BIOMED SCI 552:

STATISTICAL THINKING

LECTURE 9: CAUSAL INFERENCE AND BAYESIAN STATISTICS

QUESTIONS FROM THURSDAY?

ASSOCIATIONS VS. CAUSES

- "X is associated with Y" means that for some level of X, Y occurs more or less often than another level of X
- Association and correlation are often used interchangeably, though correlation is often assumed to be a linear association, and associations can be *much* more complex
- Association does not imply causation
- *But* all causal relationships are inherently also associations

PREDICTIVE AND DESCRIPTIVE

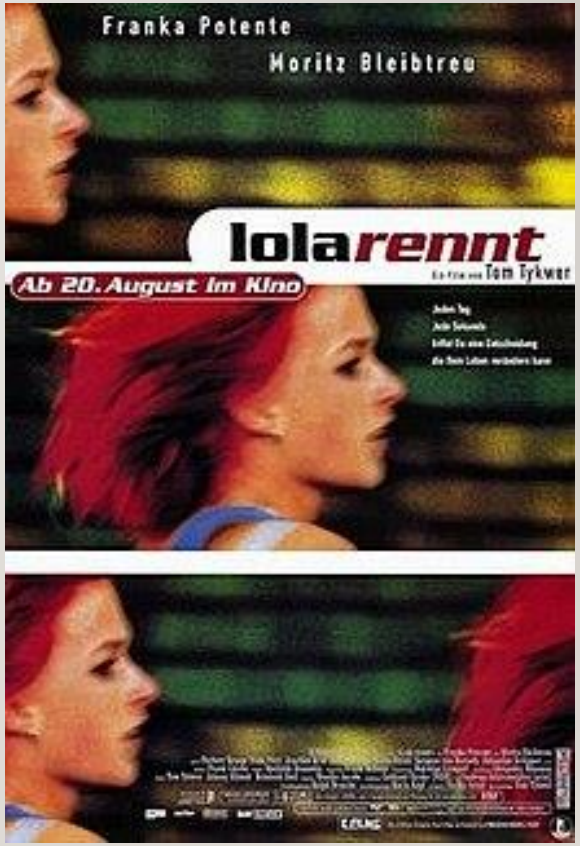
- Descriptive Epidemiology: Primarily concerned with describing associations without necessarily assigning causal arguments to them
- Predictive Epidemiology: Diagnostics and prognosis. Causal relationships are *helpful* but not necessarily required
  - Note this can lead you badly astray
  - Rotavirus and prom dresses
- Generally speaking, without one of those two qualifiers, the goal of most biomedical studies *is* to assign causal meaning, or at least go looking for it

THE PROBLEM WITH INTERVENING

- If you intervene on something that isn't a cause, it isn't going to matter
- Banning prom will not eradicate rotavirus
- To intervene effectively, we need a notion of *cause*

COUNTERFACTUALS

- Counterfactual questions are at the core of causal inference and underly most medical research.
- What if a patient had been given Treatment A instead of Treatment B
- What if someone had never smoked?
- These are impossible to answer – but important to articulate



POTENTIAL OUTCOMES

- A phrase often associated with counterfactuals
- What *can* happen
- For many health states, this is binary.
- $Y^{X=0}$ vs. $Y^{X=1}$
- If Sarah had been assigned to the treatment arm of an RCT (X=1) and *not* gotten COVID-19, that is her *factual* outcome
- Her outcome with X=0 is thus the unobservable counterfactual

WHERE THIS GETS TRICKY

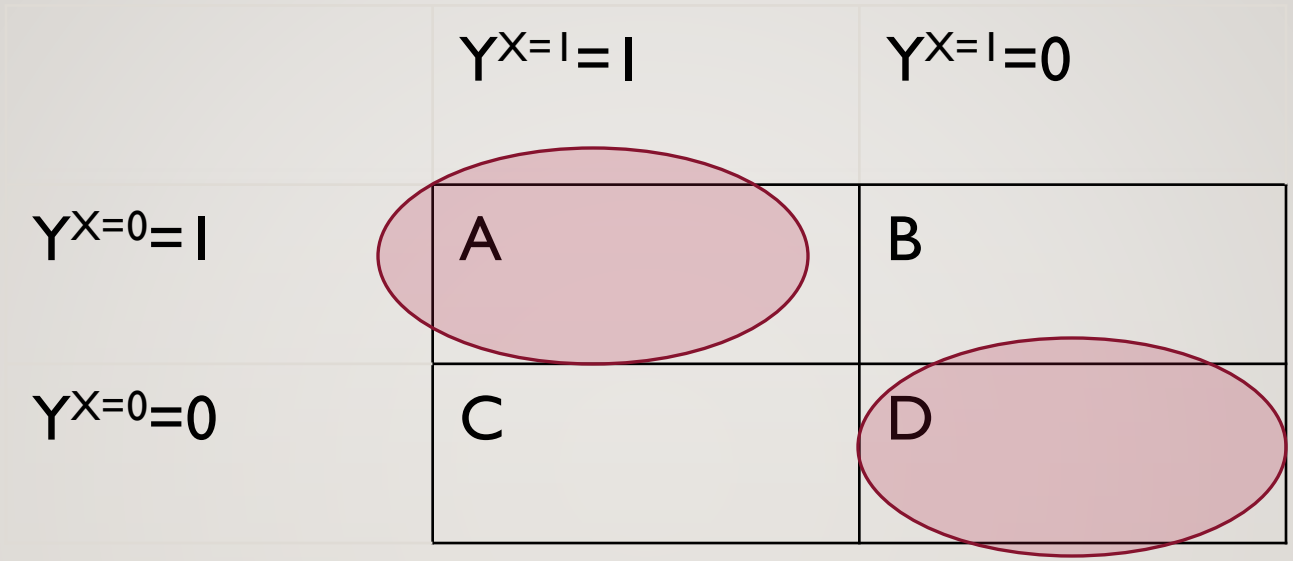
- Consider whether obesity (as measured by BMI) causes heart disease
  - Leaving aside the actual question, take this as true
- Factual outcome:
  - $Y_{obese=1}$
- What is the counter-factual?
  - $Y_{obese=0}$
- What intervention does that describe?
- Ideas?

INDIVIDUAL CAUSAL EFFECT

- This is the difference between an individual's potential outcome for X=1 and their potential outcome for X=0
- Barring a time machine, this cannot, but definition, be observed
- But it's kind of what we want, and it's *absolutely* what medicine wants

AVERAGE CAUSAL EFFECTS

- Aka "Population Average Causal Effects"
- The difference between two groups *at a population level*


$Y^{X=1}=1$
$Y^{X=1}=0$
$Y^{X=0}=1$
A
B
$Y^{X=0}=0$
C
D

	$Y^{X=1}=1$	$Y^{X=1}=0$
$Y^{X=0}=1$	A	B
$Y^{X=0}=0$	C	D

AVERAGE CAUSAL RISK DIFFERENCE

- A+C/N – A+B/N
- $P(Y^{X=1}=1)-P(Y^{X=0}=1)$
- $P(Y^{1})-P(Y^{0})$
- The average causal risk ratio is similar
- Right now, think about this conceptually – worry about how we calculate this later

TARGET POPULATION

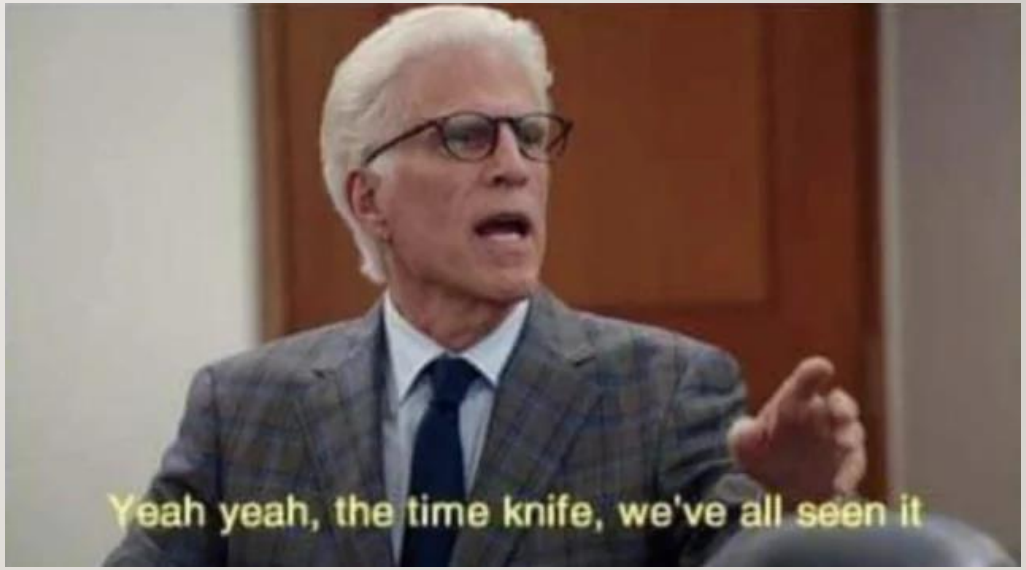
- All of this is, ideally, focused on a population we are interested in, the "target population"
- *This is not necessarily the same as the study population*
- Consider a randomized trial – we are interested in all people who might take a drug, but the trial participants are likely a subset of those

CAUSAL IDENTIFICATION CRITERIA

- These are things you need to have for causal effects to be capable of being estimated
  - Temporality
  - Consistency
  - Exchangeability
  - Positivity

TEMPORALITY

- Causes must precede effects in time
- Easiest to solve by making sure no one has the outcome at the time the study begins
- Relatively straightforward for the experimentalists, harder for observational science



CONSISTENCY

- Among people who were exposed, their outcome is the same as if they had been assigned that exposure and vice versa for unexposed
- "Treatment variation irrelevance"
  - Different doses of a drug – probably violate consistency
  - Different timing – maybe
  - Different shape of pills – probably not
- Dependent happenings (often called interference in the causal inference world) do a number on this

EXCHANGEABILITY

- Risks are equal outside the effect of exposure
- This is where RCTs draw a lot of their claims about being the only means of estimating causal effects (they aren't)
- Conditional exchangeability: Risks are equal outside the effect of the exposure *conditional on other variables*
  - This is how you can attempt to estimate causal effects in observational studies

POSITIVITY

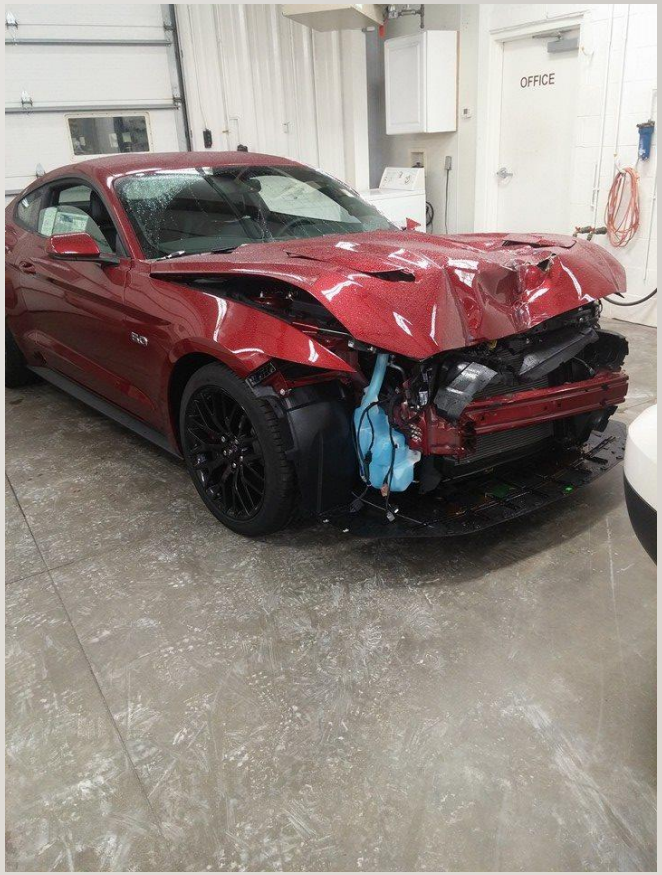
- All subjects must have a non-zero probability of being exposed or unexposed
  - There are no always or never exposed study subjects
  - Note this must be true across all values of an exposure or outcome for non-binary versions
  - Note that does not mean there has to *be* anyone in that category
  - For conditional exchangability *this includes the conditional variables*

OTHER CONCEPTIONS OF CAUSALITY

- Koch's Postulates
  - Especially important in infectious diseases
  - The microorganism must be found in abundance in all organisms suffering from the disease but should not be found in healthy organisms.
  - The microorganism must be isolated from a diseased organism and grown in pure culture
  - The cultured microorganism should cause disease when introduced into a healthy organism.
  - The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.
- Can anyone see a problem with these?

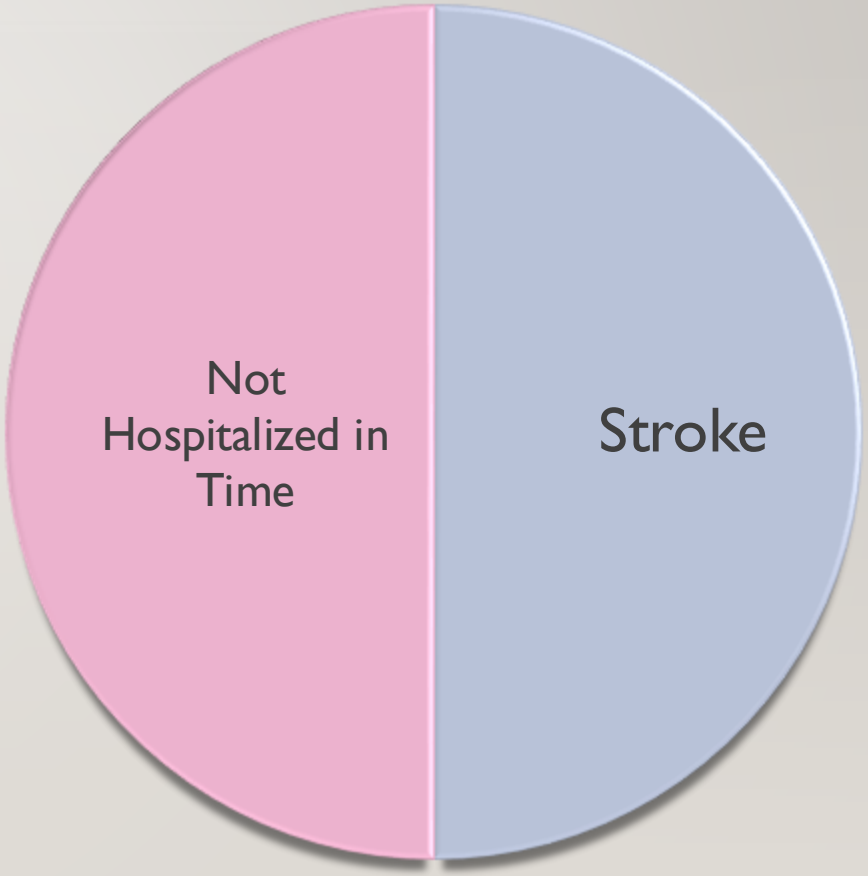
NECESSARY AND SUFFICIENT CAUSES

- Necessary Causes:
  - If X is a necessary cause of Y, the presence of Y implies the occurrence of X. X does not however imply that Y will occur.
- Sufficient Causes:
  - If X is a sufficient cause of Y, then the presence of X will cause Y. Y however does not imply X occurred.



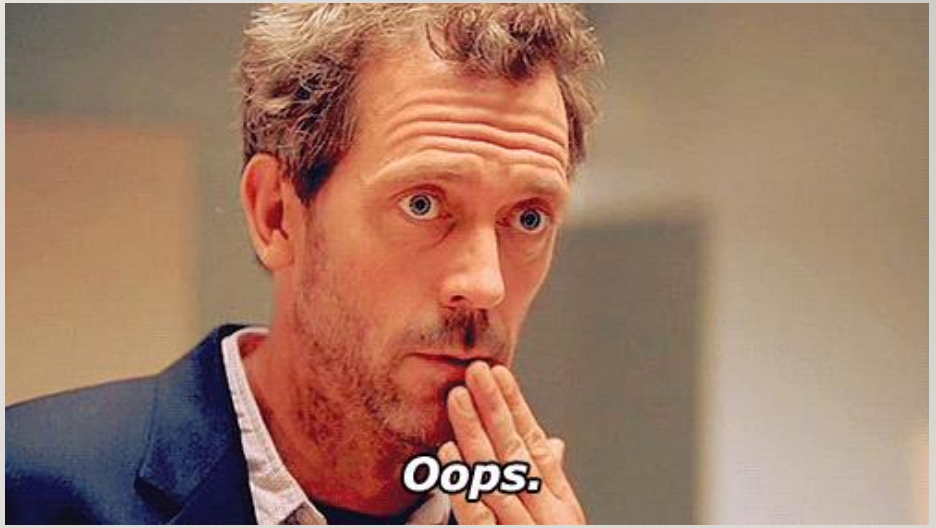
CAUSAL PIES AND CAUSAL COMPONENTS

- Introduced by Ken Rothman
- Useful for understanding multi-cause outcomes
- Probably the only good use for pie charts
- Death from Stroke
  - If either thing hadn't happened, the outcome would not have occurred
- Mostly used as a conceptual framework



BRADFORD HILL'S CAUSAL CRITERIA

- Lecture given in 1965 to the Royal Society of Medicine
- Nine Criteria
- "We can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis, and none can be required as a *sine qua non*."
  - "Essential condition"



HILL'S CRITERIA

- Strength
- Consistency
- Specificity
- Temporality
- Biological Gradient
- Plausibility
- Coherence
- Experiment
- Analogy

STRENGTH

- The larger an association, the more likely an effect is causal
  - Infectious diseases gets lots of big, double-and-triple digit relative risks because of clean causation

CONSISTENCY

- Consistent findings over different samples in different places strengthens the case that an effect is causal
- Aka Reproducibility

SPECIFICITY

- Causation is likely if an exposure leads to a specific outcome, not multiple outcomes
- You can get overly cheeky with this – "Arsenic specifically causes arsenic poisoning"


PARTICIPATION

SPECIFICITY

TEMPORALITY

- The cause occurs before the effect

BIOLOGICAL GRADIENT

- There is a dose-response relationship for continuous exposures – greater exposure should lead to greater effect (or at least, changed effect)

PLAUSIBILITY

- A plausible mechanism between cause and effect is helpful
- This is limited, but is often useful for ruling out causal arguments

COHERENCE

- Agreement between laboratory findings and epidemiological studies
- This criterion is often abused by certain industries demanding further and different sources of evidence

EXPERIMENT

- If you can show something in an experiment, you're likely on the right track
- Again, this is where RCTs draw a lot of their argument from

ANALOGY

- It is helpful if you can leverage evidence that a similar exposure causes a similar outcome

A BIT ON BAYES

MORE COMPLEX BAYESIAN ANALYSIS

- Up to this point, we have largely been using somewhat "toy" Bayesian examples – ones that can be calculated by hand, don't have distributions or uncertainty, etc.
- Clearly, we have to go beyond that
- The math gets complicated quickly

CONJUGATE PRIORS

- A prior distribution P(A) is called "conjugate" if, when combined with the likelihood function(P(B|A)), the resulting posterior distribution (P(A|B)) is of the same *family* of probability distributions
- The scope of that is best reserved for a dedicated Bayesian statistics course, but the *implications* are helpful
- Conjugate priors can result in *closed form* solutions to the posterior
- That is, you can solve for the posterior using math (albeit complex math)

WHAT HAPPENS IF YOU DON'T HAVE THIS?

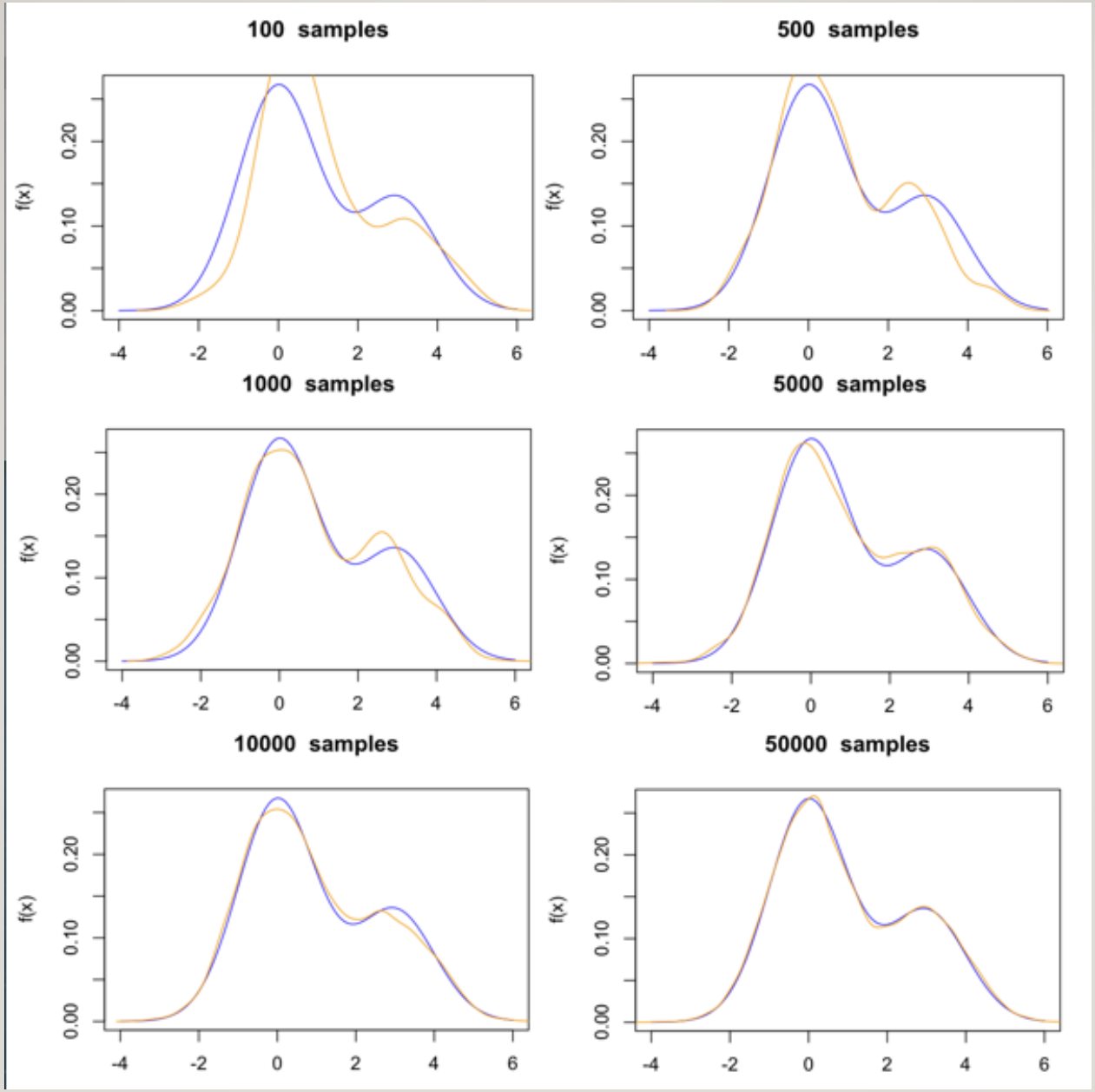
- If there's not a closed form solution to the posterior distribution, that means we don't know it mathematically
- We still potentially have a tool to deal with this
- Would anyone like to guess what it is?
  - Hint: We've seen it before…a lot

WE CAN SAMPLE!

- We use a class of algorithms called "Markov chain Monte Carlo" or MCMC to draw samples from a probability distribution – this time, the posterior, to estimate its properties
- Markov Chain: A Markov chain (or Markov process) is a stochastic process describing a sequence of possible events where the probability of each event only depends on the current state
  - "What happens next only depends on now"
  - No information is retained
  - These are *all over the place*
- Monte Carlo: Just means repeated random draws

HOW THIS WORKS

- Usually several Markov chains are allowed to randomly sample the posterior distribution
- These will, once they reach equilibrium, allow us to estimate the moments of the posterior distribution – its mean, median, what we call credible intervals, etc.
- These are approximations, but usually, quite good ones


100 samples
500 samples
1000 samples
5000 samples
10000 samples
50000 samples
f(x)
0.00
0.10
0.20
-4
-2
0
2
4
6

METROPOLIS-HASTINGS

- One algorithm you see a lot, and is often the default for MCMC software
- An example of how this algorithm works, shamelessly stolen from https://www.publichealth.columbia.edu/research/population-health-methods/markov-chain-monte-carlo

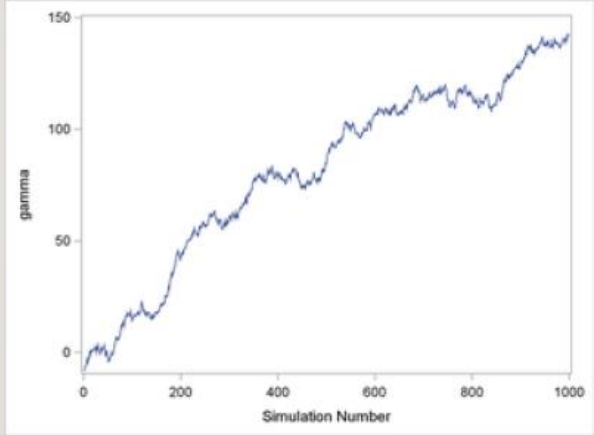
A politician is campaigning in 7 districts, one adjacent to the other. She wants to spend time in each district, but due to financial constraints, would like to spend time in each district proportional to the number of likely voters in that district. The only information available is the number of voters in the district she is currently in, and in those that are directly adjacent to it on either side. Each day, she must decide whether to campaign in the same district, move to the adjacent eastern district, or move to the adjacent western. On any given day, here's how the decision is made whether to move or not:

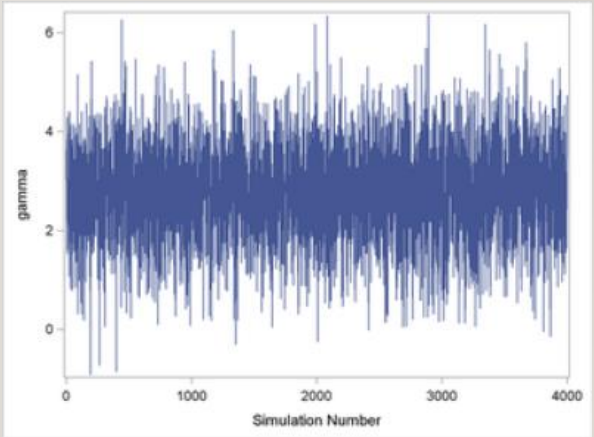
1.Flip a coin. Heads to move east, tails to move west.
2.If the district indicated by the coin (east or west) has more voters than the present district, move there.
3.If the district indicated by the coin has fewer likely voters, make the decision based on a probability calculation:
4.calculate the probability of moving as the ratio of the number of likely voters in the proposed district, to the number of voters in the current district:
5.Pr[move] = voters in indicated district/voters in present district
6.Take a random sample between 0 and 1.
7.If the value of the random sample is between 0 and the probability of moving, move. Otherwise, stay put.

Over time, this process will converge on the voter distribution in the districts

CAVEATS

- "Over time, this process will converge on the voter distribution in the districts"
- That convergence can be very slow
- There is often what's called a "burn in" period where you toss out the first N samples as the random walk…randomly walks for a bit until it starts finding equilibrium
- There's also often autocorrelation in your samples, and getting this requires "thinning" – i.e. taking only the 100th sample
- Tuning models like this is very complex, but this is meant to give you a feel





DATA PROBLEMS IN BIOMEDICAL SCIENCE

- What if we know *something* is going on, but our data isn't great?
- Examples:
  - A single sentence from a paper that is about something else
  - A summary statistic or set of summary statistics, but not the individual level data
  - Information on one scale when you need it on another
  - Like half of disease ecology
- "Data parasites" - NEJM

APPROXIMATE BAYESIAN COMPUTATION

- In the crudest form:
  - Sample a parameter from a prior distribution
  - Simulate the system using that parameter
  - Accept the parameter if the simulated results meets some criteria, usually with a tolerance e
  - Can use particle filtering or other algorithms to improve computational efficiency
- As e approaches 0, this is the Bayesian posterior
- If $e > 0$, this approximates the posterior

WHY THIS IS USEFUL

- Theoretical:
  - Likelihood-free inference
  - Acceptance may be made using summary statistics (with some loss of information), qualitative patterns (with lots of loss of information), etc. to constrain a model in ways that are difficult to capture using likelihood-based methods
  - This involves giving up some information compared to directly fitting to data, but in our use case, this has already been done for us
- Practical:
  - Relatively straightforward to think about/implement
  - Gains in the simulation engine translate to gains in fitting

MY INITIAL REACTION

- This has been an idea for a long time – simulated priors were proposed by Rubin, but computing power wasn't there yet
- Everything beyond a toy example *needs* some form of HPC/parallel coding – this is a major drawback

WHAT THIS GETS US

- Bayesian inference is a lot of work, and much more complicated than most statistical methods
- There are some upsides though:
  - You get to use prior information
  - You seem terribly sophisticated
  - Bayesian methods (mostly MCMC) can help in circumstances where some maximum likelihood based methods (what we normally use in regression) fail
- But the big one is the posterior – you can estimate whatever you like from it
- Credible intervals – i.e. the 2.5^{th} and 97.5^{th} percentiles of the posterior give you what most people want a confidence interval to be.
  - There is a 95% probability that the true estimate lies within this range

Fin