Six percent power and barely selective inference

Art B. Owen

Stanford University

Confidence intervals with control of the sign error in low power settings

arXiv:1610.10028

Six percent power

If you test $H_0 : \theta = 0$ at $\alpha = 5$% and have 6% power then

1) Significant findings grossly exaggerate $|\theta|$, and
2) the sign of $\hat{\theta}$ is often wrong.

Andrew Gelman's blog 2014.

80% power

Q: aren't we supposed to have 80% power?

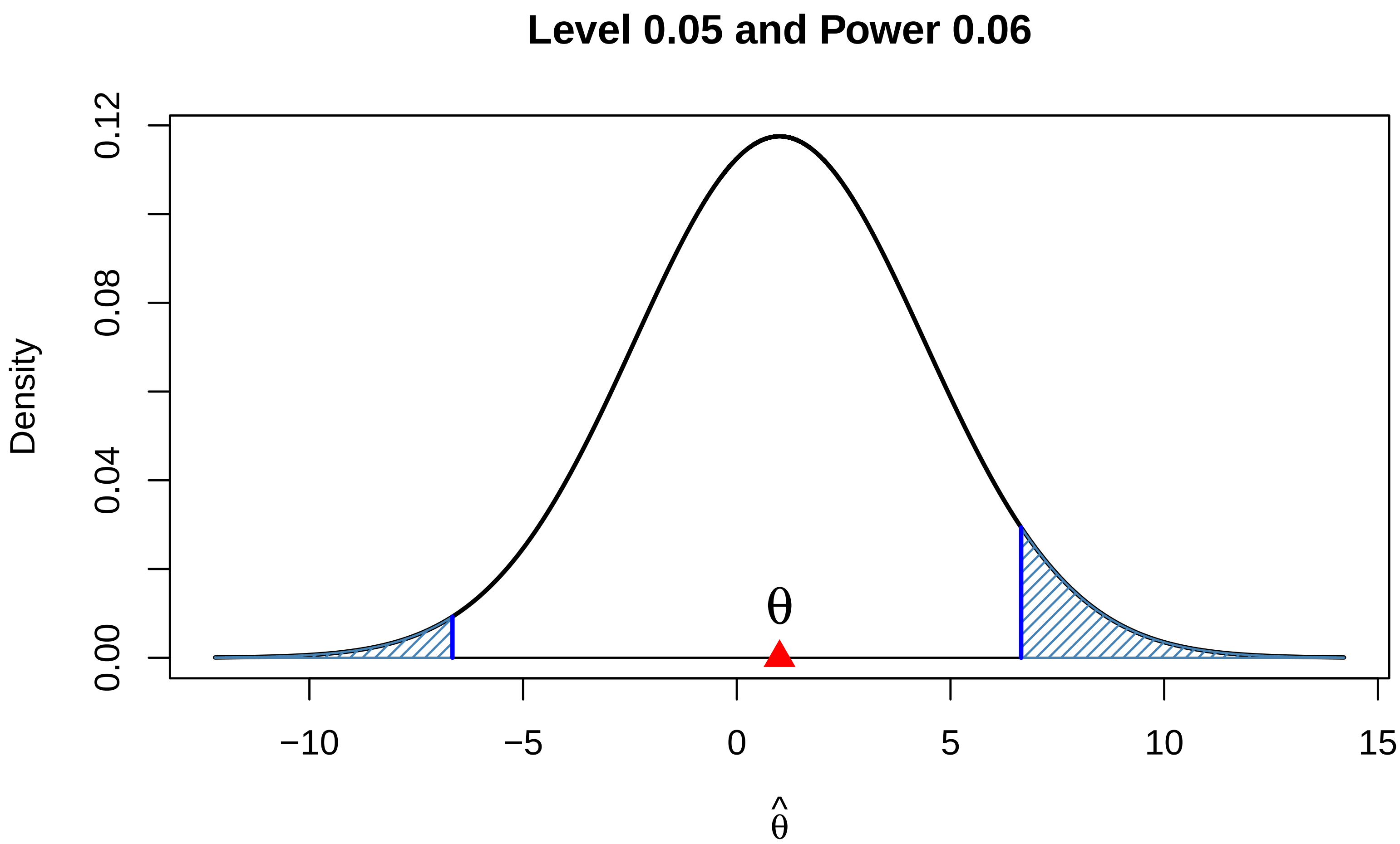
A: maybe, but we don't.

Low power settings are common

1) A/B testing on a nearly optimal product
2) Genomic or other screening
3) High dimensional models

all ***include*** cases where power is low.

Gelman's example



WLOG, true $\theta = 1$. Rejections

always have $|\hat{\theta}| \gg |\theta|$

often have $\mathrm{sign}(\hat{\theta}) \neq \mathrm{sign}(\theta)$.

Point nulls are problematic

It is almost never plausible that $\theta = 0$ exactly.

Even if the science says so, measurements are not perfect.

Then why test $\theta = 0$?

Old answer:

1) Reject $\theta = 0 \implies$ get $\operatorname{sign}(\theta)$.

2) Accept $\theta = 0 \implies$ don't get $\operatorname{sign}(\theta)$.

Gelman convinced me that part 1 is not right.

Old setup

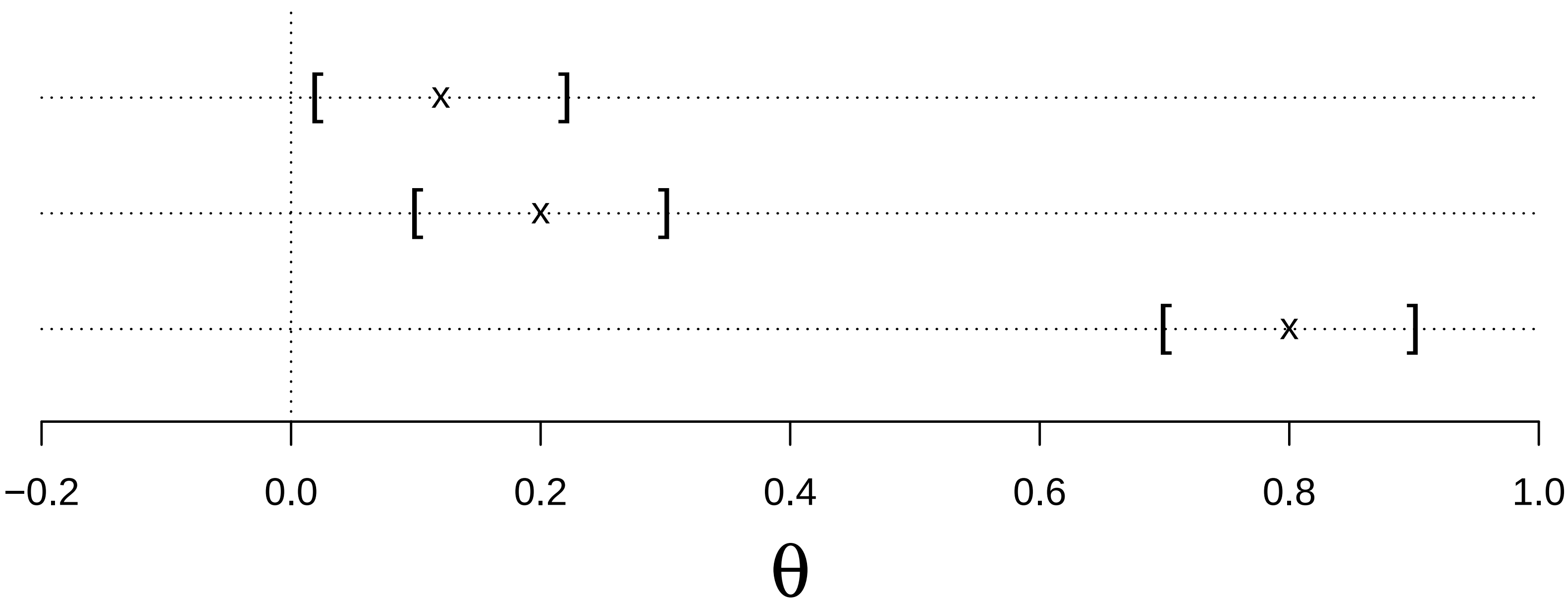
Test $\theta > 0$ versus $\theta = -\epsilon$.

Let $\epsilon \to 0$ to get hardest test.

Result is $H_0 : \theta = 0$.

Confidence intervals

If the whole CI is several CI-widths from 0 then we may confident about the sign.



Otherwise not.

Formulation for $\theta \neq 0$

- WLOG, $\theta = 1$ (change units)
- $\hat{\theta} \sim \mathcal{N}(\theta, \tau^2)$ (large sample)
- $s^2 \sim \tau^2 \chi^2_{(\nu)}/\nu$ (large sample)

Test statistic

As $\nu \to \infty$, $s^2 \xrightarrow{\mathrm{d}} \tau^2$ and $Z = \hat{\theta}/s \xrightarrow{\mathrm{d}} \mathcal{N}(1/\tau, 1)$

$$Z^2 = \frac{\hat{\theta}^2}{s^2} \xrightarrow{\mathrm{d}} \begin{cases} \chi^2_{(1)}, & \text{Null} \\ \chi'^{,2}_{(1)}(\tau^{-2}), & \text{Alternative} \end{cases}$$

Given α and $1 - \beta > \alpha$, we solve for τ and compute sign error probability.

Solving for τ

1) Define threshold based on α
2) Choose τ from β

The threshold

Reject H_0 iff $Z^2 \geqslant \chi^{2,1-\alpha}_{(1)}$ I.e., $|Z| \geqslant \Phi^{-1}(1-\alpha/2)$

The variance τ^2

Solve

$$\Pr\big(\chi'^{,2}_{(1)}(\tau^{-2}) \geqslant \chi^{2,1-\alpha}_{(1)}\big) = 1-\beta$$

for τ

R code

```
tau = function(alpha,powr){
# A Gaussian test at level alpha
# has power powr for Y ~ N( 1, tau^2 ).
# Solve for tau using noncentral chisquare.

aux = function(tau){
  1-powr - pchisq(qchisq(1-alpha,1),1,ncp=1/tau^2)
}

ur = uniroot( aux, lower=10^-9,upper=10^6,
    tol = .Machine$double.eps^0.9)
ur$root
}
```

Results

For $\alpha = 0.05$ and power $= 0.06$ we get

$$\tau \doteq 3.39$$[a]

So rejections happen if $|\hat{\theta}| \geqslant 1.96 \times 3.39 \doteq 6.65$

With low power comes great exaggeration

At least 6.65-fold. On average about 8-fold

Gelman and Carlin's (2014) type M error.

Also wrong signs

About 20% of rejections have $\hat{\theta} < 0$

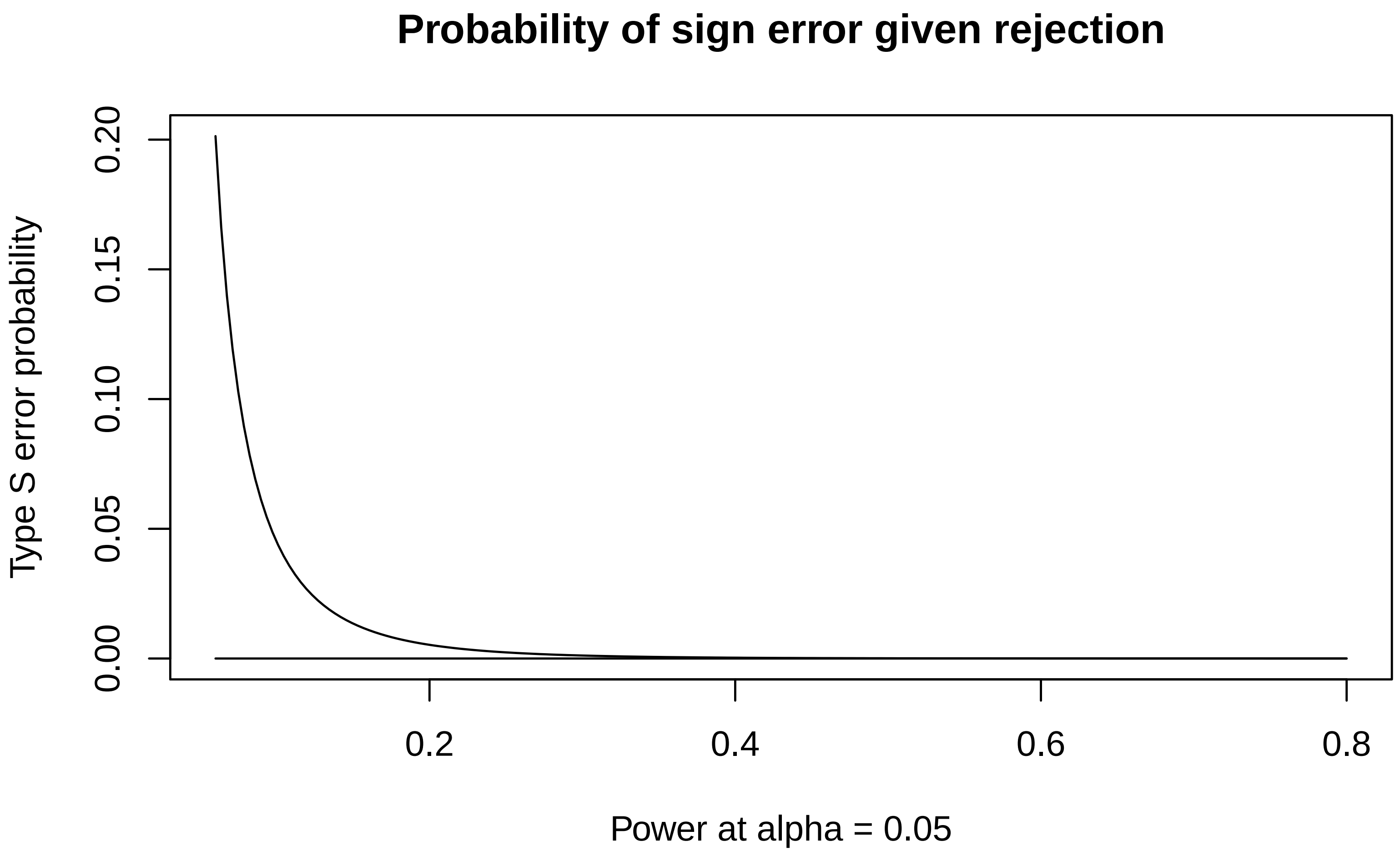
Gelman and Carlin's (2014) type S error.

Nota bene

I don't advocate $\alpha = 0.05$. Illustration only. Do not attempt.

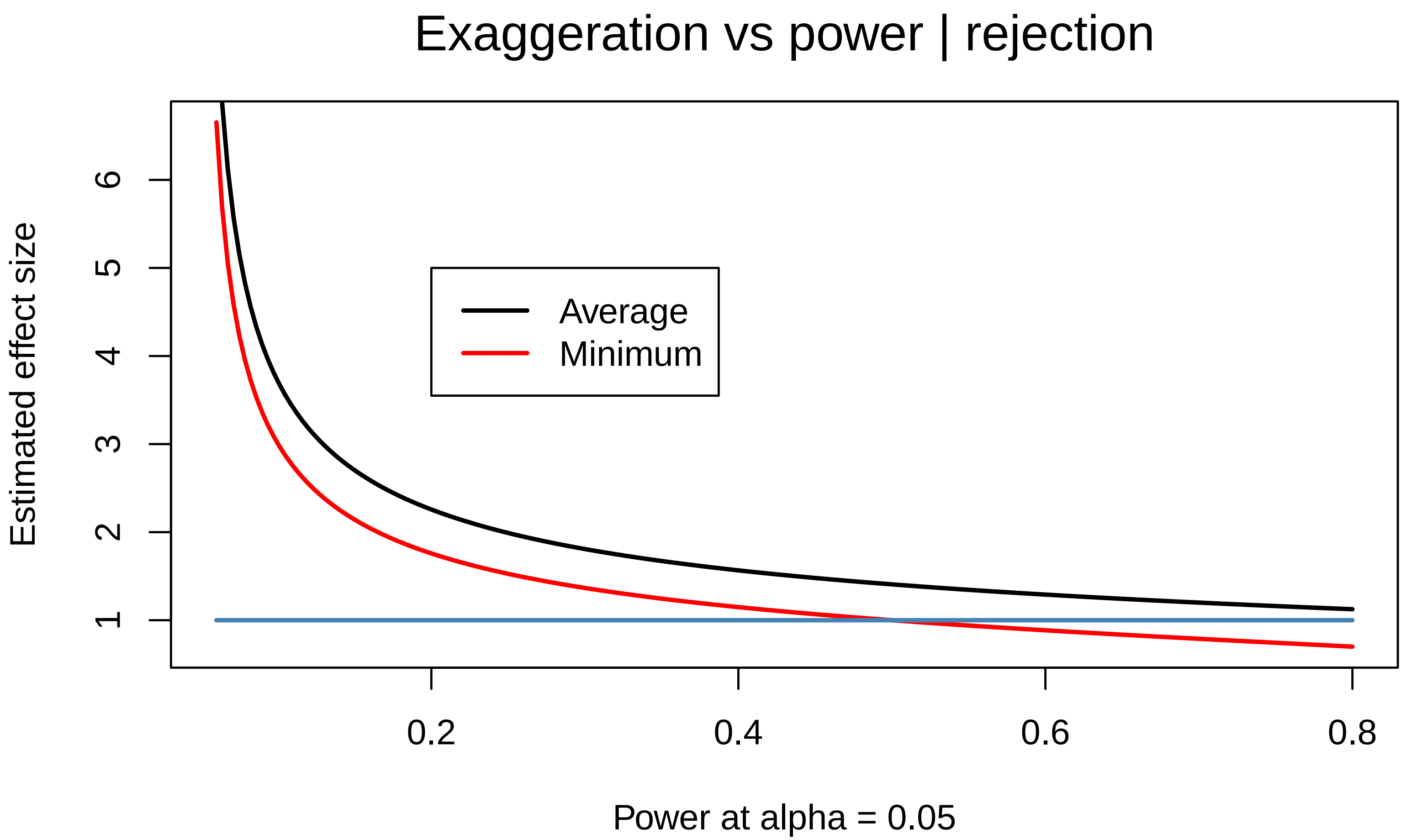
[a] Actually 3.394507. Gelman's blog really illustrated ≈ 5.5% power not 6% 24% wrong signs (not 20%) and over 9-fold exaggeration (not 6.65).

Sign error vs power



If you have 80% power at $\alpha = .05$, your signs are safe.

Exaggeration vs power



$$\mathbb{E}\left(\frac{|\hat{\theta}|}{|\theta|} \mid \text{Reject } H_0\right) \quad \text{and} \quad \min\left(\frac{|\hat{\theta}|}{|\theta|} \mid \text{Reject } H_0\right)$$

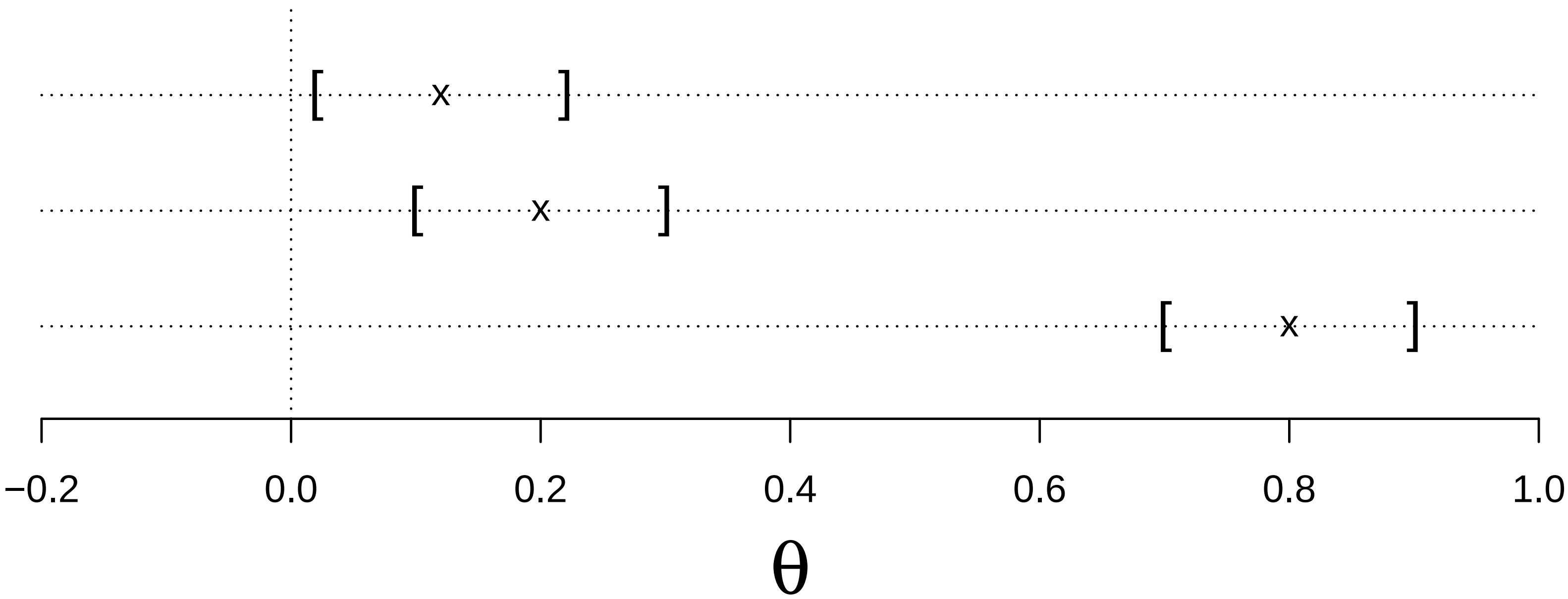
Confidence in the sign

Suppose that we only declare $\text{sign}(\theta) = \text{sign}(\hat{\theta})$ when

$$\text{dist}(\text{CI}, 0) \geqslant \text{dist}(\hat{\theta}, \text{CI}^c)$$

There is at least half an CI width between the CI and 0

Middle example below



Analysis for $\alpha = 0.05$

Event R_1: we rejected H_0 if $|\hat{\theta}/s| \geqslant 1.96$.

Event R_2: we declare sign if $|\hat{\theta}/s| \geqslant 2 \times 1.96 = 3.92$

Conditional probability of a sign error

$$
\begin{aligned}
&\Pr(\text{ called the wrong sign } \mid \text{rejected } H_0; H_0) \\
&\leqslant \Pr(R_2 \mid R_1; H_0) \\
&= \frac{\Pr(R_2 \cap R_1; H_0)}{\Pr(R_1; H_0)} \\
&= \frac{\Pr(R_2; H_0)}{\Pr(R_1; H_0)} \\
&= \frac{2\Phi(-3.92)}{\Phi(-1.96)} \\
&\doteq 0.0018.
\end{aligned}
$$

Conservatively assumes all sign calls wrong. For Gaussians, $1/2$ would be wrong under H_0. Getting $0.0018/2 = .0009 \doteq 0.001$.

Barely selective inference

Pick two thresholds $\alpha_1 > \alpha_2 > 0$

R_1: reject H_0 at level α_1

R_2: reject H_0 at level α_2

Under R_2 declare $\text{sign}(\theta) = \text{sign}(\hat{\theta})$

Probability of a sign error

$$\begin{aligned} &\Pr(\text{sign}(\hat{\theta}) \neq \text{sign}(\theta)) \\ &\leqslant \Pr(\text{sign}(\hat{\theta}) \neq \text{sign}(\theta); H_0) && |\theta| \to 0 \text{ is hardest} \\ &= \frac{1}{2}\frac{\alpha_2}{\alpha_1} && \text{Gaussian symmetry} \\ &\equiv \alpha_S \end{aligned}$$

Philosophically hard case

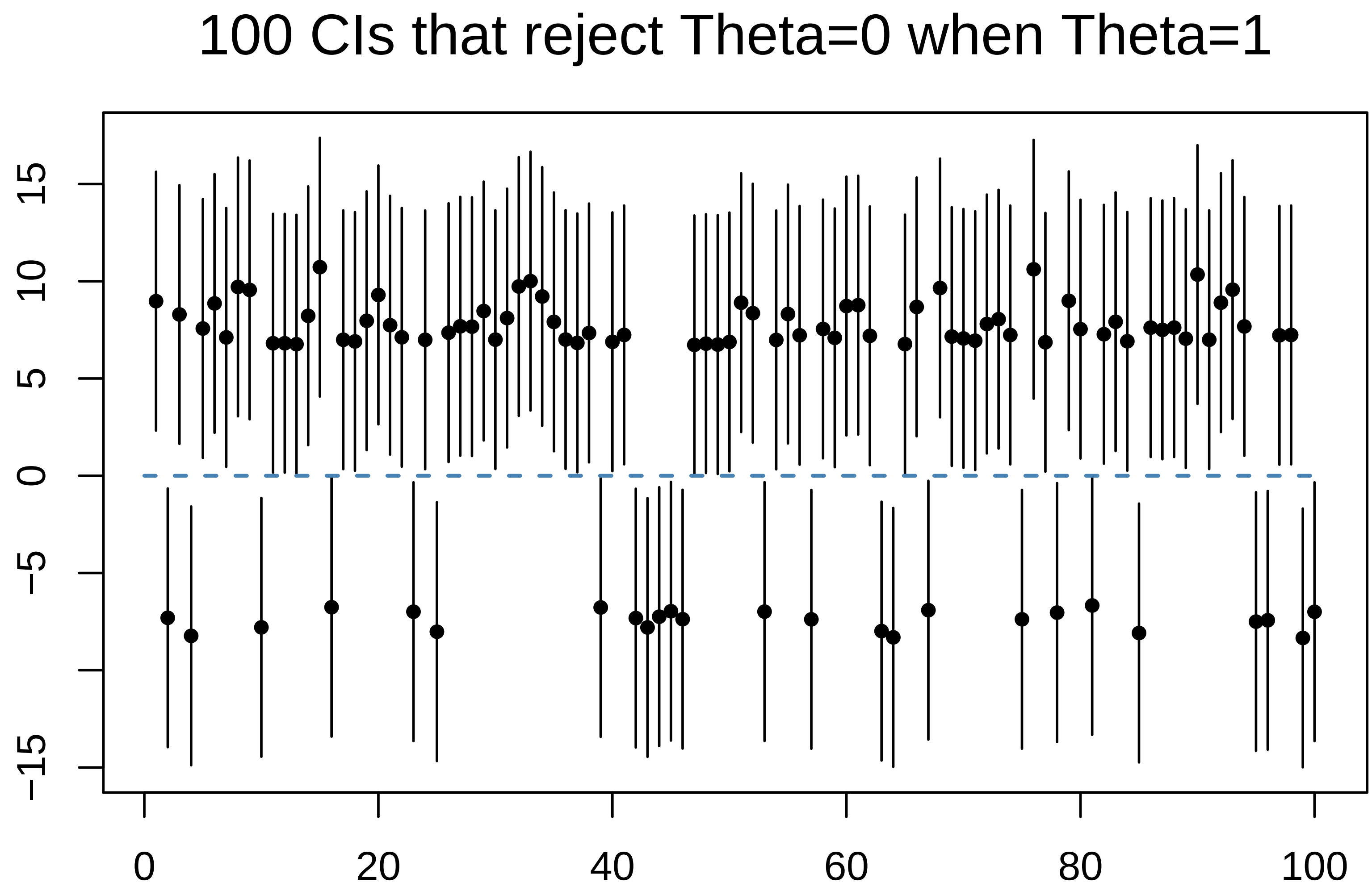
If R_1 but not R_2:

we declare $\theta \neq 0$ but do not declare $\text{sign}(\theta)$

A conversation

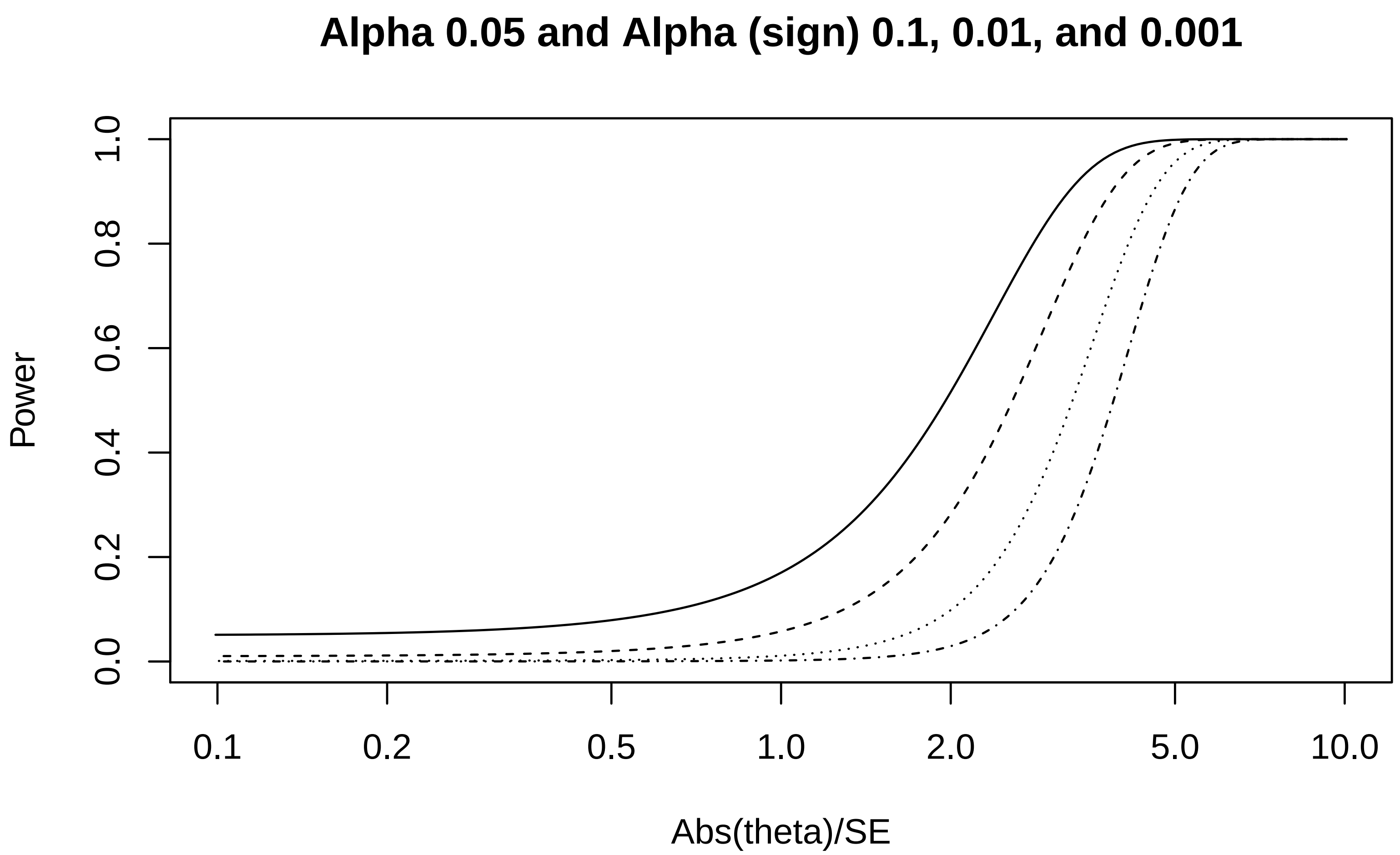
Domain person	Statistician
Here's my data, tell me about θ	
	First, $\hat{\theta} > 0$
	Also we reject $H_0 : \theta = 0$
So is $\theta > 0$?	
	Maybe
	but I'm not confident

Explaining the hard case



$\alpha = 0.05$ power 0.06 $\nu \to \infty$

Power and sign power



Top curve is $\mathrm{Pr}(\text{reject } H_0)$ vs $|\theta|/\tau$.

Others are $\mathrm{Pr}(\text{declare sign}) \doteq \mathrm{Pr}(\text{declare correct sign})$

Sign p-values

Test 1000s of hypotheses at α_1.

For each p_2 is smallest α with rejection of H_0

For those with $p < \alpha_1$, let

$$p_S = \frac{p_2}{2\alpha_1}$$

This “sign p-value” assumes α_1 is pre-specified.

One sided tests

They don't fix the sign error problem. We use them when

1) Only $\theta \geqslant 0$ is possible, or

2) Only $\theta \geqslant 0$ is consequential.

[Or similar for $\theta \leqslant 0$]

The real doubt on those is not necessarily $\ll \alpha$

As before $Z \sim \mathcal{N}(1/\tau, 1)$.

Testing in the "negative direction" for positive θ rejects when

$$\hat{\theta}/\tau \leqslant \Phi^{-1}(\alpha)$$

Some numbers

$\alpha = 0.05$ and power $= 0.06 \implies \tau \doteq 3.39$ with probability

$$\Phi(\Phi^{-1}(0.05) - 1/\tau) \doteq 0.026$$

Wrong direction power is not $\ll \alpha$

Thanks

- Jelena Markovic, Andrew Gelman
- NSF DMS-1407397, DMS-1521145, IIS-1837931
- Kuffner & Samworth & Kolassa, organization
- Todd Kuffner, invitation plus many emails

Final thoughts

$$\Pr(\theta \in \mathrm{CI} \mid \text{called the sign}) \doteq 38\%$$

$$\Pr(\theta \in \mathrm{CI} \mid \text{called the correct sign}) \doteq 47\%$$