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

Universität Stuttgart

Ioan Scheffel

October 25, 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 $\mathrm{diam}(\Theta) < \infty$.
(iii) $\lambda_0 \in \mathrm{int}(\Theta)$, where $\mathrm{int}(\cdot)$ stands for the interior of a set.

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

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

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*

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

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:*

$$\begin{aligned} v(S) :&= \max\left\{\left\|\mathbb{E}(SS^T)\right\|, \left\|\mathbb{E}(S^TS)\right\|\right\} &(3.3)\\ &= \max\left\{\left\|\sum_{k=1}^{n}\mathbb{E}(A_kA_k^T)\right\|, \left\|\sum_{k=1}^{n}\mathbb{E}(A_k^TA_k)\right\|\right\}. &(3.4)\end{aligned}$$

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

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

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). $\square$