Module 1: Introduction to Experimental Optimization

DAV-6300-1: Experimental Optimization

David Sweet // 20240829

Career trajectory

Motivation to take EO

Coin Flipping Experiment

Industrial engineered systems

Predictors in controllers

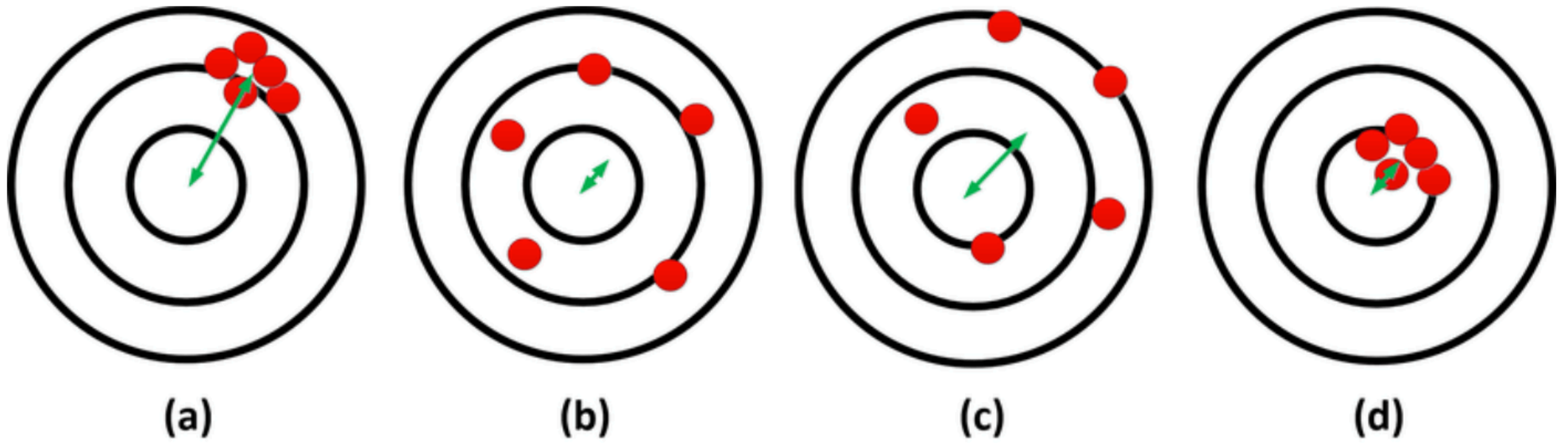
Controller	**Prediction**	**Action**	**Business Metric**
Ad server	P{click}	Show ad with highest P{click}	CPC revenue
Fraud detector	P{fraudulent}	Hold charges with high P{fraudulent} until	Avoid losing money to fraud
Trading strategy	E[return]	Buy when E[return] > 0, sell when E[return] < 0	Revenue (“PnL”)
Social media feed	P{like}	Show posts with highest P{like}	Users spend more time on feed

Evaluation

- Complexity in system, environment
  - Hard to reason about business metric
  - Hard to simulate/estimate business metric
  - Better test-set RMSE $\not\Rightarrow$ Better business metric
- Measure business metric in production

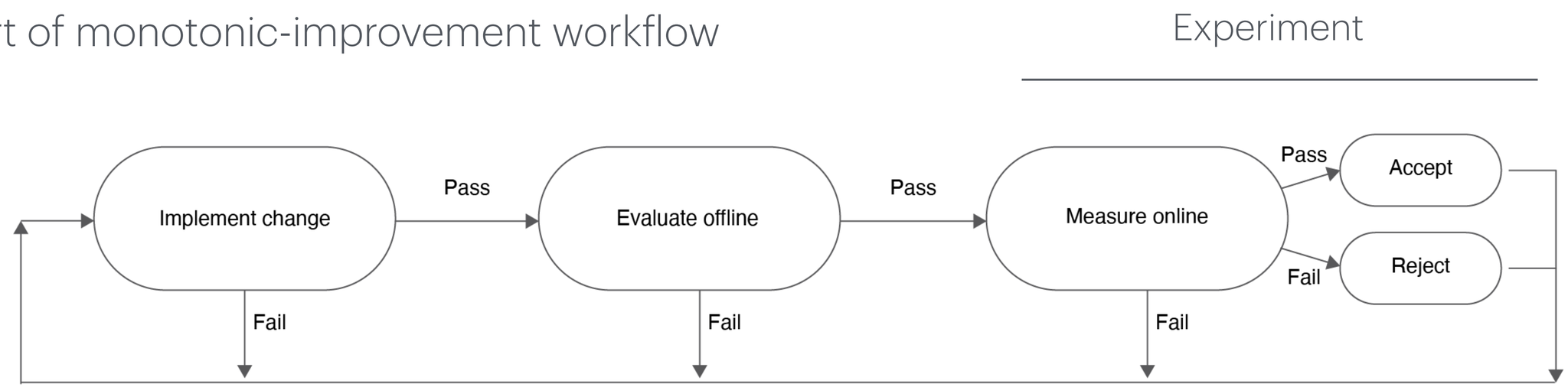
Measurement problems

- Expensive: Dollars, time, risk
- Bias and variance



Experimental methods

- Control bias and variance of measurements
- Minimize experimentation cost
- Part of monotonic-improvement workflow



Sidebar: Experimentation is everywhere

- Experimentation enables
  - Steady improvement
  - Understanding
- In diverse fields, such as
  - medicine, psychology, behavioral economics, web search, online advertising, social media, food engineering, manufacturing, food production, energy, finance

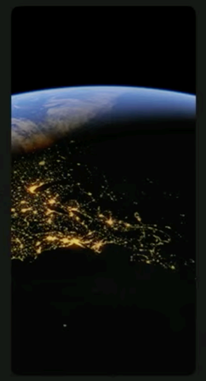
Domain Knowledge?

- Your good ideas probably won’t work. https://ai.stanford.edu/~ronnyk/ExPThinkWeek2009Public.pdf
  - Amazon reports < 50% of their A/B tests improve metrics
  - Microsoft reports only 1/3
  - Netflix reports only 10%
- My informal poll: 1/10

Just Simulate?

TESLA | Livestream

In the Real World

Ideal	Reality	Real Behavior
Simplified Bot Model	 + 	Unmodeled Dynamics ↓ Destabilization
+		
Simplified Real World Assumptions		

TESLA LIVE

Stylized History of Experimentation

(Decade precision)

- 1920: A/B testing, Fisher [agriculture]
- 1930: Bandits / sequential experiments, Thompson
- 1940: Response surface methodology (RSM), Hotelling, Box & Wilson [chemical processes]
- 1950: Bandits / sequential, Wald [manufacturing]
- 1960: Total Quality Management, Toyota [automotive manufacturing]
- 1970: Bayesian optimization (BO), Mockus; Krige [mining]
- 1980: Six Sigma, Motorola / GM; [broad dissemination, manufacturing]
- 1990: Bayesian optimization (E.G.O.), Jones

Internet redux

Same pattern, only faster

- 2000: A/B testing in web-scale systems
- 2010: Multi-armed Bandits (MAB) in web-scale systems
- 2020: Bayesian optimization (BO) in web-scale systems

Semester outline

Faster still

- A/B testing: randomization, replication, design
- RSM: continuous parameters, surrogate, offline optimization
- MAB: sequential decisions, exploration vs. exploitation
- BO: All of the above and more, automated, SOTA
  - aka: adaptive experimentation, black box optimization, model-based optimization

.

Homework

Standardized format

- One Jupyter notebook
  - Single file, .ipynb
- May include:
  - Markdown, LaTeX
  - Images of handwritten work (math, diagrams)
  - Code, Plots

Mid-term Project

Measure three versions: A, B, C

- Individual work
- Compare three versions of a system using the methods presented so far in class.
- Turn in one Jupyter notebook with measurements, analysis, and conclusion
- Deliver a 5-minute in-class presentation

Mid-term Project

Measure three versions: A, B, C

- Measurement server: http://44.198.169.94:8080
- Serves one measurement / day for each of A, B, C
- May query the server as often as you like
  - manually or via script
- Serves json

Readings for Week 2

- Expectation & Variance (brief)
  https://online.stat.psu.edu/stat500/lesson/3/3.2/3.2.1

- Expectation & Variance (long)
  A. Meyer
  https://www.cs.princeton.edu/courses/archive/fall06/cos341/handouts/variance-notes.pdf

- Central Limit Theorem and the Law of Large Numbers
  J. Orloff and J. Bloom
  https://math.mit.edu/~dav/05.dir/class6-prep.pdf

Summary

- Evaluate changes by comparing business metric values
- Measure your business metric with experimental methods
- Experimental methods
  - Control bias and variance
  - Minimize experimentation cost
- Follow a monotonic-improvement workflow