Simulating power in practice

Author: Nicholas G Reich

This material is part of the **statsTeachR** *project*

Today's Lecture

- What is statistical power?
- Why/how might we want to simulate it?
- An example

Refresher: statistical power

Definition of statistical power

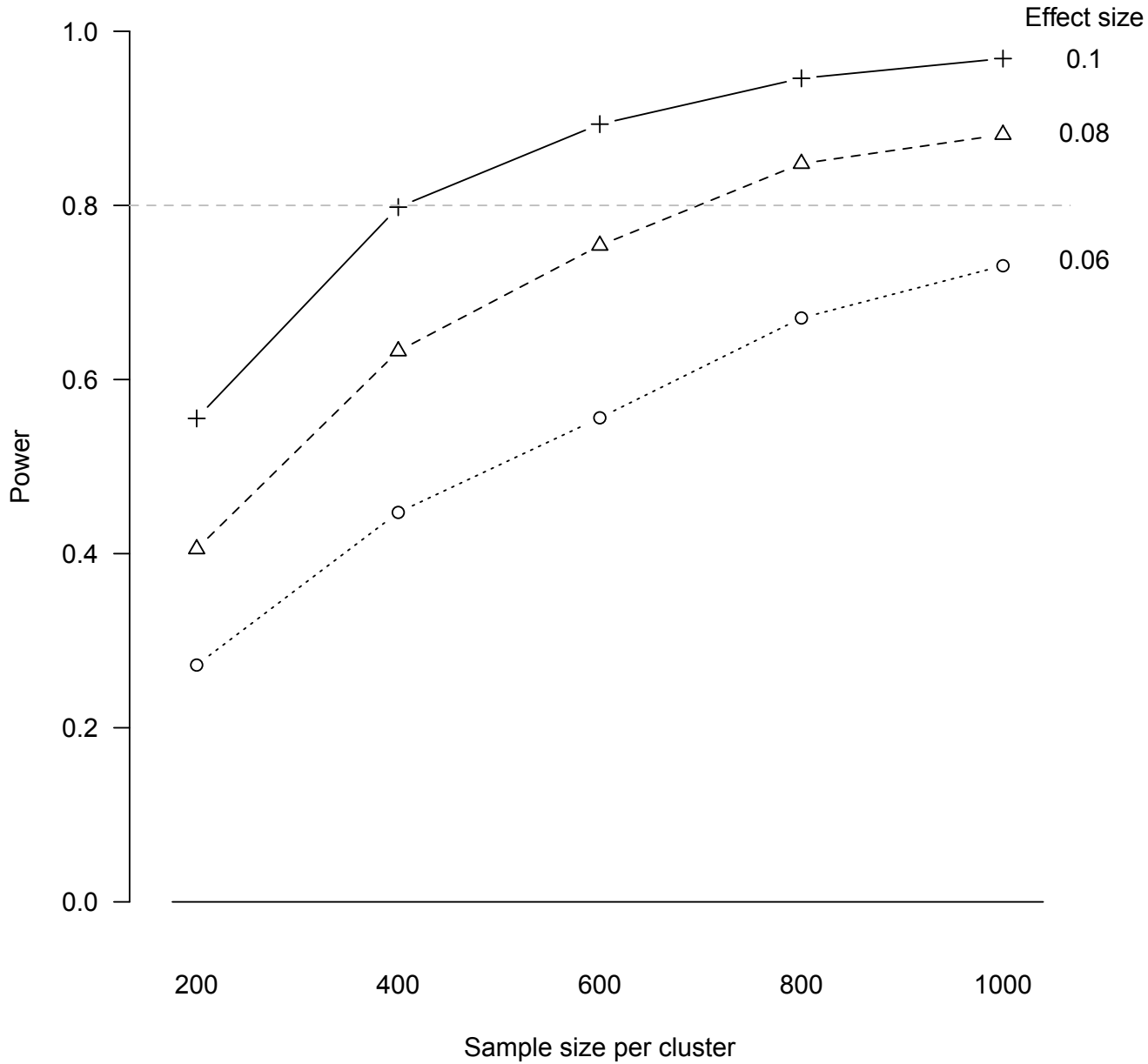
- The ability of a method/test to detect an effect, conditional on that effect actually existing.
- The probability that our test rejects the null hypothesis when the null hypothesis is not true.

Characteristics that impact power

What imapct do increases in these features have on power?

- sample size
- effect size
- variance of outcome
- variance of predictors
- number of predictors
- grouped/clustered observations

Typical power curve



Reich NG, Myers JA, Obeng D, Milstone AM, Perl TM. Empirical power and sample size calculations for cluster-randomized and cluster-randomized crossover studies. PLoS ONE. 2012. 7(4): e35564.

"Post-hoc" power calculations are controversial

It is *always* preferable to calculate power prior to running your analysis.
The Abuse of Power

Formula-based power calculation

Many simple tests have formulas for power, these ...

- are easy to use
- may require you to estimate parameters from existing data (or make up justifiable numbers to plug in)
- are often appropriate for simple tests
- assume all standard assumptions are met
- are only available for simple/standard tests

Simulation-based power calculation

Calculating power via simulation is a tradeoff: computational complexity for customization and flexibility.

Power simulations...

- are available for any setting where you can simulate data (not limited to simple scenarios)
- can be used to preserve complex correlation sturctures in predictors (resample your X'x)
- are not assumption- or parameter-free
- often require more complicated coding
- may be computationally intensive (i.e. need a long time to run)

Example: t-test power calculation "by hand"

T-test: comparing mean between two groups

- $\mu_1 = 5, \mu_2 = 7$
- $\sigma_1^2 = \sigma_2^2 = 5$ (assume known)
- $n_1 = n_2 = 20$
- Type I error rate $= \alpha = 0.05$
- H_0: $\mu_1 - \mu_2 = 0$

$$Power = 1 - \beta = Pr\left(Z > 1.96 - \frac{|\mu_1 - \mu_2|}{\sqrt{2\sigma^2/n}}\right)$$

```
pnorm(1.96 - 2/sqrt(2*5/20), lower.tail = FALSE)
```

```
## [1] 0.8074
```

Example: t-test power calculation "black box"

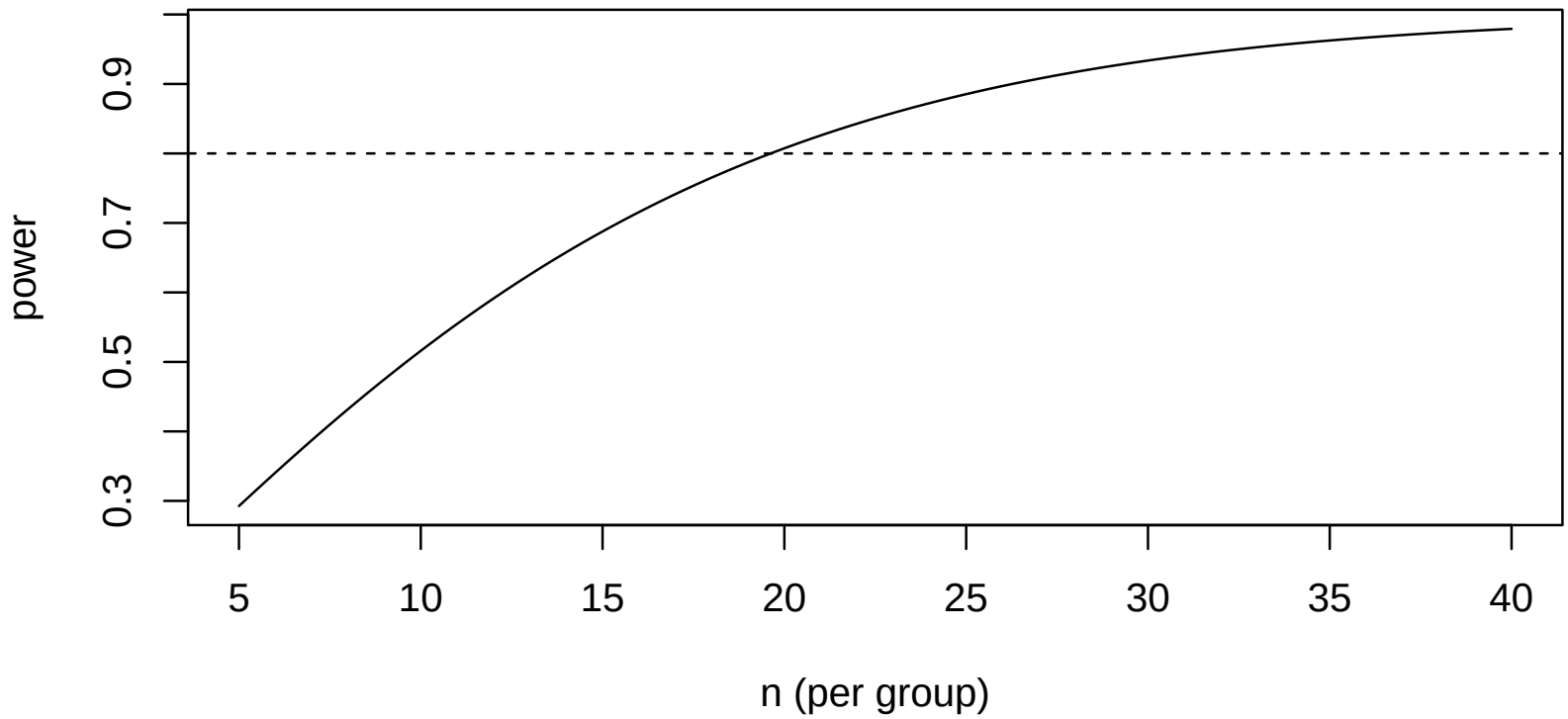
Compare to another method, which uses numerical optimization

```
pnorm(1.96 - 2/sqrt(2*5/20), lower.tail = FALSE)

## [1] 0.8074

power.t.test(n = 20, delta = 2, sd = sqrt(5), sig.level=0.05)

##
##      Two-sample t test power calculation
##
##               n = 20
##           delta = 2
##              sd = 2.236
##       sig.level = 0.05
##           power = 0.7871
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Example: t-test power calculation (graph)

Evaluate power across sample sizes

```
curve(pnorm(1.96 - 2/sqrt(2*5/x), lower.tail = FALSE),
      from=5, to=40, ylab="power", xlab="n (per group)")
abline(h=.8, lty=2)
```



Example: t-test power calculation "simulated"

```
nsim <- 1000
n <- 20
mu1 <- 5
mu2 <- 7
s2 <- 5
reject <- rep(0, nsim)
for(i in 1:nsim){
    x <- rnorm(20, mean=mu1, sd=sqrt(s2))
    y <- rnorm(20, mean=mu2, sd=sqrt(s2))
    tt <- t.test(x, y)
    reject[i] <- tt$p.value<.05
}
mean(reject)

## [1] 0.786
```

Power by simulation: different flavors

Option 1: generate all data from scratch

- will generate "clean" data
- hard to insert authentic noise: outliers, missingness, correlated predictor structure

Option 2: resample predictors from a *training* dataset, simulate outcome

- preserves structure of real predictor data
- requires a large dataset similar to the one you will be analyzing
- you should not, in general, do this type of computation on the actual dataset that you are analyzing – best to have a "training" dataset, similar to but independent from the one you will be analyzing

Power by simulation: resampling algorithm

Resampled power algorithm

1. resample (with replacement) rows of design matrix
2.

Adapted from Kleinman and Huang (2014), and Meyers et al. (2014)