Experimental optimization

Lecture 13: Experimenting without experiments

David Sweet

Review
Experimentation cost

- Experimentation costs:
    - time spent
    - risks posed to subjects, systems
    - pay for experimenters/engineers
    - opportunity cost from running suboptimal versions/configurations

Review

Experimentation cost

- Ex: Advertising
  - poor ad policies might drive users away
  - “B” version of an ad system might produce less revenue than “A”
- Ex: Health care, experimental treatment
  - might cause illness or death
  - might miss opportunity to cure
  - wastes time/money that could have been spent on better treatment

Review

Experimentation cost

- Improved / specialized experimentation methods can reduce experimentation costs
  - MAB reduces exp. costs while experimenting
  - RSM more efficient for continuous parameters
  - CB efficient for short-term rewards, many paramters
  - BO even more efficient than RSM

Observational data

What if we just looked at the logs?

- Observational data is the data you collect all the time from, ex., logs
  - even when no experiment is running
  - system still produces measurement of business metric, features describing users, ads, posts, transactions, etc.
- Just study / model the logged data?
  - Problem: spurious correlation
  - Problem: missing counterfactuals

Observational data

Spurious correlation

- Classic example
- Barometer gives low reading just before rain

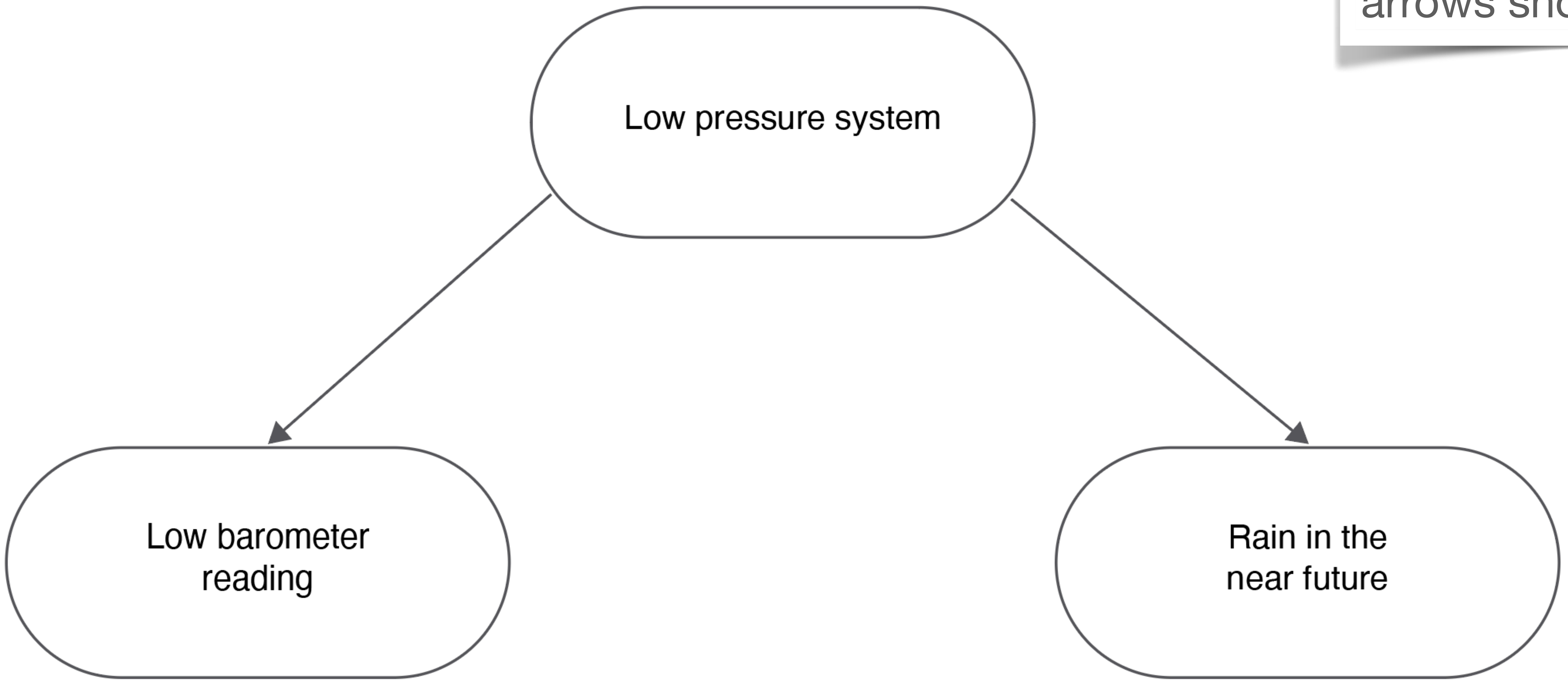
  correlation (spurious)
- You don’t like rain, so you fiddle with the barometer to keep the reading high

  causation?

Observational data

Spurious correlation

arrows show causation



Many more examples at https://tylervigen.com/spurious-correlations

Observational data

Missing counterfactuals

- But say you **do** fiddle with the barometer to keep the reading low, and it **doesn’t** rain.
- That’s more observational data
- Missing the counterfactual: What would have happened if I hadn’t fiddled?

Causation

Experiments test for causation

- A/B test *intervenes*, i.e. takes an action on the environment
  - Ex: fiddle with the barometer
- A/B test collects counterfactual
  - Ex: try both fiddling and **not** fiddling
- A/B test randomizes to break spurious correlations
  - Ex: Might fiddle (or not) before rain or sun
  - If barometer reading really caused rain, you'd get fewer samples of "fiddle before rain"

But wait: Supervised learning

Model all of the variables/factors/features

- If you really knew **all** of the factors affecting the system, this could work
  - But in the example above you were missing genuine knowledge of the low-pressure front
  - Even still, you had no observational data that contained the action “fiddle with barometer”
  - (Could you even know that you had identified **all** of the factors?)
- Contextual bandits supplement SL with exploration (randomized actions), i.e. experimentation

Natural experiments

But maybe in *my* data…

- Sometimes a randomized intervention just happens, and you can act as if you had planned it.
- Ex: Effect of Vietnam draft on later earnings
  - http://www3.nccu.edu.tw/~hmlien/pfinance/pf1/readings/draft.pdf
  - lottery (**randomization**) determined whether drafted (**intervention**)
  - Result: being drafted reduced later earnings by 15%

Natural experiments

More examples

- Ex: Oregon health insurance experiment
  - https://www.nber.org/programs-projects/projects-and-centers/oregon-health-insurance-experiment?page=1&perPage=50
  - Oregon offered 10,000 Medicare policies (**intervention**)
  - 90,000 signed up, so chose by lottery (**randomization**)
  - Result: health care caused (i) increased used of health-care services, (ii) decreased financial strain, (iii) improved depression, but no change in physical health, (iv) no effect on employment or earnings

Matching method

Example

- Ex: Effect of raising minimum wage on employment
  - https://davidcard.berkeley.edu/papers/njmin-aer.pdf
  - NJ raised minimum wage (**intervention**), PA did not
  - analyzed fast-food restaurants near the border of the two states
    - “**matching**” in lieu of randomization
    - restaurants near border exposed to similar factors
    - posit that the side of the border on which they fall is arndom

Matching method

Example

- Result: raising minimum wage did not reduce employment
- David Card awarded 2021 Nobel prize in economics for use of natural experiments like this
- more examples of natural experiments and matching methods: http://economicspsychologypolicy.blogspot.com/2015/06/list-of-19-natural-experiments.html

Causal inference

- Natural experiments and matching methods examples of *causal inference*:
  - attempt to infer causation from observational data
- save on experimentation costs
- without good domain knowledge, might miss confounders
  - not as believable as an experiment
  - but sometimes you just can’t conduct the experiment

Sequences of decisions

- Recall contextual bandit problem:
    - observe some features, the *state*; ex., user & ad features
    - take some *action*, ex., show user an ad
    - receive some *reward*; ex., user clicks on ad
- Consider a longer-term view of this problem:
    - maybe show user an ad, then show it again tomorrow, maybe skip a day…
    - and eventually they buy the product or you give up trying to sell it to them

Sequences of decisions

Reinforcement learning (RL)

- CB data was: $\{(state_i, action_i, reward_i)\}$ or $\{(s_i, a_i, r_i)\}$
  - each triple — (s_i, a_i, r_i) — represents one decision
  - there are many **independent** decisions in the data set
- RL data is: $\{(s_{1,1}, a_{1,1}, r_{1,1}, s_{1,2}, a_{1,2}, r_{1,2}, s_{1,3}, \ldots)\}$
  - each sequence contains multiple **dependent** (s_i, a_i, r_i) triples, i.e. multiple sequentially-dependent decisions
  - there are many **independent** sequences in the data set

Sequences of decisions

Reinforcement learning (RL)

- Ex: Decision making in health care
  - actions are available interventions (tests and treatments)
  - reward is health
  - high cost to taking the wrong action, so can't explore
  - well, sometimes we run RCT (ex., drug trials), so actually there is experimentation, but it's very costly and narrowly applied
  - w/o exploration would be safer, maybe more widely/often used

Sequences of decisions

Reinforcement learning (RL)

- Ex: Chatbot
  - state = recent dialog w/a person
  - action = next thing for chatbot to say
  - reward = some goal of chat; ex., help a customer return a pair of shoes
  - exploration — saying randomized things -- could turn away customers
  - prefer to learn a policy from pre-existing conversations between humans

Sequences of decisions

Reinforcement learning (RL)

- Problems (still):
  - spurious correlation
  - missing counterfactuals
- CB solution: explore (randomize) actions and refit
- RL? Could you randomize actions of whole sequences?
  - Number of possibilities grows exponentially in the length of the sequence.
  - Methods exist, but require many, many iterations of “explore, refit”

Sequences of decisions

Offline reinforcement learning

- IOW, direct (*model-free*) RL methods have very high experimentation cost
- One alternative: *Offline RL*
  - collect data from logs of existing production system, whatever version it is
  - look for sequences in that observational data
  - fit a new *policy* (mapping from s_i to a_i; a controller) from the sequences

Sequences of decisions

Offline reinforcement learning

- Offline RL solution methods
  - borrow ideas from causal inference to deal with spurious correlation & missing counterfactuals
  - stay close to existing policy to mitigate problem of missing counterfactuals
- Good read for application examples, overview of existing methods, and a focus on simplifying methods:
  - https://arxiv.org/pdf/2112.10751.pdf

Summary

- Observational data is copious and cheap. Experimental data is the opposite.
- Methods exist to tease out experiment-esque — causal —
results from observational data
- Causal inference is a set of such methods.
    - natural experiments due to randomized interventions
    - matching methods compare groups that differ in only one aspect
- Offline reinforcement learning attempts to build sequential decision-makers (controllers) from observational data