Week 7:
Response Surface Methodology

AIM-5014-1A: Experimental Optimization

David Sweet // 20230713

Review: LLN, CLT, A/B Testing

- As $N \to \infty$, $\bar{y} \to E[BM]$ (LLN)
  - CLT: $\bar{y} \sim \mathcal{N}(E[BM], \sigma^2)$, "measured BM is gaussian"
- **Design**: $N \geq \left(\frac{2.5\hat{\sigma}_\delta}{PS}\right)^2$
- **Measure**: Randomize, $\bar{\delta} = \bar{y}_B - \bar{y}_A$, $se = \sigma_\delta/\sqrt{N}$
- **Analyze**: Accept B if $\bar{\delta} > PS$ and $\frac{\bar{\delta}}{se} \geq 1.64$ (check guardrails)
- **False Positive Traps**: Early stopping, multiple comparisons (use Bonferroni)

Review: Experimental cost

- Experiments are expensive:
  - They take time to run
  - They put users at risk of bad experience
  - They cost money: engineer’s salary, lost revenue
- The only way to reduce this cost is to **take fewer measurements**.
- Research into experimental methods seeks to reduce the number of measurements required to achieve an experiment’s goal

Case: Song recommender

- In prod (A): Ranking songs by p_{listen} = P{user will listen until the end}
- In dev (B): Ranking songs by p_{like} = P{user will click song's like button}
- A/B test the two models?
- Why not use both? Rank by a score:

$$score = \alpha p_{\text{listen}} + (1 - \alpha) p_{\text{like}}$$

Case: Song recommender

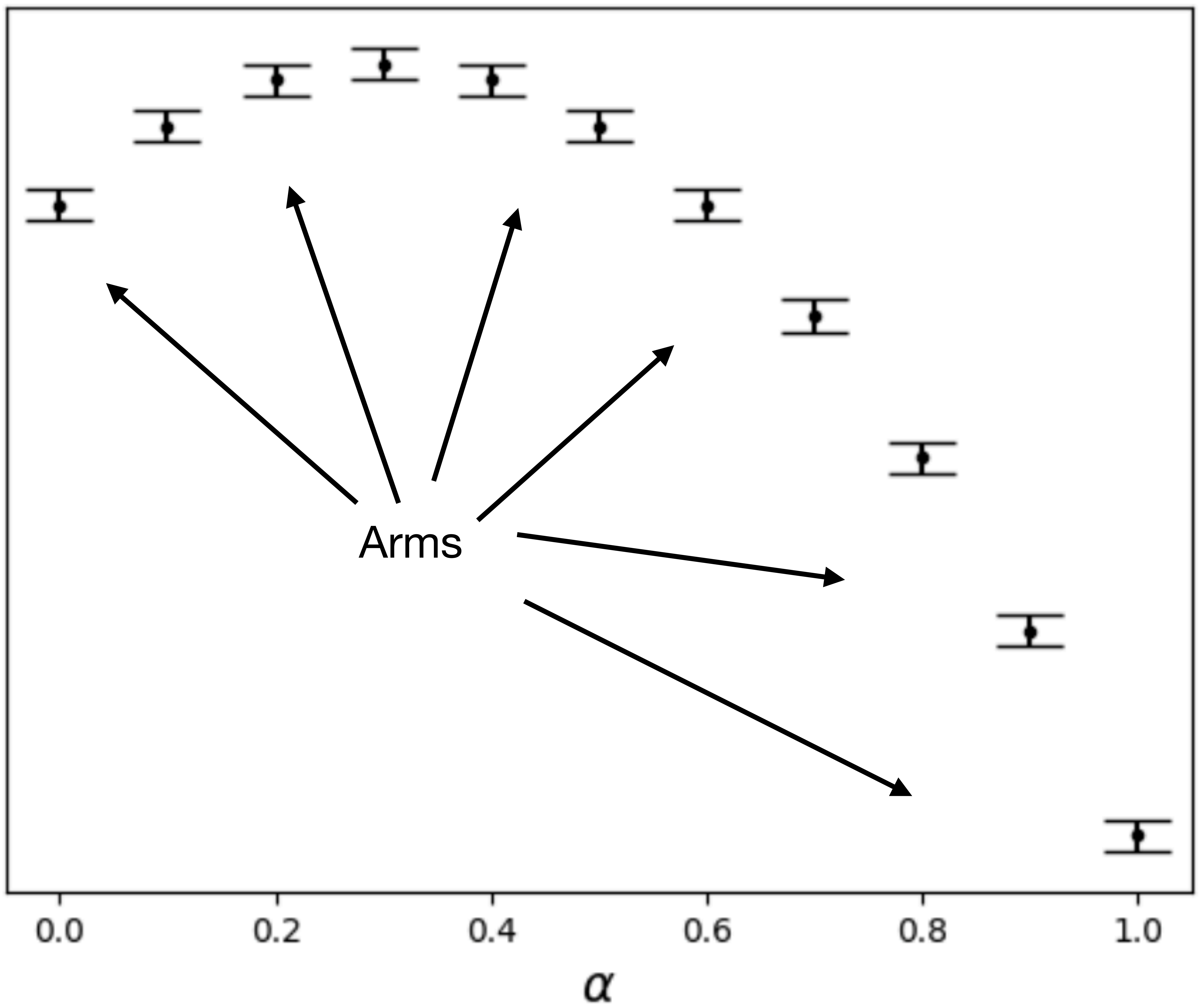
- Combine models, $\alpha \in [0,1]$

$$score = \alpha p_{\text{listen}} + (1 - \alpha) p_{\text{like}}$$

- Find α that gives highest BM
  - … via experimental optimization

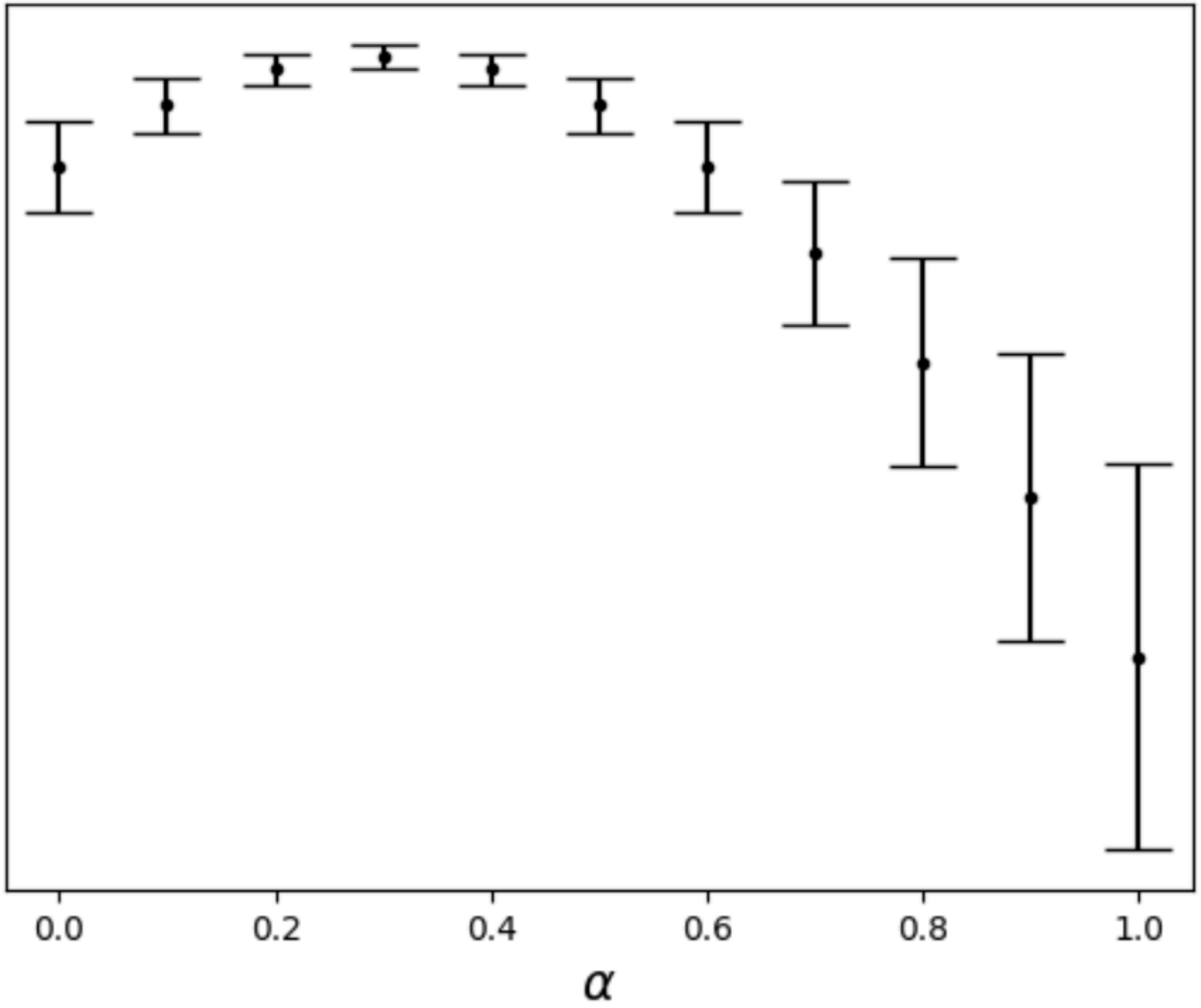
Case: Song recommender

- Approach I: A/B/n test
- Measure $\alpha \in \{0, 0.1, 0.2, \ldots, 1.0\}$
- Req. many observations:
  - Lots of capacity
  - Bonferroni



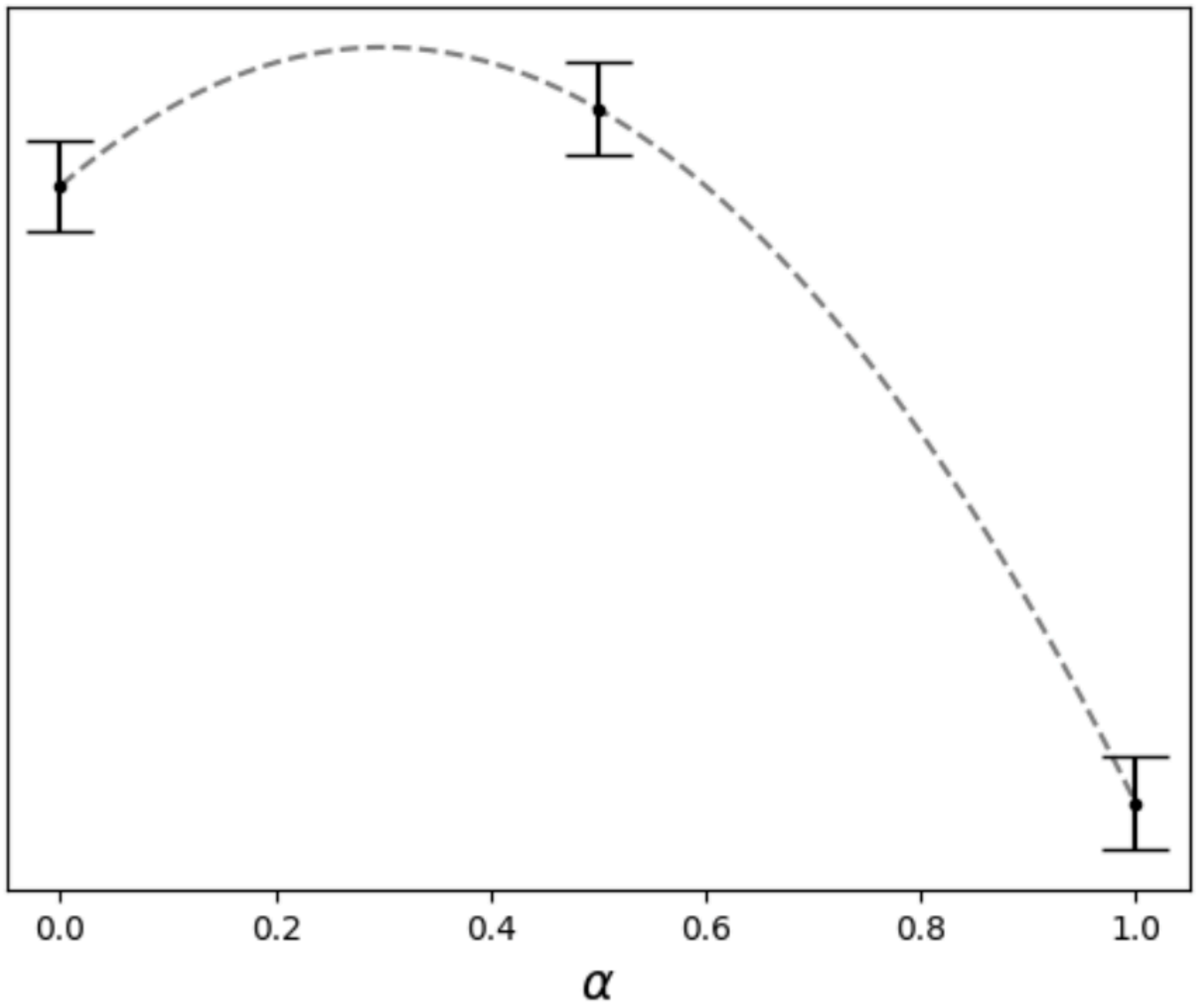
Case: Song recommender

- Approach II: Multi-armed bandit
- Same number of arms
- Fewer observations than A/B/n:
    - Worse arms are allocated fewer observations



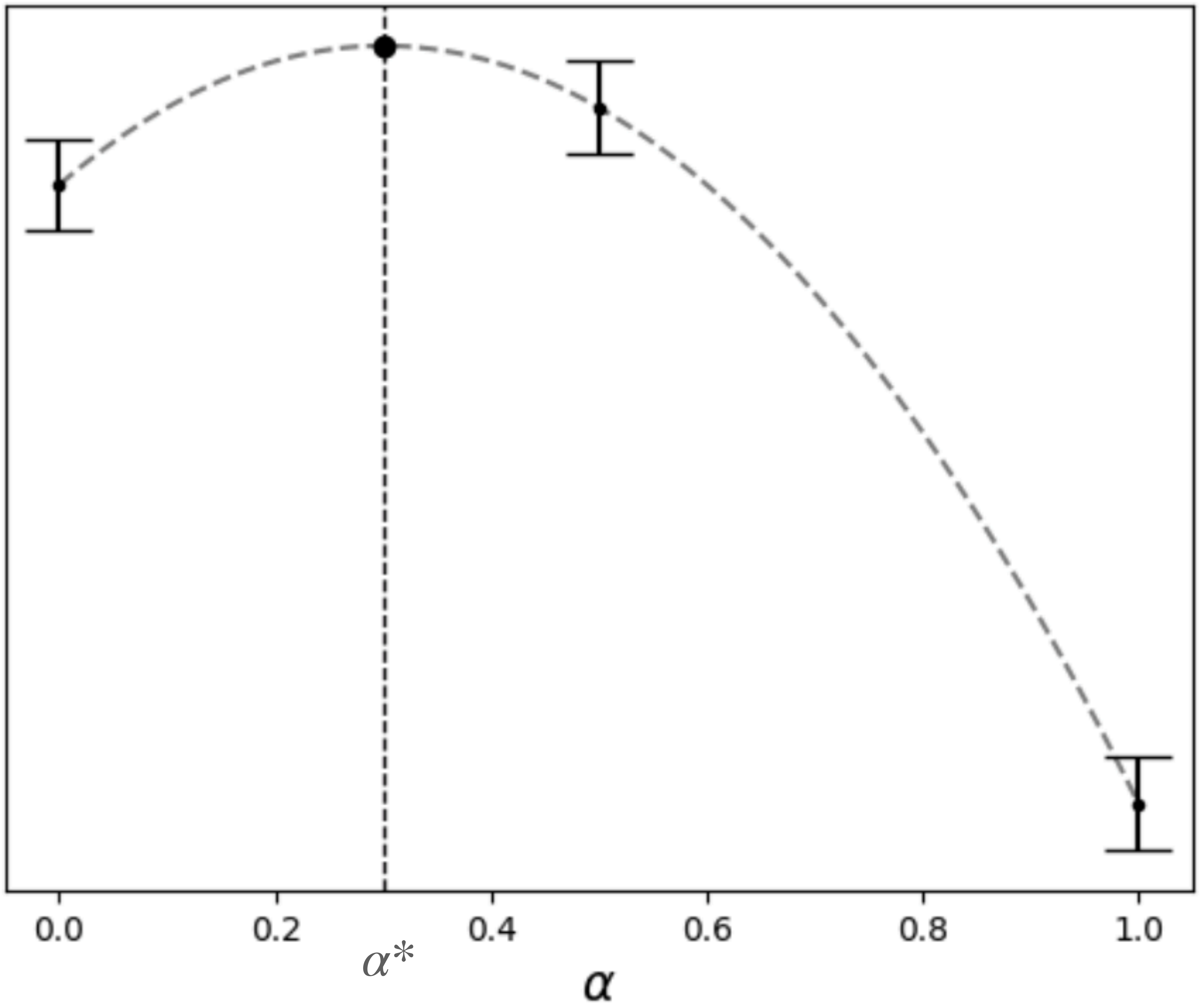
Case: Song recommender

- Measure only three arms: $\alpha \in \{0, 0.5, 1.0\}$
- Fit a parabola
- Guess/hope that max of parabola is true (expected) max



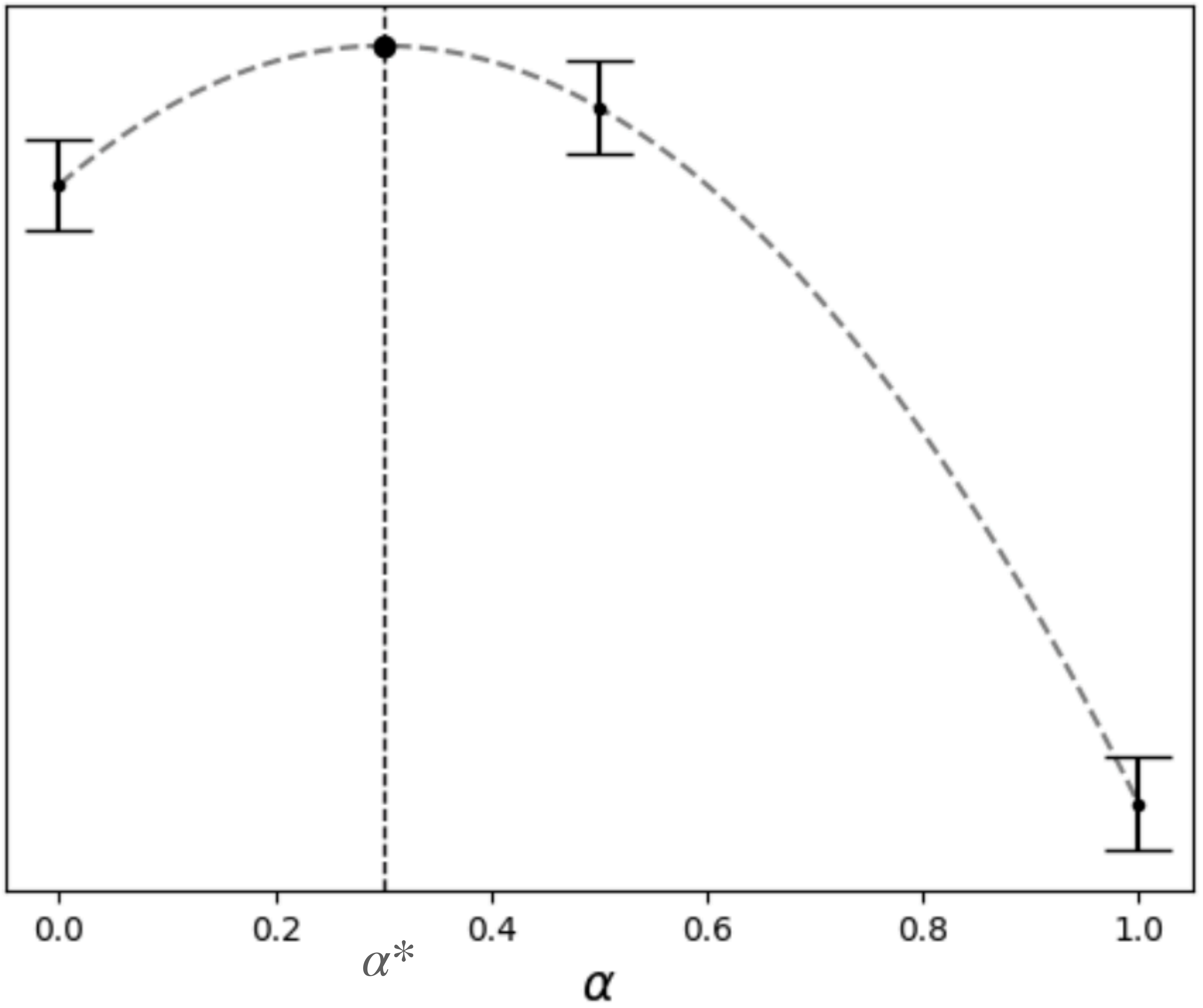
Case: Song recommender

- Max of parabola: $\alpha = 0.3$
- Run A/B test:
  A: Current prod version
  B: $\alpha = 0.3$
- A/B test validates inference (or *in*validates)



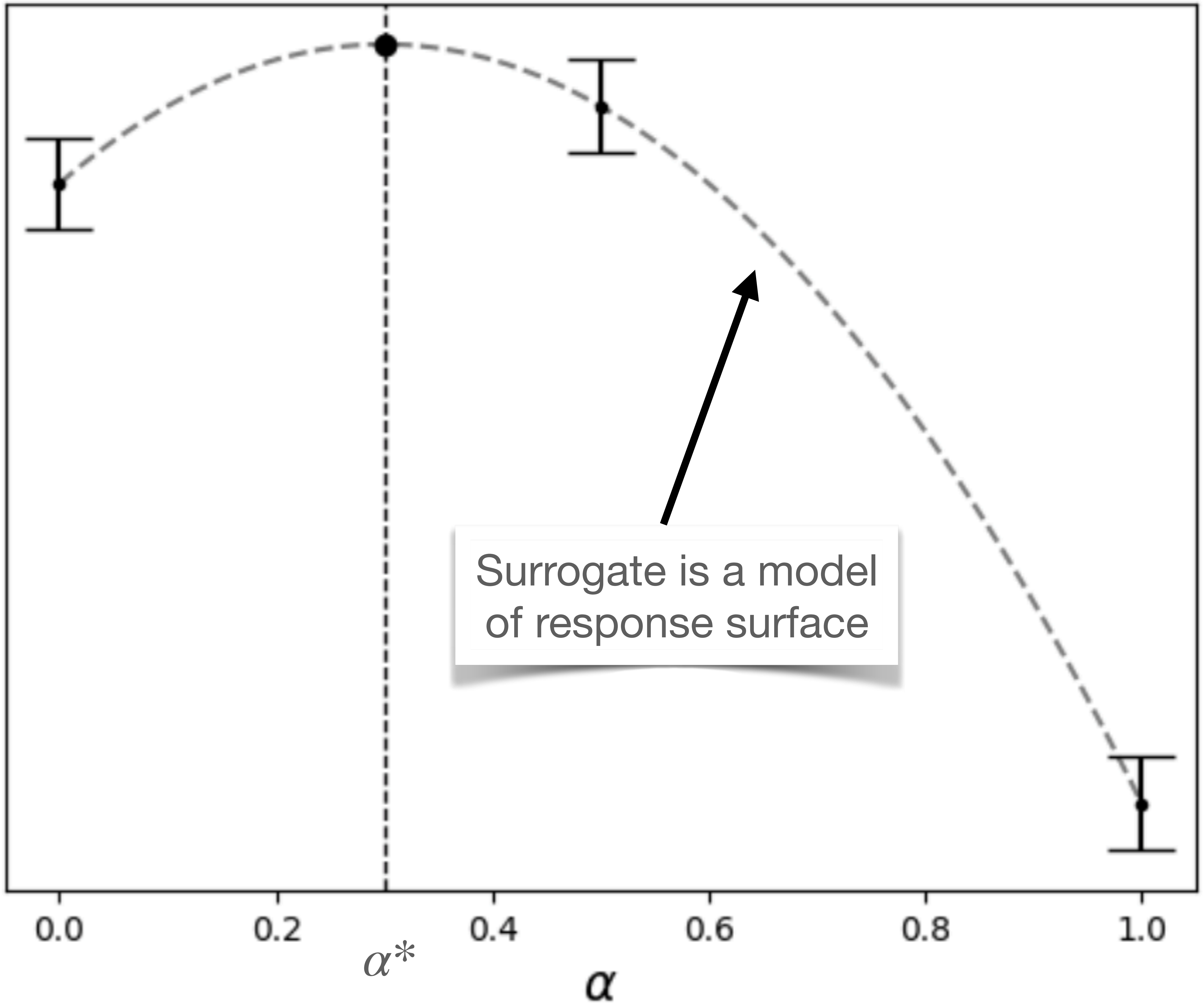
Case: Song recommender

- Function being modeled is BM vs. α
  - $y = $ BM
  - $x = \alpha$ (parameter)
- $E[y(x)]$: Expected BM as function of parameter, x



Case: Song recommender

- Unobservable, “true” BM function, $E[y(x)]$, called *response surface*
- Our fit parabola (dashed) called *surrogate function* (or just *surrogate*)
- *Response surface method:*
    - Model, optimize, validate



Compare A/B test to RSM

- A/B tests and MABs compare distinct versions of system
- RSM compares continuous family of systems
  - IOW, RSM finds optimal value of a continuous parameter

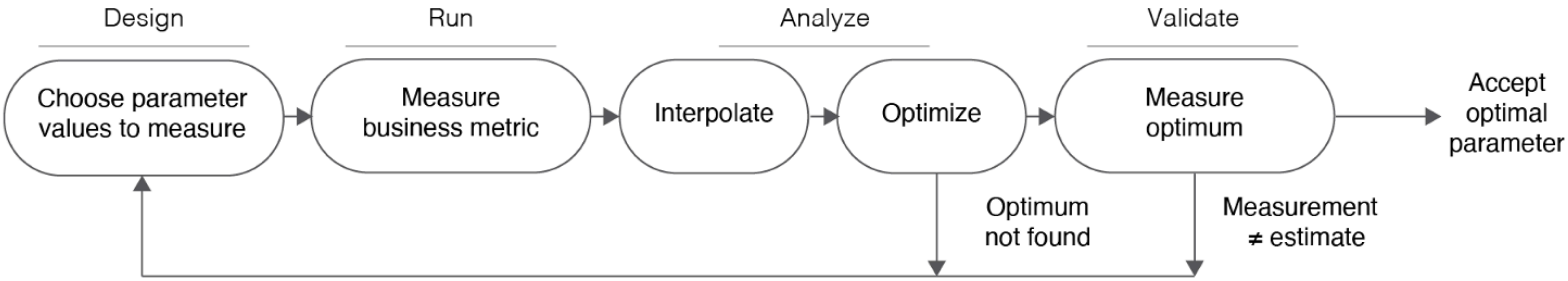
A/B Test	**RSM**
BM	BM
BM(A), BM(B)	BM(x)
$y, E[y]$	$y(x), E[y(x)]$

Compare A/B testing to RSM

- Parameter types:
    - Categorical: A, B, C, …; true/false; red/green/blue; low/medium/high
    - Ordinal: 1, 2, 3, 4, …
    - Continuous: [0.0, 1.0]; [-3.14, 3.14]; real, double, float
- Think of
    - A/B testing as optimization over a categorical parameter
    - RSM as optimization over a continuous parameter

Response-surface methodology (RSM)

Summary of method



Validate optimum

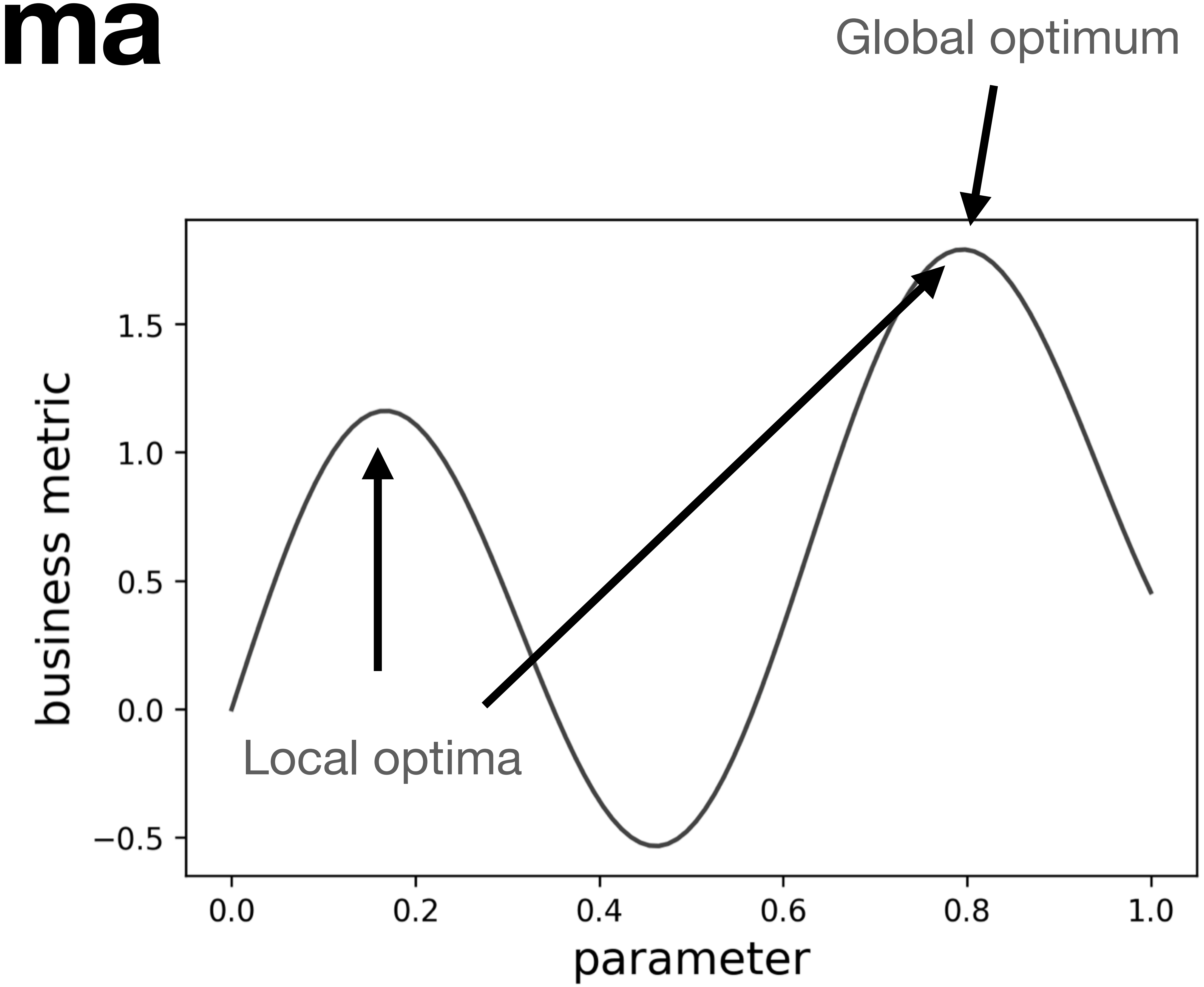
- Surrogate (model of RS) is only an approximation
- Validate by measuring at the predicted-best parameter
    - A/B test (or MAB) just best vs. old prod

Use N from A/B testing

- RSM measurements are aggregate measurements
- Use $N = (\frac{2.48\hat{\sigma}}{PS})^2$
- PS here says “If the BM of two parameters is within PS, I’ll treat them as equivalent”
- Alternatively, “I want to be within PS of the true optimum”

Local vs. global optima

- Respons surface might have multiple humps
- You want the highest hump
- RSM will only search locally
- Think hard about parameter range
- Local optimum is better than nothing

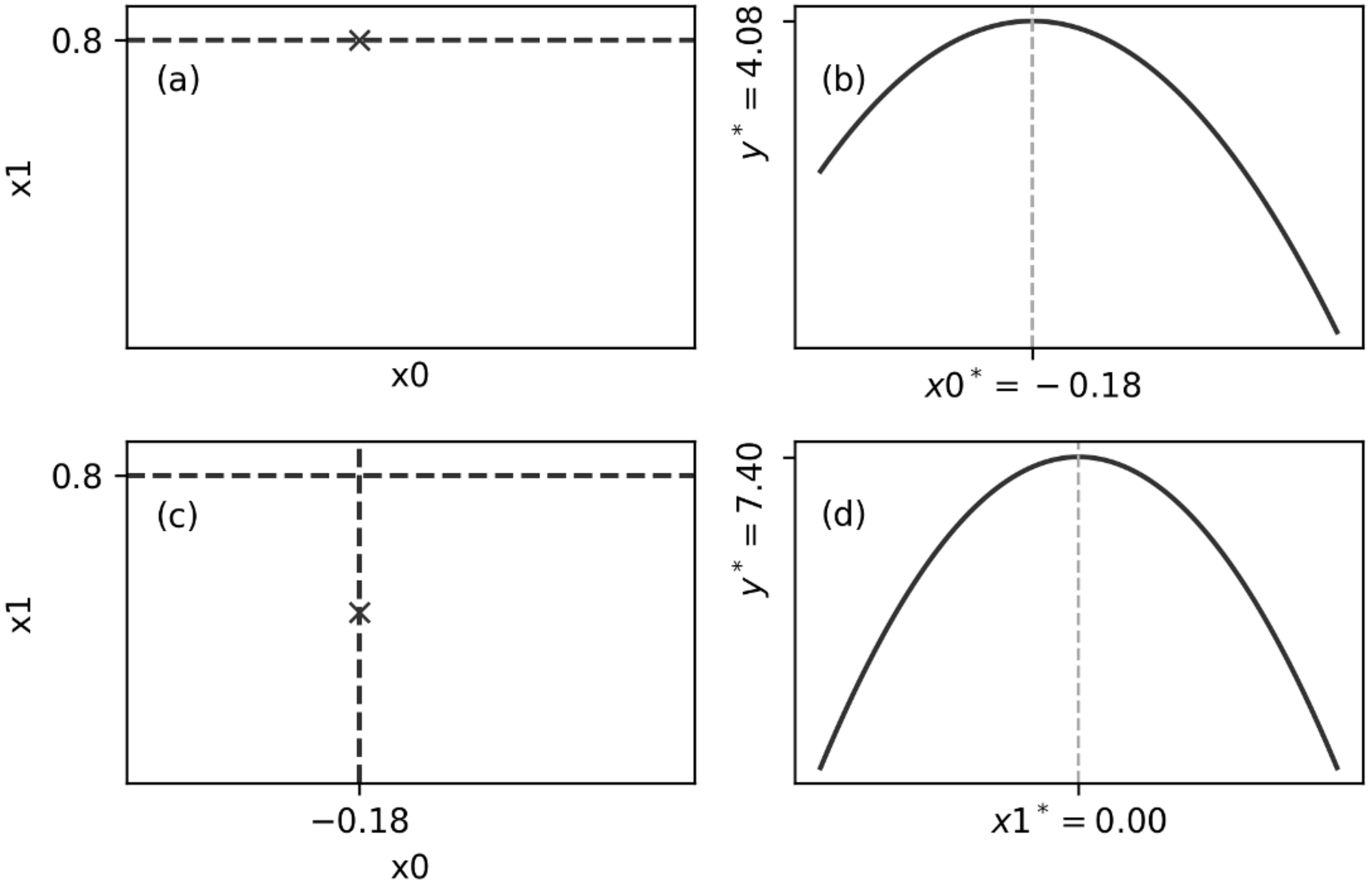


Interactive, manual process

- Engineer chooses
  - *Region of interest* (ROI): range of parameter(s) to investigate
  - Design of experiment: which specific parameter values to measure
  - Form of model — parabola? multiple parameters
- May make decisions via visualization of surrogate
- ROI “recentered” on each iteration

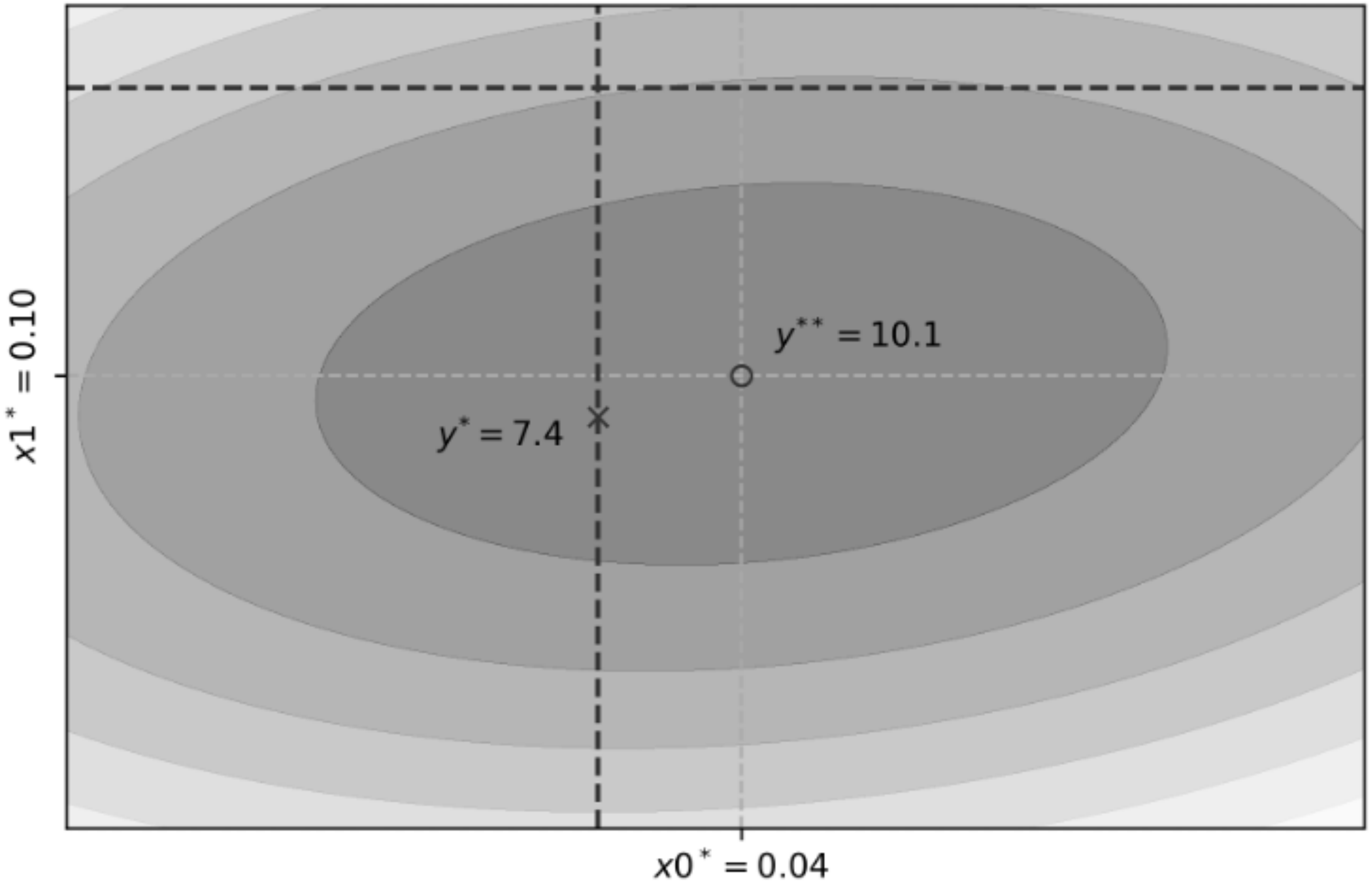
Multiple parameters

- Ex x_0, x_1
- Optimize x_0
- Optimize x_1
- *OFAT: One factor at a time*
  - Suboptimal approach



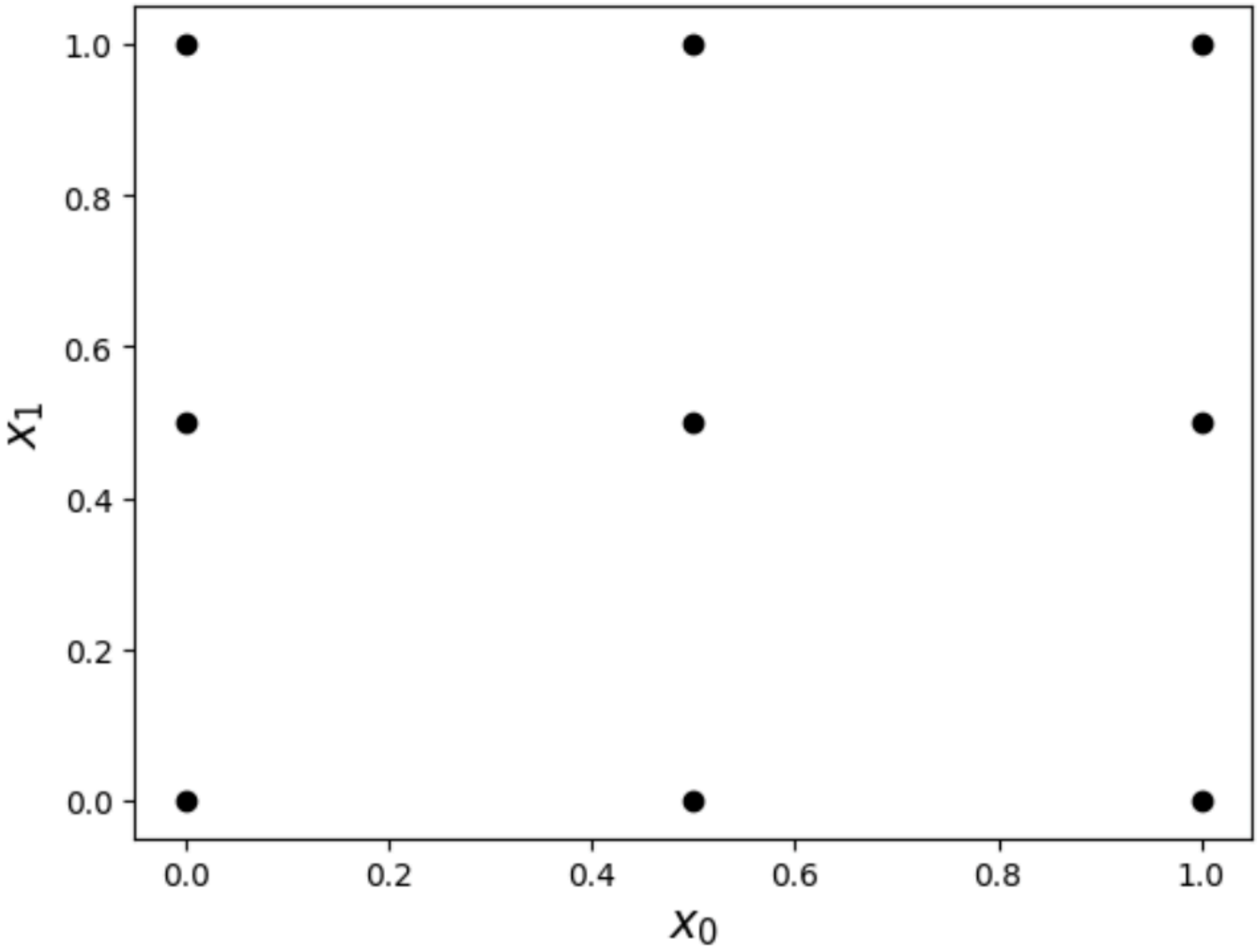
Multiple parameters

- OFAT finds $y* = 7.4$
- RSM applied simultaneously to x_0, x_1 finds $y* = 10.1$
- Realistically:
  - System has **many** parameters
  - "A few at a time" is typically as a good as it gets



Two-parameter RSM

- Two parameters (*dimensions*)
  - Take 9 measurements on a grid
  - Fit surrogate $y(x_0, x_1)$
  - Optimize to find x_0^*, x_1^*
  - A/B test
    A=current, B=x_0^*, x_1^*



Linear regression surrogate

- Surrogate model: linear regression
- Ex: $y = \beta_0 + \beta_1 x + \varepsilon$

Aggregate measurements, not observations

  - Take measurements $\{(y_0, x_0), (y_1, x_1), (y_2, x_2), \ldots (y_m, x_m)\}$
  - Fit model

$$\beta_0 = \frac{\sum_i y_i}{m}, \beta_1 = \frac{\sum_i x_i y_i}{\sum_i x_i x_i}$$

- Parabola: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$

Linear regression surrogate

- Two parameters (dimensions, 2D)
  - $y = \beta_0 + \beta_{1,0}x_0 + \beta_{1,1}x_1 + \beta_{2,0,0}x_0^2 + \beta_{2,1,1}x_1^2 + \beta_{2,0,1}x_0x_1 + \varepsilon$
  - notation: $y \sim x_0 + x_1 + x_0^2 + x_1^2 + x_0x_1$
  - even better: $y \sim \sum_i^2 x_i + \sum_i^2 \sum_j^2 x_i x_j$

Linear regression surrogate

$$y \sim \sum_i^2 x_i + \sum_i^2 \sum_j^2 x_i x_j$$

- Fit: $\vec{\beta} = (X^\top X)^{-1}(X^\top y)$

First column of X is all ones

- NumPy: `beta = np.linalg.inv(X.T @ X) @ (X.T @ y)`
- Works for any number of dimensions (parameters)

Linear regression surrogate

- More parameters
  - 3D: $y \sim \sum_{i}^{3} x_i + \sum_{i}^{3} \sum_{j}^{3} x_i x_j$
  - d dimensions: $y \sim \sum_{i}^{d} x_i + \sum_{i}^{d} \sum_{j}^{d} x_i x_j$
- Too many terms for only a few measurements
- Use automated variable selection and/or domain knowledge to limit terms

Summary

- RSM introduces
  - surrogate: model of response function
  - optimization of surrogate
- RSM is interactive/manual
  - Engineer decides ROI, design, and form of surrogate
- A/B testing : categorical :: RSM : continuous