Submitted to the Annals of Statistics

NOTES DURING PHD

BY YONGQI WANG

CONTENTS

1. LaTeX Project Management. A consistent boilerplate for LaTeXprojects, I choose AOS for regular articles.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`. 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, mac
    ├── custom-style.cls
    ├── custom-style.def
    ├── custom-style.sty
    └── custom-style.bst
```

1.1. *Figure sizes and font sizes.* Ratio, margin, font size, how to adjust accordingly...

1.2. *Reference style.*

2. 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.

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

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

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

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

This stackexchange answer https://tex.stackexchange.com/a/58124 is probably the most comprehensive answer to which fonts are shown in LaTeX.

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

TABLE 1
Macros for Letters

3. Reverse Inverse Projections (RIPr).

3.1. *Li's Algorithm.* Originally developed by **?**, the following algorithm is written based on the notes from **??**. To quote **?**, 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 Which should I use? $\mathcal{Q}$ or $\mathbb{Q}$ (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.I think

Regularity condition on alternative

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 $\mathcal{Q}$.

3.2. *Cisszar Algorithm.* Originally proposed by

4. Integrals. What is it other than the area under curve?

Interesting examples where the Riemann's integration fails is that $f(x) = 1$ if x is rational and 0 otherwise over the interval $[0, 1]$. Why this function's integral is problematic in Riemann's definition.

Lebesgue used measure theory to define integral.

> Many times, when Riemann fails, Lebesgue works.

Set being countable and uncountable rely on if we can arrange them in a 'readable' sequence, could be infinite like rational numbers.

Proving that $\mathbb{R}$ is uncountable is not so trivial: decimal expansion.

No matter how hard you try to write the numbers in $[0, 1)$, there is always some values left out.

Continuum hypothesis: How do you check which ∞ is bigger? Or which set has the higher cardinality?

Definition 1.2.2 ()

EXAMPLE 4.1 (Set of Lebesgue measure zero, 1.2.2). *A set S is zero in Lebesgue measure when it can be covered with a sequence of open intervals $(I_1, I_2, \dots)$. And the sum $\sum_0^\infty m(I_n)$ can be made arbitrarily small*

In other words, for any $\epsilon > 0$, we can always find a sequence of I_n that covers S.

THEOREM 4.2. *Any countable infinite set of S has Lebesgue measure zero.*

REMARK 4.3. *The set of measure zero would not change a damn thing on f*