incentive-compatible critical values

Adam McCloskey (Colorado)

Pascal Michaillat (Brown)

January 5, 2021

HYPOTHESIS TESTING: A TOOL FOR SCIENTIFIC RESEARCH

- to assess the validity of existing paradigms & models
  - null hypothesis = existing paradigm/model is realistic
  - rejecting the null = identifying an anomaly
- to assess the effectiveness of new treatments & policies
  - null hypothesis = new treatment/policy is ineffective
  - rejecting the null = discovering an effective treatment/policy
- key parameter: significance level α (often 5%)
  - need to enforce size of test (type I error rate) $\leq \alpha$
  - otherwise rejections are uninformative

RESEARCHERS' INCENTIVES

- researchers have incentives to publish
  - academic job & salary
  - tenure & promotion
  - grants
- journals prefer to publish rejections of H_0
  - preference for "significant results"
  - possible reason: new anomalies/discoveries are important to scientific progress

⇝ researchers have incentives to search until they reject H_0

RESEARCH INCENTIVES INDUCE OVER-REJECTIONS

- classical assumption: t-statistic drawn from random sample
- in reality: scientists search for & report t-statistic rejecting $H_0 : \beta = \beta_0$
  - form t-statistics from multiple data sets & specifications:

$$X_i = \frac{\hat{\beta}_i - \beta_0}{\mathrm{se}(\hat{\beta}_i)}$$

  - then report $\max_i |X_i|$
- but under H_0: $\max_i |X_i| \sim \max_i |\mathcal{N}_i(0,1)| > |\mathcal{N}(0,1)|$

$\rightsquigarrow$ given research incentives: test size using classical CV $> \alpha$

INCENTIVE-COMPATIBLE CRITICAL VALUES (ICCVS)

- ensure test size $\leq \alpha$ while respecting researchers' incentives
- need to solve fixed-point problem:

$$CV = 1 - \alpha \text{ quantile of } G$$

  - G: distribution of reported test statistics under H_0
  - CV $\rightsquigarrow$ research payoffs $\rightsquigarrow$ researchers' behavior $\rightsquigarrow$ G
- challenge: need to model researchers' behavior to obtain G
  - $G < \Phi$ as soon as there is "p-hacking"

EXISTING METHODS TO CONTROL SIZE

- multiple testing & sequential analysis corrections (Howard et al, 2019)
  - $\rightsquigarrow$ requires number n of conducted studies to be known
- debiased meta-analyses (Stanley 2005, Andrews & Kasy 2019)
  - $\rightsquigarrow$ need to wait until many studies are conducted

$\rightsquigarrow$ both methods fail to account for CV $\rightsquigarrow$ n and G

- pre-analysis plans (as in medical RCTs, Journal of Dev Econ)
  - prevent exploratory analysis
  - difficult to implement with observational data

RESEARCHER'S BEHAVIOR UNDER CV z

- forms a sequence of latent t-statistics $X_i = (\hat{\beta}_i - \beta_0)/\text{se}(\hat{\beta}_i)$
- decides whether to conduct n^{th} study based on
  - expected reward v from exceeding z
  - cost $c(n)$ of conducting n^{th} study
  - researcher's beliefs $\mathbb{P}_n$ after $n-1$ studies
  - results of previous studies
- researcher conducts n^{th} study iff $|X_{n-1}| < z$ and
  - expected benefits > cost
  - $v \times \mathbb{P}_n(|X_n| > z \mid X_{n-1}, X_{n-2}, \ldots, X_1) \geq c(n)$

DEFINITION OF ICCV

- "IC": researchers behave optimally, so
  - reported test statistics $\leq$ $\max\left\{|X_1|, |X_2|, \ldots, |X_{N(z)}|\right\}$
  - $N(z)$ = maximum # of profitable studies n such that $|X_{n-1}| < z$ and

$$v \times \mathbb{P}_n(|X_n| > z \mid X_{n-1}, X_{n-2}, \ldots, X_1) \geq c(n)$$

- "CV": nominal level $= \alpha$, so need z to satisfy
  - $P_{H_0}\left(\max\left\{|X_1|, |X_2|, \ldots, |X_{N(z)}|\right\} > z\right) \leq \alpha$

COMPUTING ICCV z^*

- two-step computation:
  1. determine maximum # of profitable studies $z \mapsto N(z)$
  2. given $X_i \sim \mathcal{N}(0, 1)$ under H_0, find smallest z^* such that

$$P_{H_0}\left(\max\left\{|X_1|, |X_2|, \ldots, |X_{N(z^*)}|\right\} > z^*\right) \leq \alpha$$

- research cost $\uparrow \Rightarrow N(z) \downarrow \Rightarrow$ ICCV $\downarrow$
  - costly research
- research reward $\uparrow \Rightarrow N(z) \uparrow \Rightarrow$ ICCV $\uparrow$
  - "important" question / non-risky research

IN BASIC CASE: IMPOSSIBLE TO LEARN FROM TESTS

- iid studies: $X_i|\beta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ where $\theta = \frac{\beta - \beta_0}{\mathrm{sd}(\hat{\beta})}$
- no learning: researcher's subjective distribution $\mathbb{P}$ does not depend on previous studies' outcomes
- constant study cost: $c(n) = c$

⇝ continuation condition $v \times \mathbb{P}(|X_n| > z) \geq c$ is same at every step

⇝ if there is any research, researcher continues to conduct studies until rejection and size $= 100\%$

EXISTENCE OF ICCV

- impossibility result $\Longrightarrow$ ICCV does not exist
- impossibility result is broken, and ICCV is well defined, if
    1. researchers learn about β as they conduct studies
    2. there is correlation in t-statistics across studies
    3. study cost $c(n)$ is increasing in n
- we provide methods to calculate ICCV in various settings
- we calibrate costs, benefits and subjective beliefs based upon existing literature in a lab experiment setting

ICCV WITH RESEARCHER LEARNING

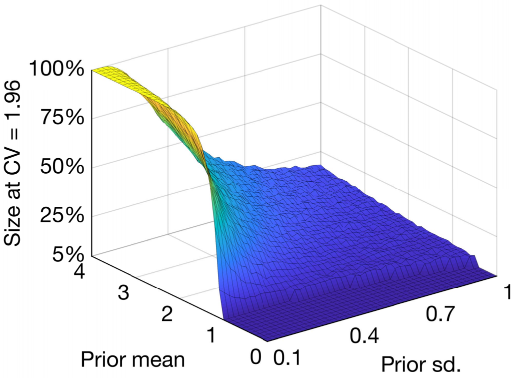
- researcher Bayesian updates beliefs about true value of β
- $\mathbb{P}_n$ depends upon outcomes of previous studies

$\rightsquigarrow$ continuation condition $v \times \mathbb{P}_n(|X_n| > z) \geq c$ depends upon outcome of previous studies via researcher's beliefs $\mathbb{P}_n$

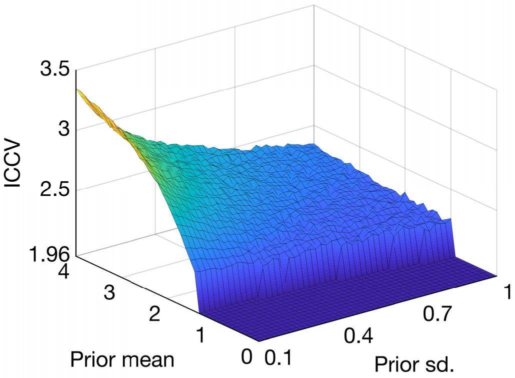
$\rightsquigarrow$ researcher eventually stops conducting studies without rejecting with probability > 0 if CV is large enough

  - ICCV exists

MONTE CARLO: RESEARCHER LEARNING



MONTE CARLO: RESEARCHER LEARNING



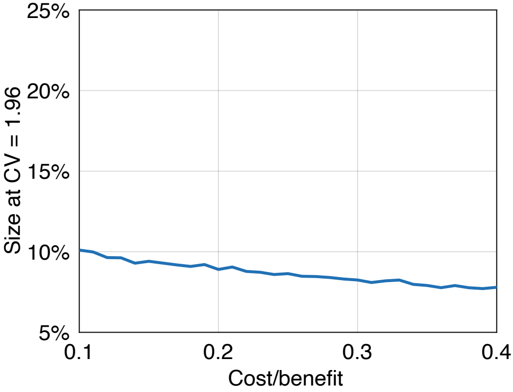
ICCV WITH DATA POOLING

- studies are correlated according to simple data pooling
  - i^{th} study's estimate of β is $\hat{\beta}_i = \frac{1}{iT}\sum_{j=1}^{iT} Y_j$

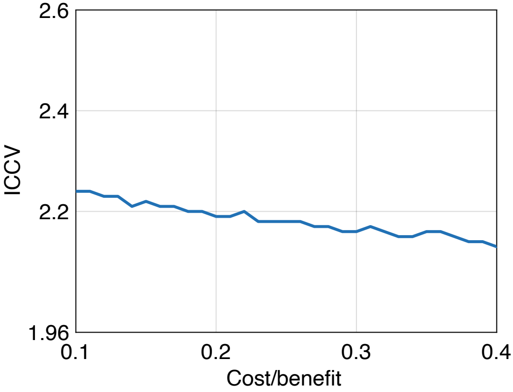
⇝ continuation condition $v \times \mathbb{P}(|X_n| > z \mid X_{n-1}, X_{n-2}, \ldots, X_1) \geq c$ depends upon outcomes of previous studies due to dependence across studies

⇝ researcher eventually stops conducting studies without rejecting with probability > 0 if CV is large enough (ICCV exists)

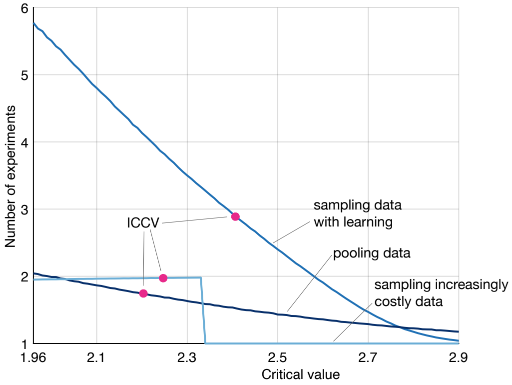
MONTE CARLO: DATA POOLING



MONTE CARLO: DATA POOLING



AVERAGE # OF STUDIES FOR DIFFERENT CVs



GENERAL CASE

- allow dependence and different means across latent t-statistics:

$$X_i|\theta_i \sim \mathcal{N}(\theta_i, 1) \text{ and } \operatorname{cov}(X_i, X_j) = \omega_{i,j}$$

- different θ_i's allows for
  - misspecification
  - studies of varying precision
- ω_{ij}'s known to researcher: large sample approximation with consistently estimable covariances between estimators
- examples: sampling/pooling datasets of different sizes, running different regression/model specifications

CONCLUSION

- if researchers report function of latent studies $f(X_1, \ldots, X_n) < \max\{X_1, \ldots, X_n\}$, ICCVs control size but are "conservative"
- flexible framework allows editor to choose directions of robustness
- given researcher behavior, ICCVs relatively insensitive to inputs
    - conservative rule-of-thumb setting CV equal to 3 controls size across wide range of researcher behavior and inputs
    - extrapolation to settings we did not study?

WORK IN PROGRESS

- current version of paper does not incorporate “continuation value” in researcher’s continuation decision
  - even if expected marginal benefit of next study does not exceed cost, there could be value in continuing due to allowing for rejection in future studies
- currently working on optimal stopping version of the problem to incorporate this