

# Checklist for effectively writing papers in stat-ML

Aaditya Ramdas (aramdas@cmu.edu), Carnegie Mellon University

Some personal suggestions via accumulated wisdom:

- **Naming, planning.** Everything should be in one file, except possibly macros. Name the file intelligently, not `paper.tex`. Plan papers in a top-down fashion: what is the paper’s purpose, the point of each section, etc.
- **Use macros** for all symbols, in case they need to be changed. Name macros or labels so that you recognize them a year later. Use a `~` when referring to equations/sections/figures: write `Section~\ref{sec:power}`. Familiarize yourself with commands like `bmatrix`, `align`, `cases`, `subequations`, `mbox`, `hspace`, `vspace`, etc.
- **Abstracts** need not detail every single advance made: summarize the setup and main results succinctly. Avoid citations unless necessary. Avoid future tense in abstract and paper (we will provide  $\rightarrow$  we provide).
- **Divide related work** into two parts: (a) those who have studied aspects of the same problem and those whose work your paper directly builds on or uses insights from, (b) other somewhat orthogonal but complementary papers. I prefer to include (a) in the introduction, and possibly (b) either in the discussion or in an “other related work” section at the end of the paper. Be concise and to the point, and give fair credit when it is due.
- **Figures.** Any text inside a figure, such as in a legend, must have a readable font size, almost matching but not exceeding the size of the other text. Any lines in a figure must have multiple complementary sources of identifiable information: for example, use color, texture and symbols (red vs. blue, dashed vs. dotted, triangles vs. circles). Figures should have error bars; if negligible, explain why explicitly in the caption.
- **Equations** are often part of a sentence, so follow them with the appropriate punctuation (commas or full-stops). Don’t number every equation in a proof, only those that a reviewer/reader may want to refer to.
- **Theorems versus rest.** Leave the annotation “theorem” for a handful of powerful central results: trivial theorems reduce trust in your judgment. Some may be propositions (interesting results of independent interest), or lemmas (a packaged technical result within a longer proof), or facts (known theorems from classical papers) or corollaries (interesting implications of theorems), or remarks. Number each type separately. Statements must be crisp and yet self-contained. Consider defining technical assumptions and necessary notation before the theorem and discussing what they mean and why they are needed. Even so, refer to all assumptions, by number or phrase, in the theorem statement so that it is self-contained.
- **Interpret** any theorem, proposition, corollary or lemma for the reader to provide intuition, and link to its proof. Avoid superlatives like “extremely general” or “very powerful” or “huge improvement”. Organize proofs modularly, give proof outlines, and don’t end them abruptly.
- **Mathematical writing.** If possible, don’t use equations/references as nouns. Change “(12) suggests” or “[12] claimed” to “Definition/condition (12) suggests” or “Maryam [12] claimed”. Don’t say “This implies...” or “So, ...”; instead try “Invoking the constraint (4), we see that inequality (8) implies...”. Avoid  $\forall, \exists$  and say “for all, some” if possible. Change phrases like “we get” to “we derive/conclude/infer/...”. Distinguish between a function  $f$  and its value  $f(x)$  at point  $x$ . A function  $f$  can be monotone, but  $f(x)$  cannot. Similarly, distinguish between random variables  $Y$  and a particular instantiation  $y$ .
- **Flow.** Sections should not start or end abruptly. The last line of a section should serve as a natural connector to the next section. Don’t have sections with just one subsection. If you have one important point in a section, don’t make a subsection, if you have two important and separate points, have two subsections.
- **Check bibliography** for capitalization, authors, spurious “et al.”, see if arXiv papers have been published.
- **Appendices** must be organized logically, preferably in order of their first reference.
- **Spelling and grammar** should be checked by software (eg: Copy text into Google Docs or Microsoft Word).

When you sign off on a paper, you should be proud of it, and ready to defend its details. Integrity is paramount.