High-dimensional Multivariate Mediation with Application to Neuroimaging Data

Supplementary Materials

Oliver Y. Chén[1], Elizabeth L. Ogburn[1], Ciprian M. Crainiceanu[1],
Brian S. Caffo[1], Tor D. Wager[2], Martin A. Lindquist[1]

[1] Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

[2] Department of Psychology and Neuroscience
University of Colorado Boulder

Proofs

Lemma 1 (Consistency Theorem): Suppose that $Q(\boldsymbol{\theta}; \mathbf{Z}_n)$ is continuous in $\boldsymbol{\theta}$ and there exists a function $Q_0(\boldsymbol{\theta})$ such that: $Q_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$; $\boldsymbol{\Theta}$ is compact; $Q_0(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$; and $Q(\boldsymbol{\theta}; \mathbf{Z}_n)$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$. Then $\hat{\boldsymbol{\theta}}(\mathbf{Z}_n)$ defined as the value of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ which for each $\mathbf{Z}_n = \mathbf{z}_n$ maximizes the objective function $Q(\boldsymbol{\theta}; \mathbf{Z}_n)$ satisfies $\hat{\boldsymbol{\theta}}(\mathbf{Z}_n) \xrightarrow{p} \theta_0$.

PROOF. See Theorem 2.1 in (Newey and McFadden, 1994). ■

Lemma 2: Consider a compact space $\boldsymbol{\Theta}$. Let

$$L(z, \mathbf{w}, \boldsymbol{\theta}, \lambda) = f^{\text{obj}}(z, \mathbf{w}, \boldsymbol{\theta}) + f^{\text{pen}}(z, \mathbf{w}, \lambda),$$

where $f^{\text{obj}}(z, \mathbf{w}, \boldsymbol{\theta})$ is an objective function and $f^{\text{pen}}(z, \mathbf{w}, \lambda) = \dfrac{\lambda\{f^{\text{cons}}(\mathbf{w}) - c\}}{n}$ is a penalization function, for some constant c. If both the objective function and the penalization function can be profiled by $\boldsymbol{\theta}$, defined as $f^{\text{obj}}(z, \boldsymbol{\theta})$ and $f^{\text{pen}}(\boldsymbol{\theta}) := \dfrac{\lambda(\boldsymbol{\theta})\{f^{\text{cons}}(\boldsymbol{\theta}) - c\}}{n}$; the objective function is a $\log$ likelihood function; both $f^{\text{obj}}(z, \boldsymbol{\theta})$ and $f^{\text{pen}}(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$; and there exists a function $d_0(z)$ such that $\mid L(z, \boldsymbol{\theta}) \mid := \mid f^{\text{obj}}(z, \boldsymbol{\theta}) + f^{\text{pen}}(\boldsymbol{\theta}) \mid \leq d_0(z)$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $z \in \mathcal{Z}$, and $\mathbb{E}_{\boldsymbol{\theta}_0}[d_0(x)] < \infty$, then

i. $q_0(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}_0}[L(z, \boldsymbol{\theta})]$ is continuous in $\boldsymbol{\theta}$;

ii. $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mid q(\boldsymbol{\theta}; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}) \mid \xrightarrow{p} 0$, where $q(\boldsymbol{\theta}; \mathbf{Z}_n) := \dfrac{1}{n} L(\mathbf{Z}_n, \boldsymbol{\theta})$.

Note: the above Lemma can be stated in a more general case where there are multiple sets of parameters and several constraint functions.

PROOF. Consider the regularity conditions stated in the Appendix.
$\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, choose a sequence $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}$, such that $\boldsymbol{\theta}_k \to \boldsymbol{\theta}$. By (N-3), we have $L(x; \boldsymbol{\theta}_k) \to L(x; \boldsymbol{\theta})$. By (N-4) and the dominated convergence theorem (DCT), $q_0(\boldsymbol{\theta}_k) := \mathbb{E}_{\boldsymbol{\theta}_0}(L(Z, \boldsymbol{\theta}_k)) \to \mathbb{E}_{\boldsymbol{\theta}_0}(L(Z, \boldsymbol{\theta})) = q_0(\boldsymbol{\theta})$. Hence, $q_0(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$.

$L(z, \boldsymbol{\theta})$ is uniformly continuous since $f^{\text{obj}}(z, \boldsymbol{\theta})$ and $f^{\text{prof}}(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$. Hence,

$$\Delta(z, \delta) = \sup_{\{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) : \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < \delta\}} \mid L(z, \boldsymbol{\theta}_1) - L(z, \boldsymbol{\theta}_2) \mid \to 0$$

as $\delta \to 0$. By (N-4), $\Delta(z, \delta) \leq 2 d_0(z), \forall \delta$. By DCT, $\mathbb{E}_{\boldsymbol{\theta}_0}[\Delta(Z, \delta)] \to 0$ as $\delta \to 0$.

Define $B(\boldsymbol{\theta}_j, \delta) = \{\tilde{\boldsymbol{\theta}} : \| \tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \| < \delta\}$. Since $\boldsymbol{\Theta}$ is compact, for every fixed δ, $\exists$ a subcover $\{B(\boldsymbol{\theta}_j, \delta), j = 1, \ldots, J\}$ such that $\bigcup_{j=1}^{J<\infty} B(\boldsymbol{\theta}_j, \delta) \supset \boldsymbol{\Theta}$. Then, we have:

$$\begin{aligned} \mid q(\boldsymbol{\theta}; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}) \mid \quad & \leq \quad \mid q(\boldsymbol{\theta}; \mathbf{Z}_n) - q(\boldsymbol{\theta}_j; \mathbf{Z}_n) \mid & (1) \\ & + \mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid & (2) \\ & + \mid q_0(\boldsymbol{\theta}_j) - q_0(\boldsymbol{\theta}) \mid . & (3) \end{aligned}$$

Choose $\boldsymbol{\theta}_j$ such that $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_j; \delta)$. Since $\| \boldsymbol{\theta} - \boldsymbol{\theta}_j \| < \delta$, then:

(1) $= \mid \frac{1}{n} \sum_{i=1}^n \{L(Z_i, \boldsymbol{\theta}) - L(Z_i, \boldsymbol{\theta}_j)\} \mid \leq \frac{1}{n} \sum_{i=1}^n \mid L(Z_i, \boldsymbol{\theta}) - L(Z_i, \boldsymbol{\theta}_j) \mid \leq \frac{1}{n} \sum_{i=1}^n \Delta(Z_i, \delta)$.

Next, (2) $< \max_{j \in \{1,\ldots,J\}} \mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid$; and choose δ to be small, then (3) $\leq \sup_{\{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2): \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| < \delta\}} \mid q_0(\boldsymbol{\theta}_1) - q_0(\boldsymbol{\theta}_2) \mid \leq \epsilon^*(\delta)$, where $\epsilon(\delta) \to 0$ as $\delta \to 0$.

Combining (1) - (3), we have:

$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mid q(\boldsymbol{\theta}; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}) \mid \leq \frac{1}{n} \sum_{i=1}^n \Delta(Z_i, \delta) + \max_{j \in \{1,\ldots,J\}} \mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid + \epsilon^*(\delta)$.
Choose $\delta_1 \in \{\delta : \epsilon^*(\delta) \leq \frac{\epsilon}{3}\}$. Then for any $\delta < \delta_1$, we have:

$$
\begin{aligned}
&P_{\boldsymbol{\theta}_0}[\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mid q(\boldsymbol{\theta}; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}) \mid > \epsilon] \\
&\quad \leq P_{\boldsymbol{\theta}_0}[\frac{1}{n} \sum_{i=1}^n \Delta(Z_i, \delta) + \max_{j \in \{1,\ldots,J\}} \mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid > \frac{2\epsilon}{3}] \\
&\quad \leq P_{\boldsymbol{\theta}_0}[\frac{1}{n} \sum_{i=1}^n \Delta(Z_i, \delta) > \frac{\epsilon}{3}] + \qquad \text{(4a)} \\
&\qquad\quad P_{\boldsymbol{\theta}_0}[\max_{j \in \{1,\ldots,J\}} \mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid > \frac{\epsilon}{3}] \qquad \text{(4b)}
\end{aligned}
$$

Note that (4a) $= P_{\boldsymbol{\theta}_0}[\frac{1}{n} \sum_{i=1}^n \{\Delta(Z_i, \delta) - \mathbb{E}_{\boldsymbol{\theta}_0}[\Delta(Z; \delta)]\} + \mathbb{E}_{\boldsymbol{\theta}_0}[\Delta(Z; \delta)] > \frac{\epsilon}{3}]$, where $\mathbb{E}_{\boldsymbol{\theta}_0}[\Delta(Z; \delta)] \to 0$ as $\delta \to 0$. Choose $\delta_2 \in \{\delta : \mathbb{E}_{\boldsymbol{\theta}_0}[\Delta(Z; \delta)] < \frac{\epsilon}{6}\}$. Take $\delta < \min(\delta_1, \delta_2)$. Then:

$$P_{\boldsymbol{\theta}_0}[\frac{1}{n} \sum_{i=1}^n \{\Delta(Z_i, \delta) - \mathbb{E}_{\boldsymbol{\theta}_0}[\Delta(Z; \delta)]\} > \frac{\epsilon}{6}] := (4)'$$

By the Weak Law of Large Numbers (WLLN), $\exists\ N_1(\epsilon, \xi)$ such that $\forall\ n > N_1(\epsilon, \xi)$, $(4) < (4)' < \frac{\xi}{2}$.

Consider the finite subcover $\{B(\boldsymbol{\theta}_j, \delta), j = 1, \ldots, J\}$ for δ considered above. Note that:

$$(5) = P_{\boldsymbol{\theta}_0}[\bigcup_{j=1}^J \{\mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid > \frac{\epsilon}{3}\}] \leq \sum_{j=1}^J P_{\boldsymbol{\theta}_0}[\mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid > \frac{\epsilon}{3}].$$

By the WLLN, $\forall \boldsymbol{\theta}_j$ and $\forall \epsilon, \xi > 0$, $\exists\ N_{2j}(\epsilon, \xi)$ such that $\forall n > N_{2j}(\epsilon, \xi)$: $P_{\boldsymbol{\theta}_0}[\mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid > \frac{\epsilon}{3}] \leq \frac{\xi}{2J}$. Let $N_2(\epsilon, \xi) = \max_{j \in \{1,\ldots,J\}} \{N_{2j}\}$. Then, $\forall n > N_2(\epsilon, \xi)$, we have: $\sum_{j=1}^J P_{\boldsymbol{\theta}_0}[\mid q(\boldsymbol{\theta}_j; \mathbf{Z}_n) - q_0(\boldsymbol{\theta}_j) \mid > \frac{\epsilon}{3}] < \frac{\xi}{2}$. Hence, (4b) $< \frac{\xi}{2}$.

(4a) and (4b) show that $\exists$ an $N(\epsilon,\xi) = \max\big(N_1(\epsilon,\xi), N_2(\epsilon,\xi)\big)$ such that $\forall n > N(\epsilon,\xi)$, $P_{\boldsymbol{\theta}_0}[\sup_{\boldsymbol{\theta}\in\Theta} \mid q(\boldsymbol{\theta};\mathbf{Z}_n) - q_0(\boldsymbol{\theta}) \mid] < \xi$. ■

Proof of Theorem 1:

Define $Q(\boldsymbol{\theta};\mathbf{Z}_n) := \dfrac{1}{n}L(\boldsymbol{\theta};\mathbf{Z}_n)$ and $Q_0(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}_0}(L(\boldsymbol{\theta}_0;\mathbf{Z}_n))$, $L(\boldsymbol{\theta};z) := f^{\text{obj}}\big(\boldsymbol{\theta};z\big) + \dfrac{\lambda(\boldsymbol{\theta})(f^{\text{cons}} - c)}{n}$, and $\hat{\boldsymbol{\theta}}(\mathbf{Z}_n) := \operatorname{argmax}_{\boldsymbol{\theta}}\{L(\mathbf{Z}_n;\boldsymbol{\theta})\}$, henceforth $\hat{\boldsymbol{\theta}}$.

I. Consistency

To show

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0, \tag{5}$$

by Lemma 1, it suffices to show: (a) $Q(\boldsymbol{\theta};\mathbf{Z}_n)$ is continuous in $\boldsymbol{\theta}$; (b) $Q_0(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$; and (c) $\sup_{\boldsymbol{\theta}\in H} \mid Q(\boldsymbol{\theta};\mathbf{Z}_n) - Q_0(\boldsymbol{\theta}) \mid \xrightarrow{p} 0$. Note that (a) is implied by (N-3); (N-3), (N-4), and Lemma 2 give (d) $q_0(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}_0}(L(z,\boldsymbol{\theta}))$ is continuous in $\boldsymbol{\theta}$; and (e) $\sup_{\boldsymbol{\theta}\in H} \| \hat{q}(\boldsymbol{\theta};\mathbf{Z}_n) - q_0(\boldsymbol{\theta}) \| \xrightarrow{p} 0$, where $\hat{q}(\boldsymbol{\theta};\mathbf{Z}_n) = \dfrac{1}{n}\sum_{i=1}^{n} L(Z_i,\boldsymbol{\theta})$. Then (d) implies (b); (e) implies (c). ■

II. Asymptotic Normality

Let $\ell(z;\boldsymbol{\theta}) := \dfrac{\partial f^{\text{obj}}(z;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \dfrac{\nabla^{\boldsymbol{\theta}}\lambda(\boldsymbol{\theta})(f^{\text{cons}}(\boldsymbol{\theta}) - c) + \lambda(\boldsymbol{\theta})\nabla^{\boldsymbol{\theta}} f^{\text{cons}}(\boldsymbol{\theta})}{n}$, $\hat{q}(\mathbf{Z}_n;\boldsymbol{\theta}) = \dfrac{1}{n}\sum_{i=1}^{n} L(Z_i;\boldsymbol{\theta})$, $q_0(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}_0}(L(z;\boldsymbol{\theta}))$, $\hat{\dot{q}}(\mathbf{Z}_n;\boldsymbol{\theta}) = \dfrac{1}{n}\sum_{i=1}^{n} \ell(Z_i;\boldsymbol{\theta})$, and $\dot{q}_0(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}_0}(\ell(z;\boldsymbol{\theta}))$.

$\hat{\boldsymbol{\theta}}$ satisfies: $\dfrac{\partial L(\mathbf{Z}_n,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}/_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$. Hence, $\hat{\dot{q}}(\mathbf{Z}_n;\hat{\boldsymbol{\theta}}) = 0$. (N-2) and (N-6) give,

$$\dot{q}_0(\boldsymbol{\theta}) = o_p(n^{-1/2}). \tag{6}$$

Expanding $\hat{\dot{q}}(\mathbf{Z}_n;\hat{\boldsymbol{\theta}})$, we have:

$$0 = \hat{\dot{q}}(\mathbf{Z}_n;\hat{\boldsymbol{\theta}}) = \hat{\dot{q}}(\mathbf{Z}_n;\boldsymbol{\theta}_0) + D_n^*(\mathbf{Z}_n)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{7}$$

where $D_n^*(\mathbf{Z}_n) = \dfrac{\partial\hat{\dot{q}}(\mathbf{Z}_n;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}/_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$, for some $\boldsymbol{\theta}^*$ in the open interval bounded by $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$.

(N-9), (N-10) and Lemma 2 give: $\sup \parallel \hat{D}(\mathbf{Z}_n;\boldsymbol{\theta}) - D_0(\boldsymbol{\theta}) \parallel \xrightarrow{p} 0$, where $D_0(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}_0}\big(\nabla^{\boldsymbol{\theta}}\ell(z;\boldsymbol{\theta})\big)$ and $D_0(\boldsymbol{\theta}_0) = \mathbb{E}_{\boldsymbol{\theta}_0}\big(\nabla^{\boldsymbol{\theta}}\ell(z;\boldsymbol{\theta}_0)\big)$.

Since $\boldsymbol{\theta}^* \in \mathcal{N}_r(\boldsymbol{\theta}_0)$ w.p.1., then $D_n^*(\mathbf{Z}_n) \xrightarrow{p} D_0(\boldsymbol{\theta}_0)$. From (7), we have $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\{D_n^*(\mathbf{Z}_n)\}^{-1}\{\sqrt{n}\hat{\dot{q}}(\mathbf{Z}_n;\boldsymbol{\theta}_0)\}$; From (6), we have $\sqrt{n}\hat{\dot{q}}(\mathbf{Z}_n;\boldsymbol{\theta}_0) = \sqrt{n}\Bigg(\dfrac{\sum_{i=1}^n \ell(Z_i;\boldsymbol{\theta}_0)}{n} - \dot{q}_0(\boldsymbol{\theta})\Bigg) + \sqrt{n}\dot{q}_0(\boldsymbol{\theta}) \rightarrow N\big(0, V(\boldsymbol{\theta}_0)\big)$, where $V(\boldsymbol{\theta}_0) = \mathbb{E}_{\boldsymbol{\theta}_0}\Bigg(\ell(z;\boldsymbol{\theta})\ell^{\intercal}(z;\boldsymbol{\theta})\Bigg)$. It follows:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow N(0, \Sigma(\boldsymbol{\theta}_0))$$

where $\Sigma(\boldsymbol{\theta}_0) = D_0^{-1}(\boldsymbol{\theta}_0)V(\boldsymbol{\theta}_0)\big)[D_0^{-1}(\boldsymbol{\theta}_0)]^{\intercal}$. ■

Proof of Theorem 2:

I. Consistency

(N-0), and Eqs. (5)-(7) in Section 3.3 show that, for every fixed estimate $\hat{\boldsymbol{\theta}}$, there is an unique $\mathbf{w} : \mathbb{R}^{\dim(\hat{\boldsymbol{\theta}})} \longmapsto \mathbb{R}^{\dim(\mathbf{w})}$, such that $\mathbf{w} = \mathbf{w}(\hat{\boldsymbol{\theta}})$. Hence, under (N-13), (5) gives:

$$\mathbf{w}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{w}(\boldsymbol{\theta}_0).$$

II. Asymptotic Normality

Under (N-12) and (N-13), and by the Multivariate Delta Method,

$$\sqrt{n}(\mathbf{w}(\hat{\boldsymbol{\theta}}) - \mathbf{w}(\boldsymbol{\theta}_0)) \rightarrow N\Bigg(0, [\nabla\mathbf{w}(\boldsymbol{\theta})]^{\intercal}\Sigma(\boldsymbol{\theta}_0)\nabla\mathbf{w}(\boldsymbol{\theta})\Bigg). \qquad ■$$

References and Notes

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.