

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

Universität Stuttgart



Universität Stuttgart

Ioan Scheffel

October 25, 2022

Contents

1	Chapter One Title	2
2	Convex Analysis	3
3	Random Matrix Inequality	5
4	Simple yet useful Calculations	6

Chapter 1

Chapter One Title

Assumption 1. *Assume, the following conditions hold:*

(i) *The minimizer $\lambda_0 = \arg \min_{\lambda \in \Theta} \mathbb{E} [-Tn\rho(B(X)^T\lambda) + B(X)^T\lambda]$ 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 $\text{diam}(\Theta) < \infty$.*

(iii) *$\lambda_0 \in \text{int}(\Theta)$, where $\text{int}(\cdot)$ stands for the interior of a set.*

the 1(ii)

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* $\text{epi}(f) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}$ 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 $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ 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

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (2.1)$$

Definition 2.2. proper convex function

Definition 2.3. convex conjugate

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

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

The Fenchel dual problem is then

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

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

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

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*

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

Introduce the random matrix

$$S := \sum_{k=1}^n A_k. \quad (3.2)$$

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

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

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

Then

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

Furthermore, for all $t \geq 0$,

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

Chapter 4

Simple yet useful Calculations

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

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

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

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

the continuous function $f(y + \cdot)$ has a minimum in $\overset{\circ}{\mathcal{C}} := \{\|\Delta\| < C\}$. Since x^* is the unique minimum of f there exists $\Delta^* \in \overset{\circ}{\mathcal{C}}$ 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*

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

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*

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

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 $\xi \in [0, 1]$

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

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

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

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