Robust Weighting and Matching Techniques for Causal Inference in Observational Studies with Continuous Treatment

Universität Stuttgart

Universität Stuttgart

Ioan Scheffel

October 26, 2022

Contents

Chapter 1

Chapter One Title

Assumption 1. Assume, the following conditions hold:
(i) The minimizer $\lambda_0 = \arg\min_{\lambda\in\Theta} \mathbb{E}\left[-Tn\rho\left(B(X)^T\lambda\right) + B(X)^T\lambda\right]$ is unique, where $\Theta \subseteq \mathbb{R}^n$ is the parameter space for λ .
(ii) The parameter space $\Theta \subseteq \mathbb{R}^n$ is compact compact with diameter $\operatorname{diam}(\Theta) < \infty$.
(iii) $\lambda_0 \in \operatorname{int}(\Theta)$, where $\operatorname{int}(\cdot)$ stands for the interior of a set.
(iv) There existi $\lambda_1^* \in \Theta$ such that $\left\|m^*(\cdot) - B(\cdot)^T\lambda_1^*\right\|_\infty \leq \varphi_{m^*}$, where $m^*(\cdot) := \left(\rho'\right)^{-1}\left(\frac{1}{n\pi(\cdot)}\right)$.
(v) There exists a constant $\varphi_{\rho'\vee\pi} \in \left(0, \frac{1}{2}\right)$ such that $n\rho(v) \in (\varphi_{\rho'\vee\pi}, 1-\varphi_{\rho'\vee\pi})$ for $v = B(x)^T\lambda$ with $\lambda \in \operatorname{int}(\Theta)$ **or** $\pi(x) \in (\varphi_{\rho'\vee\pi}, 1-\varphi_{\rho'\vee\pi})$.
(vi) There exists $\varphi_{\rho''} > 0$ such that $-\rho^{''} \geq \varphi_{\rho''} > 0$
(vii) There exists $\varphi_{B(x)B(x)^T} > 0$ such that $B(x)B(x)^T \succcurlyeq \varphi_{B(x)B(x)^T} I$
$(viii)$ There exists $\varphi_{\|B\|} > 0$ such that $\sup_{x\in\mathcal{X}} \|B(x)\|_2 \leq \varphi_{\|B\|}$.

the $1(ii)$

We study the following problem:

$$\begin{aligned} &\underset{w\in\mathbb{R}^n}{\text{minimize}} \quad && \sum_{i=1}^{n} T_i f(w_i) \\ &\text{subject to} && \left|\sum_{i=1}^{n} w_i T_i B_k(X_i) - \frac{1}{n}\sum_{i=1}^{n} B_k(X_i)\right| \leq \delta_k,\ k = 1, \ldots, K \end{aligned} \tag{1.1}$$

Proposition 1.1. The dual of Problem (1.1) is equivalent to the unconstrained optimization problem

$$\underset{\lambda\in\mathbb{R}^K}{\text{minimize}} \quad \frac{1}{n}\sum_{j=1}^{n}\left[-T_j n\rho\left(B(X_j)^T\lambda\right) + B(X_j)^T\lambda\right] + |\lambda|^T\delta \tag{1.2}$$

Proposition 1.2. There exists a solution $\lambda^\dagger$ to (1.2) such that

$$\mathbb{P}\left(\left\|\lambda^\dagger - \lambda_1^*\right\|_2 \le C_{\mathbb{P}} C_\tau \varepsilon_n\right) \ge 1-\tau. \tag{1.3}$$

1.1 Plan of proof

We employ Theorem 2.2 together with the box constraints in Problem (1.1) to obtain Proposition 1.1.

To prove Proposition 1.2 we employ Proposition 4.1 and Corollary 4.1.1 to get

$$\begin{aligned}
&G(\lambda_1^*+\Delta) - G(\lambda_1^*)\\
&\ge \frac{1}{n}\sum_{j=1}^{n}\left[-T_j n\rho'\left(B(X_j)^T\lambda_1^*\right)+1\right]\Delta^T B(X_j)\\
&+\frac{1}{2}\sum_{j=1}^{n} -T_j\rho''\left(B(X_j)^T(\lambda_1^*+\xi\Delta)\right)\Delta^T\left(B(X_j)B(X_j)^T\right)\Delta\\
&-|\Delta|^T\delta\\
&\ge -\|\Delta\|_2\left(\left\|\frac{1}{n}\sum_{j=1}^{n}\left[-T_j n\rho'\left(B(X_j)^T\lambda_1^*\right)+1\right]B(X_j)\right\|_2 + \|\delta\|_2\right)\\
&+n\|\Delta\|_2^2\,\varphi_{\rho''}\underline{\varphi_{aa^T}}
\end{aligned} \tag{1.4}$$

Next we employ Bernstein inequality 3.1 to bound

$$\left\|\frac{1}{n}\sum_{j=1}^{n}\left[-T_j n\rho'\left(B(X_j)^T\lambda_1^*\right)+1\right]B(X_j)\right\|_2 \le C_{\mathbb{P}}C_\tau\varepsilon_n \tag{1.5}$$

with probability $1-\tau$. Then for $\|\Delta\|_2$ large enough it holds

$$G(\lambda_1^*+\Delta) - G(\lambda_1^*) > 0 \tag{1.6}$$

with probability $1-\tau$. Thus by Proposition 4.1

$$\mathbb{P}\left(\left\|\lambda^{\dagger}-\lambda_1^*\right\|_2 \leq \|\Delta\|_2\right) \geq 1-\tau. \tag{1.7}$$

It is then straightforward to prove

Theorem 1.1. *Let $\lambda^{\dagger}$ be the solution to Problem 1.2 and $w^*(x)=\rho^{'}\left(B(x)^T\lambda^{\dagger}\right)$. Then under the conditions in Assumption 1 it holds*

$$\left\|w^*(\cdot)-\frac{1}{n\pi(\cdot)}\right\|_{\mathbb{P},2} \leq \textit{stuff} \tag{1.8}$$

and

$$\mathbb{P}\left(\left\|w^*(\cdot)-\frac{1}{n\pi(\cdot)}\right\|_{\infty} \leq \textit{stuff}\right) \geq 1-\tau. \tag{1.9}$$

Proof. Motivated by Proposition 4.1 we set $\|\Delta\|_2 = C$ and consider

$$G(\lambda) := \frac{1}{n}\sum_{j=1}^{n}\left[-T_j n\rho\left(B(X_j)^T\lambda\right)+B(X_j)^T\lambda\right]+|\lambda|^T\delta. \tag{1.10}$$

Since $\rho\in C^2(\mathbb{R})$ we can employ Proposition 4.1, Corollary 4.1.1 and Proposition 4.2 to get

$$\begin{aligned}
&G(\lambda_1^*+\Delta)-G(\lambda_1^*)\\
&\geq \frac{1}{n}\sum_{j=1}^{n}\left[-T_j n\rho^{'}\left(B(X_j)^T\lambda_1^*\right)+1\right]\Delta^T B(X_j)\\
&+\frac{1}{2}\sum_{j=1}^{n}-T_j\rho^{''}\left(B(X_j)^T(\lambda_1^*+\xi\Delta)\right)\Delta^T\left(B(X_j)B(X_j)^T\right)\Delta\\
&-|\Delta|^T\delta\\
&\geq -\|\Delta\|_2\left(\left\|\frac{1}{n}\sum_{j=1}^{n}\left[-T_j n\rho^{'}\left(B(X_j)^T\lambda_1^*\right)+1\right]B(X_j)\right\|_2+\|\delta\|_2\right)\\
&+n\|\Delta\|_2^2\,\varphi_{\rho''}\underline{\varphi_{aa^T}}\\
&:= -\|\Delta\|_2\left(I_1+\|\delta\|_2\right)+\|\Delta\|_2^2 I_2.
\end{aligned} \tag{1.11}$$

The second inequality is due to the Cauchy-Schwarz-Inequality and Assumptions $1(vi)$ and $1(vii)$.

Analysis of I_1

We want to use Assumption 1(iii). Thus we perform the following split:

$$I_1 \leq \left\| \sum_{j=1}^{n} T_j \left[\rho^{'} \left(B(X_j)^T \lambda_1^* \right) - \frac{1}{n\pi(X_j)} \right] B(X_j) \right\|_2 \tag{1.12}$$

$$+ \left\| \frac{1}{n} \sum_{j=1}^{n} \left[\frac{T_j}{\pi(X_j)} - 1 \right] B(X_j) \right\|_2 \tag{1.13}$$

$$=: J_1 + J_2 \tag{1.14}$$

Analysis of J_1

By the Lipschitz-continuity of $\rho^{'}$, Assumption 1($viii$) and Assumption 1(iv), $T \in \{0, 1\}$ and the triangle inequality we have

$$J_1 \leq n L_{\rho'} \varphi_{\|B(x)\|} \varphi_{m^*} \tag{1.15}$$

Analysis of J_2

We employ Bernstein Inequality for matrices $\square$

Chapter 2

Convex Analysis

We begin by defining convex sets

Definition 2.1. *A subset $\Omega \subseteq \mathbb{R}^n$ is called CONVEX if we have $\lambda x + (1 - \lambda)y \in \Omega$ for all $x, y \in \Omega$ and $\lambda \in (0, 1)$.*

Clearly, the line segment $[a, b] := \{\lambda a + (1 - \lambda)b \mid \lambda \in [0, 1]\}$ is contained in Ω for all $a, b \in \Omega$ if and only if Ω is a convex set.

Next we define convex functions.

The concept of convex functions is closely related to convex sets.

The line segment between two points on the graph of a convex function lies on or above and does not intersect the graph.

In other words: The area above the graph of a convex function f is a convex set, i.e. the *epigraph* $\operatorname{epi}(f) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}$ is a convex set in $\mathbb{R}^{n+1}$.

Often an equivalent characterisation of convex functions is more useful.

Theorem 2.1. *The convexity of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ on $\mathbb{R}^n$ is equivalent to the following statement:*

For all $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$ we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{2.1}$$

Definition 2.2. proper convex function

Definition 2.3. convex conjugate

Given proper convex functions $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a matrix $A \in \mathbb{R}^{n\times n}$, we define the primal minimization problem as follows:

$$\text{minimize} \quad f(x) + g(Ax) \quad \text{subject to} \quad x \in \mathbb{R}^n. \tag{2.2}$$

The Fenchel dual problem is then

$$\text{maximize} \quad -f^*\left(A^T y\right) - g^*(-y) \quad \text{subject to} \quad y \in \mathbb{R}^n. \tag{2.3}$$

Theorem 2.2. *Let* $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be proper convex functions and* $0 \in ri(dom(g) - A(dom(f)))$ *. Then the optimal values of* (2.2) *and* (2.3) *are equal, i.e.*

$$\inf_{x\in\mathbb{R}^n} \{f(x) + g(Ax)\} = \sup_{y\in\mathbb{R}^n} \left\{-f^*\left(A^T y\right) - g^*(-y)\right\}. \tag{2.4}$$

Chapter 3

Random Matrix Inequality

Theorem 3.1. *Let $(A_k)_{1\leq k\leq n} \subseteq \mathbb{R}^{d_1\times d_2}$ be a finite sequence of independent, random matrices. Assume that*

$$\mathbb{E}(A_k) = 0 \quad \textit{and} \quad \|A_k\| \leq L \quad \textit{for each} \quad k \in \{1, \dots, n\}. \tag{3.1}$$

Introduce the random matrix

$$S := \sum_{k=1}^{n} A_k. \tag{3.2}$$

Let $v(S)$ be the matrix variance statistic of the sum:

$$v(S) := \max\left\{\left\|\mathbb{E}(SS^T)\right\|, \left\|\mathbb{E}(S^T S)\right\|\right\} \tag{3.3}$$

$$= \max\left\{\left\|\sum_{k=1}^{n} \mathbb{E}(A_k A_k^T)\right\|, \left\|\sum_{k=1}^{n} \mathbb{E}(A_k^T A_k)\right\|\right\}. \tag{3.4}$$

Then

$$\mathbb{E}\,\|S\| \leq \sqrt{2v(S)\log(d_1+d_2)} + \frac{1}{3}L\log(d_1+d_2). \tag{3.5}$$

Furthermore, for all $t \geq 0$,

$$\mathbb{P}(\|S\| \geq t) \geq (d_1+d_2)\exp\left(\frac{-t^2/2}{v(S)+Lt/3}\right). \tag{3.6}$$

Chapter 4

Simple yet useful Calculations

Proposition 4.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous such that a minimum x^* exists and is unique. Then for all $y \in \mathbb{R}^n$ and $C > 0$ it follows

$$\inf_{\|\Delta\|=C} f(y+\Delta) - f(y) > 0 \qquad \Rightarrow \qquad \|x^* - y\| \leq C. \tag{4.1}$$

Proof. Since $\mathcal{C} := \{\|\Delta\| \leq C\}$ is compact and

$$f(x^*) \leq f(y) < \inf_{\|\Delta\|=C} f(y+\Delta),$$

the continious function $f(y + \,\cdot\,)$ has a minimum in $\operatorname{int}(\mathcal{C}) := \{\|\Delta\| < C\}$. Since x^* is the unique minimum of f there exists $\Delta^* \in \operatorname{int}(\mathcal{C})$ such that $x^* - y = \Delta^*$. We conclude that $\|x^* - y\| \leq C$. $\square$

Theorem 4.1. (Multivariate Taylor Theorem) *Let $f \in C^2(\mathbb{R}^n, \mathbb{R})$. Then for all $x, \Delta \in \mathbb{R}^n$ there exists $\xi \in [0,1]$ such that it holds*

$$\begin{aligned} f(x+\Delta) = f(x) &+ \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i}\Delta_i + \sum_{\substack{i,j=1 \\ i\neq j}} \frac{\partial^2 f(x+\xi\Delta)}{\partial x_i \partial x_j}\Delta_i\Delta_j \\ &+ \frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2 f(x+\xi\Delta)}{\partial x_i^2}\Delta_i^2 \end{aligned} \tag{4.2}$$

Corollary 4.1.1. Let $f \in C^2(\mathbb{R})$. Then for all $a, x, \Delta \in \mathbb{R}^n$ there exist $\xi \in [0,1]$ such that it holds

$$f(a^T(x+\Delta)) - f(a^Tx) = f^{'}(a^Tx)\,\Delta^T a + \frac{1}{2} f^{''}(a^T(x+\xi\Delta))\,\Delta^T A\;\Delta, \tag{4.3}$$

where $A := aa^T \in \mathbb{R}^{n\times n}$.

Proof. By the chain rule we have for all $a, x, \Delta \in \mathbb{R}^n$ and $\xi \in [0, 1]$

$$\frac{\partial^2 f(a^T(x + \xi\Delta))}{\partial x_i \partial x_j} = f^{''}(a^T(x + \xi\Delta))\, a_i a_j. \tag{4.4}$$

Since $A := aa^T$ is symmetric we have

$$\Delta^T A\ \Delta = 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} a_i a_j \Delta_i \Delta_j + \sum_{i=1}^{n} a_i^2 \Delta_i^2. \tag{4.5}$$

Plugging (4.4) and (4.5) into (4.2) yields (4.3). □

Proposition 4.2. For all $x, y \in \mathbb{R}$ it holds

$$|x + y| - |x| \geq -|y| \tag{4.6}$$

Proof. Checking all 6 combinations of $x+y$, x, y being nonnegative or negative yields the result. □