Notes During PhD

Yongqi Wang

31 March 2025

This document is an ongoing collection of my writing during my PhD. Inspired by the Tufte-Handout Style [1], this handout is built using `tufte-latex`[2].

[1] Edward R. Tufte!

[2] `https://github.com/Tufte-LaTeX/tufte-latex`

Contents

LaTeX Project Management

A consistent boilerplate for LaTeXprojects, I choose AOS for regular article.[3]

[3] https://vtex-soft.github.io/texsupport.ims-aos/

The references are managed externally by Zotero and BBT, exported to BibTex format, then included via `natbib`.

- `chapters/01-*.tex`: individual files
- `fig`: figures to reproduce
  - External figures by R or Python DPS ≥ 300 `.pdf`
  - TikZ: `.tex` and `.pdf`
  - Asymptote: `.asy` and `.pdf`

  Below is an example project hosted on Github or Overleaf:

```
├── chapters
│   ├── 01-blabla.tex            % each chapter/section
│   ├── ...
├── fig
│   ├── R/Python.pdf
│   ├── TikZ.tex
│   ├── TikZ.pdf
│   ├── Asymptote.asy
│   ├── Asymptote.pdf
├── latexmkrc
├── main.bib                     % References
├── main.pdf                     % Main output
├── main.tex                     % Main document
└── tex
    ├── macro.tex                % All my collected macros
    ├── setup.tex                % project-specific: package, layout, macro
    ├── custom-style.cls
    ├── custom-style.def
    ├── custom-style.sty
    └── custom-style.bst
```

Mathematical Notation

It has always been a hassle to organise mathematical notation across different sources, in fact, I would go so far as to argue that it is the most annoying thing when one starts reading a book or an article.

However, there *must be* some notational conflicts beyond primary school simply due to the fact that the limited number of alphabets (**26**). For example, "$\mathbb{E}$" might be energy in physics while it could refer to expectation or scores in probability.

Another difficulty is that the authors often assume some familiarity in the topics *also* I am expected to read in some logical or chronological order. In reality, I am constantly jumping back and forth between one literature to another.

$$\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}, \mathbb{E}, \mathbb{F}, \mathbb{G}, \mathbb{H}, \mathbb{J}, \mathbb{K}, \mathbb{L}, \mathbb{M}, \mathbb{N}, \mathbb{O}, \mathbb{P}, \mathbb{Q}, \mathbb{R}, \mathbb{S}, \mathbb{T}, \mathbb{U}, \mathbb{V}, \mathbb{W}, \mathbb{X}, \mathbb{Y}, \mathbb{Z}$$
$$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{G}, \mathcal{H}, \mathcal{I}, \mathcal{J}, \mathcal{K}, \mathcal{L}, \mathcal{M}, \mathcal{N}, \mathcal{O}, \mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{S}, \mathcal{T}, \mathcal{U}, \mathcal{V}, \mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$$
$$\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}, \mathfrak{E}, \mathfrak{F}, \mathfrak{G}, \mathfrak{H}, \mathfrak{I}, \mathfrak{J}, \mathfrak{K}, \mathfrak{L}, \mathfrak{M}, \mathfrak{N}, \mathfrak{O}, \mathfrak{P}, \mathfrak{Q}, \mathfrak{R}, \mathfrak{S}, \mathfrak{T}, \mathfrak{U}, \mathfrak{V}, \mathfrak{W}, \mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$$

$$\arg\inf, \arg\sup, \arg\max, \arg\min, \operatorname{conv}$$

This stackexchange answer [4] is probably the most comprehensive answer to which fonts are shown in LaTeX.

[4] https://tex.stackexchange.com/a/58124

Table 1: Macros for Letters

Symbol	Usage	Comments
$\mathbb{B}$	`\*f`	blackboard bold except `\If` due to conflict
$\mathcal{B}$	`\*c`	calligraphic font
$\mathfrak{B}$	`\*k`	fraktur font

Reverse Inverse Projections (RIPr)

Li's Algorithm

Originally developed by @liEstimationMixtureModels1999, the following algorithm is written based on the notes from [GdK24] @haoEvalues*k*sampleTests2024. To quote Brinda, Li's inequality requires the family $\mathcal{Q}$ to have a uniformly bounded density ratio.

Li obtains the RIPr in a greedy manner where the KL divergence between distribution Q onto the convex hull of a set of distributions $\mathcal{Q}$ (composite null) is minimized. It is assumed that the KL divergence between Q and any distribution $Q \in \mathcal{Q}$ is finite (often call "nondegenerate" condition).

Algorithm 1: Li's Algorithm

$Q_{(1)} = \arg\min_{Q\in\mathcal{Q}} D(P\|Q)$
for $m = 2, 3, \ldots, K$ **do**
  $Q := \alpha Q_{(m-1)} + (1-\alpha)Q'$
  $\alpha, Q' \leftarrow \arg\min_{\alpha, Q'} D(P\|Q)$

Here, the distribution Q' and α is chosen (coupled) such that the divergence is minimized. The minimizer need not be unique.[5]

[5] I think

Regularity condition on $\mathcal{Q}$

Li's algorithm is apparently greedy with high fluctuation in initial steps. Additionally, this task is computationally expensive, and it is not clear of the convexity.

Is the returned mixture in the convex hull?

The returned $Q_{(m)}$ is in the convex hull. The first step returned a single element in $\mathcal{Q}$ with the smallest KL divergence. Iteratively, the linear combination is still in the convex hull, i.e.

$$\begin{aligned} Q_{(2)} &= \alpha_1 \cdot Q_{(1)} + (1-\alpha_1)\cdot Q'_{(1)} \\ Q_{(2)} &= \alpha_2 \cdot Q_{(2)} + (1-\alpha_2)\cdot Q'_{(2)} \\ &= \alpha_2\cdot\alpha_1\cdot Q_{(1)} + \alpha_2\cdot(1-\alpha_1)\cdot Q'_{(1)} + (1-\alpha_2)\cdot Q'_{(2)}. \end{aligned}$$

It is clear that $Q_{(2)}$ or $Q_{(m)}$ would still be a convex combination of elements in $\mathcal{Q}$.

Cisszar Algorithm

Originally proposed by

References

[GdK24] Peter Grünwald, Rianne de Heide, and Wouter Koolen. Safe testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5):1091–1128, November 2024.

[Hao25] Yunda Hao. *E-Values for Anytime-Valid Inference with Exponential Families*. PhD thesis, Leiden University, Leiden, February 2025.

[TG23] Rosanne J. Turner and Peter D. Grünwald. Exact anytime-valid confidence intervals for contingency tables and beyond. *Statistics & Probability Letters*, 198:109835, July 2023.

[TLG24] Rosanne J. Turner, Alexander Ly, and Peter D. Grünwald. Generic E-variables for exact sequential k-sample tests that allow for optional stopping. *Journal of Statistical Planning and Inference*, 230:106116, May 2024.