SOMETIMES THE SNOW COMES DOWN IN JUNE SOMETIMES THE SUN GOES 'ROUND THE MOON

Daniel Simpson
Department of Statistical Sciences
University of Toronto

"I AM NOT SOMEONE WHO DOES A LOT OF WORK IN CAUSAL INFERENCE"

Daniel Simpson
Department of Statistical Sciences
University of Toronto

WHAT REALLY IS THE POINT OF ALL THIS?

CAUSAL INFERENCE IS EASY

- Imagine you have you have a vector of outcomes Y, a treatment assignment vector T, and a matrix of confounders X.
- And the following two key assumptions + no interference:
  - $Y \perp T \mid X$
  - $0 < \Pr(T = 1 \mid X) < 1$
- Then you can do a pretty good job at causal inference by modelling the response surface

$$Y_i = f(X_i, T_i) + \epsilon_i$$

- Flexible things like BART or Gaussian Processes can be used

WHY IS THIS MODELLING USEFUL?

- Well, if $Y_i = f(X_i, T_i) + \epsilon_i$, then (assuming the errors have mean zero), the conditional ATE is

$$E(Y(1) - Y(0) \mid X) = f(X, 1) - f(X, 0)$$

- The population ATE can be computed with an empirical average over the observed confounders.
- This actually works quite well in practice:
  - It's pretty easy to implement
  - In a recent (and fairly fair) competition*, it tied with Targeted Maximum Likelihood as the best method

**Dorie, Hill, Shalit, Scott, Cervone. ArXiv:1707.02641 (2017)*

I LIKE BIG BUTS

But...

THOSE ASSUMPTIONS ONLY SOMETIMES HOLD

- Probably, Ephraim's example with Malaria nets will satisfy ignorability, overlap, and SUTVA and we could probably just shove Long/Lat into a non-parametric model and nothing will break.
- In some cases (eg air pollution), SUTVA probably does not hold (ie we need to know about non-local interventions)
- In other situations, we may have problems with overlap

REASONABLE PEOPLE MAY DISAGREE, BUT BEING REASONABLE IS PROBABLY TOO LOW A BAR

IT DOESN'T MATTER IF YOU LOVE JIM, OR CAPITAL J-I-M*






- On Monday Jim asked a reasonable question: What do you do if two different Bayesians come up with two different priors?
- His suggestion (reference priors) basically are terrible for complicated models, but it's a question we need to ask
- (Especially while we are free-styling causal assumptions…)

**(with apologies to Lady Gaga [and Jim])*

A SHORT PLAY (SET IN NORTH CAROLINA)

- If you make a wrong turn in North Carolina and end up at Duke, you may meet someone who says "It's my prior belief. Live with it."
- If you make a very wrong turn in North Carolina and end up in, say, the University of Washington ten years ago, you may meet someone who says "Bayes is bullshit. Use my estimating equations."
- If you make an extremely wrong turn in North Carolina and end up talking to someone who wants a damn answer rather than a holy war, they may say "Which method works."

WHATEVER WORKS

- As with all statistical methods, the only thing that really counts is whether or not it answers the question at hand.
- So the right question isn't "What happens if someone else has a prior?"
- The right question is "Does the posterior distribution answer the question?"
- Or, to put it another way, does the model fit?

If you can justify your posterior, you don't need to justify your prior.

AS MARTA SUGGESTED

- Stuff is easier if it's done on the same scale as the data
- There's a strong argument to be made that a model that can correctly predict the data is a good model to use.
- This is particularly relevant for causal inference

CAUSAL INFERENCE IS JUST PREDICTIVE INFERENCE APPLIED TO UN-SEEN SCENARIOS

LET'S BREAK THAT DOWN

- At the heart of it, Causal questions are of the form: "How would an observable change under a hypothetical intervention?"
- This is a predictive question.
- So we should be using predictive model comparison and evaluation tools
- But not just arbitrary ones (like AIC), ones that are adapted to the problem at hand!

I CONTINUE TO LIKE THE WORD "BUT" IN LARGE PRINT

But...

THERE'S AN OBVIOUS PROBLEM...

- We really want to compare the output at a single location (with the same confounders) under both arms of the trial.
- This is impossible.
- SUTVA++ allows us to avoid this quagmire, but SUTVA may not hold
- Maybe it's enough to have nearby sites with different treatments and similar confounders
- (So using matching to evaluate a model rather than using it for inference)
- Here, as in all places, replicates (or temporal structure) really would help a lot!

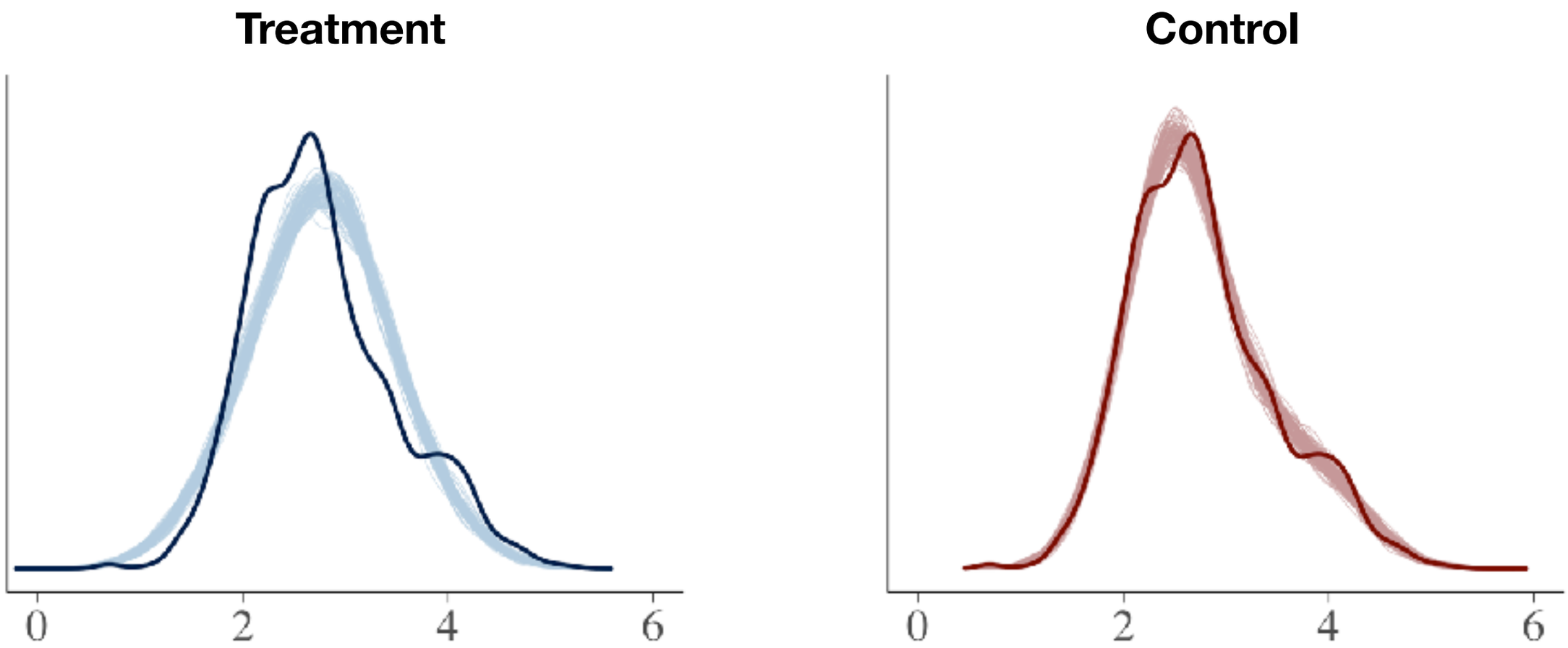
A NICE BAYESIAN TOOL FOR NICE BAYESIANS

- An under-used part of the Bayesian toolbox is the posterior predictive distribution

$$p(\tilde{y}) = \int_{\theta} p(\tilde{y} \mid \theta) p(\theta \mid y) \, d\theta$$

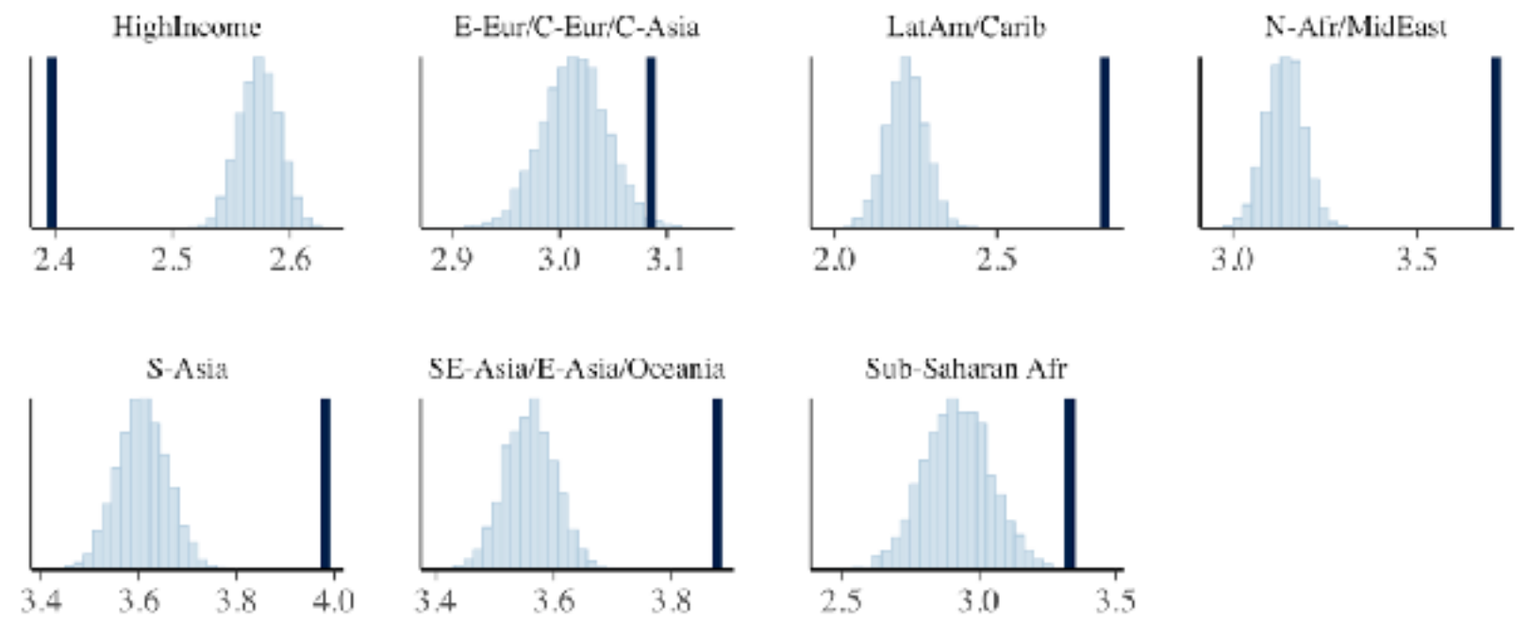
- If we seek to evaluate our posteriors rather than justify our priors, this is the thing we want.
- It is easy to compute from posterior samples.
  - Also, LOO posterior predictive distributions are easy
  - If you need to leave more structured things out (hello space!) then you just need your inference to run fast

WE CAN COMPARE POSTERIOR PREDICTIVE DISTRIBUTIONS TO DATA



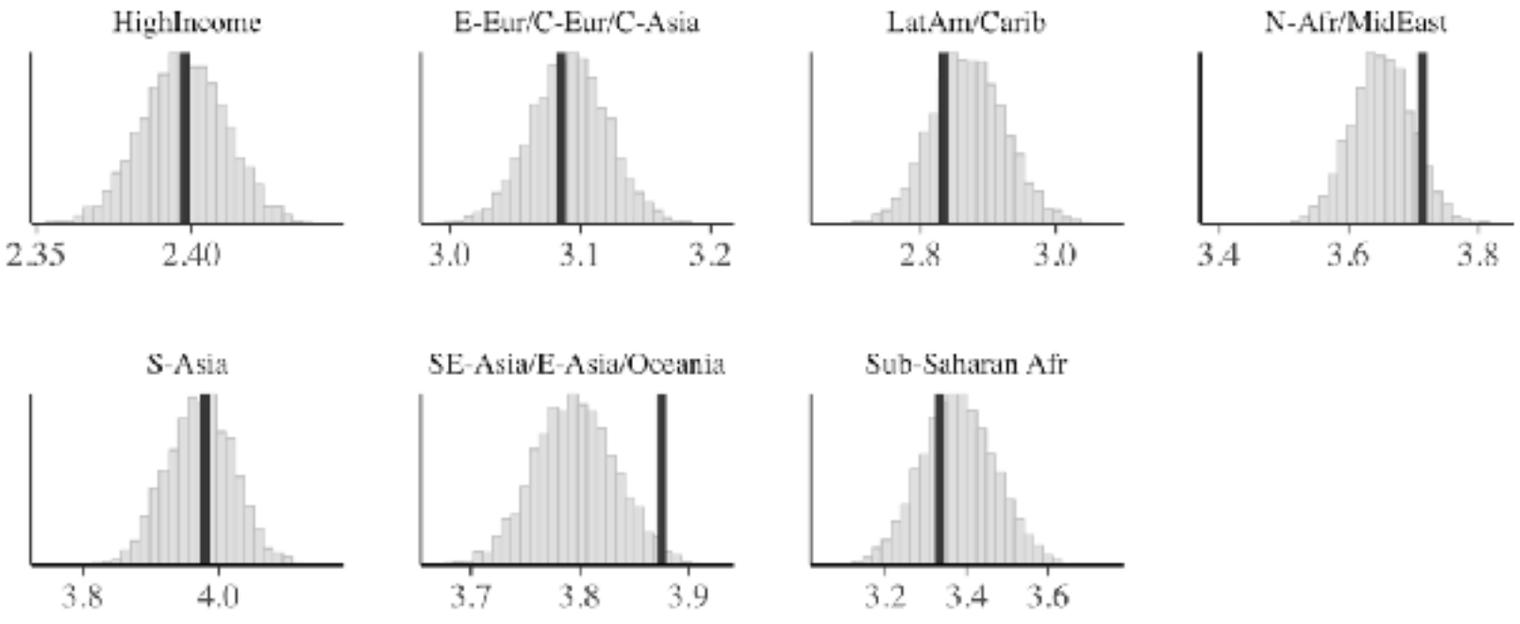
MAMMA MIGHT'VE SAID KNOCK YA OUT, BUT SHE ALSO WANTED YOU TO KNOW YOU CAN STRATIFY THESE PLOTS

Model 1 (single level)



$$T(y) = \text{med}(y|\text{region})$$

Model 2 (multilevel)



IN SEARCH OF LOST COUNTERFACTUALS

GAVIN AND JOE TALKED ABOUT OTHER THINGS

- So all that was the clean causal case where we had a control and an intervention
- But PREFERENTIAL SAMPLING isn't like that
- We have no controls (or at least no obvious ones)
- It's really the "presence only" version of Causal Inference
- Unsurprisingly, people like to use point processes here…

THE IMAGINARY DESIGNER (WHO HAPPENS TO BE REALLY INTO POINT PROCESSES)

- One option is to imagine there is an magical designer who constructs k observation locations according to some unknown strategy that depends on the confounders $\boldsymbol{X}$.
- And the probability of selecting an ensemble is

$$\Pr(sites = S \mid \mathbf{X}) \propto \ell(S; \mathbf{X})$$

- This loss function would take into account the need for diversity and agglomeration
- Janossy density for a point process ie Pr(k points in the pattern and they're within infinitesimal neighbourhoods of S)
- (side point: if these things are well mixing, then maybe this is a way towards joint propensity scores)

KILLER JOE

- Joe used a logistic regression, which is a pseudolikelihood
- he added constructed covariates to mimic repulsion and excess attraction
- There's a real research programme here on building estimating equations for repulsive point processes and using them in this context

WELL I'D LIKE TO HANG OUT,
BUT I CAN TELL THAT YOU'RE NOT A DRINKING CROWD

THE VISITORS (CRACKING UP)

- There is so much work to do here
- Existing methods in spatial statistics are clearly inadequate when it comes to answering causal questions
- We need to think carefully about what a counterfactual is
- We need to strip the biostats causality literature for parts and get something useful out of it (this is hard–they write as if they're scared of being understood)

I DON'T DANCE NOW, I MAKE MONEY MOVES

- The one thing that I hope we keep a careful eye on is model specification and checking
- It is **not** enough to simple *declare* that a method is good
- Model comparison tools (WAIC, DIC, LCPO, etc) are not adapted to this problem
- One subtlety with complex models (which Data Scientists deeply understand, but statisticians don't appear to) is that the estimates we construct come from a complicated interaction of the model and the computational method.