A Rigorous Framework for Automated Design Assessment and Type I Error Control

James Yang[1,2] T. Ben Thompson[2] Michael Sklar[2]

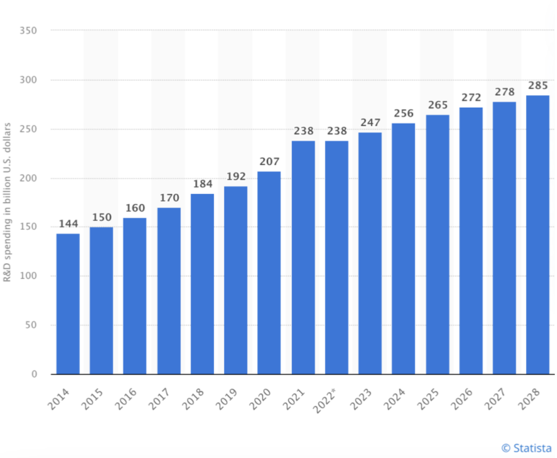
[1]Stanford University

[2]Confirm Solutions

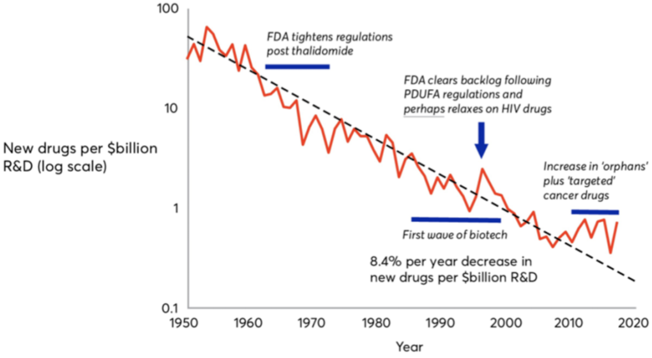
February 22, 2023

Table of Contents

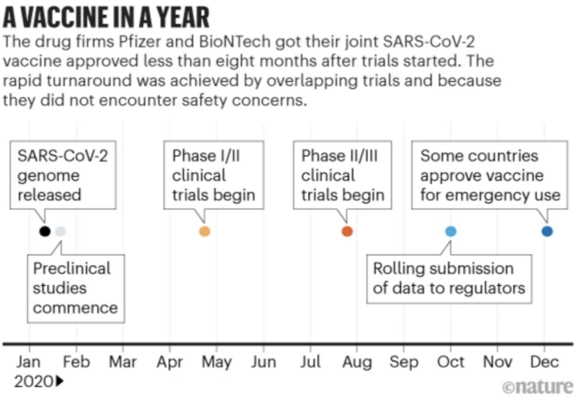
Pharma R&D is Growing



Eroom's Law: Efficiency Down

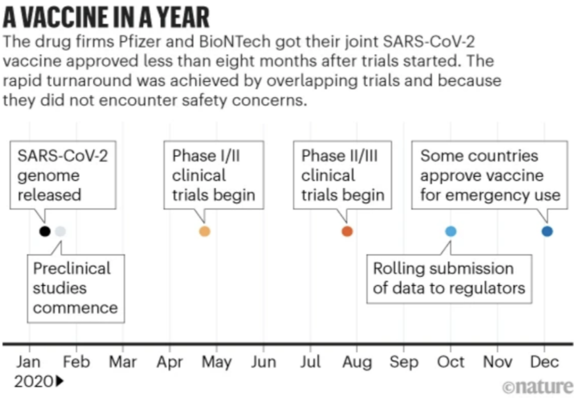


Covid Put a Focus on Shortening Clinical Trials



Sources: BioNTech/Pfizer; *Nature* analysis

Covid Put a Focus on Shortening Clinical Trials



Sources: BioNTech/Pfizer; *Nature* analysis

How can statisticians speed up the clinical trials system?

Add Features to Improve Trial Efficiency

- Smoothly combine studies (e.g. Phase I/II, or II/III).
- Stop early for success (efficacy), or failure (futility).
- Compare multiple treatments or doses to select the best.
- Adaptive sample sizing.
- Use of outside data.

Problem: Analytic Control goes Out the Window!

Adaptive T-Test:

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ (unknown μ, σ).
- $H_0 : \mu = 0$.
- Total of 6 analyses.
- Before each analysis, add 10 i.i.d. samples.
- At each analysis i, reject if

$$T_i := \frac{\sqrt{N_i}\bar{X}_i}{\hat{\sigma}_i} > 2 \quad \text{and} \quad \bar{X}_i > 0.1$$

Problem: Analytic Control goes Out the Window!

Adaptive T-Test:

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ (unknown μ, σ).
- $H_0 : \mu = 0$.
- Total of 6 analyses.
- Before each analysis, add 10 i.i.d. samples.
- At each analysis i, reject if

$$T_i := \frac{\sqrt{N_i}\bar{X}_i}{\hat{\sigma}_i} > 2 \quad \text{and} \quad \bar{X}_i > 0.1$$

What is the Type I Error?

Problem: Analytic Control goes Out the Window!

Adaptive T-Test:

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ (unknown μ, σ).
- $H_0 : \mu = 0$.
- Total of 6 analyses.
- Before each analysis, add 10 i.i.d. samples.
- At each analysis i, reject if

$$T_i := \frac{\sqrt{N_i}\bar{X}_i}{\hat{\sigma}_i} > 2 \quad \text{and} \quad \bar{X}_i > 0.1$$

What is the Type I Error?

Classical toolkit breaks even with Gaussian data.

Adaptive T-Test Non-Trivial Null Distribution

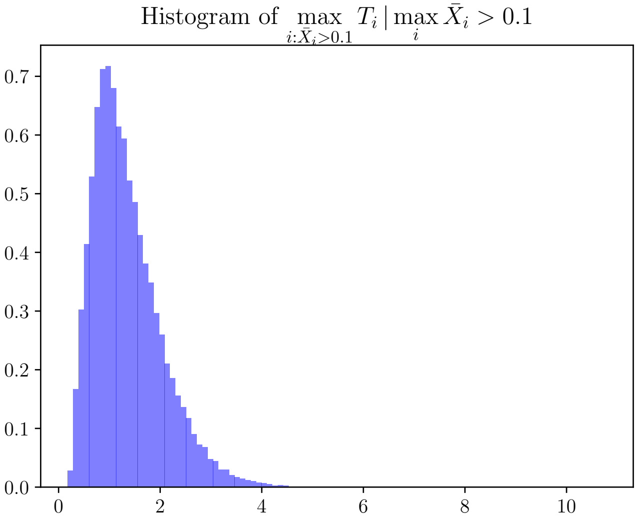
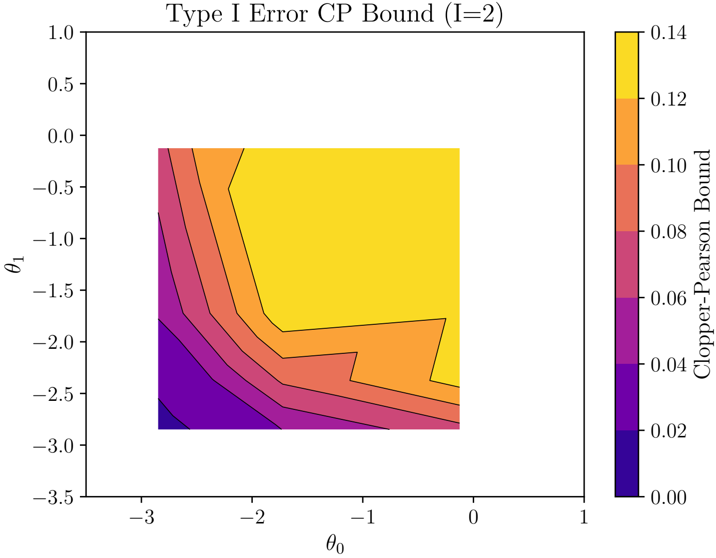


Figure: Adaptive T-Test test statistic distribution for $\sigma \equiv 1$.

Intermediate Techniques Fail

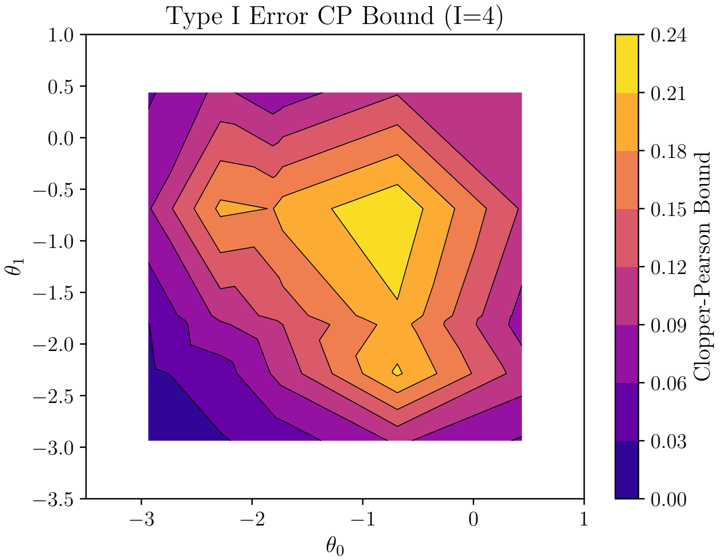
- Composite null with nuisance parameters (noise levels).
  Simulating on single null point $\not\Longrightarrow$ Type I Error control.
- Sharp null hypothesis (exact zero causal effect) is usually false ("null" treatments often increase the variability of outcomes).
  Breaks permutation methods.
- Adaptive sampling renders the test statistic to be non-pivotal.
  Breaks the bootstrap.

Simulation to the Rescue?



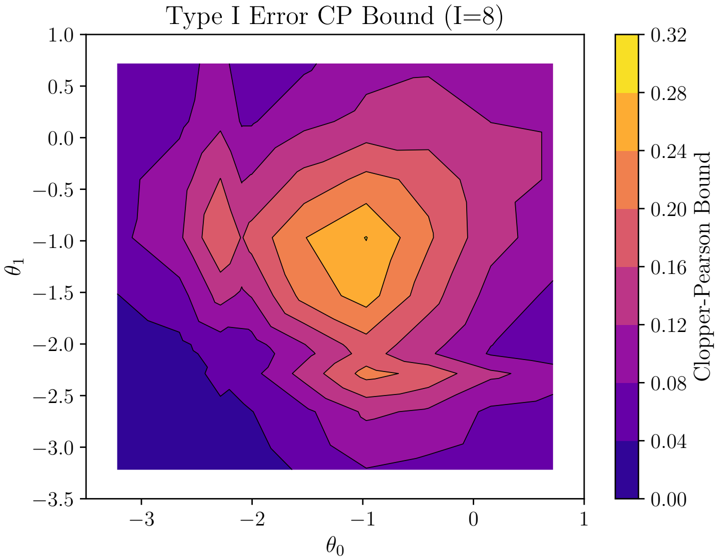
To accept or not to accept?

Simulation to the Rescue?



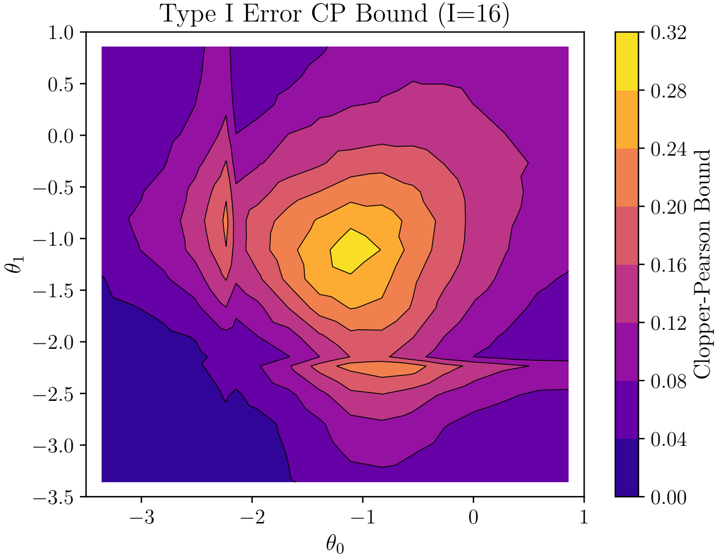
To accept or not to accept?

Simulation to the Rescue?



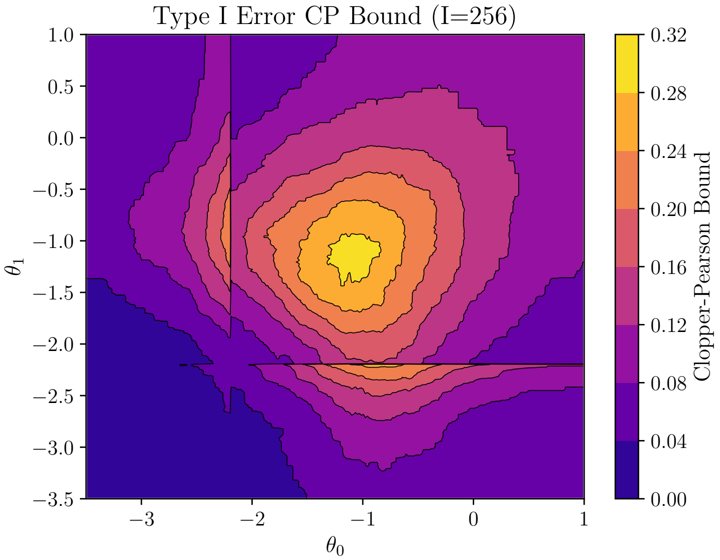
To accept or not to accept?

Simulation to the Rescue?



To accept or not to accept?

Simulation to the Rescue?



To accept or not to accept?

Simulation Raises New Challenges

- Simulation constrained to **finite** number of null points.

  How do we deal with composite nulls?
- Simulation has Monte Carlo error.

  How do we deal with Monte Carlo error?
- Bounded computing power.

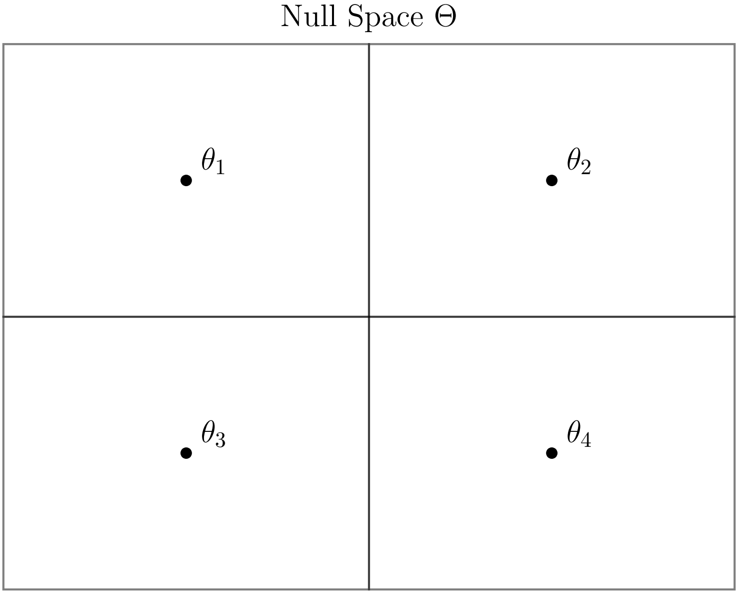
  How many points in the null space to simulate?

  Are the simulations even tractable?

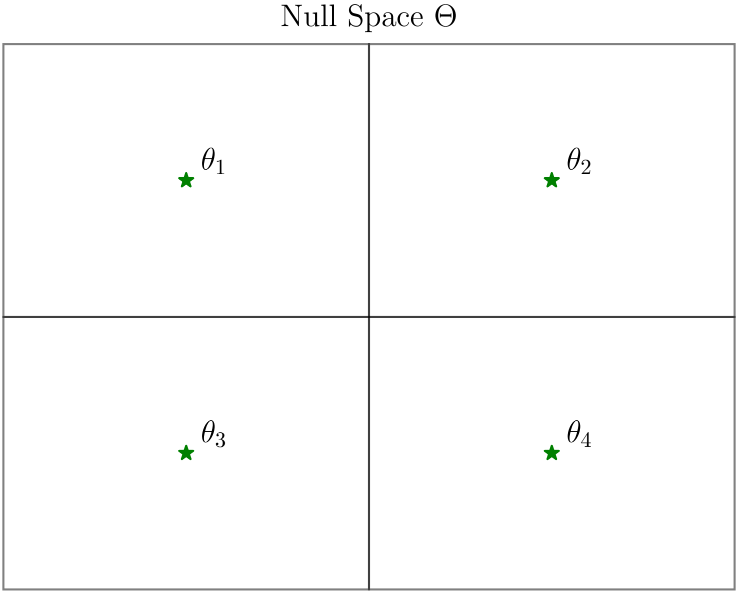
Intuition of Our Approach

Null Space Θ

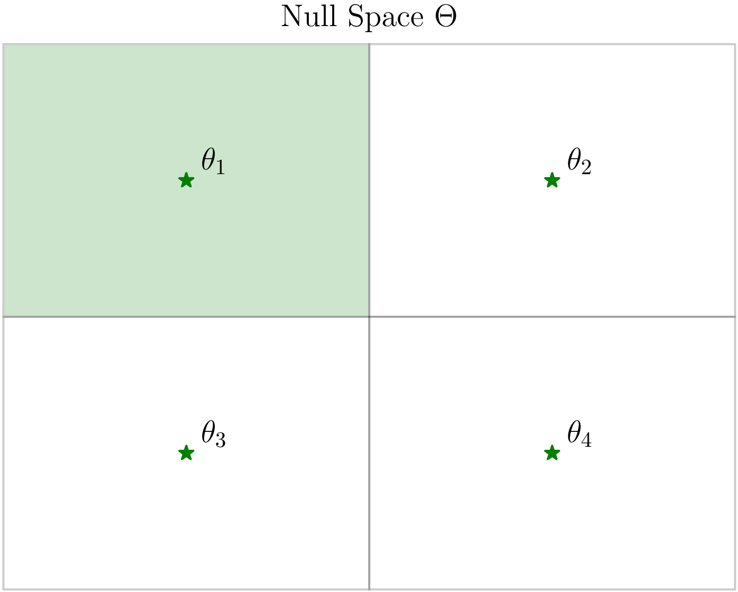
Partition Θ into Tiles with Representatives



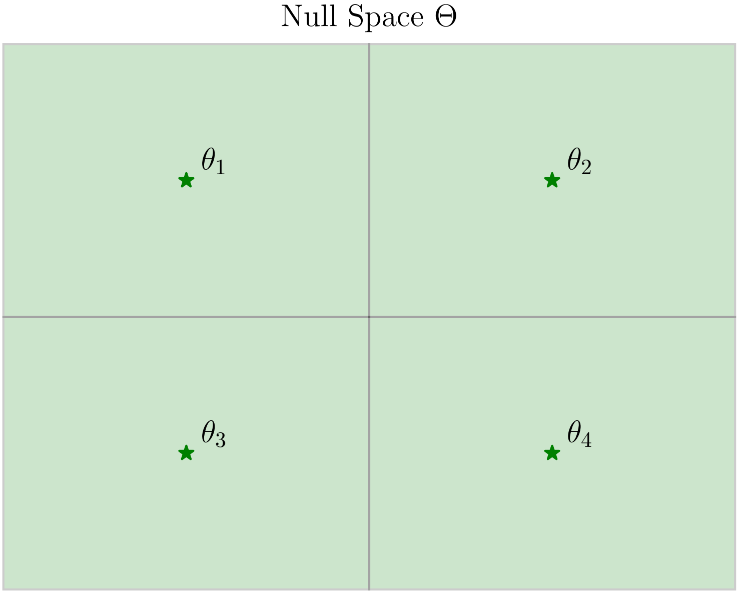
Simulate on each Representative



Extend Simulation Information to Tile



Divide-and-Conquer for Guarantees on All of Θ

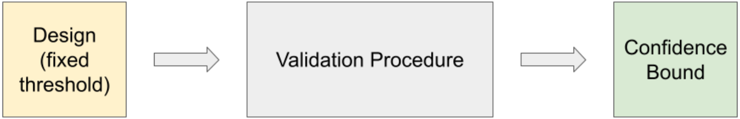


Our Approach: Proof-by-Simulation

General Workflow:

- Let Θ be a (bounded) null hypothesis space.
- Partition Θ into tiles $\{\Theta_i\}_{i=1}^{I}$ with representatives $\{\theta_i\}_{i=1}^{I}$.
- Simulate the design on each θ_i and output test statistics.
- Use our method **Continuous Simulation Extension** (CSE) to *extend* information at each θ_i to any other point in Θ_i.
- Divide-and-conquer to get guarantees on *all of* Θ.

Method 1: Validation for Point-wise Confidence



Method 1: Validation for Point-wise Confidence

- **Validation**: Construct bounds $(\hat{l}(\cdot), \hat{u}(\cdot))$ for the true Type I Error, $f(\cdot)$, with confidence $1-\delta$:

$$\forall \theta \in \Theta, \mathbb{P}\left(\hat{l}(\theta) \leq f(\theta)\right) \geq 1-\delta \text{ and}$$
$$\mathbb{P}\left(\hat{u}(\theta) \geq f(\theta)\right) \geq 1-\delta$$

- Point-wise guarantee is appropriate since there is only one *true* value of θ.

Method 2: Calibration for Type I Error Proof

Method 2: Calibration for Type I Error Proof

- **Calibration**: Select a (random) critical threshold, $\hat{\lambda}^*$, such that

$$\forall \theta \in \Theta, \mathbb{E}\left[f_{\hat{\lambda}^*}(\theta)\right] \leq \alpha$$

where $f_\lambda(\theta)$ is the Type I Error at θ using threshold λ.

Random $\hat{\lambda}^*$ is acceptable

- Guarantee is **overall** valid (regulators want this!).
- Practitioners **already use** simulations to evaluate designs.
- Our approach is **strictly stronger** because we can give guarantees.

Main Task: Find Type I Error at θ

- $X \sim P_\theta$ (known distribution), null space Θ.
- Any arbitrary design $\mathcal{D}$.
- $f(\theta) := \mathbb{P}_\theta(\mathcal{D} \text{ rejects})$.

Main Task: Find **Upper Bound** of Type I Error at θ

- $X \sim P_\theta$ (known distribution), null space Θ.
- Any arbitrary design $\mathcal{D}$.
- $f(\theta) := \mathbb{P}_\theta(\mathcal{D} \text{ rejects})$.

Main Task: Find **Upper Bound** of Type I Error at θ

- Assume further that P_θ is an **exponential family**.
- Does this help?

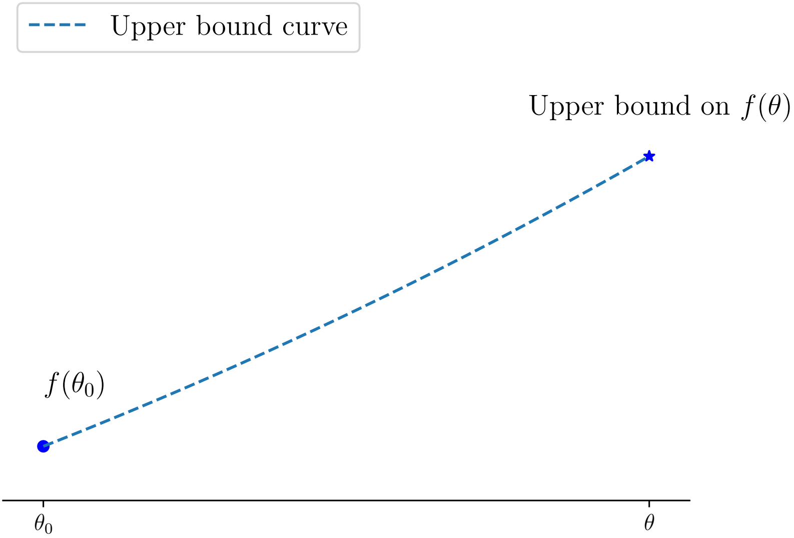
Main Task: Find **Upper Bound** of Type I Error at θ

- Assume further that P_θ is **Gaussian**.
- Does this help?

Intuition for Upper Bounding the Type I Error

- **Morally**, distribution assumptions *should* help!
- Use **“curvature”** information in distribution.
- **Restrict** the possible values for $f(\theta)$.

Claim: Upper Bound on the Type I Error



Derivation: Begin with a Change of Measure

Let $A := \{x : \mathcal{D}(x) \text{ rejects}\}$.
Then,

$$f(\theta) = \mathbb{E}_\theta \left[\mathbb{1}_{X \in A} \right] = \mathbb{E}_{\theta_0} \left[\mathbb{1}_{X \in A} \frac{p_\theta(X)}{p_{\theta_0}(X)} \right]$$

Use Hölder's Inequality!

For any $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$
\begin{aligned}
f(\theta) &\leq \|\mathbb{1}_{X \in A}\|_{L^p(P_{\theta_0})} \left\| \frac{p_\theta(X)}{p_{\theta_0}(X)} \right\|_{L^q(P_{\theta_0})} \\
&= f(\theta_0)^{1-\frac{1}{q}} \left\| \frac{p_\theta(X)}{p_{\theta_0}(X)} \right\|_{L^q(P_{\theta_0})}
\end{aligned}
$$

Introduce Distributional Assumptions

Let P_θ have a density of the form:

$$p_\theta(x) = \exp\{g_\theta(x) - A(\theta)\}$$

By a simple calculation, one can show that

$$\left\| \frac{p_\theta(X)}{p_{\theta_0}(X)} \right\|_{L^q(P_{\theta_0})} = \exp\left\{ \frac{\psi(\theta_0, \theta - \theta_0, q)}{q} - \psi(\theta_0, \theta - \theta_0, 1) \right\}$$

$$\psi(\theta_0, v, q) := \log \mathbb{E}_{\theta_0}\left[\exp\left\{q\left(g_{\theta_0+v}(X) - g_{\theta_0}(X)\right)\right\}\right]$$

We did it!

For any $q \geq 1$,

$$f(\theta) \leq f(\theta_0)^{1-\frac{1}{q}} \exp\left\{ \frac{\psi(\theta_0, \theta - \theta_0, q)}{q} - \psi(\theta_0, \theta - \theta_0, 1) \right\}$$

Tilt-Bound and Special Cases

Tilt-Bound ($q \geq 1$):

$$U(\theta_0, v, q, f(\theta_0)) := \underbrace{f(\theta_0)^{1-\frac{1}{q}}}_{\theta_0 \textbf{ info}} \underbrace{\exp\left\{\frac{\psi(\theta_0, v, q)}{q} - \psi(\theta_0, v, 1)\right\}}_{\textbf{Curvature info}}$$

Exponential family:

$$\psi(\theta_0, v, q) := A(\theta_0 + qv) - A(\theta_0)$$

Normal family $\{\mathcal{N}(\theta, 1) : \theta \in \Theta\}$:

$$U(\theta_0, v, q, f(\theta_0)) := f(\theta_0)^{1-\frac{1}{q}} \exp\left\{\frac{(q-1)v^2}{2}\right\}$$

Optimize over q!

$$f(\theta_0 + v) \leq U(\theta_0, v, q, f(\theta_0)) \quad \forall q \geq 1$$

$$\implies f(\theta_0 + v) \leq \underbrace{\inf_{q \geq 1} U(\theta_0, v, q, f(\theta_0))}_{\textbf{Optimized Tilt-Bound}}$$

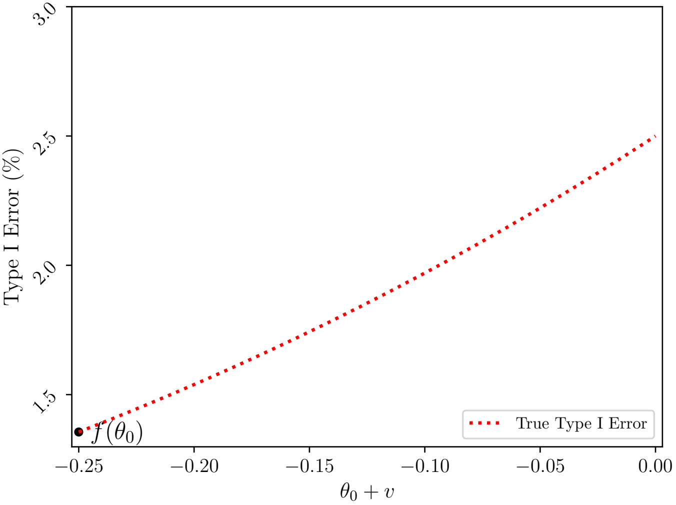
How to optimize over q?

- Tilt-Bound is **quasi-convex** in q!
- Very **simple, fast** $O(\log(\epsilon^{-1}))$ algorithm with **guaranteed convergence**.

Theorem (Quasi-convex in q)

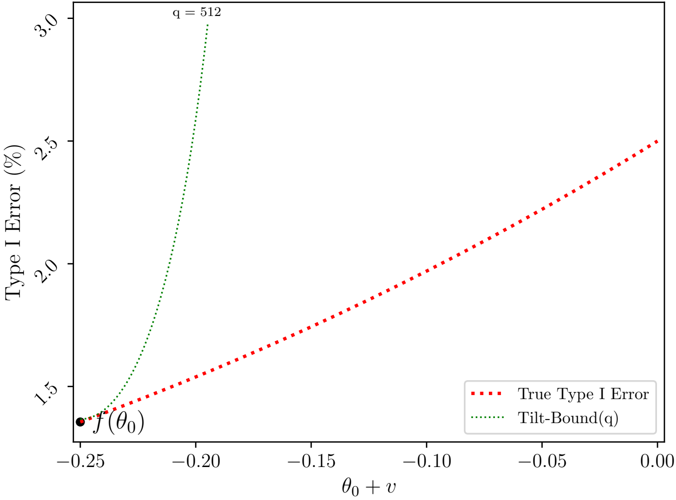
Fix any $\theta_0 \in \Theta \subseteq \mathbb{R}^d$, a set $S \subseteq \mathbb{R}^d$, and $a \geq 0$. Assume that for all $v \in S$, $\Delta(v, X) := g_{\theta_0+v}(X) - g_{\theta_0}(X)$ is not constant P_{θ_0}-a.s.. Then, $q \mapsto \sup_{v \in S} U(\theta_0, v, q, a)$ is quasi-convex. Moreover, it is strict if $a > 0$, S is finite, and not identically infinite, respectively.

Demonstrating the Tilt-Bound on the One-Sided Z-Test



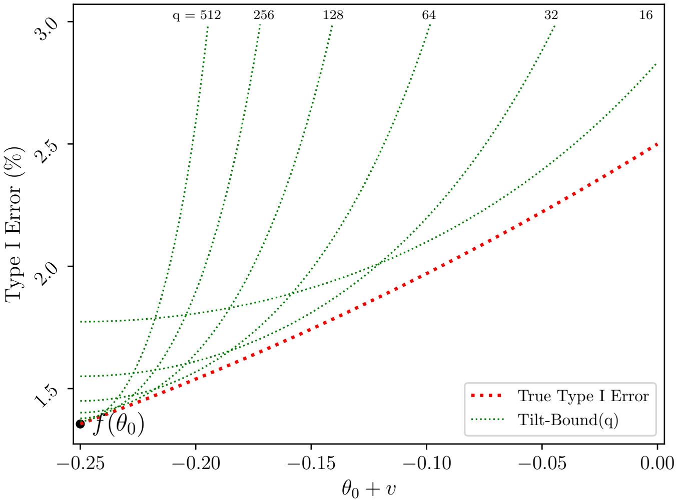
- $X \sim \mathcal{N}(\theta, 1)$, $\Theta = [-0.25, 0]$.
- $\mathcal{D}(X)$: reject if $X > z_{1-\alpha}$.

The Tilt-Bound for a Particular q



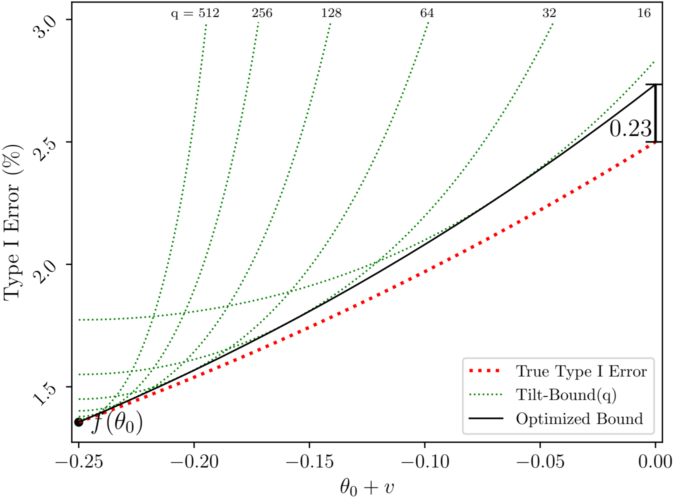
$$U(\theta_0, v, q, f(\theta_0)) = f(\theta_0)^{1-\frac{1}{q}} \exp\left\{\frac{(q-1)v^2}{2}\right\}$$

The Tilt-Bound for Many qs



$$U(\theta_0, v, q, f(\theta_0)) = f(\theta_0)^{1-\frac{1}{q}} \exp\left\{\frac{(q-1)v^2}{2}\right\}$$

The Optimized Tilt-Bound is Tight



$$\inf_{q \geq 1} U(\theta_0, v, q, f(\theta_0))$$

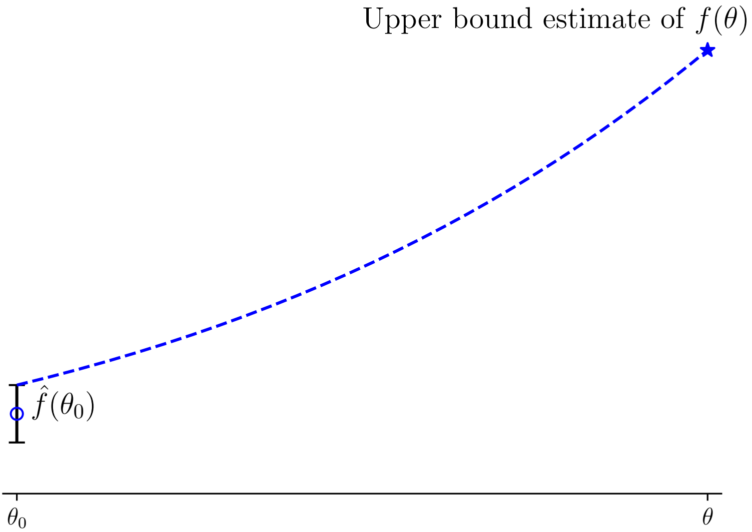
Tilt-Bound Summary

- Tilt-Bound is a **deterministic** bound.
- Tight over small to medium distances.
- Valid for **any rejection set**.
- Depends on Type I Error at the **initial point** θ_0 and the **distributional family** P_θ (which implicitly accounts for the sample size).

Main Task: Point-wise Valid Upper Bound on Type I Error

- $X \sim P_\theta$ (known distribution), null space Θ.
- Any arbitrary design $\mathcal{D}$.
- Clopper-Pearson bound using Monte Carlo estimate $\hat{f}(\theta_0)$.

Claim: Valid Upper Bound on the Type I Error



Use Tilt-Bound on Upper Bound Estimate!

Monotone Property:

- $a \mapsto U(\theta_0, v, q, a)$ is **non-decreasing**.

Validation Proof:

- $\hat{\eta}$ be a $1 - \delta$ upper bound of $f(\theta_0)$.
- $\hat{u} := U(\theta_0, v, q, \hat{\eta})$ for any $q \geq 1$.

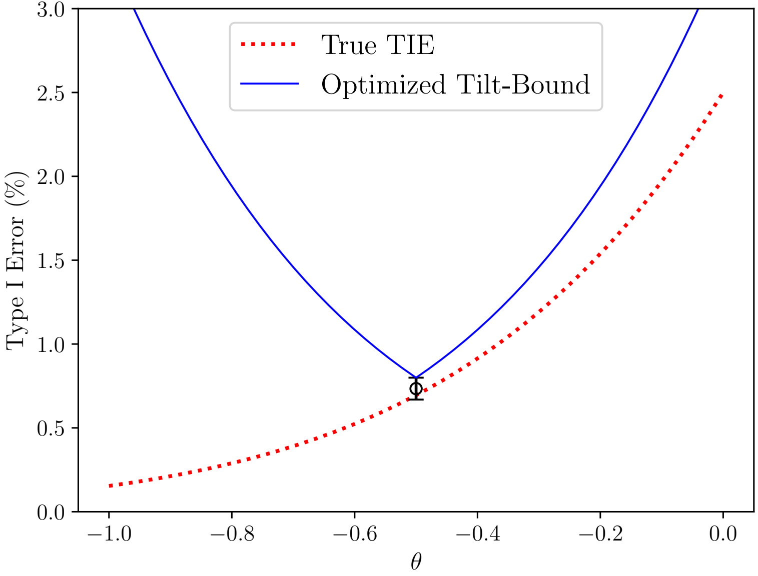
Recall,

$$f(\theta_0 + v) \leq U(\theta_0, v, q, f(\theta_0))$$

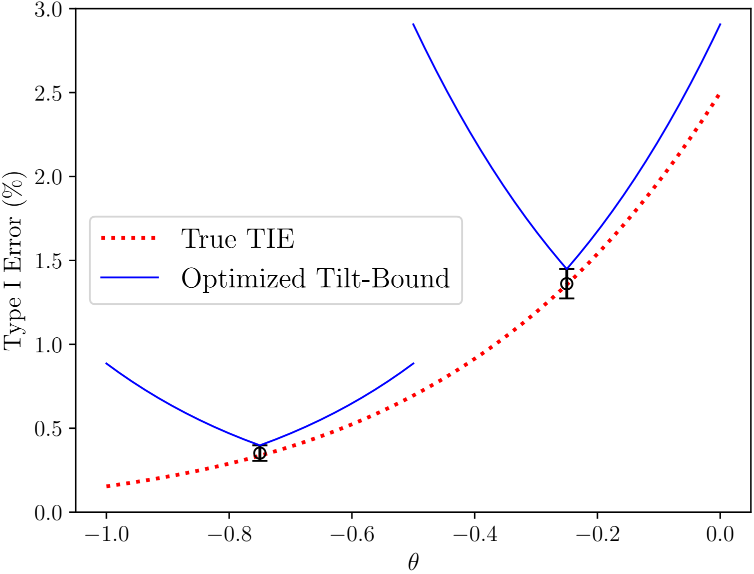
Then,

$$\mathbb{P}\left(f(\theta_0 + v) \leq \hat{u}\right) \geq \mathbb{P}\left(f(\theta_0) \leq \hat{\eta}\right) \geq 1 - \delta$$

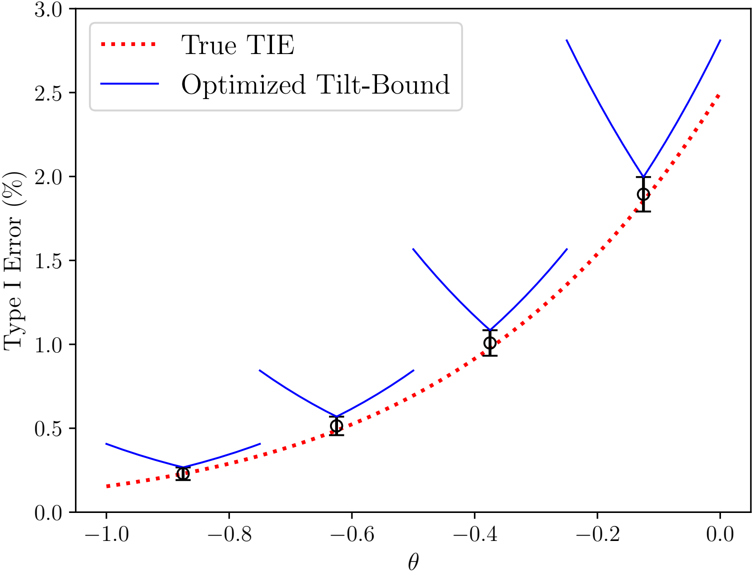
Use Tilt-Bound on a Grid for the Z-Test



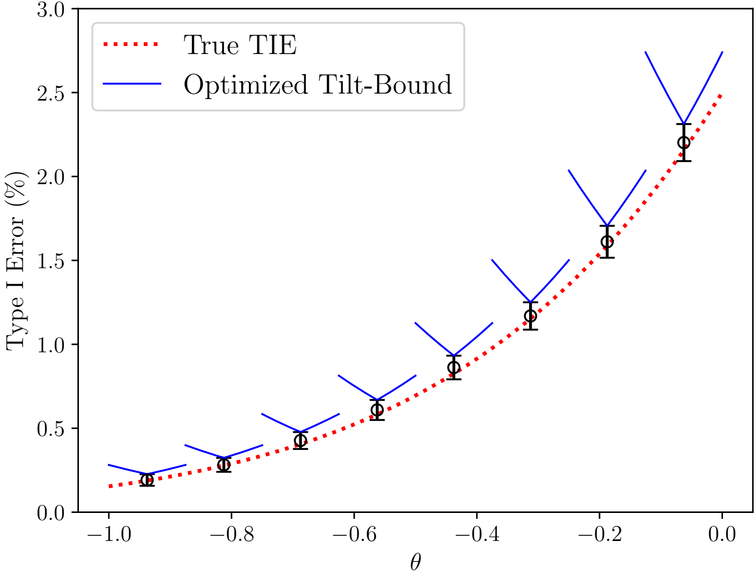
Use Tilt-Bound on a Grid for the Z-Test



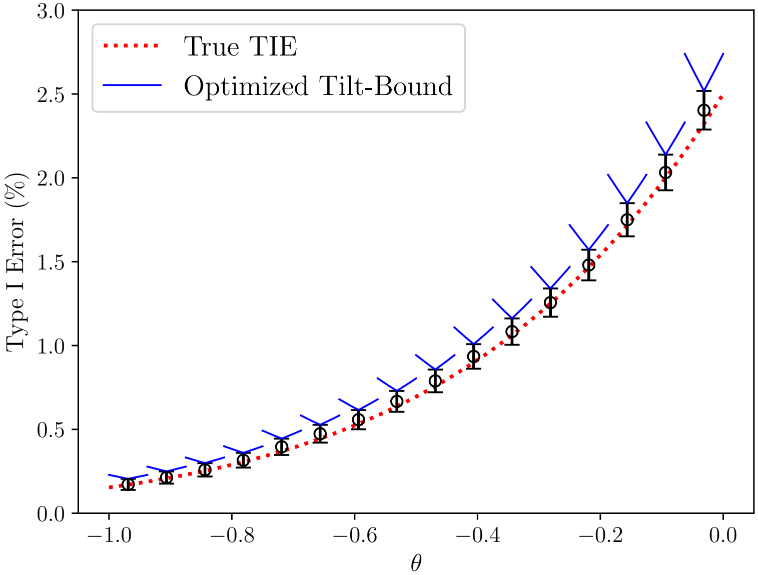
Use Tilt-Bound on a Grid for the Z-Test



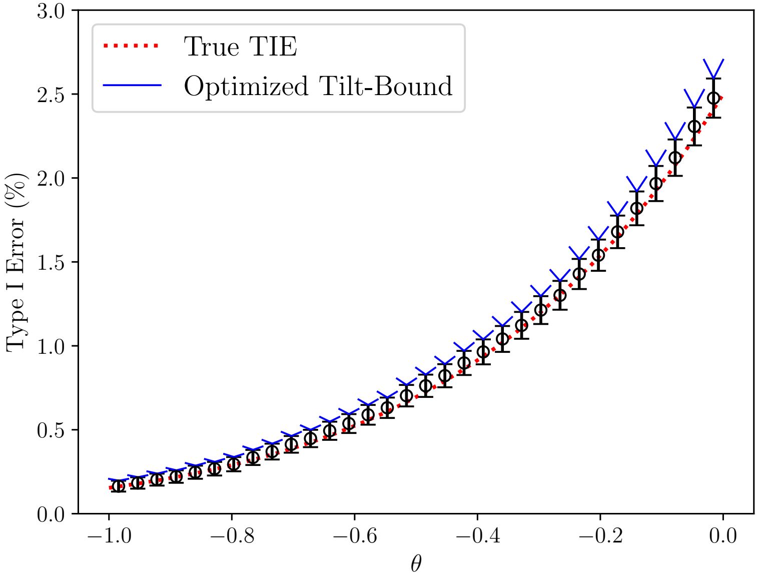
Use Tilt-Bound on a Grid for the Z-Test



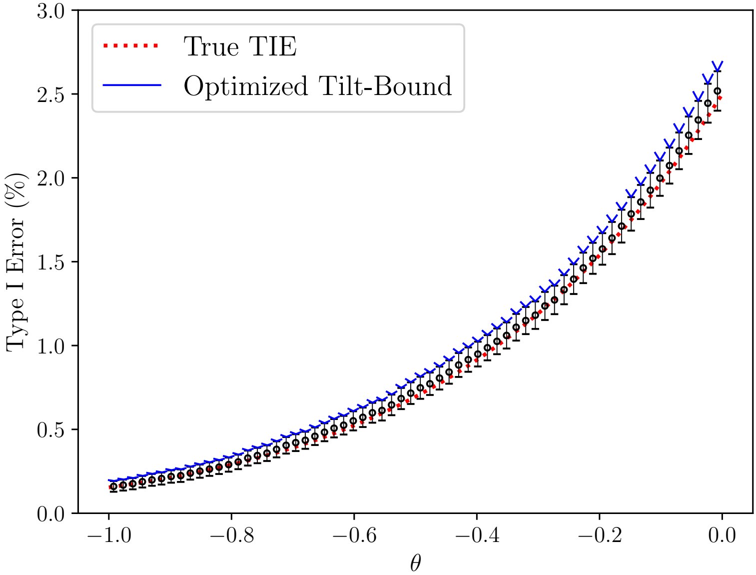
Use Tilt-Bound on a Grid for the Z-Test



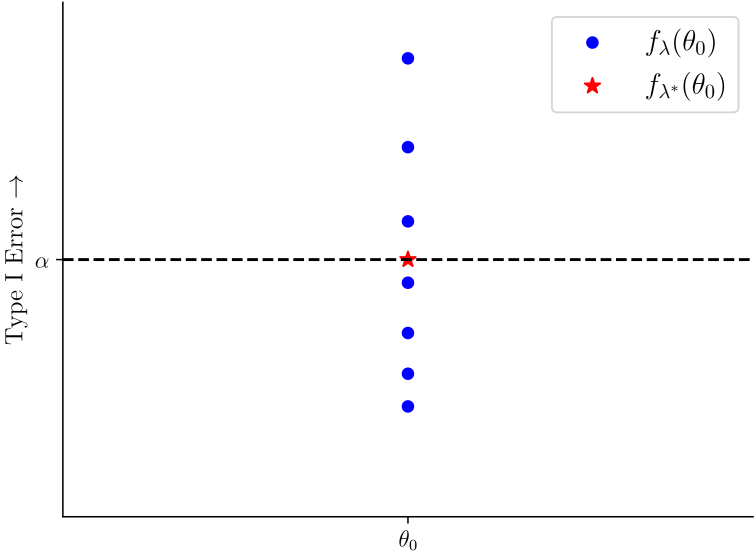
Use Tilt-Bound on a Grid for the Z-Test



Use Tilt-Bound on a Grid for the Z-Test



Main Task: Find Critical Threshold with Level α at θ_0



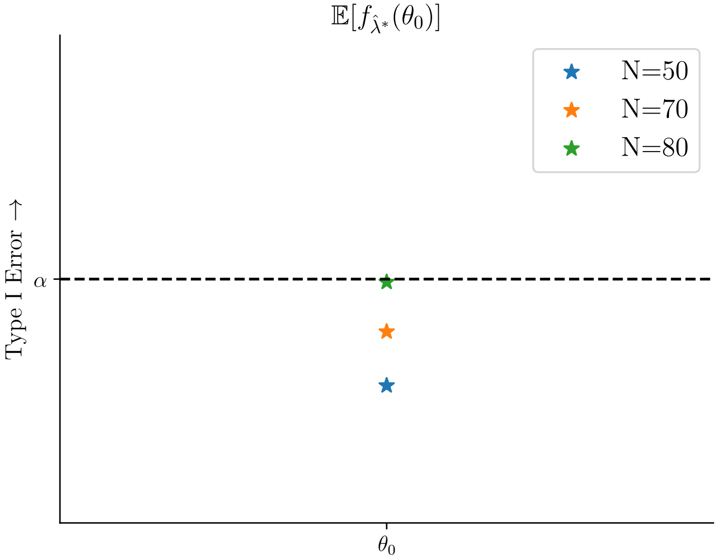
Straightforward for a Single Point

- Let $S(X)$ be the test statistic with data X.
- Design $\mathcal{D}$: rejects if $S(X) < \lambda$.
- Given $S_1, \ldots, S_N$ i.i.d. test statistics,

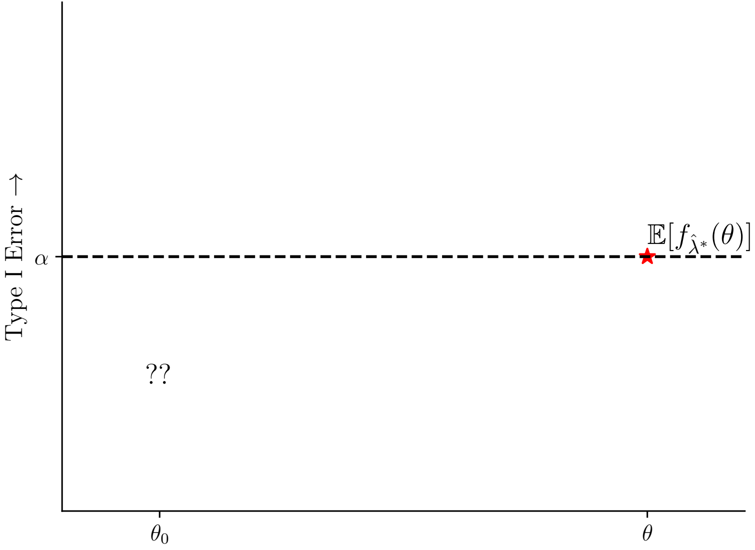
$$\hat{\lambda}^* := S_{\lfloor (N+1)\alpha \rfloor} \implies \mathbb{E}\left[f_{\hat{\lambda}^*}(\theta_0)\right] \leq \alpha$$

- Easy to show using Beta distribution.

Calibration Results for a Single Point

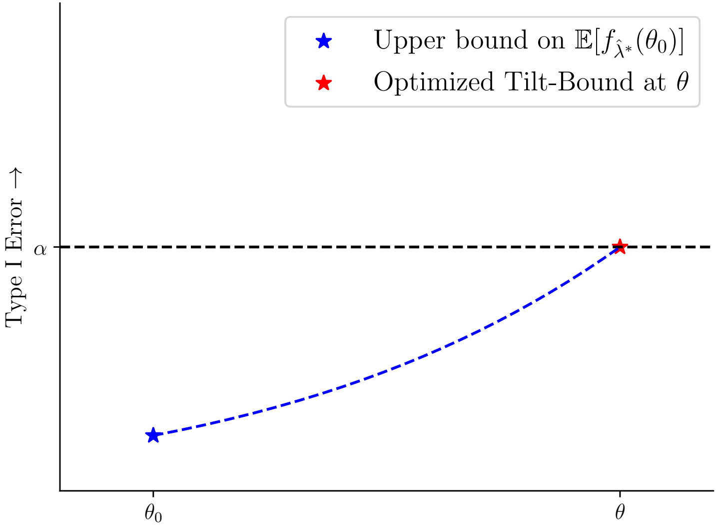


Main Task: Find Critical Threshold with Level α at θ



Inverted Tilt-Bound

Claim: Calibrate at θ_0 to Control Tilt-Bound at θ



Use CSE to Control Type I Error at θ

$$\mathbb{E}\left[f_\lambda(\theta_0 + v)\right] \leq U(\theta_0, v, q, \mathbb{E}\left[f_\lambda(\theta_0)\right])$$

- Back-solve to hit level α:

$$U(\theta_0, v, q, \mathbb{E}\left[f_\lambda(\theta_0)\right]) \leq \alpha \iff \mathbb{E}\left[f_\lambda(\theta_0)\right] \leq U^{-1}(\theta_0, v, q, \alpha)$$

Inverted Tilt-Bound:

$$U^{-1}(\theta_0, v, q, \alpha) := \left(\alpha \exp\left[-\frac{\psi(\theta_0, v, q)}{q} + \psi(\theta_0, v, 1)\right]\right)^{\frac{q}{q-1}}$$

Use Point-Null Case to Find Critical Threshold

- Let $\alpha' := U^{-1}(\theta_0, v, q, \alpha)$.
- Find $\hat{\lambda}^*$ such that

$$\mathbb{E}\left[f_{\hat{\lambda}^*}(\theta_0)\right] \leq \alpha'$$

- Maximize α' over $q \geq 1$ for free to get least-conservative threshold.

Then,

$$\mathbb{E}\left[f_{\hat{\lambda}^*}(\theta_0 + v)\right] \leq U(\theta_0, v, q, \mathbb{E}\left[f_{\hat{\lambda}^*}(\theta_0)\right]) \leq \alpha$$

Control Type I Error in a Region

- Back-solve with **worst-case** Tilt-Bound:

$$\begin{aligned} &\sup_{v \in \Theta - \theta_0} U(\theta_0, v, q, \mathbb{E}[f_\lambda(\theta_0)]) \leq \alpha \\ \iff \quad &\mathbb{E}[f_\lambda(\theta_0)] \leq \inf_{v \in \Theta - \theta_0} U^{-1}(\theta_0, v, q, \alpha) \end{aligned}$$

- Let $\alpha' := \inf_{v \in \Theta - \theta_0} U^{-1}(\theta_0, v, q, \alpha)$.
- Maximize α' over $q \geq 1$.
- Find $\hat{\lambda}^*$ such that

$$\mathbb{E}\left[f_{\hat{\lambda}^*}(\theta_0)\right] \leq \alpha'$$

How to optimize over v?

$$\inf_{v \in \Theta - \theta_0} U^{-1}(\theta_0, v, q, \alpha)$$

- **Assume a linearity condition**!
- Inverted Tilt-Bound is **quasi-concave** in v, respectively!
- If Θ is a polytope, suffices to compute only at the vertices!

How to optimize over v?

Theorem (Quasi-convex in v)

Let $\{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^d\}$ *be a family of distributions:*

$$dP_\theta(x) = \exp\{g_\theta(x) - A(\theta)\}\, d\mu(x)$$

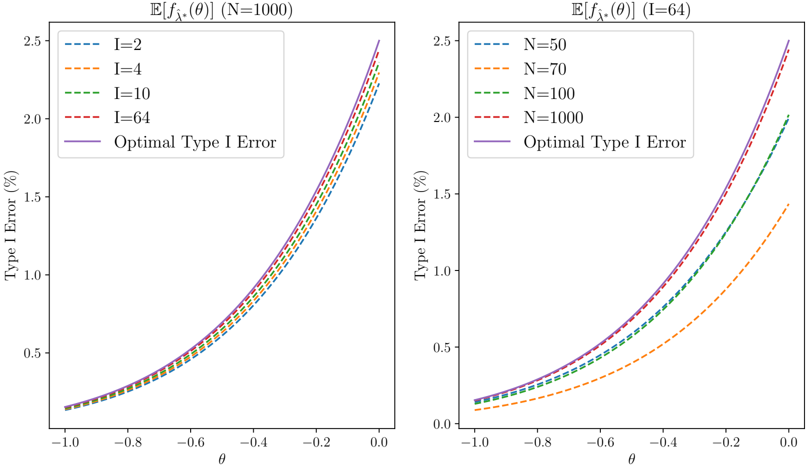
Fix any $\theta_0 \in \Theta$*. Suppose that* $\Delta(v, x) \equiv g_{\theta_0+v}(x) - g_{\theta_0}(x)$ *is linear in* v*, i.e.* $\Delta(v, x) = W(x)^\top v$ *for some* $W(x) \in \mathbb{R}^d$*. Then, the Inverted Tilt-Bound is quasi-concave as a function of* v*.*

Divide-and-Conquer yields a Global Guarantee

- Assume Θ any bounded space.
- Partition Θ into tiles $\{\Theta_i\}_{i=1}^{I}$ with representatives $\{\theta_i\}_{i=1}^{I}$.
- Calibrate to get $\hat{\lambda}_i^*$ for each tile Θ_i.
- Take the most conservative, $\hat{\lambda}^* := \min\limits_{i=1,\ldots,I} \hat{\lambda}_i^*$.
- Then,

$$\begin{aligned}
\sup_{\theta\in\Theta} \mathbb{E}\left[f_{\hat{\lambda}^*}(\theta)\right] &= \max_{i=1,\ldots,I} \sup_{\theta\in\Theta_i} \mathbb{E}\left[f_{\hat{\lambda}^*}(\theta)\right] \\
&\leq \max_{i=1,\ldots,I} \sup_{\theta\in\Theta_i} \mathbb{E}\left[f_{\hat{\lambda}_i^*}(\theta)\right] \\
&\leq \alpha
\end{aligned}$$

Type I Error Tightens with More Simulations and Tiles

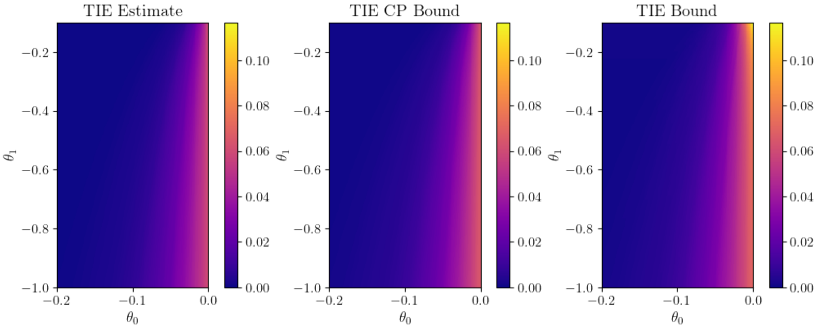


Revisiting the Adaptive T-Test

- Data $X_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$ with unknown μ, σ^2.
- $H_0 : \mu \leq \mu_0$.
- Initial sample size of n_0.
- K interims where each interim stops if the t-statistic given observed data up to now is above the threshold t^*.
- If continue, add n_i more data.
- Final analysis: reject if t-statistic is above t^*.

$$\theta_0 := \frac{\mu}{\sigma^2} \quad \theta_1 := -\frac{1}{2\sigma^2}$$

Tight Analysis Despite Lack of Exact Theory



Adaptive T-Test Computation and Configuration

- **Computation:**
  - 327 million simulations.
  - Runtime: 2s.
  - M1 Macbook Pro.
- **Configuration:**
  - $K = 3$ interims.
  - $n_0 = 100$.
  - $n_i = 50$ for $1 \leq i \leq K$.
  - $\mu_0 = 0$.
  - $\alpha = 0.025$.

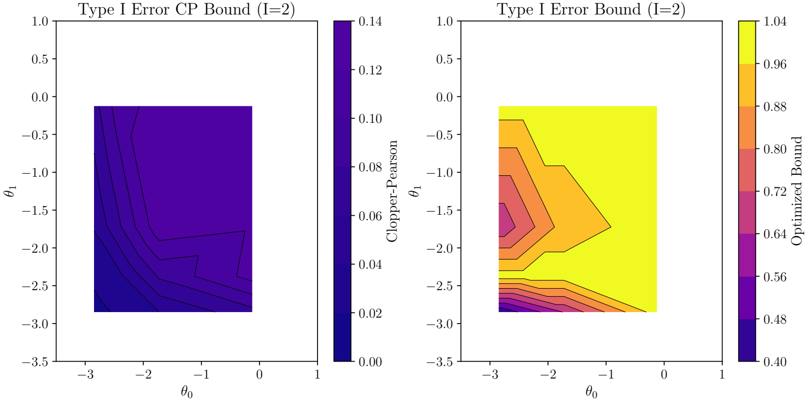
Bayesian Basket Trial from Berry et al. [2013]

- Design:

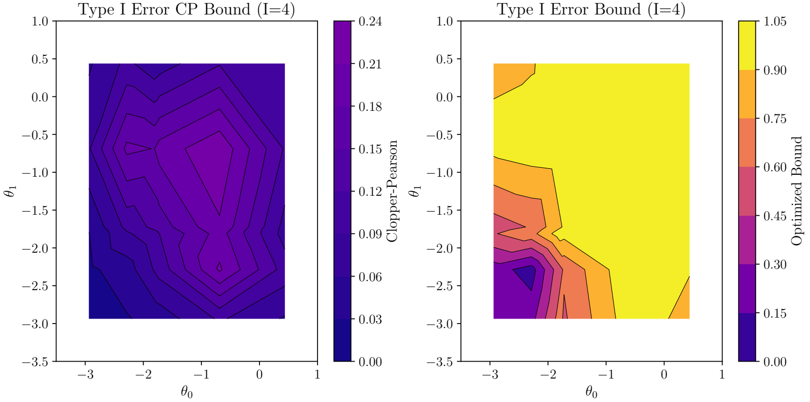
$$
\begin{aligned}
Y_j &\sim \text{Binom}(n_j, p_j) \quad j = 1, \ldots, d \\
\theta_j &= \theta_{0j} + \text{logit}(p_j) \\
\theta_j &\sim \mathcal{N}\left(\mu, \sigma^2\right) \\
\mu &\sim \mathcal{N}\left(\mu_0, \sigma_0^2\right) \\
\sigma^2 &\sim \Gamma^{-1}\left(\alpha_0, \beta_0\right)
\end{aligned}
$$

- Let $c \in [0,1]^{d-1}$ be a vector of fixed thresholds and $d \equiv 4$.
- Reject if $\mathbb{P}\left[p_i > p_0 | Y\right] > c_i$ for some null (treatment) arm i.

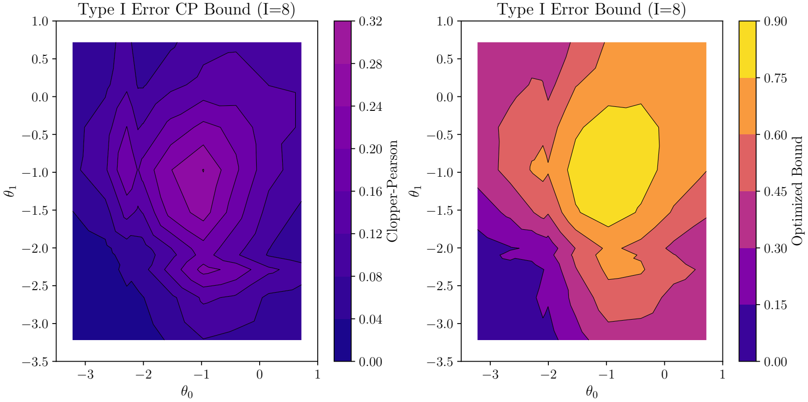
Validation shows Type I Error Surface for Bayesian Design



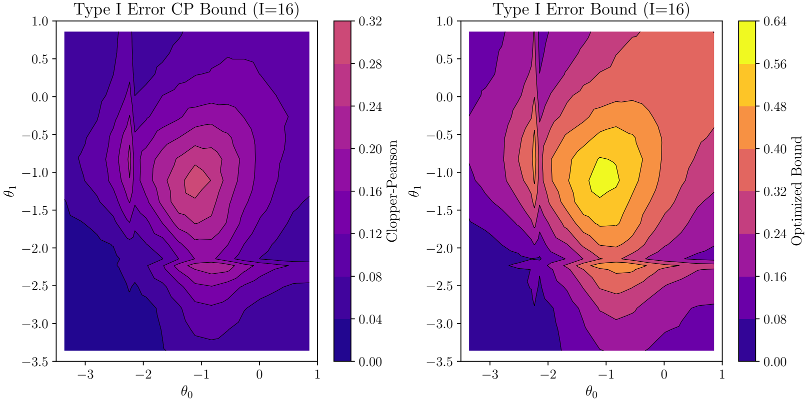
Validation shows Type I Error Surface for Bayesian Design



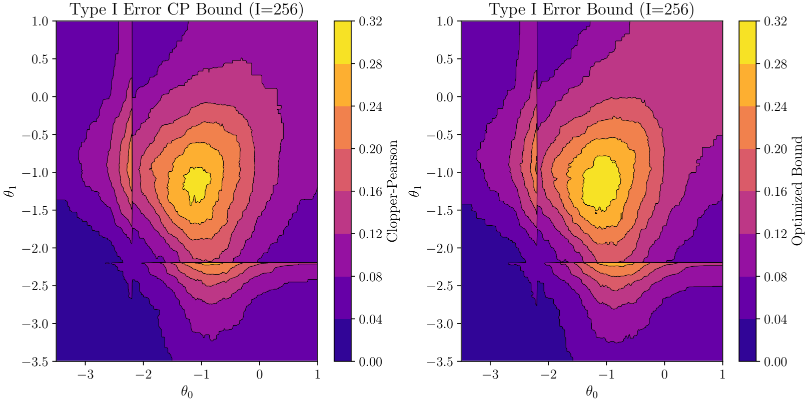
Validation shows Type I Error Surface for Bayesian Design



Validation shows Type I Error Surface for Bayesian Design



Validation shows Type I Error Surface for Bayesian Design



Berry et al. [2013] Computation and Configuration

- **Computation**:
  - 7.34 trillion simulations.
  - Runtime: 4 hours.
  - Nvidia V100 GPU.
- **Configuration:**
  - $n_i = 35$ for all $i = 1, \ldots, d$.
  - $\mu_0 = -1.34$, $\sigma_0 = 10$, $\alpha_0 = 0.0005$, $\beta_0 = 0.000005$.
  - $c_i = 0.85$ for all $i = 1, \ldots, d$.

A Complicated Phase II/III Selection Design

- 3 treatment and 1 control arm with binary outcomes.
- Trial decisions using the Bayesian hierarchical model as in Berry et al. [2013].
- Stage 1: select the "best" treatment arm against control with interim analyses.
- Each of 3 interim analyses can stop for futility, drop one or more poorly performing treatments, or accelerate an arm to move to stage 2.
- Stage 2: one interim and one final analysis.
- The total number of patients across all arms and stages is at most 800 with at most 350 in any single arm.

Phase II/III Selection Design Calibrated Successfully

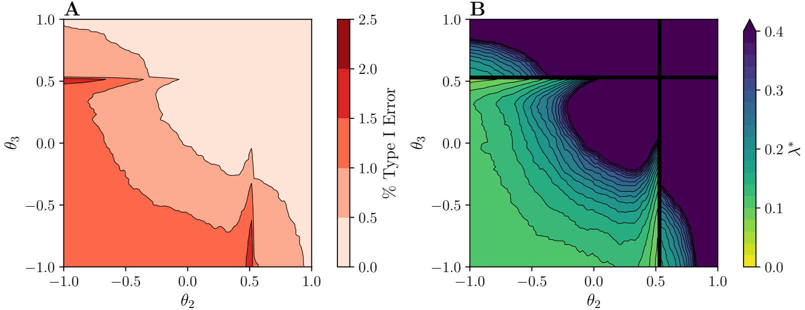


Figure: Both plots slice the domain by fixing 2 parameters, $\theta_0 = \theta_1 = 0.533$. Figure **A** shows the Tilt-Bound profile for the selected threshold $\hat{\lambda}^* = 0.06253$. Figure **B** shows the critical value $\hat{\lambda}_i^*$ separately for each tile such that its Tilt-Bound is 2.5%.

Phase II/III Selection Design Computation and Configuration

- **Computation:**
  - 960 billion simulations.
  - Runtime: 5 days.
  - Nvidia V100 GPU.
- **Configuration:**
  - $H_0 : \theta_i \leq \theta_0$ for all $i = 1, \ldots, d-1$.
  - Restrict to $\theta_i \in [-1, 1]$ for all i.

Phase II/III Selection Design Remarks (Optional)

- Max Tilt-Bound occurs at the tile with center $\theta_0 = (0.4925, 0.4925, 0.4925, -1.0)$.
- Paradox: worst Type I Error **does not** occur at the global null (where all treatments perform equally to control), but when **one** treatment performs poorly.

Further Results of CSE

- Can study **power, False Discovery Rate (FDR)**, and **bias of bounded estimators**.
- Theory also holds for **Generalized Linear Models (GLMs)** after conditioning on covariates.
  - E.g. logistic regression.
- **Quasi-convexity** results for the Tilt-Bound/Inverted Tilt-Bound simplify computations to checking vertices.
- See pre-print for details.

Computational Tricks

- Adaptive simulation/grid sizing (dramatic overall cost reduction!).
- Correlated simulations (dramatic sampling reduction!).
  - **BoTorch** uses a similar (more advanced) trick.
  - Thanks to **Prof. Art Owen** for the idea!
- How to perform **1 trillion simulations** of a complex Bayesian design?
  - **Integrated Nested Laplace Approximation (INLA)**.
  - Our INLA code is **1 million times faster** than standard MCMC packages.
  - Similar accuracy in most cases.

Remarks

- Proof-by-simulation is **general, powerful, and robust**.
- Continuous Simulation Extension converts simulations at finite points into guarantees over **regions**.
- Practical advantage: CSE analyzes the design **as represented in code**. Robust to:
  - Approximations.
  - Theoretical uncertainties with convergence of algorithms
- With the right software, method is **tractable**!

End Goals

- Streamline innovation in trial design.
- Improve regulatory consistency with objective proofs.
- Reduce time and human capital cost of validating new procedures.
- Speculatively: enable new “black-box” statistical procedures.

References I

Scott M. Berry, Kristine R. Broglio, Susan Groshen, and Donald A. Berry. Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase ii oncology clinical trials. *Clinical Trials*, 10(5):720–734, 2013.