STAT 539/518 Writing research papers

Marina Meilă

Department of Statistics
University of Washington

April 7, 2015

You will do a lot of writing

- In academia, research involves a lot of writing
  - PhD thesis
  - Papers, lectures notes, proposals
- Writing well helps a lot
  - Gets your papers published
  - Makes your work more credible, more convincing, more read

Writing style I. Use your own words

These points are derived from the most common mistakes I've seen in students' papers.

- (when you are writing about someone else's paper, e.g in a report or a review)
- this shows you have digested/interpreted the material
- shows you can relate to your audience (by "translating" the paper for them)
- simplifying a complicated paper/idea often requires rephrasing
- when someone[1] doesn't understand a paper, they stay close to the text[2]
- rephrase whole theorems and mathematical notation when you need to simplify (e.g. "if n is large enough" is better (in a report) than "if there is a constant C so that $n > \lambda\sigma^{2/C}\log\frac{1}{\delta}$)")

- This advice sometimes applies when you prepare a talk from a paper.

---

[1]Read: "a student"
[2]Read: "copy the text verbatim"

Writing style I. Be precise

These points are derived from the most common mistakes I've seen in students' papers.

- Define all terms
- Support all statements
- Make precise statements
    - "worked quite well" → "had 30% better accuracy than [competing method]", "converged in all cases"
    - "didn't work" → "failed to converge", "failed to find the correct model", "program did not fit in memory"
    - it is fine to say "large n", "high dimensions" (i) after you defined what "large/high" mean (e.g $p \gg n$) (ii) when in the experiments you have chosen some specific n values as examples of "large" n
- Technical writing is formal
    - (like talking to a stranger, not like talking to your friends)
    - should look polished
    - should be precise (see above)
    - impersonal/objective
    - don't try to be funny (and be original in the contents not the style)

    - but don't overdo it – occasional remarks to spur curiosity and enthusiasm are healthy (e.g. "this remarkably simple result", "a careful reader may have noticed that")

Writing style II

These is general advice (a small summary of it)

- Think why you are writing this paper. What is the core contribution that you will claim?
- Organize your ideas with the reader in mind. This may not be the order in which you first thought of these ideas.
- A paper tells a story – organize your ideas sequentially with the story in mind.
- Make the reader curious
    - Don't use the devices of fiction
    - Achieve it by carefully organizing the material (e.g start by describing the problem)
- Try to make the paper accessible without sacrificing accuracy
    - have a running example
    - give informal, simplified examples of the main theorems before the full statements (this is more or less the rule with mathematical papers)
    - put words around formulas

Why the literature review matters

- it's required (you can't publish a paper without it)
- gives credit where credit is due
- demonstrates understanding and breadth (scholarship)
    - especially if you the relation to your work is not obvious
    - or if you explain a related method in your own words, different from the original description
- puts your paper in context
- helps other better understand this area of research
    - a good literature review will be read/cited/recommended/and copied for its own sake
- helps you

Paper organization at a higher level I

- Abstract: "should I read this paper?" (Is this paper relevant to me?)
- Introduction: informal problem description, places the problem in context, related work
- Problem formulation
- Methodology/Proposed solution
- Discussion (placing the work in perspective)
  - what is achieved and what is not
  - related work again
  - what else could be done with this method
- Experimental evaluation
  - demonstrate empirically that method works
  - verify provable properties
  - learn about unknown/conjectured properties
  - compare with other methods

  - demonstrate that it can solve real problems (better than competition)

  - discuss the results!
- Conclusion
  - a summary of achievements
  - more discussion
  - a final advertisement of your method

What is reproducible research?

- The purpose of research is not just to solve a problem for someone, but to also teach a larger group of people how to solve similar problem.
  Otherwise it's called *development*.
- The research (results) you publish must be reproducible
- A research paper must contain enough information to allow a reader to reproduce/verify its results
- Verify that your assigned papers are reproducible, and how they achieve this

The research (results) you publish must be reproducible I

- Theorems/methods have proofs
- Algorithms
    - describe in enough detail that they can be implemented by an informed reader
    - balance between what described and what referenced only
- Data
    - make data available (whenever possible)
    - use benchmark data sets (wide and indirect comparisons possible)
    - give sources for data sets not your own
- Parameters of algorithms
    - describe methods to set the parameters
    - for each experiment: give all parameters settings, for all algorithms used

Figures and experimental results

- Show only relevant results
- General principles for displaying information (figures vs. tables, etc), see Tufte
- Make clear annotations, as close to data as possible. Spare comments (e.g "loweris better")

Rule Figure/Table caption explains **what** is in the figure/table; comments are in the main text

- Discussing the experimental results
  - don't repeat what's in the tables/figures (but refer to them)
  - appeal to reader's intuition
  - but be precise

Reproducible research one step further

(From Buckheit and Donoho)

we . . . distill Claerbout's insight into a slogan

An article about computational science in a scientic publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.

Do reproducible research and increase your impact too

- Make code available (and easily usable by others)
- Make a "project webpage"
  - publications
  - code, data, instructions, figures
  - other resources that you find useful

Planning your project

- what the final writeup should contain
- what you will implement
- what extensions you will make to the assigned paper
- the schedule

- Remember this is the first draft. Will be revised in the next weeks.

Plan – first draft

- the paper summary
- what parts of the paper you will implement
- what programming language will you use for implementation.
  if you think there is a reason to use code not written by you, which package and why (it is understood that you will use graphics/plotting software, and optimization software like LP, QP solvers, LBFGS as needed)
- are there any theoretical/methodological parts that you want to extend/develop?
- are there any problems with the paper that you will try to fix?
- data sets you will use (from the paper or from elsewhere) (optional, for now) a plan of the experimental evaluation
- extensions
- a schedule of the implementation stages by week (including week 2)
- a list of other related papers (background, followup research, competing methods, applications) that are relevant to your project