Science-Wise False Discovery Rate

Nathan (Nat) Goodman

June 1, 2017

```
out=knitr::opts_knit$get("rmarkdown.pandoc.to");
paste(sep='','output format=',out);
```

```
## [1] "output format=latex"
```

A collection of R scripts exploring the statistical concept of science-wise false discovery rate (SWFDR) and illustrating software design choices in R. The base script is ***`swfdr_base.R`****, a simple implementation that uses base R capabilities only. Other scripts extend the base implementation by providing solutions to some of the Exercises for the Reader.*

`swfdr_base.R`

This script 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 script produces graphs of FDR for a range of parameter values.

Installation and Usage

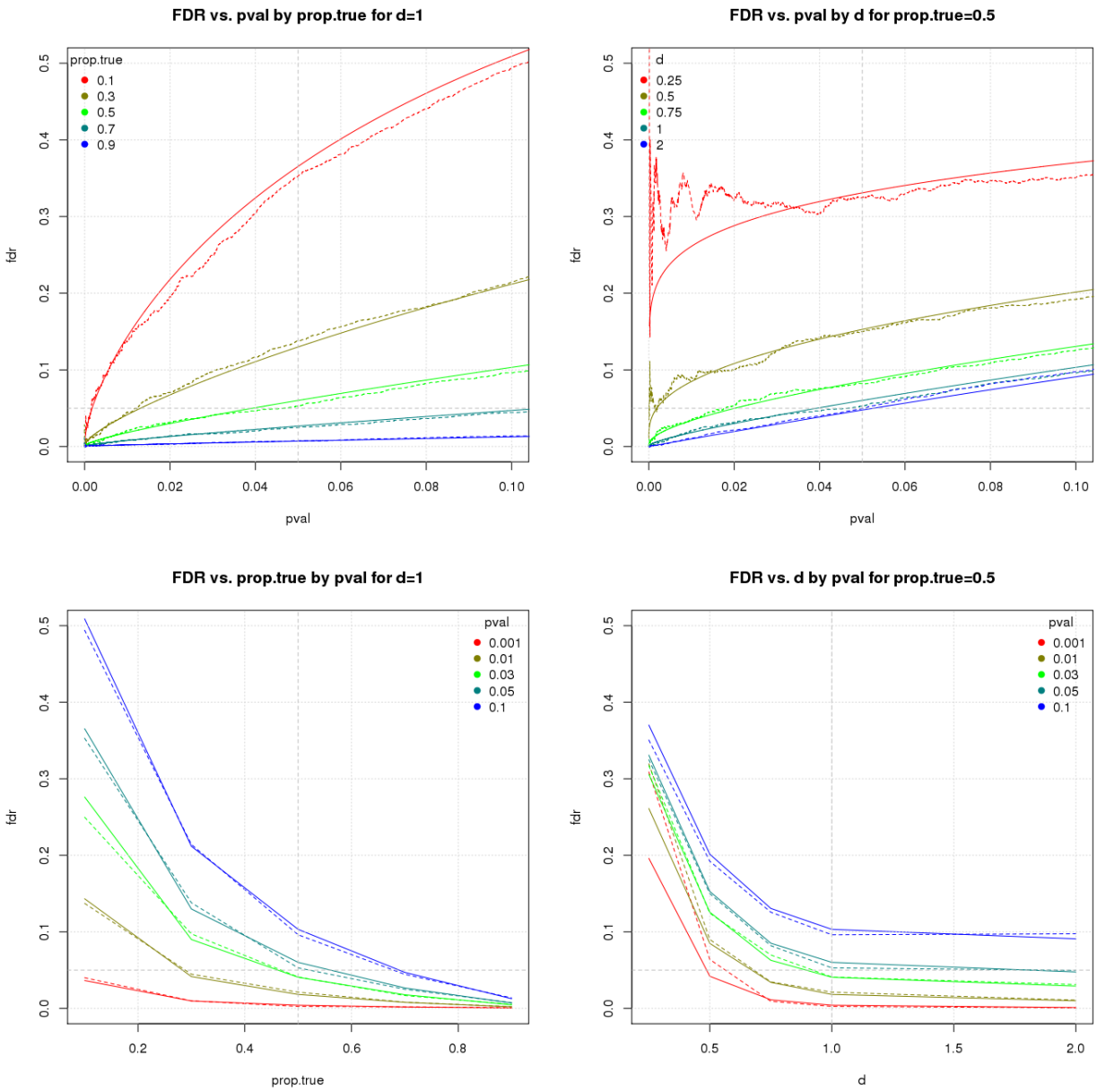
The simplest way to get the software is to download the entire repository. The software is in the `script` subdirectory. At present, the only available script is `swfdr_base.R`. This program uses base R capabilities only and will run "out of the box" on any (reasonably modern) R installation.

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

```
# this code block assumes your working directory is the root of the repository

source("script/swfdr_base.R");
run();
```

This runs the programs with default parameters produce four graphs similar to the ones below. The default process 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 User Guide shows how to run the program with non-default parameters and briefly describes the rest of the distribution
- Science: Science-Wise False Discovery Rate explains the scientific concepts underlying the software
- Software: Science-Wise False Discovery Rate discusses the software design and design choices

- Exercises: Science-Wise False Discovery Rate provides exercises for the reader to improve the program
- NEWS lists major differences between releases
- R-style function-level documentation

The documentation is also in the `doc/` subdirectory.

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.

`swfdr_base.R` is pretty basic. See the User Guide and Exercises for a list of known limitations, and other scripts in the `script/` subdirectory for solutions to some of these problems.

Copyright & License