Science-Wise False Discovery Rate

Nathan (Nat) Goodman

October 2, 2017

An R script and short article exploring the statistical concept of science-wise false discovery rate (SWFDR). Some authors use SWFDR and its complement, positive predictive value, to argue that most (or, at least, many) published scientific results must be wrong unless most hypotheses are a priori true. I disagree. While SWFDR is valid statistically, the real cause of bad science is "Publish or Perish".

Overview

The SWFDR software reimplements the core idea in David Colquhoun's fascinating paper, "An investigation of the false discovery rate and the misinterpretation of p-values" and further discussed in Felix Schönbrodt's blog post, "What's the probability that a significant p-value indicates a true effect?" and related ShinyApp. The term *science-wise false discovery rate* is from Jager and Leek's paper, "An estimate of the science-wise false discovery rate and application to the top medical literature". John Ioannidis's landmark paper, "Why most published research findings are false", is the origin of it all.

The *false discovery rate* (*FDR*) is the probability that a significant p-value indicates a false positive, or equivalently, the proportion of significant p-values that correspond to results without a real effect. The *science-wise false discovery rate* (*SWFDR*) is the probability that a significant p-value in the *published literature* is a false positive, or equivalently, the proportion of published significant results that are wrong. The scripts simulate a large number of problem instances and produce graphs of SWFDR for a range of parameter values.

Installation and Usage

The simplest way to get the software is to download the script swfdr.R from the R subdirectory of the repository. The script uses base R capabilities only and will run "out of the box" on any (reasonably modern) R installation.

The recommended way to run the script is to `source` it into your R session and run the statement `run();` as shown below.

```
## This code block assumes your working directory is the root of the distribution.

source('R/swfdr.R');
run();
```

This runs the program with default parameters producing four graphs similar to the ones below. The default computation performs 2.5×10^5 simulations (taking about 3 minutes on my small Linux server).
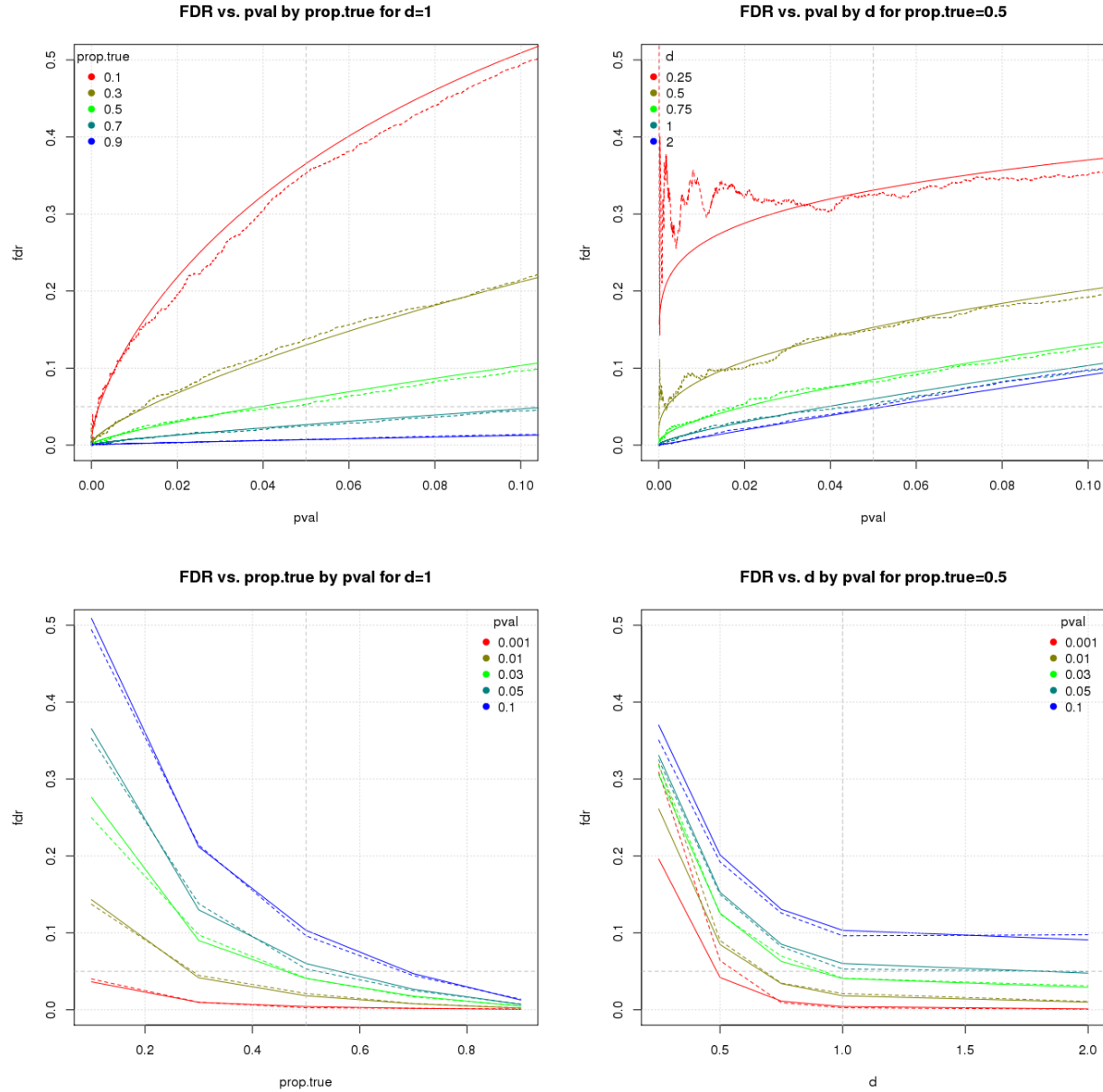
The notation is

- solid lines show theoretical results; dashed lines are empirical results from the simulation
- *fdr.* false discovery rate
- *pval.* p-value cutoff for significance
- *prop.true.* proportion of simulated cases that have a real effect
- *d.* standardized effect size, aka *Cohen's d*

See Also

The article discussing the results in html and pdf is on the GitHub Pages site associated with this repository. It's also in the repository as files swfdr.html and swfdr.pdf. (But note that GitHub, unlike GitHub Pages, renders html files as raw text).

Author

Nathan (Nat) Goodman, (natg at shore.net)

Bugs and Caveats

Please report any bugs, other problems, and feature requests using the GitHub Issue Tracker. I will be notified, and you'll be apprised of progress.

Copyright & License

Copyright (c) 2017 Nathan Goodman

The software is **open source and free**, released under the MIT License. The documentation is **open access**, released under the Creative Commons Attribution 4.0 International License.