2}{n}\int_{-\infty}^{\infty}\int_{-Cn^{1/4-\eta}}^{Cn^{1/4-\eta}} I\{\log n \geq u(x_1,x_2)\}\left\{2 \vee \log\frac{(2\pi\sigma_1\sigma_2)^{-1}n}{u(x_1,x_2)}\right\} u(x_1,x_2)dx_1dx_2 \\
&\leq \frac{4\sigma_1^2 \log^{1/2} n}{n}\left\{2 \vee \log\frac{(2\pi\sigma_1\sigma_2)^{-1}n}{\log n}\right\}\int_{-\infty}^{\infty}\int_{-Cn^{1/4-\eta}}^{Cn^{1/4-\eta}} u^{1/2}(x_1,x_2)dx_1dx_2 \\
&\leq \frac{4\sigma_1^2 \log^{1/2} n}{n}\left\{2 \vee \log\frac{(2\pi\sigma_1\sigma_2)^{-1}n}{\log n}\right\}\frac{2Cn^{1/4-\eta}}{(2\pi\sigma_1\sigma_2)^{1/2}}\times \\
&\quad \int_{-\infty}^{\infty}\left[\sum_j \exp\left\{-\frac{1}{2\sigma_2^2}(x_2-\theta_{j2})^2\right\}\right]^{1/2} dx_2.
\end{aligned}$$

To finish bounding $\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_2)$, integrate over $x_2 \leq -Cn^{1/4-\eta}$, $-Cn^{1/4-\eta} < x_2 \leq Cn^{1/4-\eta}$, and $Cn^{1/4-\eta} < x_2$, as was done when bounding $\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_1)$. Using Assumption 1, it can be shown that

$$\int_{-\infty}^{\infty}\left[\sum_j \exp\left\{-\frac{1}{2\sigma_2^2}(x_2-\theta_{j2})^2\right\}\right]^{1/2} dx_2 \leq n^{1/2}(4\pi^{1/2}\sigma_2 + 2Cn^{1/4-\eta}),$$

which means that

$$\Delta_n(\mathcal{L}_n^c \cap \mathcal{A}_2) \leq \frac{4\sigma_1^2 \log^{1/2} n}{n^{1/2}}\left\{2 \vee \log\frac{(2\pi\sigma_1\sigma_2)^{-1}n}{\log n}\right\}\frac{2Cn^{1/4-\eta}}{(2\pi\sigma_1\sigma_2)^{1/2}}(4\pi^{1/2}\sigma_2 + 2Cn^{1/4-\eta}) \to 0.$$

S4 Proof of Theorem 2

S4.1 Outline

For the parameter vector $\mathbf{t} = (t_{11}, \ldots, t_{n1}, t_{12}, \ldots, t_{n2})$, define the function

$$f(x_1, x_2, \mu; \mathbf{t}) = \frac{\sum_j (t_{j1} - x_1)^2 p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_1, t_2)} - \sigma_1^2 \frac{\sum_j p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_{j1}, t_{j2})} - \left\{ \frac{\sum_j (t_{1j} - x_1) p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_{j1}, t_{j2})} \right\}^2 - (x_1 - \mu) \frac{\sum_j (t_{1j} - x_1) p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_{j1}, t_{j2})}. \tag{S4.1}$$

Then it is straightforward to show that

$$\textsc{sure}(\mathbf{t}) - \ell_n(\mathbf{t}) = \sigma_1^2 - \frac{1}{n} \sum_{i=1}^n (X_{i1} - \theta_{i1})^2 + \frac{2}{n} \sum_{i=1}^n f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}).$$

By Jensen's inequality,

$$\begin{aligned} \left\{ E \left| \sigma_1^2 - \frac{1}{n} \sum_i (X_{i1} - \theta_{i1})^2 \right| \right\}^2 \leq & E \left\{ \sigma_1^2 - \frac{1}{n} \sum_i (X_{i1} - \theta_{i1})^2 \right\}^2 \\ = & \sigma_1^4 - \frac{2\sigma_1^2}{n} \sum_i E(X_{i1} - \theta_{i1})^2 + \frac{1}{n^2} E \left\{ \sum_i (X_{i1} - \theta_{i1})^2 \right\}^2 \\ = & -\sigma_1^4 + \frac{\sigma_1^4}{n^2} E \left\{ \sum_i \left(\frac{X_{i1} - \theta_{i1}}{\sigma_1} \right)^2 \right\}^2. \end{aligned}$$

Next, the variable $\sum_i (X_{i1} - \theta_{i1})^2 / \sigma_1^2$ has a χ_n^2 distribution, which has mean n and variance $2n$. Therefore

$$\left\{ E \left| \sigma_1^2 - \frac{1}{n} \sum_i (X_{i1} - \theta_{i1})^2 \right| \right\}^2 \leq -\sigma_1^4 + \frac{\sigma_1^4}{n^2}(n^2 + 2n) = \frac{2\sigma_1^4}{n} \to 0.$$

The remainder of the proof uses empirical process techniques to show that

$$E \sup_{\mathbf{t} \in \mathcal{T}} \left| \frac{1}{n} \sum_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) \right| \to 0.$$

S4.2 Symmetrization

The terms $f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t})$ are independent but not identically distributed. Nevertheless, the proof of the symmetrization Lemma 2.3.1 of van der Vaart and Wellner (1996) still holds and implies that

$$E \sup_{\mathbf{t} \in \mathcal{T}} \left| \frac{1}{n} \sum_i \{f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) - Ef(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t})\} \right| \leq 2E \sup_{\mathbf{t} \in \mathcal{T}} \left| \frac{1}{n} \sum_i \varepsilon_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) \right|,$$

where the ε_i are Rademacher random variables independent of the X_{i1} and X_{i2}. Using Stein (1981), it can be shown that $Ef(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) = 0$. Therefore it only remains to show that

$$E \sup_{\mathbf{t} \in \mathcal{T}} \left| \frac{1}{n} \sum_i \varepsilon_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) \right| \to 0. \tag{S4.2}$$

S4.3 Truncation

This section first constructs a function $F_n(x_1, \mu)$ that satisfies $|f(x_1, x_2, \mu; \mathbf{t})| \leq F_n(x_1, \mu)$ as well as $E\{F_n(X_{i1}, \mu)\} < \infty$, for f defined in (S4.1). By the triangle inequality,

$$
\begin{aligned}
&|f(x_1, x_2, \mu; \mathbf{t})| \\
&\leq \frac{\sum_j (t_{j1} - x_1)^2 p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_1, t_2)} + \sigma_1^2 + \\
&\quad \left\{ \frac{\sum_j (t_{1j} - x_1) p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_{j1}, t_{j2})} \right\}^2 + |x_1 - \mu| \left| \frac{\sum_j (t_{1j} - x_1) p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_{j1}, t_{j2})} \right|.
\end{aligned}
$$

Using Jensen's inequality as in (S3.2), it follows that

$$
\frac{\sum_j (t_{j1} - x_1)^2 p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_1, t_2)} \leq \frac{\sum_j p(x_1, x_2; t_1, t_2)}{\rho + \sum_j p(x_1, x_2; t_1, t_2)} 2\sigma_1^2 \log \frac{(2\pi\sigma_1\sigma_2)^{-1} n}{\sum_j p(x_1, x_2; t_1, t_2)}.
$$

Let $y = p(x_1, x_2; t_1, t_2)$ and $C = (2\pi\sigma_1\sigma_2)^{-1} n$ such that this upper bound can be written as

$$
u(y) = \frac{y}{\rho + y} 2\sigma_1^2 \log \frac{C}{y},
$$

where $0 \leq y \leq C$ by the definition of $p(x_1, x_2; t_1, t_2)$ in (3.3). Since

$$
u'(y) = \frac{2\sigma_1^2}{\rho + y} \left(\frac{\rho}{\rho + y} \log \frac{C}{y} - 1 \right)
$$

and the function $\{\rho/(\rho + y)\} \log(C/y) - 1$ is monotone decreasing, $u(y)$ attains a unique maximum at a $y^\star$ that satisfies $u'(y^\star) = 0$. This implies that

$$
u(y) \leq \frac{2\sigma_1^2 y^\star}{\rho} \frac{\rho}{\rho + y^\star} \log \frac{C}{y^\star} = \frac{2\sigma_1^2 y^\star}{\rho}
$$

for all $0 \leq y \leq C$. The exact value of $y^\star$ is difficult to determine exactly, by assumption $0 < \rho \leq 1$, so for every $n \geq 2\pi\sigma_1\sigma_2 e^{1/e}$, $\rho + \log C \geq 0$,

$\log\log C \geq -1$, and

$$u'(\log C) = \frac{2\sigma_2^2}{\rho + \log C}\frac{\rho\log C - \rho\log\log C - \rho - \log C}{\rho + \log C} \leq \frac{2\sigma_2^2}{\rho + \log C}\frac{(\rho-1)\log C}{\rho + \log C} \leq 0.$$

Since $u'(y^\star) = 0$, it follows that $y^\star \leq \log C$ and $u(y) \leq (2\sigma_1^2 \log C)/\rho$, or in other words,

$$\frac{\sum_j (t_{j1} - x_1)^2 p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_1, t_2)} \leq \frac{2\sigma_1^2}{\rho}\log\frac{n}{2\pi\sigma_1\sigma_2}. \tag{S4.3}$$

Furthermore, by Jensen's inequality

$$\left\{\frac{\sum_j (t_{j1} - x_1) p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_1, t_2)}\right\}^2 \leq \frac{\sum_j (t_{j1} - x_1)^2 p(x_1, x_2; t_{j1}, t_{j2})}{\rho + \sum_j p(x_1, x_2; t_1, t_2)}.$$

Therefore $|f(x_1, x_2, \mu; \mathbf{t})| \leq F_n(x_1, \mu)$ for the function

$$F_n(x_1, \mu) = \sigma_1^2 + \frac{4\sigma_1^2}{\rho}\log\frac{n}{2\pi\sigma_1\sigma_2} + \frac{2^{1/2}\sigma_1}{\rho^{1/2}}|x_1 - \mu|\log^{1/2}\frac{n}{2\pi\sigma_1\sigma_2}. \tag{S4.4}$$

The next step to truncate f (S4.1) in expression (S4.2), which will be useful when applying a maximal inequality later in the proof. Define

$$C_n = (2\sigma_1^2 \log\log n)^{1/2}. \tag{S4.5}$$

By the triangle inequality and the construction of the function F_n (S4.4),

$$E\sup_{\mathbf{t}\in\mathcal{T}}\left|\frac{1}{n}\sum_i \varepsilon_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t})\right| \leq E\sup_{\mathbf{t}\in\mathcal{T}}\left|\frac{1}{n}\sum_i \varepsilon_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) I(|X_{i1} - \theta_{i1}| \leq C_n)\right| + \frac{1}{n}\sum_i E\{F_n(X_{i1}, \theta_{i1}) I(|X_{i1} - \theta_{i1}| > C_n)\}.$$

For each i, $E\{F_n(X_{i1}, \theta_{i1}) I(|X_{i1} - \theta_{i1}| > C_n)\}$ equals

$$\left(\sigma_1^2 + \frac{4\sigma_1^2}{\rho} \log \frac{n}{2\pi\sigma_1\sigma_2}\right) P(|X_{i1} - \theta_{i1}| > C_n) +$$
$$\frac{2^{1/2}\sigma_1}{\rho^{1/2}} E\{|X_{i1} - \theta_{i1}| I(|X_{i1} - \theta_{i1}| > C_n)\} \log^{1/2} \frac{n}{2\pi\sigma_1\sigma_2}$$

By Mill's inequality,

$$P(|X_{i1} - \theta_{i1}| > C_n) = P(\sigma_1^{-1}|X_{i1} - \theta_{i1}| > \sigma_1^{-1} C_n) \leq \frac{2^{1/2}\sigma_1}{\pi^{1/2} C_n \log n}$$

and furthermore,

$$\begin{aligned} E\{|X_{i1} - \theta_{i1}| I(|X_{i1} - \theta_{i1}| > C_n)\} &= \int |z| I(|z| > C_n) \frac{1}{(2\pi\sigma_1^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_1^2} z^2\right) dz \\ &= -\frac{2\sigma_1}{(2\pi)^{1/2}} \int_{C_n}^{\infty} \left(-\frac{z}{\sigma_1^2}\right) \exp\left(-\frac{1}{2\sigma_1^2} z^2\right) dz \\ &= \frac{2\sigma_1}{(2\pi)^{1/2} \log n}. \end{aligned}$$

This implies that $n^{-1} \sum_i E\{F_n(X_{i1}, \theta_{i1}) I(|X_{i1} - \theta_{i1}| > C_n) = O(C_n^{-1}) \to 0$, and it remains to show that

$$E \sup_{\mathbf{t} \in \mathcal{T}} \left| \frac{1}{n} \sum_i \varepsilon_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) I(|X_{i1} - \theta_{i1}| \leq C_n) \right| \to 0. \qquad \text{(S4.6)}$$

S4.4 Approximation and covering number

For $Cn^{1/4-\eta}$ as in Assumption 1 and some integer K_n that can grow with n, define

$$\mathcal{U}(K_n) = \left\{(u_{11},\ldots,u_{K_n1},u_{12},\ldots,u_{K_n2}) : |u_{kd}| \le Cn^{1/4-\eta}, k=1,\ldots,K_n, d=1,2\right\},$$
$$\mathcal{W}(K_n) = \left\{(w_1,\ldots,w_{K_n}) : w_k > 0, \sum_k w_k = 1\right\}. \tag{S4.7}$$

Next define functions

$$\begin{aligned}
&g(x_1,x_2,\mu;\mathbf{u},\mathbf{w})\\
=&\frac{\sum_{k=1}^{K_n}(u_{k1}-x_1)^2 p(x_1,x_2;u_{k1},u_{k2})w_k}{\rho n^{-1}+\sum_{k=1}^{K_n}p(x_1,x_2;u_{k1},u_{k2})w_k} - \sigma_1^2\frac{\sum_{k=1}^{K_n}p(x_1,x_2;u_{k1},u_{k2})}{\rho n^{-1}+\sum_{k=1}^{K_n}p(x_1,x_2;u_{k1},u_{k2})w_k} -\\
&\left\{\frac{\sum_{k=1}^{K_n}(u_{k1}-x_1)p(x_1,x_2;u_{k1},u_{k2})w_k}{\rho n^{-1}+\sum_{k=1}^{K_n}p(x_1,x_2;u_{k1},u_{k2})w_k}\right\}^2 - |x-\mu|\frac{\sum_{k=1}^{K_n}(u_{k1}-x_1)p(x_1,x_2;u_{k1},u_{k2})w_k}{\rho n^{-1}+\sum_{k=1}^{K_n}p(x_1,x_2;u_{k1},u_{k2})w_k}
\end{aligned} \tag{S4.8}$$

indexed by $\mathbf{u} \in \mathcal{U}(K_n)$ and $\mathbf{w} \in \mathcal{W}(K_n)$. This section constructs discrete subsets $\mathbb{U}(K_n) \subset \mathcal{U}(K_n)$ and $\mathbb{W}(K_n) \subset \mathcal{W}(K_n)$ such that for all $\mathbf{t} \in \mathcal{T}$ (4.10), there exists some $(\mathbf{u},\mathbf{w}) \in \mathbb{U}(K_n) \times \mathbb{W}(K_n)$ such that conditional on the (X_{i1}, X_{i2}), for any $\epsilon > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} |f(X_{i1},X_{i2},\theta_{i1};\mathbf{t}) - g(X_{i1},X_{i2},\theta_{i1};\mathbf{u},\mathbf{w})| \le \epsilon \tag{S4.9}$$

for n sufficiently large, with C_n defined in (S4.5).

The ϵ-covering number $\mathcal{N}(\mathcal{T})$ of the set of functions $f(x_1,x_2,\mu;\mathbf{t})$ indexed by $\mathbf{t} \in \mathcal{T}$ is defined to be the cardinality of the smallest set $\mathbb{U}(K_n) \times$

$\mathbb{W}(K_n)$ such that (S4.9) holds. This section establishes an upper bound for this covering number, which will be a function of the (X_{i1}, X_{i2}) and K_n. The proof makes use of the following lemmas.

Lemma 1. *For any $\gamma_n \leq e^{-e}$ and any integer $q > 0$, define*

$$
\begin{aligned}
M_n &= \max_{j,d} \frac{|X_{jd}| + Cn^{1/4-\eta}}{\sigma_d}, \\
L_n &= \left(1 + \frac{M_n^2 e}{2}\right) \log \frac{1}{\gamma_n}, \\
K_n &= (2L_n - 2 + q)(2L_n - 2) + 1.
\end{aligned}
\tag{S4.10}
$$

Then for each $\mathbf{t} \in \mathcal{T}$ (4.10) there exist K_n support points $\boldsymbol{\mu} \in \mathcal{U}(K_n)$ and $\boldsymbol{\omega} \in \mathcal{W}(K_n)$, with $\mathcal{U}(K_n)$ and $\mathcal{W}(K_n)$ defined in (S4.7)*, such that for each (X_{i1}, X_{i2}),*

$$
\left| \frac{1}{n} \sum_{j=1}^{n} (t_{j1} - X_{i1})^q p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) - \sum_{k=1}^{K_n} (\mu_{k1} - X_{i1})^q p(X_{i1}, X_{i2}; \mu_{k1}, \mu_{k2}) \omega_k \right|
$$
$$
\leq \frac{\gamma_n + 2}{\pi \sigma_1 \sigma_2} \sigma_1^q M_n^q \gamma_n.
$$

Proof. By Taylor expansion and because $L_n! \geq L_n^{L_n} e^{-L_n}$,

$$
\left| \exp\left\{ -\frac{1}{2\sigma_d^2}(X_{id} - t_{jd})^2 \right\} - \sum_{l=0}^{L_n - 1} \frac{(-1)^l (X_{id} - t_{jd})^{2l}}{(2\sigma_d^2)^l l!} \right| \leq \frac{|X_{id} - t_{jd}|^{2L_n}}{(2\sigma_d^2)^{L_n} L_n!} \leq \left(\frac{|X_{id} - t_{jd}|^2 e}{2\sigma_d^2 L_n} \right)^{L_n}
$$

for $d = 1, 2$. By assumption, $\mathbf{t} \in \mathcal{T}$, so

$$
\frac{|X_{id} - t_{jd}|^2 e}{2\sigma_d^2 L_n} \leq \frac{M_n^2 e}{2L_n}.
$$

Then choosing L_n as in (S4.10) guarantees that

$$
\log \left(\frac{M_n^2 e}{2\sigma_d^2 L_n} \right)^{L_n} \leq \left(1 + \frac{M_n}{2}\right) \log \frac{1}{\gamma_n} \left(-\log \log \frac{1}{\gamma_n} \right) \leq \log \gamma_n,
$$

where the last inequality follows because since $\gamma_n < e^{-e}$ by assumption, $\log \gamma_n < 0$ while $\log \log \gamma_n^{-1} \geq 1$. Therefore

$$\left| \exp\left\{ -\frac{1}{2\sigma_d^2}(X_{id} - t_{jd})^2 \right\} - \sum_{l=0}^{L_n - 1} \frac{(-1)^l (X_{id} - t_{jd})^{2l}}{(2\sigma_d^2)^l l!} \right| \leq \gamma_n,$$

which also implies that

$$\left| \sum_{l=0}^{L_n - 1} \frac{(-1)^l (X_{id} - t_{jd})^{2l}}{(2\sigma_d^2)^l l!} \right| \leq \gamma_n + 1.$$

This imply that

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{j=1}^{n} (t_{j1} - X_{i1})^q p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) - \frac{1}{n}\sum_{j=1}^{n} \frac{(t_{j1} - X_{i1})^q}{2\pi\sigma_1\sigma_2} \prod_{d=1}^{2} \sum_{l=0}^{L_n - 1} \frac{(-1)^l (X_{id} - t_{jd})^{2l}}{(2\sigma_d^2)^l l!} \right| \\
\leq& \frac{1}{n}\sum_{j=1}^{n} \frac{|t_{j1} - X_{i1}|^q}{2\pi\sigma_1\sigma_2} \left| \exp\left\{ -\frac{1}{2\sigma_1^2}(X_{i1} - t_{j1})^2 \right\} - \sum_{l=0}^{L_n - 1} \frac{(-1)^l (X_{i1} - t_{j1})^{2l}}{(2\sigma_1^2)^l l!} \right| + \\
&\frac{1}{n}\sum_{j=1}^{n} \frac{|t_{j1} - X_{i1}|^q (\gamma_n + 1)}{2\pi\sigma_1\sigma_2} \left| \exp\left\{ -\frac{1}{2\sigma_2^2}(X_{i2} - t_{j2})^2 \right\} - \sum_{l=0}^{L_n - 1} \frac{(-1)^l (X_{i2} - t_{j2})^{2l}}{(2\sigma_2^2)^l l!} \right| \\
\leq& \frac{\gamma_n + 2}{2\pi\sigma_1\sigma_2} \sigma_1^q M_n^q \gamma_n. \qquad \text{(S4.11)}
\end{aligned}$$

Next, by binomial expansion,

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n} \frac{(t_{j1} - X_{i1})^q}{2\pi\sigma_1\sigma_2} \prod_{d=1}^{2} \sum_{l=0}^{L_n - 1} \frac{(-1)^l (X_{id} - t_{jd})^{2l}}{(2\sigma_d^2)^l l!} = \frac{1}{2\pi\sigma_1\sigma_2} \times \\
&\sum_{\ell, \ell' = 0}^{L_n - 1} \frac{(-1)^\ell (-1)^{\ell'}}{(2\sigma_1^2)^\ell \ell! (2\sigma_2^2)^{\ell'} \ell'!} \sum_{m=0}^{2\ell + q} \sum_{m'=0}^{2\ell'} \binom{2\ell + q}{m} (-X_{i1})^{2\ell + q - m} \binom{2\ell'}{m'} (-X_{i2})^{2\ell' - m'} \sum_{j=1}^{n} t_{j1}^m t_{j2}^{m'} n^{-1}.
\end{aligned}$$

Lemma A.1 of Ghosal and Van Der Vaart (2001) can be used to show that for K_n as in (S4.10), there exist $\boldsymbol{\mu} \in \mathcal{U}(K_n)$ and $\boldsymbol{\omega} \in \mathcal{W}(K_n)$ (S4.7) such that $\sum_{j=1}^{n} t_{j1}^m t_{j2}^{m'} n = \sum_{k=1}^{K_n} \mu_{k1}^m \mu_{k2}^{m'} \omega_k$ for every $0 \leq m \leq 2\ell + q$ and

$0 \le m' \le 2\ell'$. Therefore

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\frac{(t_{j1}-X_{i1})^q}{2\pi\sigma_1\sigma_2}\prod_{d=1}^{2}\sum_{l=0}^{L_n-1}\frac{(-1)^l(X_{id}-t_{jd})^{2l}}{(2\sigma_d^2)^l l!} \\
&=\sum_{k=1}^{K_n}\frac{(\mu_{k1}-X_{i1})^q}{2\pi\sigma_1\sigma_2}\prod_{d=1}^{2}\sum_{l=0}^{L_n-1}\frac{(-1)^l(X_{id}-u_{kd})^{2l}}{(2\sigma_d^2)^l l!}\omega_k.
\end{aligned} \tag{S4.12}$$

Finally, by the same reasoning that led to (S4.11),

$$\begin{aligned}
&\left|\sum_{k=1}^{K_n}\frac{(\mu_{k1}-X_{i1})^q}{2\pi\sigma_1\sigma_2}\prod_{d=1}^{2}\sum_{l=0}^{L_n-1}\frac{(-1)^l(X_{id}-u_{kd})^{2l}}{(2\sigma_d^2)^l l!}\omega_k-\sum_{k=1}^{K_n}(u_{k1}-X_{i1})^q p(X_{i1},X_{i2};u_{k1},u_{k2})\omega_k\right| \\
&\le\frac{\gamma_n+2}{2\pi\sigma_1\sigma_2}\sigma_1^q M_n^q\gamma_n
\end{aligned} \tag{S4.13}$$

Combining (S4.11), (S4.12), and (S4.13) concludes the proof. $\square$

Next, for any $\gamma_n > 0$, define the discrete sets

$$\begin{aligned}
&\mathbb{U}(K_n)\subset\mathcal{U}(K_n)\text{ such that for any }\boldsymbol{\mu}\in\mathcal{U}(K_n),\ \inf_{\mathbf{u}\in\mathbb{U}(K_n)}\max_{k,d}|u_{kd}-\mu_{kd}|\le\gamma_n, \\
&\mathbb{W}(K_n)\subset\mathcal{W}(K_n)\text{ such that for any }\boldsymbol{\omega}\in\mathcal{W}(K_n),\ \inf_{\mathbf{w}\in\mathbb{W}(K_n)}\sum_{k=1}^{K_n}|w_k-\omega_k|\le\gamma_n.
\end{aligned} \tag{S4.14}$$

The set $\mathbb{U}(K_n)$ can be constructed by specifying a grid of equally-spaced points on the interval $[-Cn^{1/4-\eta}, Cn^{1/4-\eta}]$, where any two neighboring points are separated by at most a distance γ_n, for each of the $2K_n$ dimensions of the vector $\boldsymbol{\mu} \in \mathcal{U}(K_n)$. Therefore the cardinality of the smallest possible $\mathbb{U}(K_n)$ satisfies

$$|\mathbb{U}(K_n)| \le \left(\lceil 2Cn^{1/4-\eta}/\gamma_n\rceil + 1\right)^{2K_n}. \tag{S4.15}$$

Arguments from Jiang et al. (2009) and Zhang (2009) imply that the cardinality of the smallest possible $\mathbb{W}(K_n)$ satifies

$$|\mathbb{W}(K_n)| \leq (2/\gamma_n + 1)^{K_n}. \tag{S4.16}$$

Lemma 2. *For any* $(\boldsymbol{\mu}, \boldsymbol{\omega}) \in \mathcal{U}(K_n) \times \mathcal{W}(K_n)$ (S4.7) *and any integer* $q > 0$, *there exists* $(\mathbf{u}, \mathbf{w}) \in \mathbb{U}(K_n) \times \mathbb{W}(K_n)$ (S4.14) *and a constant* D_q *such that for each* (X_{i1}, X_{i2}),

$$\left|\sum_{k=1}^{K_n}(\mu_{k1} - X_{i1})^q p(X_{i1}, X_{i2}; \mu_{k1}, \mu_{k2})\omega_k - \sum_{k=1}^{K_n}(u_{k1} - X_{i1})^q p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k\right|$$
$$\leq \left(M_n^q + \frac{D_q}{\sigma_1} + \frac{D_q M_n^q}{\sigma_2}\right)\frac{\sigma_1^q}{2\pi\sigma_1\sigma_2}\gamma_n.$$

Proof. For any $(\boldsymbol{\mu}, \boldsymbol{\omega}) \in \mathcal{U}(K_n) \times \mathcal{W}(K_n)$, by construction of $\mathbb{W}(K_n)$ (S4.14) there exists a $\mathbf{w} \in \mathbb{W}(K_n)$ such that

$$\left|\sum_{k=1}^{K_n}(\mu_{k1} - X_{i1})^q p(X_{i1}, X_{i2}; \mu_{k1}, \mu_{k2})(\omega_k - w_k)\right| \leq \frac{\sigma_1^q M_n^q}{2\pi\sigma_1\sigma_2}\gamma_n, \tag{S4.17}$$

and for any $\mathbf{u} \in \mathcal{U}(K_n)$,

$$\left|\sum_{k=1}^{K_n}\{(\mu_{k1} - X_{i1})^q p(X_{i1}, X_{i2}; \mu_{k1}, \mu_{k2}) - (u_{k1} - X_{i1})^q p(X_{i1}, X_{i2}; u_{k1}, u_{k2})\} w_k\right|$$
$$\leq \sum_{k=1}^{K_n}\left|\frac{(\mu_{k1} - X_{i1})^q}{\sigma_1^q}\exp\left\{-\frac{(X_{i1} - \mu_{k1})^2}{2\sigma_1^2}\right\} - \frac{(u_{k1} - X_{i1})^q}{\sigma^q}\exp\left\{-\frac{(X_{i1} - u_{k1})^2}{2\sigma_1^2}\right\}\right|\frac{\sigma_1^q w_k}{2\pi\sigma_1\sigma_2} +$$
$$\sum_{k=1}^{K_n}\left|\exp\left\{-\frac{(X_{i2} - \mu_{k2})^2}{2\sigma_2^2}\right\} - \exp\left\{-\frac{(X_{i2} - u_{k2})^2}{2\sigma_2^2}\right\}\right|\frac{\sigma_1^q M_n^q w_k}{2\pi\sigma_1\sigma_2}.$$

Define the function

$$h_q(x) = x^q \exp(-x^2/2)$$

for $q \geq 0$. Now if $\mathbf{u}$ is also in $\mathbb{U}(K_n)$ (S4.14), then by construction

$$
\begin{aligned}
&\left|\sum_{k=1}^{K_n}\{(\mu_{k1}-X_{i1})^q p(X_{i1},X_{i2};\mu_{k1},\mu_{k2})-(u_{k1}-X_{i1})^q p(X_{i1},X_{i2};u_{k1},u_{k2})\}w_k\right| \\
&\leq \sum_{k=1}^{K_n}\sup_x |h_q'(x)|\left|\frac{\mu_{k1}-X_{i1}}{\sigma_1}-\frac{u_{k1}-X_{i1}}{\sigma_1}\right|\frac{\sigma_1^q w_k}{2\pi\sigma_1\sigma_2}+ \\
&\quad \sum_{k=1}^{K_n}\sup_x |h_0'(x)|\left|\frac{\mu_{k2}-X_{i2}}{\sigma_2}-\frac{u_{k2}-X_{i2}}{\sigma_2}\right|\frac{\sigma_1^q M_n^q w_k}{2\pi\sigma_1\sigma_2} \\
&\leq \sup_x |h_q'(x)|\frac{\sigma_1^q}{2\pi\sigma_1^2\sigma_2}\gamma_n+\sup_x |h_0'(x)|\frac{\sigma_1^q M_n^q}{2\pi\sigma_1\sigma_2^2}\gamma_n.
\end{aligned}
$$

Since

$$
\begin{aligned}
h_q'(x) &= (q-x^2)x^{q-1}\exp(-x^2/2), \\
h_q''(x) &= \{q(q-1)-(2q+1)x^2+x^4\}x^{q-2}\exp(-x^2/2),
\end{aligned}
$$

the first derivative of $\{h_q'(x)\}^2$ is given by

$$
\begin{aligned}
&\frac{d}{dx}\{h_q'(x)\}^2 = 2h_q'(x)h_q''(x) \\
&= (q-x^2)\left\{x^2-\frac{2q+1+(8q+1)^{1/2}}{2}\right\}\left\{x^2-\frac{2q+1-(8q+1)^{1/2}}{2}\right\}x^{2q-3}\exp(-x^2/2).
\end{aligned}
$$

Therefore $\{h_q'(x)\}^2$ has at most seven stationary points, is increasing for x smaller than the smallest negative stationary point, and is decreasing for x larger than the largest positive stationary point. Therefore the maximum value of $|h_q'(x)|$ can only occur at one of the seven stationary points, at which $|h_q'(x)|$ is finite. Therefore there is some constant D_q such that

$$\left|\sum_{k=1}^{K_n}\{(\mu_{k1}-X_{i1})^q p(X_{i1},X_{i2};\mu_{k1},\mu_{k2})-(u_{k1}-X_{i1})^q p(X_{i1},X_{i2};u_{k1},u_{k2})\}w_k\right|$$
$$\leq D_q\left(\frac{\sigma_1^q}{2\pi\sigma_1^2\sigma_2}+\frac{\sigma_1^q M_n^q}{2\pi\sigma_1\sigma_2^2}\right)\gamma_n. \tag{S4.18}$$

Combining (S4.17) and (S4.18) concludes the proof. □

For K_n (S4.10) large enough, the result (S4.9), that f (S4.1) can be in a sense approximated by g (S4.8), can now be established using Lemmas 1 and 2. For each (X_{i1}, X_{i2}),

$$|f(X_{i1},X_{i2},\theta_{i1};\mathbf{t})-g(X_{i1},X_{i2},\theta_{i1};\mathbf{u},\mathbf{w})|\leq$$
$$\left|\frac{\sum_j^n(t_{j1}-X_{i1})^2p(X_{i1},X_{i2};t_{j1},t_{j2})n^{-1}}{\rho n^{-1}+\sum_j^n p(X_{i1},X_{i2};t_1,t_2)n^{-1}}-\frac{\sum_k^{K_n}(u_{k1}-X_{i1})^2p(X_{i1},X_{i2};u_{k1},u_{k2})w_k}{\rho n^{-1}+\sum_k^{K_n}p(X_{i1},X_{i2};u_{k1},u_{k2})w_k}\right|+ \tag{S4.19}$$

$$\sigma_1^2\left|\frac{\sum_j^n p(X_{i1},X_{i2};t_{j1},t_{j2})n^{-1}}{\rho n^{-1}+\sum_j^n p(X_{i1},X_{i2};t_{j1},t_{j2})n^{-1}}-\frac{\sum_k^{K_n}p(X_{i1},X_{i2};u_{k1},u_{k2})}{\rho n^{-1}+\sum_k^{K_n}p(X_{i1},X_{i2};u_{k1},u_{k2})w_k}\right|+ \tag{S4.20}$$

$$\left|\left\{\frac{\sum_j^n(t_{1j}-X_{i1})p(X_{i1},X_{i2};t_{j1},t_{j2})n^{-1}}{\rho n^{-1}+\sum_j^n p(X_{i1},X_{i2};t_{j1},t_{j2})n^{-1}}\right\}^2-\left\{\frac{\sum_k^{K_n}(u_{k1}-X_{i1})p(X_{i1},X_{i2};u_{k1},u_{k2})w_k}{\rho n^{-1}+\sum_k^{K_n}p(X_{i1},X_{i2};u_{k1},u_{k2})w_k}\right\}^2\right|+ \tag{S4.21}$$

$$|X_{i1}-\theta_{i1}|\left|\frac{\sum_j^n(t_{1j}-X_{i1})p(X_{i1},X_{i2};t_{j1},t_{j2})n^{-1}}{\rho n^{-1}+\sum_j^n p(X_{i1},X_{i2};t_{j1},t_{j2})n^{-1}}-\frac{\sum_k^{K_n}(u_{k1}-X_{i1})p(X_{i1},X_{i2};u_{k1},u_{k2})w_k}{\rho n^{-1}+\sum_k^{K_n}p(X_{i1},X_{i2};u_{k1},u_{k2})w_k}\right|. \tag{S4.22}$$

The terms (S4.19)–(S4.22) can be bounded using (S4.3) and Lemmas 1 and 2. For L_n as defined in (S4.10) and

$$K_n = 4L_n^2 + 1, \tag{S4.23}$$

there exist $(\mathbf{u}, \mathbf{w}) \in \mathbb{U}(K_n) \times \mathbb{W}(K_n)$ (S4.14) such that for any $\gamma_n \leq e^{-e}$,

$$\begin{aligned}
\text{(S4.19)} \leq & \left| \frac{\sum_j^n (t_{j1} - X_{i1})^2 p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_1, t_2) n^{-1}} - \frac{\sum_k^{K_n} (u_{k1} - X_{i1})^2 p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} \right| + \\
& \left| \frac{\sum_k^{K_n} (u_{k1} - X_{i1})^2 p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} - \frac{\sum_k^{K_n} (u_{k1} - X_{i1})^2 p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k} \right| \\
\leq & \frac{n}{\rho} \left| \sum_j^n (t_{j1} - X_{i1})^2 p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1} - \sum_k^{K_n} (u_{k1} - X_{i1})^2 p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k \right| + \\
& \frac{n}{\rho} \frac{\sum_k^{K_n} (u_{k1} - X_{i1})^2 p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k} \times \\
& \left| \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1} - \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k \right|,
\end{aligned}$$

so by Lemmas 1 and 2,

$$\begin{aligned}
\text{(S4.19)} \leq & \frac{n}{\rho} \left\{ \frac{(\gamma_n + 2)\sigma_1^2 M_n^2}{\pi \sigma_1 \sigma_2} + \left(M_n^2 + \frac{D_2}{\sigma_1} + \frac{D_2 M_n^2}{\sigma_2} \right) \frac{\sigma_1^2}{2\pi \sigma_1 \sigma_2} \right\} \gamma_n + \\
& \frac{2\sigma_1^2 n}{\rho^2} \log \frac{n}{2\pi \sigma_1 \sigma_2} \left\{ \frac{\gamma_n + 2}{\pi \sigma_1 \sigma_2} + \left(1 + \frac{D_0}{\sigma_1} + \frac{D_0}{\sigma_2} \right) \frac{1}{2\pi \sigma_1 \sigma_2} \right\} \gamma_n.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
(\text{S4.20}) \le\ & \sigma_1^2 \left| \frac{\sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} - \frac{\sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} \right| + \\
& \sigma_1^2 \left| \frac{\sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} - \frac{\sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k} \right| \\
\le\ & \frac{\sigma_1^2 n}{\rho} \left| \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1} - \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k \right| + \\
& \frac{\sigma_1^2 n}{\rho} \frac{\sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k} \times \\
& \left| \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1} - \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k \right| \\
\le\ & \frac{2\sigma_1^2 n}{\rho} \left\{ \frac{\gamma_n + 2}{\pi \sigma_1 \sigma_2} + \left(1 + \frac{D_0}{\sigma_1} + \frac{D_0}{\sigma_2} \right) \frac{1}{2\pi \sigma_1 \sigma_2} \right\} \gamma_n.
\end{aligned}
$$

Next, using Jensen's inequality and the bound (S4.3),

$$
\begin{aligned}
(\text{S4.21}) =\ & \left| \frac{\sum_j^n (t_{1j} - X_{i1}) p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} + \frac{\sum_k^{K_n} (u_{k1} - X_{i1}) p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k} \right| \times \\
& \left| \frac{\sum_j^n (t_{1j} - X_{i1}) p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} - \frac{\sum_k^{K_n} (u_{k1} - X_{i1}) p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k} \right| \\
\le\ & 2 \left(\frac{2\sigma_1^2}{\rho} \log \frac{n}{2\pi \sigma_1 \sigma_2} \right)^{1/2} \times \\
& \left| \frac{\sum_j^n (t_{1j} - X_{i1}) p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}}{\rho n^{-1} + \sum_j^n p(X_{i1}, X_{i2}; t_{j1}, t_{j2}) n^{-1}} - \frac{\sum_k^{K_n} (u_{k1} - X_{i1}) p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k}{\rho n^{-1} + \sum_k^{K_n} p(X_{i1}, X_{i2}; u_{k1}, u_{k2}) w_k} \right|.
\end{aligned}
$$

By similar reasoning as in the bounding of (S4.19), it follows that

$$
\begin{aligned}
(\text{S4.21}) \le\ & 2 \left(\frac{2\sigma_1^2}{\rho} \log \frac{n}{2\pi \sigma_1 \sigma_2} \right)^{1/2} \frac{n}{\rho} \left\{ \frac{(\gamma_n + 2) \sigma_1 M_n}{\pi \sigma_1 \sigma_2} + \left(M_n + \frac{D_1}{\sigma_1} + \frac{D_1 M_n}{\sigma_2} \right) \frac{\sigma_1^1}{2\pi \sigma_1 \sigma_2} \right\} \gamma_n + \\
& \frac{4\sigma_1^2 n}{\rho^2} \log \frac{n}{2\pi \sigma_1 \sigma_2} \left\{ \frac{\gamma_n + 2}{\pi \sigma_1 \sigma_2} + \left(1 + \frac{D_0}{\sigma_1} + \frac{D_0}{\sigma_2} \right) \frac{1}{2\pi \sigma_1 \sigma_2} \right\} \gamma_n,
\end{aligned}
$$

and that

$$
\begin{aligned}
(\text{S4.22}) \leq \frac{\sigma_1 M_n n}{\rho} & \left\{ \frac{(\gamma_n + 2)\sigma_1 M_n}{\pi \sigma_1 \sigma_2} + \left(M_n + \frac{D_1}{\sigma_1} + \frac{D_1 M_n}{\sigma_2} \right) \frac{\sigma_1^1}{2\pi\sigma_1\sigma_2} \right\} \gamma_n + \\
& \frac{\sigma_1 M_n n}{\rho} \left(\frac{2\sigma_1^2}{\rho} \log \frac{n}{2\pi\sigma_1\sigma_2} \right)^{1/2} \left\{ \frac{\gamma_n + 2}{\pi\sigma_1\sigma_2} + \left(1 + \frac{D_0}{\sigma_1} + \frac{D_0}{\sigma_2} \right) \frac{1}{2\pi\sigma_1\sigma_2} \right\} \gamma_n.
\end{aligned}
$$

Finally, choose

$$
\gamma_n = \{n(M_n + 1)^2 \log^2 n\}^{-1}, \tag{S4.24}
$$

so that as long as $n \log^2 n > e^e$, $\gamma_n \leq n^{-e}$ and Lemmas 1 and 2 can be applied to obtain the bounds on (S4.19)–(S4.22). These bounds then imply that for C_n as in (S4.5), M_n and L_n as in (S4.10), K_n as in (S4.23), and $\mathbb{U}(K_n)$ and $\mathbb{W}(K_n)$ constructed as in (S4.14), for any $\mathbf{t} \in \mathcal{T}$ (4.10) there exists some $(\mathbf{u}, \mathbf{w}) \in \mathbb{U}(K_n) \times \mathbb{W}(K_n)$ such that conditional on the (X_{i1}, X_{i2}),

$$
\frac{1}{n} \sum_{i=1}^{n} |f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) - g(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{u}, \mathbf{w})| \leq O\left\{ n(M_n + 1)^2 \log n \right\} \gamma_n.
$$

With γ_n as in (S4.24), for any $\epsilon > 0$ there exists a sufficiently large n such that the approximation result (S4.9) holds. The cardinality bounds (S4.15) and (S4.16) imply that the ϵ-covering number $\mathcal{N}(\mathcal{T})$ of the $f(x_1, x_2, \mu; \mathbf{t})$ indexed by $\mathbf{t} \in \mathcal{T}$ obeys

$$
\begin{aligned}
\log \mathcal{N}(\mathcal{T}) & \leq 2K_n \log(2Cn^{1/4-\eta}/\gamma_n + 1) + K_n \log(2/\gamma_n + 1) \\
& \leq K_n \log C_1 + K_n \log(n^{1/2-2\eta}/\gamma_n^3),
\end{aligned}
$$

where C and η are defined in Assumption 1 and C_1 and C_2 are positive constants. Since $K_n = 4L_n^2 + 1$ from (S4.23) and $L_n = (1 + M_n^2 e/2)\log(1/\gamma_n) > 1$ from (S4.10), $K_n \le 5L_n^2$ and there is a constant C_2 such that

$$\begin{aligned}\log \mathcal{N}(\mathcal{T}) \le\, & C_2(M_n+1)^4 \log^2\{n(M_n+1)^2 \log^2 n\}\times \\ & \left[\log C_1 + \log\{n^{1/2-2\eta} n^3 (M_n+1)^6 \log^6 n\}\right].\end{aligned} \tag{S4.25}$$

S4.5 Maximal inequality

This section concludes the proof of Theorem 2 by showing (S4.6). First, the supremum over the infinite set $\mathcal{T}$ can be converted to a maximum over the discrete set $\mathbb{U}(K_n) \times \mathbb{W}(K_n)$ constructed in (S4.14). For any $\mathbf{t} \in \mathcal{T}$, using the approximation result (S4.9) there exists a $(\mathbf{u}, \mathbf{w})$ inside a suitably constructed set $\mathbb{U}(K_n) \times \mathbb{W}(K_n)$ (S4.14) such that for any $\epsilon > 0$, there exists a sufficiently large n such that

$$\begin{aligned}& \left|\frac{1}{n}\sum_i \varepsilon_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) I(|X_{i1} - \theta_{i1}| \le C_n)\right| \\ \le & \left|\frac{1}{n}\sum_i \varepsilon_i g(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{u}, \mathbf{w}) I(|X_{i1} - \theta_{i1}| \le C_n)\right| + \epsilon,\end{aligned}$$

where ε_i are Rademacher variables and $C_n = (2\sigma_1^2 \log\log n)^{1/2}$ from (S4.5). Since $\mathbb{U}(K_n) \times \mathbb{W}(K_n)$ is discrete, it follows that

$$\begin{aligned}& E \sup_{\mathbf{t}\in\mathcal{T}} \left|\frac{1}{n}\sum_i \varepsilon_i f(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{t}) I(|X_{i1} - \theta_{i1}| \le C_n)\right| \\ \le\, & E \max_{(\mathbf{u},\mathbf{w})\in\mathbb{U}(K_n)\times\mathbb{W}(K_n)} \left|\frac{1}{n}\sum_i \varepsilon_i g(X_{i1}, X_{i2}, \theta_{i1}; \mathbf{u}, \mathbf{w}) I(|X_{i1} - \theta_{i1}| \le C_n)\right| + \epsilon.\end{aligned}$$

To bound the expectation of the maximum over $\mathbb{U}(K_n)\times\mathbb{W}(K_n)$, define the convex function $\psi(x)=\exp(x^2)-1$. Using Lemma 2.2.2 of van der Vaart and Wellner (1996), it can be shown that there is a constant A such that

$$\begin{aligned}&E_\varepsilon \max_{(\mathbf{u},\mathbf{w})\in\mathbb{U}(K_n)\times\mathbb{W}(K_n)} \left|\frac{1}{n}\sum_i \varepsilon_i g(X_{i1},X_{i2},\theta_{i1};\mathbf{u},\mathbf{w})I(|X_{i1}-\theta_{i1}|\leq C_n)\right|\\ &\leq A\{1+\log\mathcal{N}(\mathcal{T})\}^{1/2}\max_{(\mathbf{u},\mathbf{w})\in\mathbb{U}(K_n)\times\mathbb{W}(K_n)}\left\|\frac{1}{n}\sum_i \varepsilon_i g(X_{i1},X_{i2},\theta_{i1};\mathbf{u},\mathbf{w})I(|X_{i1}-\theta_{i1}|\leq C_n)\right\|_\psi,\end{aligned}$$

where E_ε is the expectation over the ε_i conditional on the (X_{i1},X_{i2}) and $\|X\|_\psi$ is the Orlicz norm of the random variable X with respect to the function $\psi(x)$. It is straightforward to show that $|g(x_1,x_2,\mu;\mathbf{u},\mathbf{w})|\leq F_n(x_1,\mu)$ for the function F_n defined in (S4.4), so by Hoeffding's inequality and Lemma 2.2.7 of van der Vaart and Wellner (1996),

$$\begin{aligned}&\left\|\frac{1}{n}\sum_i \varepsilon_i g(X_{i1},X_{i2},\theta_{i1};\mathbf{u},\mathbf{w})I(|X_{i1}-\theta_{i1}|\leq C_n)\right\|_\psi\\ &\leq\left(\frac{6}{n}\right)^{1/2}\left(\sigma_1^2+\frac{4\sigma_1^2}{\rho}\log\frac{n}{2\pi\sigma_1\sigma_2}+\frac{2^{1/2}\sigma_1}{\rho^{1/2}}C_n\log^{1/2}\frac{n}{2\pi\sigma_1\sigma_2}\right)=O\left(\frac{\log n}{n^{1/2}}\right)\end{aligned}$$

for any $(\mathbf{u},\mathbf{w})\in\mathbb{U}(K_n)\times\mathbb{W}(K_n)$. Therefore, using Jensen's inequality,

$$E\sup_{\mathbf{t}\in\mathcal{T}}\left|\frac{1}{n}\sum_i \varepsilon_i f(X_{i1},X_{i2},\theta_{i1};\mathbf{t})I(|X_{i1}-\theta_{i1}|\leq C_n)\right|\leq\left[A\{1+E\log\mathcal{N}(\mathcal{T})\}^{1/2}\right]O\left(\frac{\log n}{n^{1/2}}\right).$$

To show (S4.6), it remains to show that $E\log\mathcal{N}(\mathcal{T})=o(n/\log^2 n)$, where the expectation is now over the (X_{i1},X_{i2}). From (S4.25) and the inequality $\log(x)\leq mx^{1/m}$ for any $x>0$ and integer $m>0$, for n sufficiently large there is some constant A_2 such that

$$E\log\mathcal{N}(\mathcal{T}) \le A_2(n^{11/2m-2\eta/m}\log^{10/m} n)E\{(M_n+1)^{4+10/m}\}.$$

By the definition of M_n (S4.10) and the bounds on θ_{id} from Assumption 1,

$$M_n = \max_{j,d}\frac{|X_{jd}|+Cn^{1/4-\eta}}{\sigma_d} \le \max_{j,d}\left|\frac{X_{jd}-\theta_{jd}}{\sigma_d}\right| + \frac{2Cn^{1/4-\eta}}{\sigma_1\wedge\sigma_2}.$$

Therefore letting Z_i, $i=1,\dots,2n$ denote independent standard normal random variables,

$$\begin{aligned} E\{(M_n+1)^{4+10/m}\} \le\, & E\left\{(M_n+1)^{4+10/m} I\left(\max_{j,d}\left|\frac{X_{jd}-\theta_{jd}}{\sigma_d}\right| \le \frac{2Cn^{1/4-\eta}}{\sigma_1\wedge\sigma_2}+1\right)\right\} + \\ & E\left\{(M_n+1)^{4+10/m} I\left(\max_{j,d}\left|\frac{X_{jd}-\theta_{jd}}{\sigma_d}\right| > \frac{2Cn^{1/4-\eta}}{\sigma_1\wedge\sigma_2}+1\right)\right\} \\ \le\, & \left(\frac{4Cn^{1/4-\eta}}{\sigma_1\wedge\sigma_2}+2\right)^{4+10/m} + 2^{4+10/m}E(\max_{i=1,\dots,2n}|Z_i|^{4+10/m}). \end{aligned}$$

Now define the function $\psi(x) = x^{p/(4+10/m)}$, which is convex for $p \ge 4 + 10/m$. Then by Lemma 2.2.2 of van der Vaart and Wellner (1996), there is some constant A_3 such that

$$E(\max_i |Z_i|^{4+10/m}) \le \psi^{-1}(1)\left\|\max_{i=1,\dots,2n}|Z_i|^{4+10/m}\right\|_\psi \le A_3(2n)^{(4+10/m)/p}\max_{i=1,\dots,2n}\left\||Z_i|^{4+10/m}\right\|_\psi,$$

where $\|X\|_\psi$ is the Orlicz norm of X with respect to the function $\psi(x)$. But

$$\left\||Z_i|^{4+10/m}\right\|_\psi = E(|Z_i|^p)^{(4+10/m)/p},$$

which is a constant that does not depend on n, so choosing p such that $p \ge 5$ and $1/p \le 1/4-\eta$ gives

$$E\{(M_n+1)^{4+10/m}\} \le O(n^{1-4\eta+10/(4m)-12\eta/m}).$$

Then

$$E \log \mathcal{N}(\mathcal{T}) \leq O(n^{1-4\eta+32/(4m)-12\eta/m} \log^{10/m} n).$$

Choosing m large enough such that $4\eta - 32/(4m) + 12\eta/m > 0$ implies that $E \log \mathcal{N}(\mathcal{T}) = o(n/log^2 n)$, concluding the proof. $\square$

S5 Proof of Theorem 3

Let $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{n1}, \theta_{12}, \dots, \theta_{n2})$. Then when $\mathbf{t} = \boldsymbol{\theta}$, the estimator $\boldsymbol{\delta}^{\mathbf{t}}_{\rho}$ (4.7) equals the oracle regularized estimator $\boldsymbol{\delta}^{\star}_{\rho}$. Therefore with $\textsc{sure}(\mathbf{t})$ defined as in (4.8),

$$\begin{aligned} E\ell_n(\hat{\mathbf{t}}) - R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^{\star}) &= E\ell_n(\hat{\mathbf{t}}) - E\textsc{sure}(\hat{\mathbf{t}}) + E\textsc{sure}(\hat{\mathbf{t}}) - E\textsc{sure}(\theta) + E\textsc{sure}(\theta) - R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^{\star}) \\ &\leq |E\{\ell_n(\hat{\mathbf{t}}) - \textsc{sure}(\hat{\mathbf{t}})\}| + R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^{\star}_{\rho}) - R_n(\boldsymbol{\theta}, \boldsymbol{\delta}^{\star}) \\ &\leq E|\ell_n(\hat{\mathbf{t}}) - \textsc{sure}(\hat{\mathbf{t}})| + o(1), \end{aligned}$$

where the second line follows because $E\textsc{sure}(\hat{\mathbf{t}}) \leq E\textsc{sure}(\theta)$, since $\hat{\mathbf{t}}$ (4.11) is defined to minimize $\textsc{sure}(\mathbf{t})$, and the third line follows by Theorem 1. Finally, since $\hat{\mathbf{t}} \in \mathcal{T}$ by construction, by Theorem 2,

$$E|\ell_n(\hat{\mathbf{t}}) - \textsc{sure}(\hat{\mathbf{t}})| \leq E \sup_{\mathbf{t} \in \mathcal{T}} |\ell_n(\mathbf{t}) - \textsc{sure}(\mathbf{t})| = o(1).$$

Bibliography

Ghosal, S. and A. W. Van Der Vaart (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Annals of Statistics 29*(5), 1233–1263.

Jiang, W., C.-H. Zhang, et al. (2009). General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics 37*(4), 1647–1684.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics 9*, 1135–1151.

van der Vaart, A. W. and J. Wellner (1996). *Weak convergence and empirical processes: with applications to statistics*. New York: Springer Science+Business Media.

Zhang, C.-H. (2009). Generalized maximum likelihood estimation of normal mixture densities. *Statistica Sinica 19*, 1297–1318.