Nat's Website

Nathan (Nat) Goodman

December 1, 2019

Welcome to my incipient website! It's just a landing page with links to recent working papers and blog posts. Clicking on a title takes you to the latest released version of the article on GitHub Pages. For articles published on blogs elsewhere, there's also a link to the posting.

About Me

I am a retired computer scientist. I spent the first half of my career doing mainstream CS and the second half in bioinformatics. I split my years pretty evenly between academia and industry. My academic positions included faculty spots at Harvard and Boston University, and research scientist positions at the Whitehead Institute Genome Center (led by Eric Lander), the Jackson Laboratory, and the Institute for Systems Biology (in Lee Hood's lab). In industry, I mostly worked in startups and other small companies with positions running the gamut from programmer to president. As a retiree, I'm working on an eclectic mix of projects that interest me, stopping from time-to-time to write papers and posts on aspects that might interest others. I'm also doing a little consulting for select clients developing innovative, important products.

Blog Posts (see Working Papers below)

- Ladies And Gentlemen, I Introduce To You, "Plausibility Limits" by Nat, December 1, 2019, with editorial assistance by Bob Reed (thanks!) kindly posted by Bob Reed on The Replication Network (TBD) *Confidence intervals get top billing as the alternative to significance. But beware: confidence intervals rely on the same math as significance and share the same shortcominings. Confidence intervals don't tell where the true effect lies even probabilistically. What they do is delimit a range of true effects that are broadly consistent with the observed effect.*
- What's the True Effect Size? It Depends What You Think by Nat, October 2, 2019, with edits by Bob Reed (thanks!) kindly posted by Bob Reed on The Replication Network *What's the true effect size? That's my bottom line question when doing a study or reading a paper. I don't expect an exact answer, of course. What I want is a probability distribution telling where the true effect size probably lies. I used to think confidence intervals answered this question, but they don't except under artificial conditions. A better answer comes from Bayes's formula. But beware of the devil in the priors.*
- A Friendly Debate About Pre-Registration by Nat and Bob Reed, June 19, 2019 *I'm generally pessimistic about the benefits of pre-registration, while Bob Reed is generally optimistic. The post is a back-and-forth dialogue between us. My main point is that pre-registration is essential for confirmatory work but irrelevant for exploratory research. And moreover, most published research is exploratory even when the authors claim the work was hypothesis-driven. Bob argues that pre-registration is valuable for exploratory research, too, by reducing the temptation for investigators to graze through data looking for patterns. Good arguments on both sides!*
- Your P-Values are Too Small! And So Are Your Confidence Intervals! by Nat, May 1, 2019, with edits by Bob Reed (thanks!) kindly posted by Bob Reed on The Replication Network *An oft-overlooked detail in the significance debate is the challenge of calculating correct p-values and confidence intervals, the favored statistics of the two sides. Standard methods rely on assumptions about how the data were generated and can be way off when the assumptions don't hold. Papers on heterogenous effect sizes by Kenny and Judd and McShane and Böckenholt present a compelling scenario where the standard calculations are highly optimistic. Even worse, the errors grow as the sample size increases, negating the usual heuristic that bigger samples are better.*

- When You Select Significant Findings, You're Selecting Inflated Estimates by Nat, February 15, 2019, with edits by Bob Reed (thanks!) kindly posted by Bob Reed on The Replication Network *Replication researchers cite inflated effect sizes as a major cause of replication failure. It turns out this is an inevitable consequence of significance testing. The reason is simple. The p-value you get from a study depends on the observed effect size, with more extreme observed effect sizes giving better p-values; the true effect size plays no role. Significance testing selects studies with good p-values, hence extreme observed effect sizes. This selection bias guarantees that on average, the observed effect size will inflate the true effect size. The overstimate is large, 2-3x, under conditions typical in social science research. Possible solutions are to increase sample size or effect size or abandon significance testing.*
- Preregistration: Hold the Bus! by Nat, January 1, 2019, with edits by Bob Reed (thanks!) kindly posted by Bob Reed on The Replication Network *A short rejoinder to a recent news piece in Nature that reported in glowing terms on the "first analysis of 'pre-registered' studies". The study in question is a preprint Open Science challenges, benefits and tips in early career and beyond. My assessment is not so rosy.*
- Systematic Replication May Make Many Mistakes. Short version of Systematic Replication Has Limited Power to Detect Bad Science below by Nat, September 23, 2018 kindly posted by Bob Reed on The Replication Network *Replication seems a sensible way to assess whether a scientific result is right. The intuition is clear: if a result is right, you should get a significant result when repeating the work; if it it's wrong, the result should be non-significant. I test this intuition across a range of conditions using simulation. For exact replications, the intuition is dead on, but when replicas diverge from the original studies, error rates increase rapidly. Even for the exact case, false negative rates are high for small effects unless the samples are large. These results bode ill for large, systematic replication efforts, which typically prioritize uniformity over fidelity and limit sample sizes to run lots of studies at reasonable cost.*
- Science-Wise False Discovery Rate Does Not Explain the Prevalence of Bad Science by Nat, October 16, 2017 kindly posted by Daniel Lakens on The 20% Statistician *This article explores 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".*

Working Papers

- Bayesian Basics for Estimating True Effect Size by Nat, September 15, 2019 *This working paper shows an R implementation of Bayes's formula as used in the blog post What's the True Effect Size? It Depends What You Think to estimate true effect size in two group difference-of-mean studies. The main point is that the implementation is really, really simple. With all the fuss and bother about Bayesian methods, I imagined it would be incredibly hard to code. Happily not.*
- Supplementary Material for Mistakes of Significance repo by Nat, June 12, 2019 *Contains supplementary material for the README document associated with this repo. It's not meant to be read as a standalone document. It's more of a "usage test" demonstrating that all major functions work in the limited setting of this document. It proved it's worth in this regard, driving out several bugs and leading to software improvements. The document exercises all capabilities of the program: simulation and other data generation, plotting of figures, and saving of summary results. It has figures showing each plot function applied to each kind of data that can be sensibly displayed using that function.*
- Measuring Replication Success by Nat, December 5, 2018 *Statisticians have devised numerous statistical tests for deciding whether a replication passes or fails, thus validating of refuting the original result. I call these tests "measures". Elsewhere I report results from simulating these tests across a range of conditions. Here I describe the measures themselves and provide implementation details for ones that aren't obvious.*

- Noncentral d2t-Distribution Applied to Measures of Replication Success by Nat, December 3, 2018 *Statisticians have devised numerous statistical tests for deciding whether a replication passes or fails, thus validating of refuting the original result. The noncentral t-distribution underlies many of these tests. I developed software that repackages R's built-in functions for the noncentral t-distribution to operate on sample size (instead of degrees of freedom), population effect size (instead of the noncentrality parameter), and observed effect size (instead of the t-statistic). I call this the d2t-distribution.*
- Supplementary Material for Systematic Replication May Make Many Mistakes above by Nat, November 7, 2018 *Contains results that didn't fit in the blog post as published for reasons of space or pedagogy. It's not terribly coherent as a standalone document. Sorry.*
- Systematic Replication Has Limited Power to Detect Bad Science by Nat, August 28, 2018 *Replication seems a sensible way to assess whether a scientific result is right. The intuition is clear: if a result is right, you should get a similar answer when repeating the work; if it it's wrong, your answer should be quite different. Statisticians have devised numerous statistical tests for deciding whether a replication passes or fails thus validating of refuting the original result. I simulate many of these tests across a range of conditions. For exact replications, simple significance testing works fine as a validation test, but when replicas diverge from the original studies no test works well. Much of the replication literature focuses, perhaps unwittingly, on methods for testing whether the studies are similar; these tests work quite poorly under all conditions analyzed here. Many caveats apply, but if correct, my results bode ill for large, systematic replication efforts, which typically prioritize uniformity over fidelity to run lots of studies at reasonable cost.*
  - version 1.00 html, pdf

Comments Please!

Please post comments on Twitter or Facebook, or contact me by email natg@shore.net.