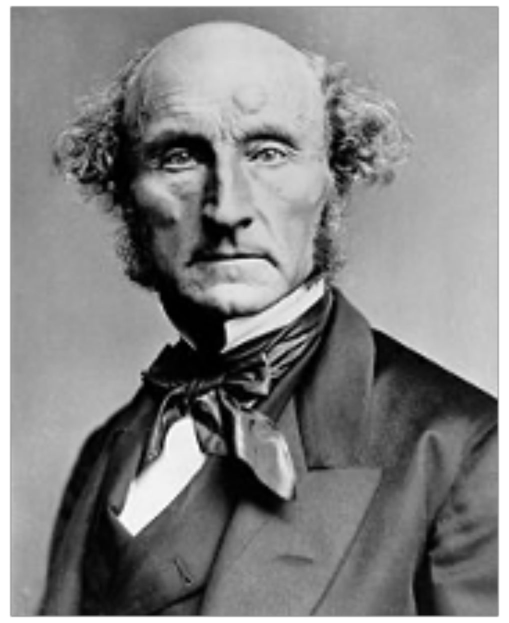
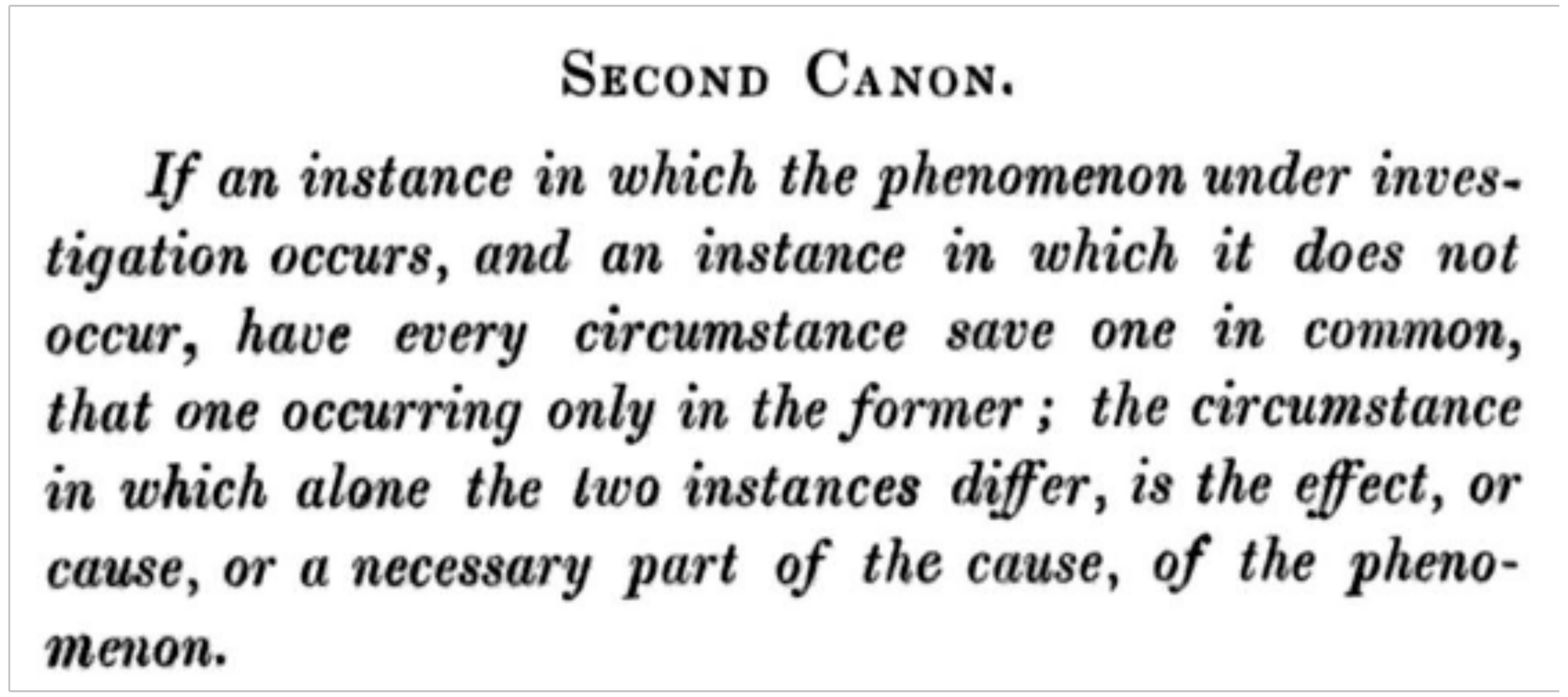
How can we give scientific explanations for observed data?

We need **a model** to explain what we observed and tested.

Munafo & Davey Smith, "Repeating Experiments is not enough" Nature (2018)



John Stuart Mill



SECOND CANON.

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.

JS Mill, A system of Logic (1843)

Peter Lipton

Contrastive Explanation & causal triangulation, Philosophy of Science (1991)

George Davey Smith

Today’s lecture: Bayesian, PGM, Causality

- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

Statistical Methods for High-dimensional Biology

Bayesian Inference & Probabilistic Graphical Models, Causal Inference Tools

Yongjin Park, UBC Path&Lab, STAT, BC Cancer

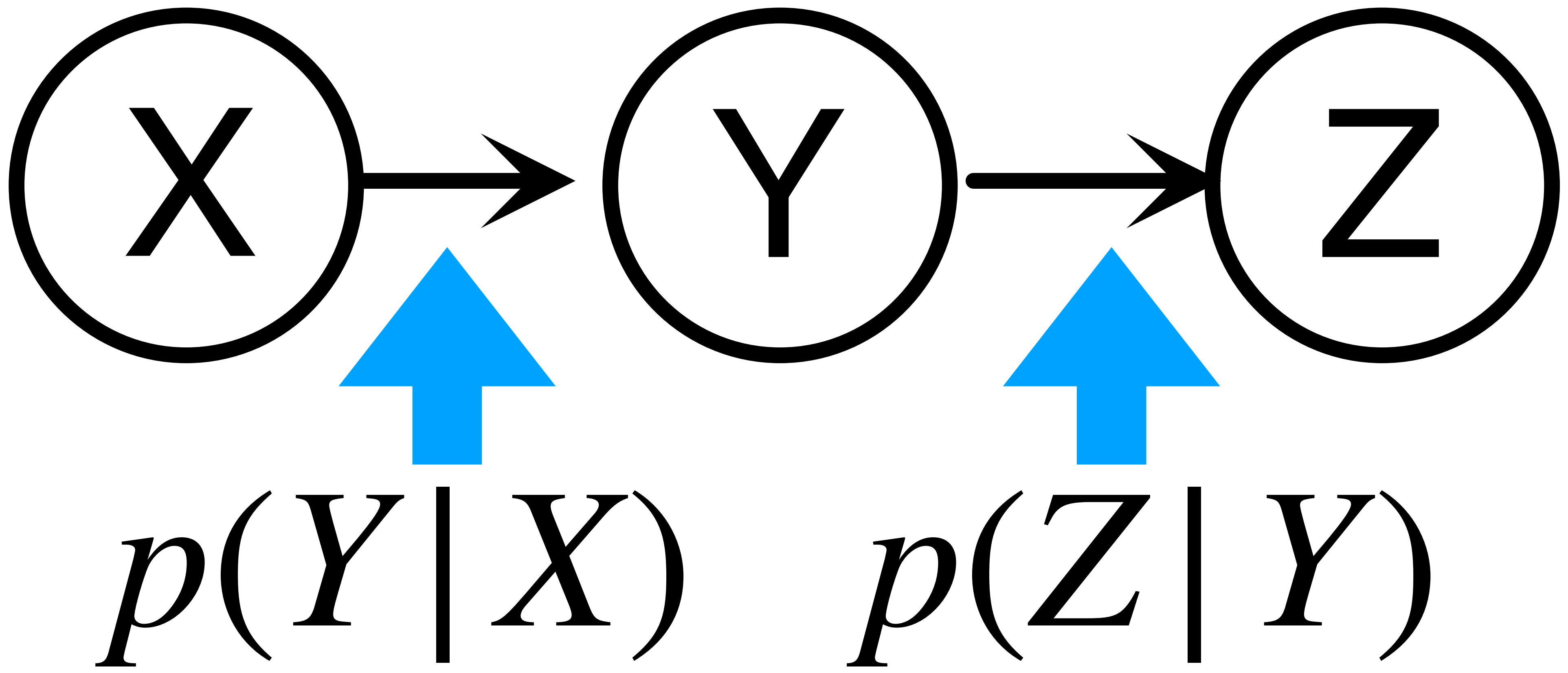
What will be a good language to describe our model, hypothesis, belief, or assumptions?

- It must be easy to understand (non-mathematical audience should be able to grasp the meaning intuitively)
- It must generalize well across many different scientific fields
- It can represent the notion of uncertainty
- It must allow a general-purpose computer algorithm to simulate and compute.

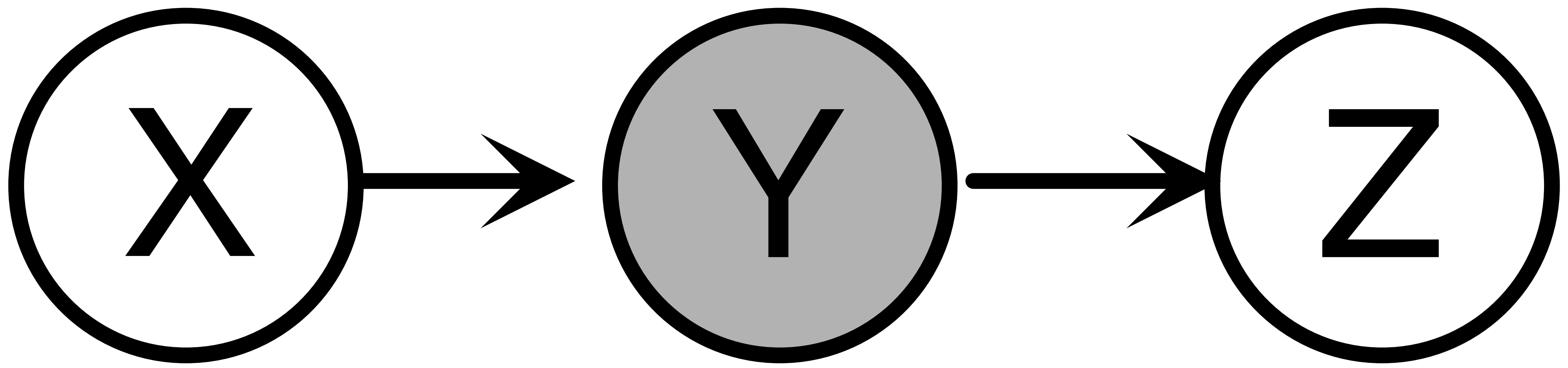
Today's lecture: Bayesian, PGM, Causality

- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

A graphical language for a probabilistic model

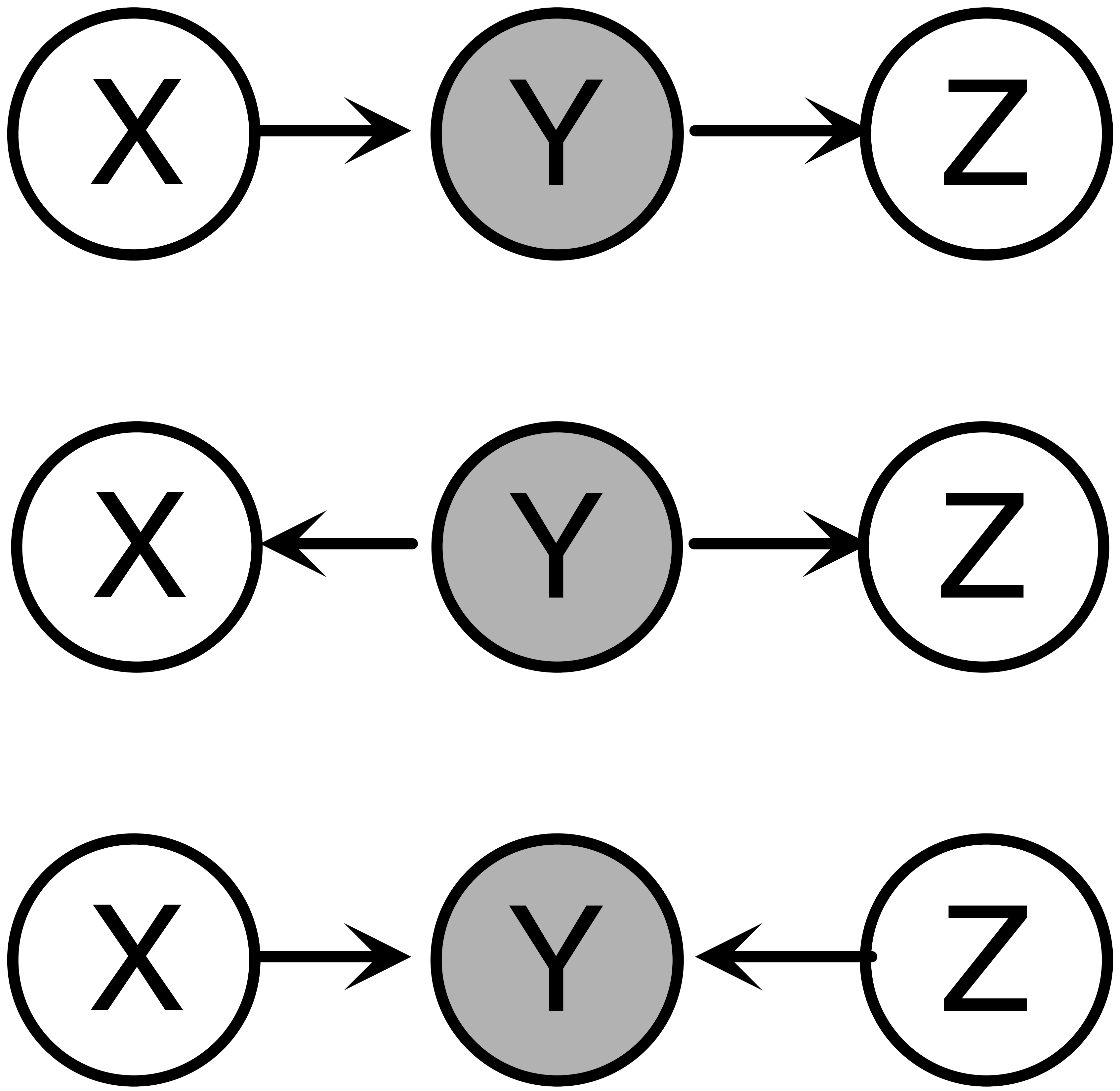


- Each node: a random variable
- Arrow/edge: dependency, conditional probability

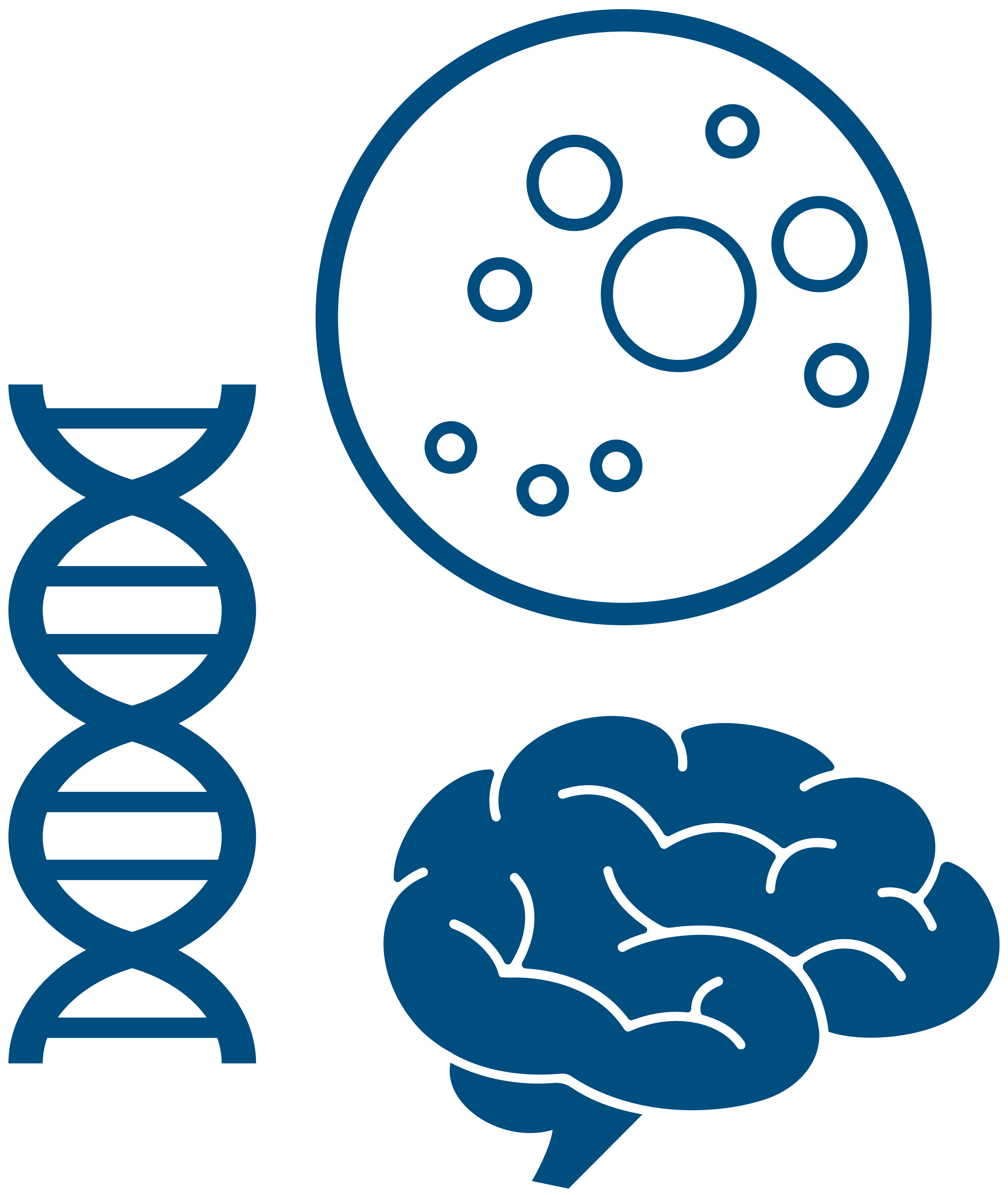


- Shaded: conditioned/observed
- Open: unobserved/unknown (yet)

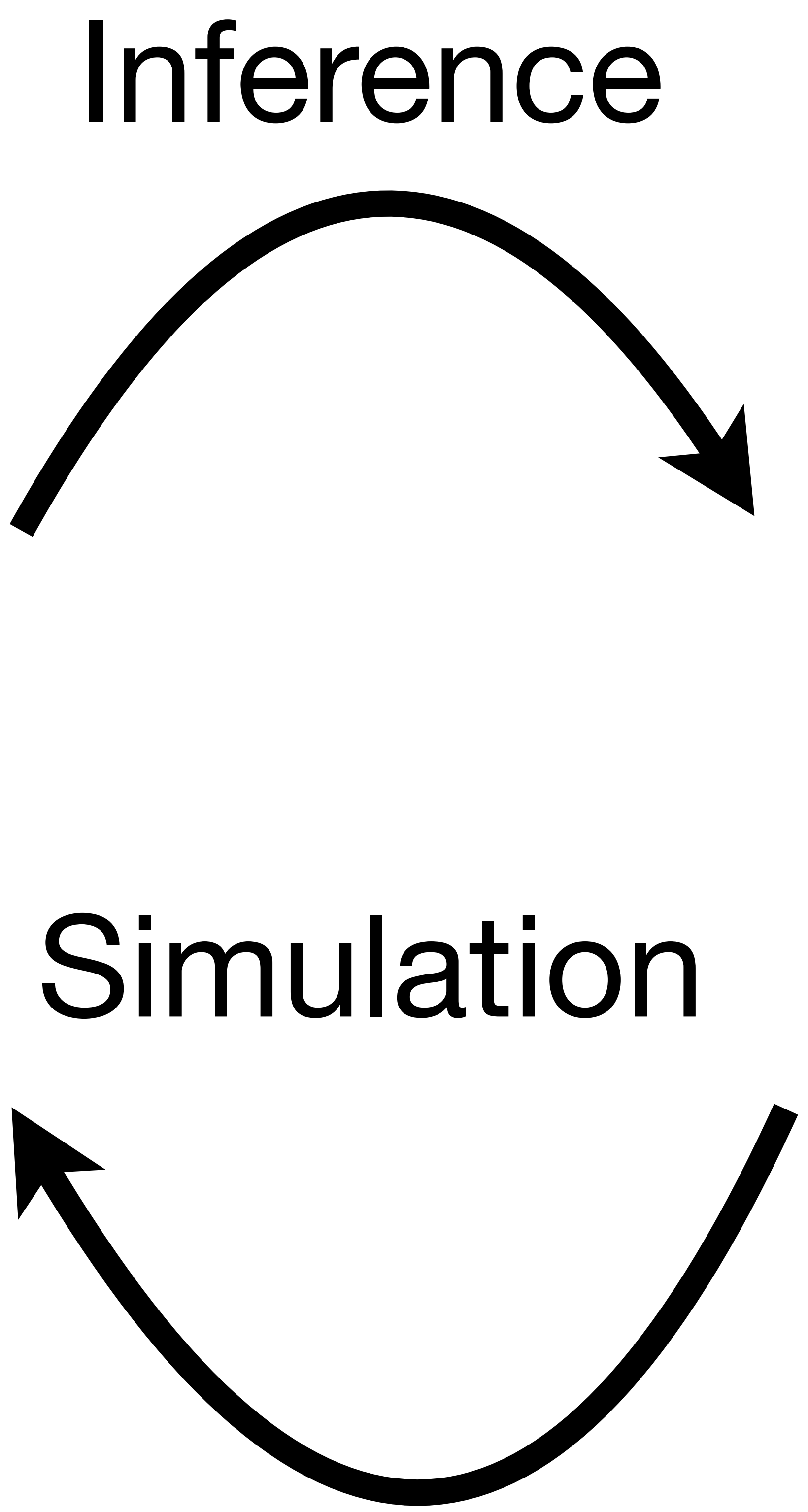
Almost all the joint probabilities (parametric models) can be described by a graph language

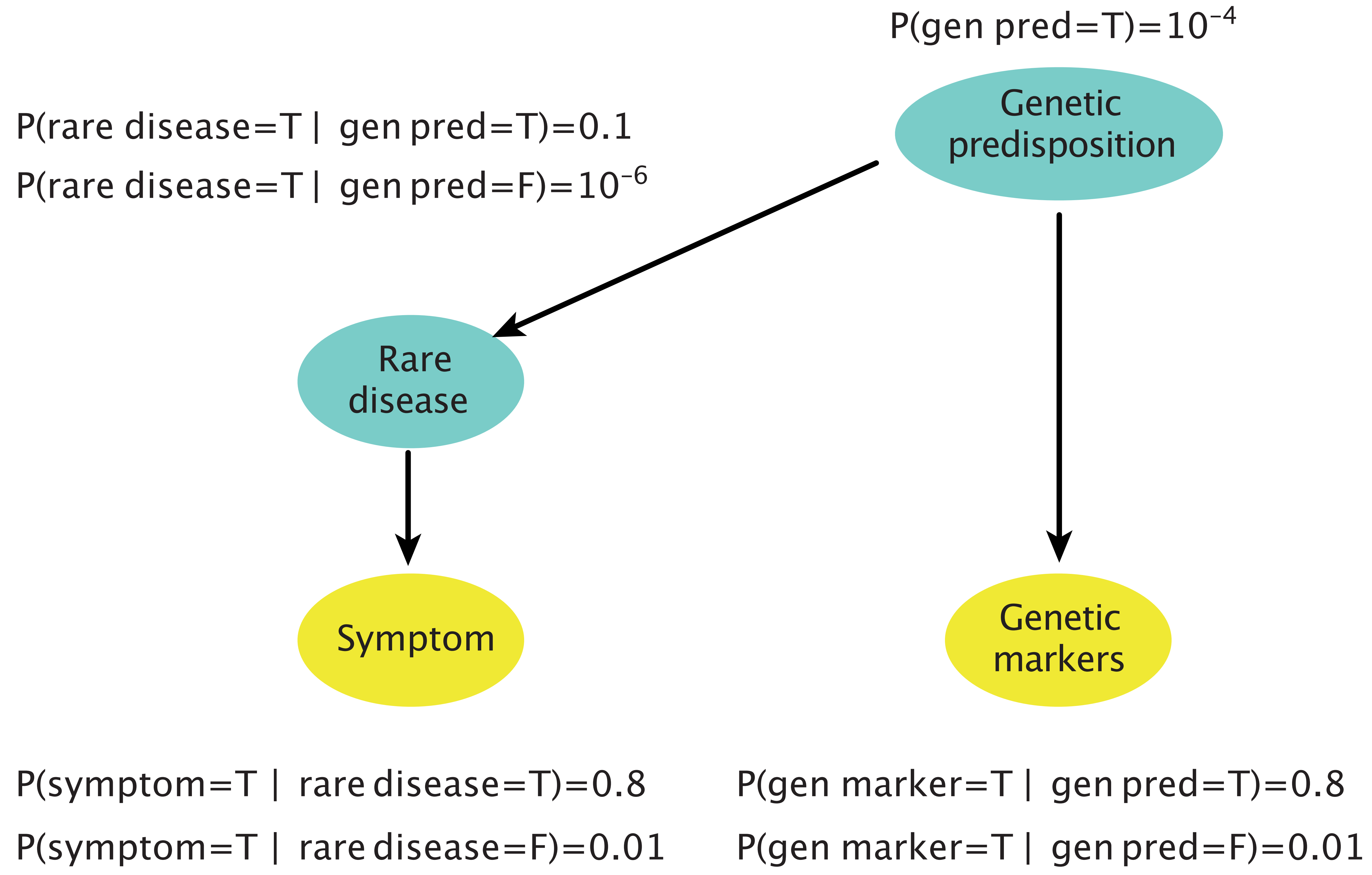


Modelling = synthesizing conditional probabilities!



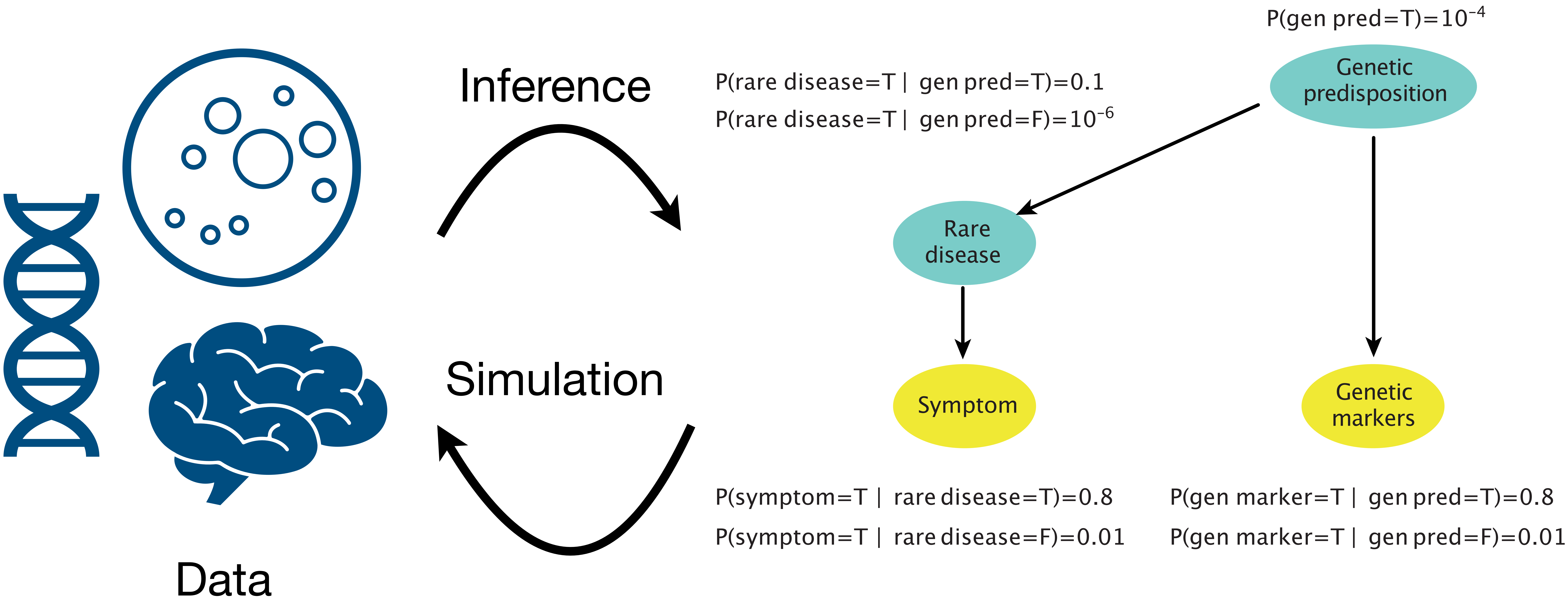
Data





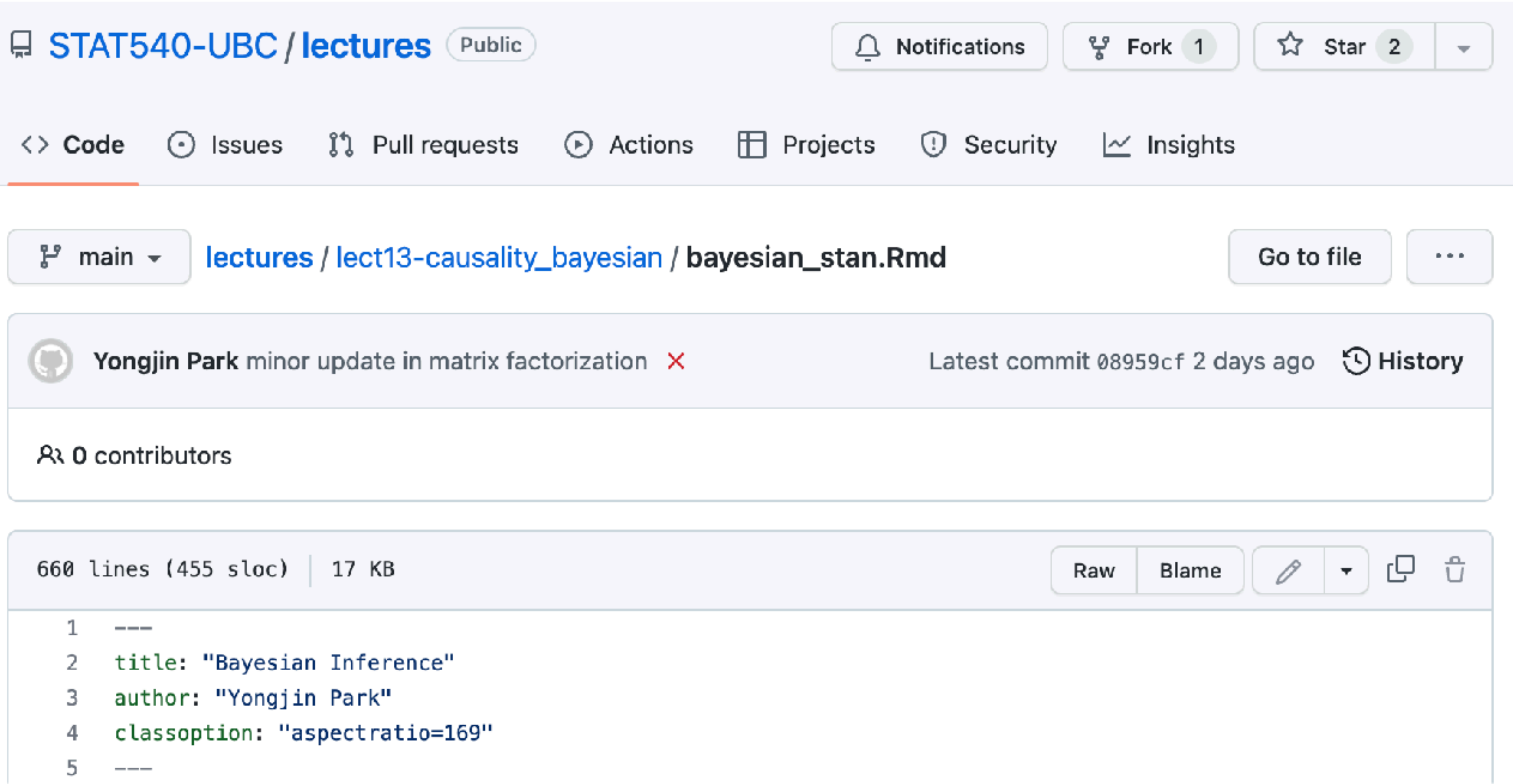
Z. Ghahramani, *Nature* (2015)

Bayesian inference is a model-based approach



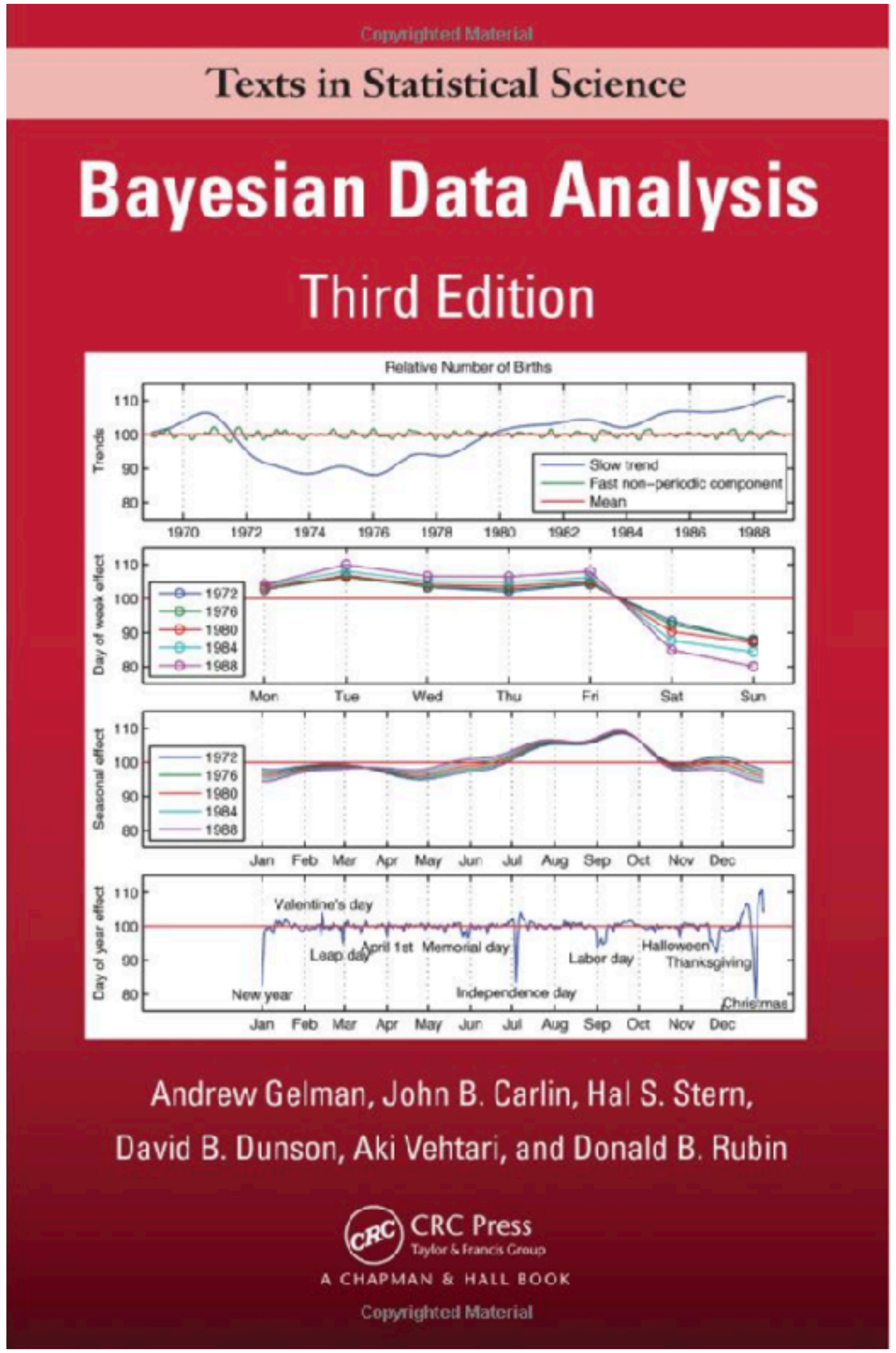
Z. Ghahramani, *Nature* (2015)

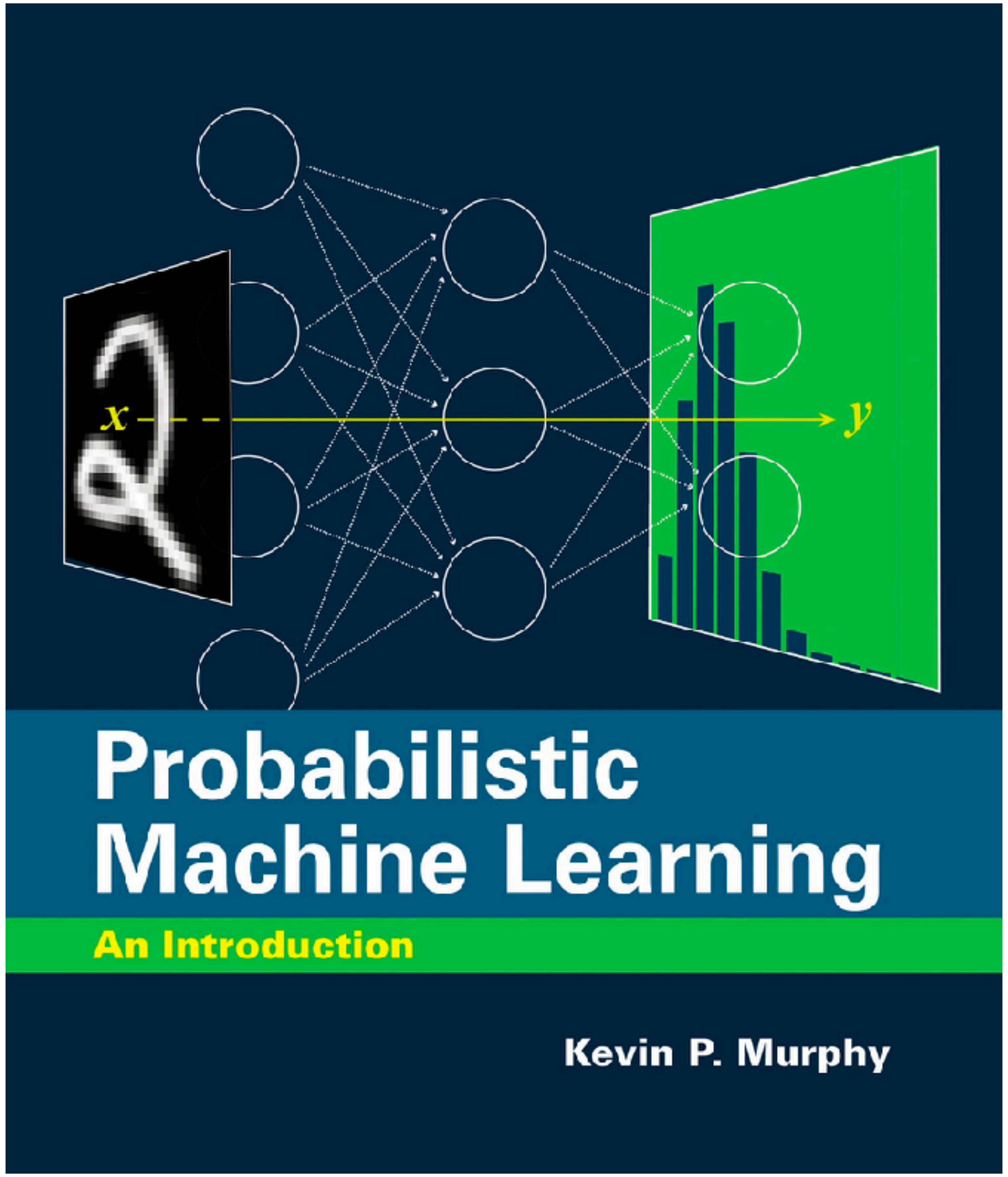
Source code available!



https://github.com/STAT540-UBC/lectures/blob/main/lect13-causality_bayesian/bayesian_stan.Rmd

If you're more interested in Bayesian statistics and probabilistic graphical models





Today’s lecture: Bayesian, PGM, Causality

- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

A warm-up example: the mean and variance of log-Normal

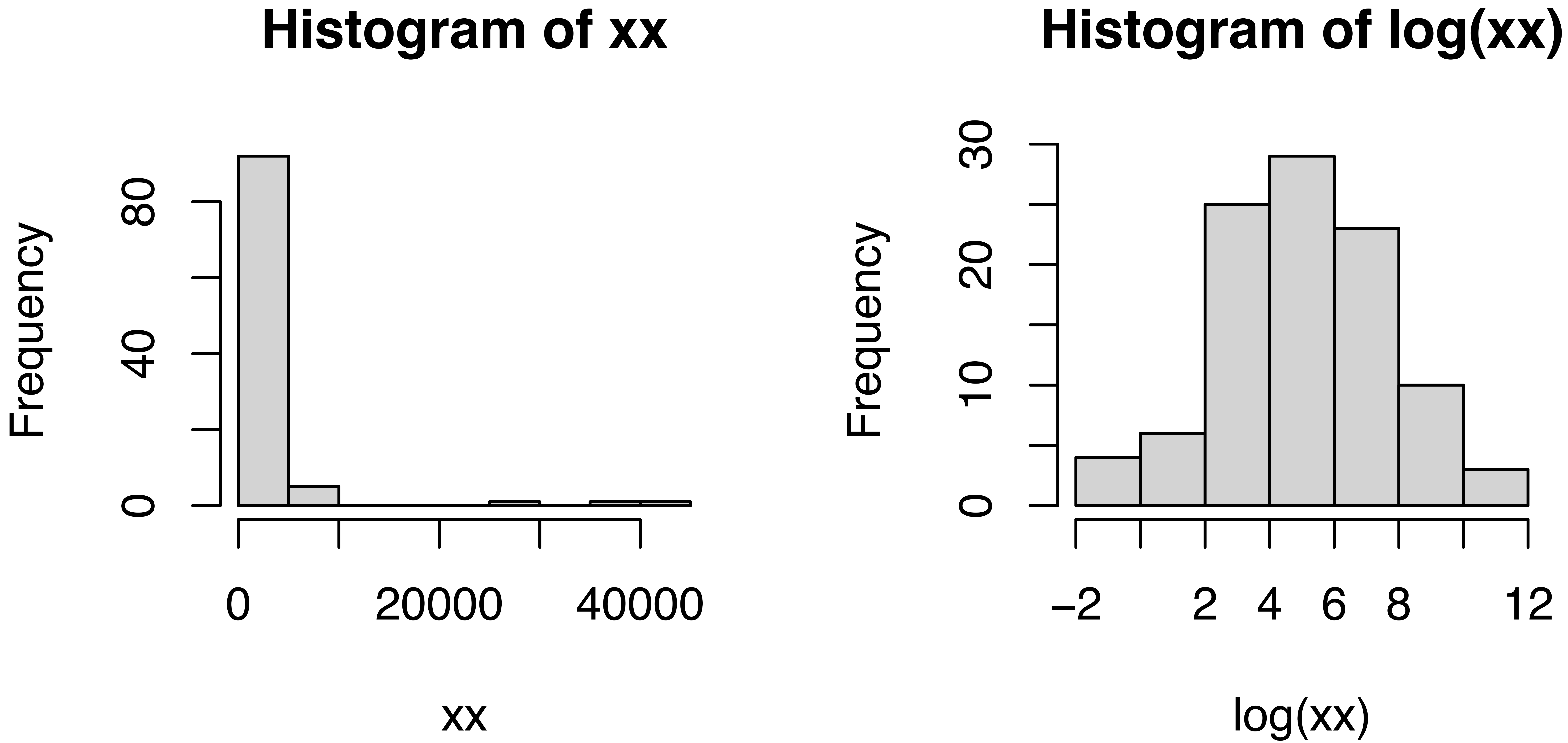
Suppose we have a sequence of real-numbered experimental data observed: $x_1, x_2, \ldots$

- A previous graduate student told me that she calibrated the device, so that the numbers generally follow log-Normal distribution.
- However, she rushed to join some cool biotech start-up and did not inform us what were the desired mean and error parameters.
- Your PI desperately wants to know about the parameters to complete the Aim 3 in recent grant applications.

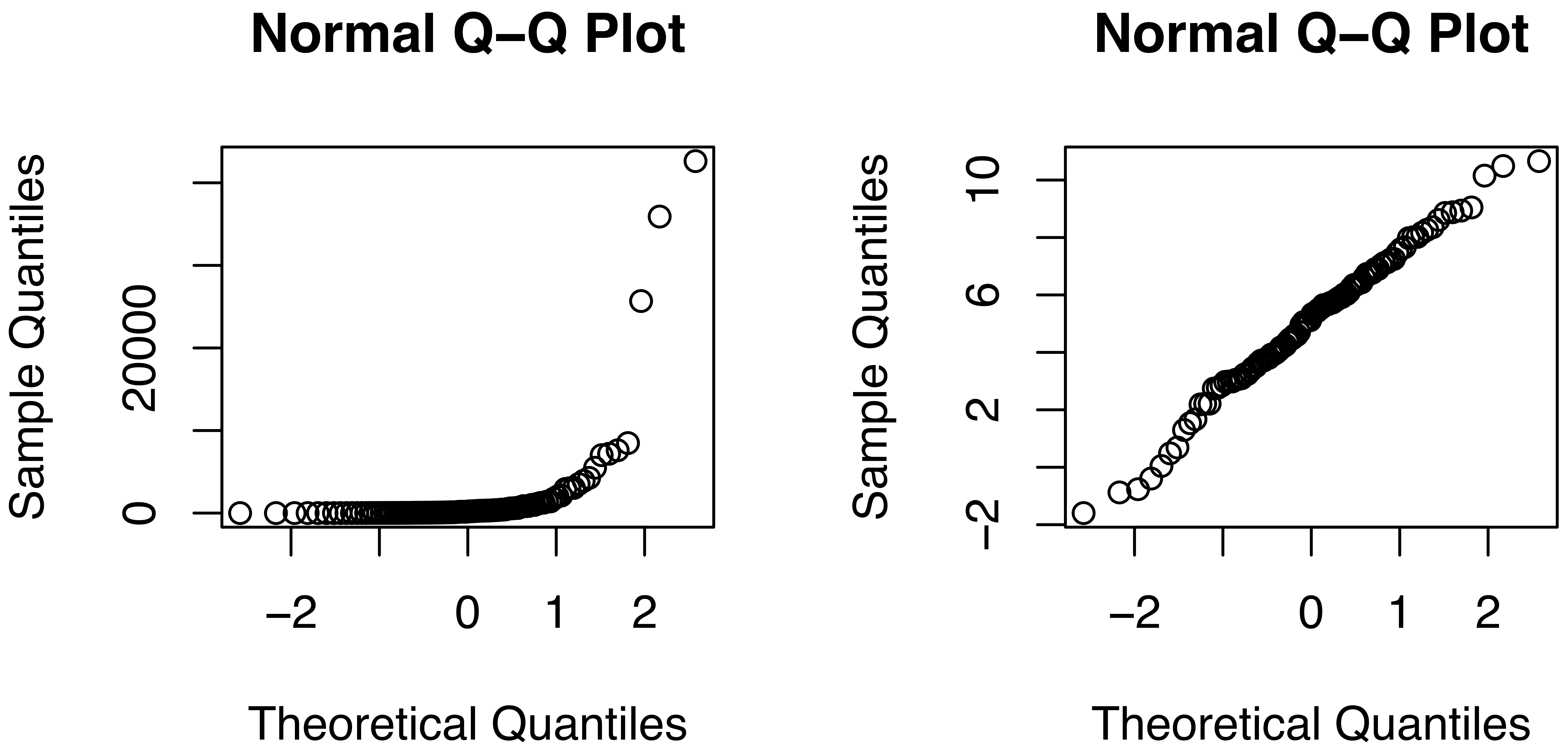
The actual generative scheme

```r
set.seed(1331)
nn <- 100
xx <- rlnorm(nn, 5, sqrt(7))
```

First, what do they look like?



First, what do they look like?



Can we infer the mean and variance of the observed data?

Now we are wondering what will be the suitable parameters...

To answer this question, we can turn to statistics.

Let's define data likelihood:

$$P(\{x_i\}|\mu,\gamma) = \prod_{i=1}^{n} \log \mathcal{N}\left(x_i \middle| \mu, \gamma^{-1}\right)$$

A **Frequentist** Question/Approach:

- What are the **most likely** μ and γ values?
- **How confident** are we about the estimated $\hat{\mu}$ and $\hat{\gamma}$?
  - Standard error (error bars)
  - p-value of some hypothesis test (rejecting the null)

Slightly different world views: Frequentist vs. Bayesian

Frequentist inference

- If $n \to \infty$, how confident are you about the estimated parameter $\hat{\theta}$?
- Will a true parameter θ (theoretical quantity) be consistently captured around a confidence interval around the estimated parameter $\hat{\theta}$?
- If so, how "frequently" the constructed CI will include the true parameter as $n \to \infty$?

$$p\left(\theta \in (\hat{\theta}_n - 2\hat{\text{se}}_n, \hat{\theta}_n + 2\hat{\text{se}}_n)\right)$$

Bayesian inference

- Given the data observed ($n < \infty$), what is the distribution of the unknown parameter?
- What was the underlying generative model?
- Can we take into account uncertainty across multiple types of models?
- How likely is this new observation $x^\star$ given the other observed data points?

$$p\left(\theta | \{i \in [n] : x_i\}\right)$$

Frequentist Inference

What can I say about
this parameter (model) $\boldsymbol{\theta}$,
given plenty of
unseen data?

https://unsplash.com/photos/aebPbwAWjDs

Bayesian Inference

https://unsplash.com/photos/77AW8rM9KGg

Based on what I’ve seen so far,
the parameter **θ** is about here…

A Bayesian way to know about the "how confident" question (given data)

- **Given** data $\{x_i\}$, what is the probability of the parameters (posterior probability)?

$$p(\underbrace{\mu,\gamma}_{\text{unknown parameter}} \,|\, \underbrace{\{x_i\}}_{\text{data}})$$

- Eventually we also want to **predict** future outcomes (posterior prediction),

$$p(\underbrace{x^\star}_{\text{future}} \,|\, \underbrace{\{x_i\}}_{\text{data}}) = \int d\mu\, d\gamma\, \overbrace{p(\underbrace{x^\star}_{\text{new observation}} \,|\, \mu,\gamma)}^{\text{data generating}}\, \overbrace{p(\mu,\gamma \,|\, \underbrace{\{x_i\}}_{\text{training data}})}^{\text{posterior probability}}$$

- Moreover, we want to find a better model **comparing** the model evidence:

$$p(\text{data} \,|\, \text{model 1}) \text{ vs. } p(\text{data} \,|\, \text{model 2})$$

A Bayesian way to know about the "how confident" question (given data)

- **Given** data $\{x_i\}$, what is the probability of the parameters (posterior probability)?

$$p(\underbrace{\mu,\gamma}_{\text{unknown parameter}} \,|\, \underbrace{\{x_i\}}_{\text{data}})$$

- Eventually we also want to **predict** future outcomes (posterior prediction),

$$p(\underbrace{x^\star}_{\text{future}} | \underbrace{\{x_i\}}_{\text{data}}) = \int d\mu\, d\gamma\, p(x^\star \,|\, \underbrace{\mu,\gamma)\; p(\mu,\gamma}_{\text{averaging over uncertainty}} \,|\{x_i\})$$

- Moreover, we want to find a better model **comparing** the model evidence:

$$p(\text{data} \,|\, \text{model 1}) \text{ vs. } p(\text{data} \,|\, \text{model 2})$$

Side note: Bayesian inference often involves heavy computational work

In the order of "easiness" in statistical learning (generally):

- Model parameter estimation/optimization
- Classification/categorical value prediction
- Real-valued prediction
- Model averaging $\leftarrow$ Bayesian inference
- Probability/density estimation $\leftarrow$ Bayesian inference
- Evidence computation (aka, partition function in physics) $\leftarrow$ Bayesian inference

The first step in Bayesian inference is to build a generative model

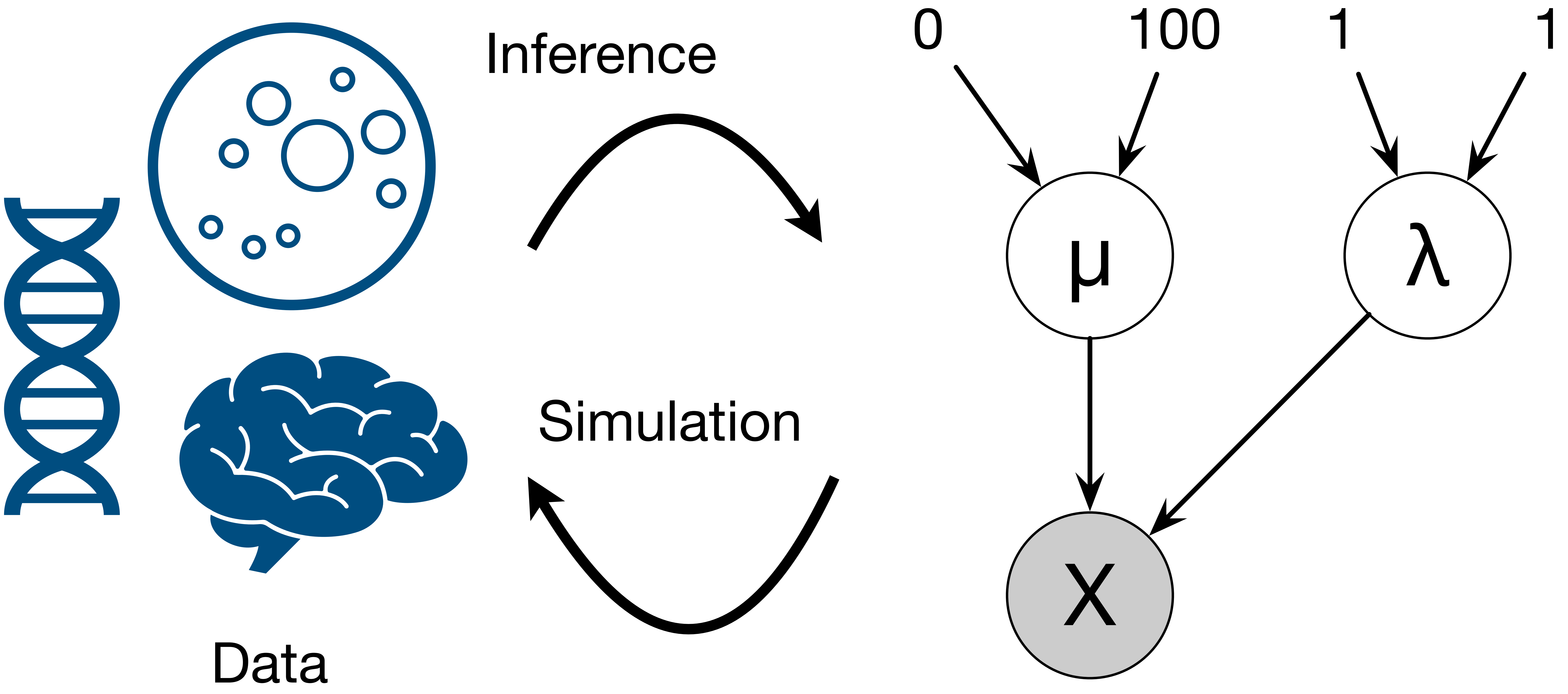
In our example:

1. Sample $\mu \sim \mathcal{N}(0, 10^2)$
2. Sample $\gamma \sim \text{Gamma}(1, 1)$
3. Sample $X \sim \text{log-}\mathcal{N}(\mu, \gamma^{-1})$

- A prior generative scheme of μ
- A prior generative scheme of γ
- A data-generating of x given μ, γ

- “What I cannot create, what I cannot understand” - Richard Feynman
- How do we recover (reverse-engineer) these unknown parameters?

The goal is to infer μ, λ

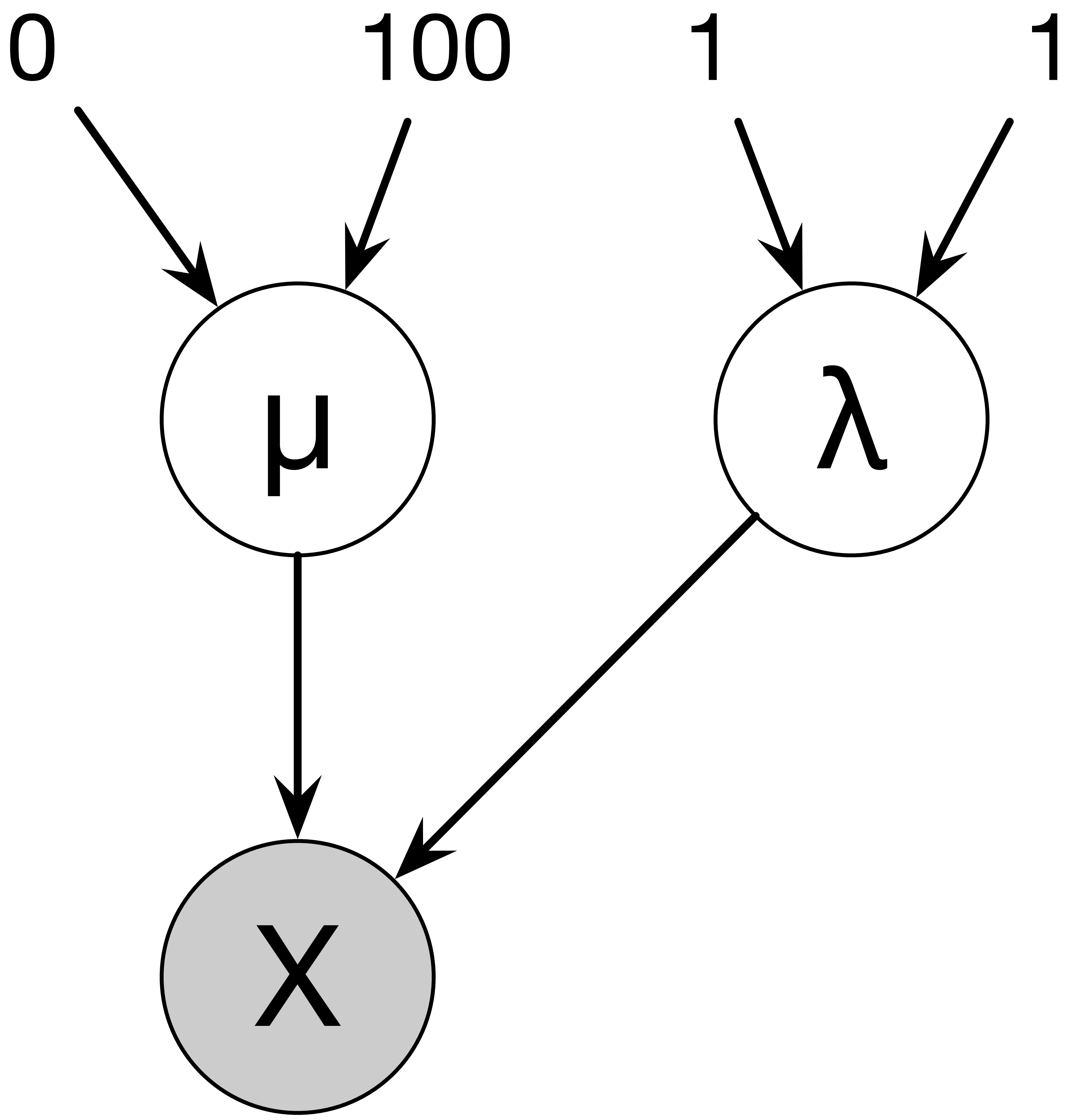


The goal is to infer μ, λ

$$\mu \sim \mathcal{N}(0,10^2)$$

$$\lambda \sim \text{Gamma}(1,1)$$

$$X \sim \log\text{-}\mathcal{N}(\mu,\lambda)$$



We will use *stan* to infer/simulate the "posterior" distribution of μ and λ

Stan

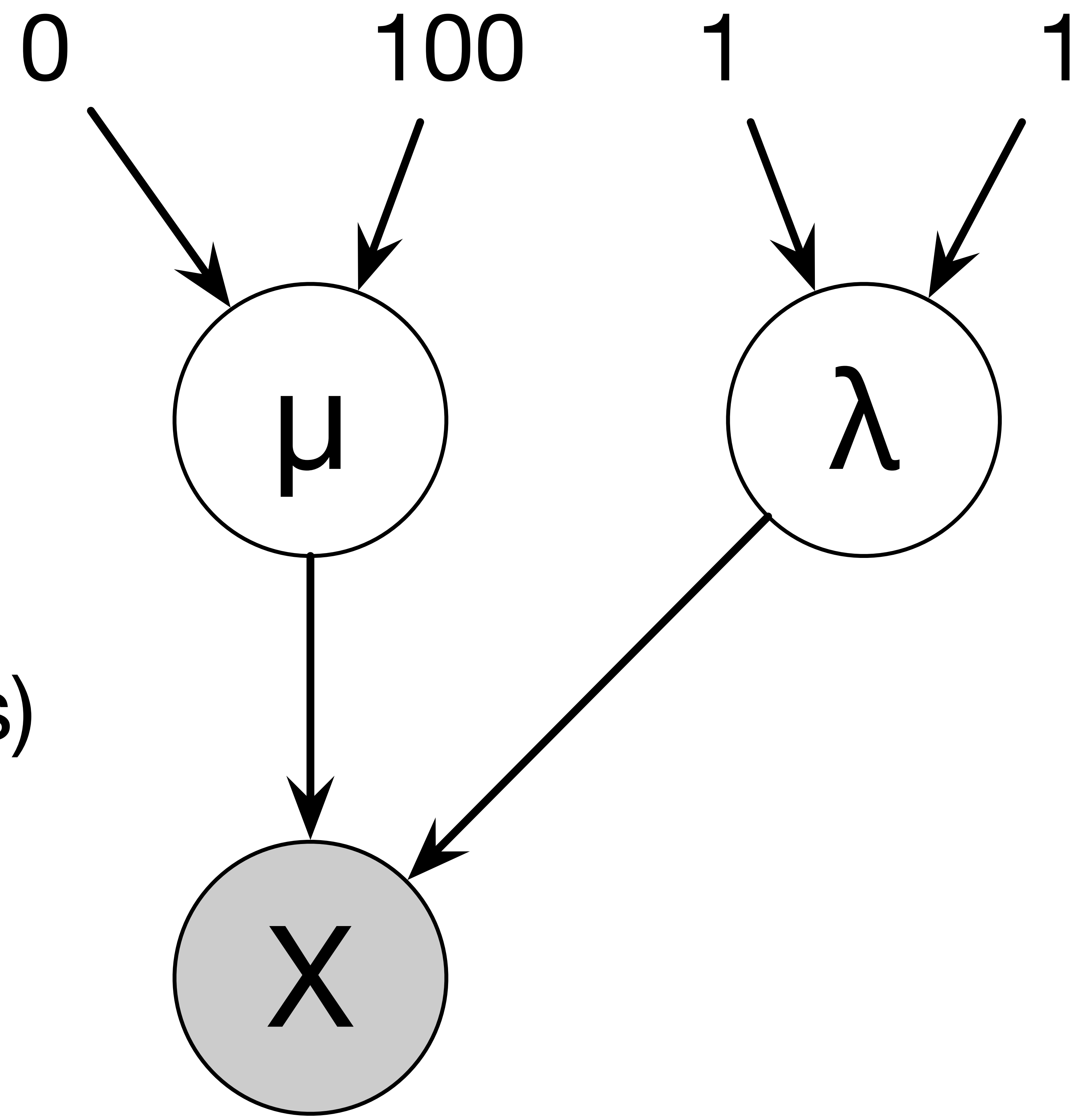
https://mc-stan.org/

Simulation of posterior sampling:
What we need and what *Stan* does for us

Stan

For each Markov Chain (simulation process)

1. Simulate μ given current λ and X
2. Simulate λ given current μ and X
3. Repeat 1-2



`stan`: A good news! All you need is to complete the model description!

```
data {
 int n;                     // sample size
 real x[n];                 // n data points
}
parameters {
 real mu;                   // mean parameter
 real<lower=1e-8> lambda; // precision
}
model {
 mu ~ normal(0, 1e2);          // prior
 lambda ~ gamma(1e-2, 1e-2); // prior
 for(i in 1:n){                // data generating
   x[i] ~ lognormal(mu, sqrt(1/lambda));
 }
}
```

We can have a separate file with model description in `stan` language:

1. `data` block to describe expected data dimensions and types
2. `parameters` block to describe (unknown) model parameters to infer
3. `model` block to describe data-generating schemes (a probabilistic graph model)

rstan incorporates existing stan codes into R scripts to run Bayesian inference algorithm

In R, it can be a simple string

```
.code <- "
data {
 int n;
 real x[n];
}
parameters {
 real mu;
 real<lower=1e-8> lambda;
}
model {
 mu ~ normal(0, 1e2);
 lambda ~ gamma(1e-2, 1e-2);
 for(i in 1:n){
   x[i] ~ lognormal(mu, sqrt(1/lambda));
 }
}
"
```

We can copy/paste stan's model code in some R string.

1. data block to describe expected data dimensions and types
2. parameters block to describe (unknown) model parameters to infer
3. model block to describe data-generating schemes (a probabilistic graph model)

Running stan is as easy as calling other R functions

```r
library(rstan)
options(mc.cores = parallel::detectCores())

if.needed("example_stan_lgaussian.rds", {       ## don't re-reun

    ## Run MCMC inference algorithm
    .fit <- stan(model_code = .code,                ## code
                 data = list(n = nn, x = xx),       ## a list of data
                 pars = c("mu", "lambda"),          ## parameters of interest
                 chains = 5,                        ## number of parallel MCMC chains
                 iter=1000)                         ## how many iterations?

    saveRDS(.fit, "example_stan_lgaussian.rds") ## save the results
})

## Extract the sampled parameters
.fit <- readRDS("example_stan_lgaussian.rds")
.mu <- rstan::extract(.fit, pars="mu", inc_warmup=TRUE, permuted=FALSE)
.lambda <- rstan::extract(.fit, pars="lambda", inc_warmup=TRUE, permuted=FALSE)
```

As a result of MCMC...

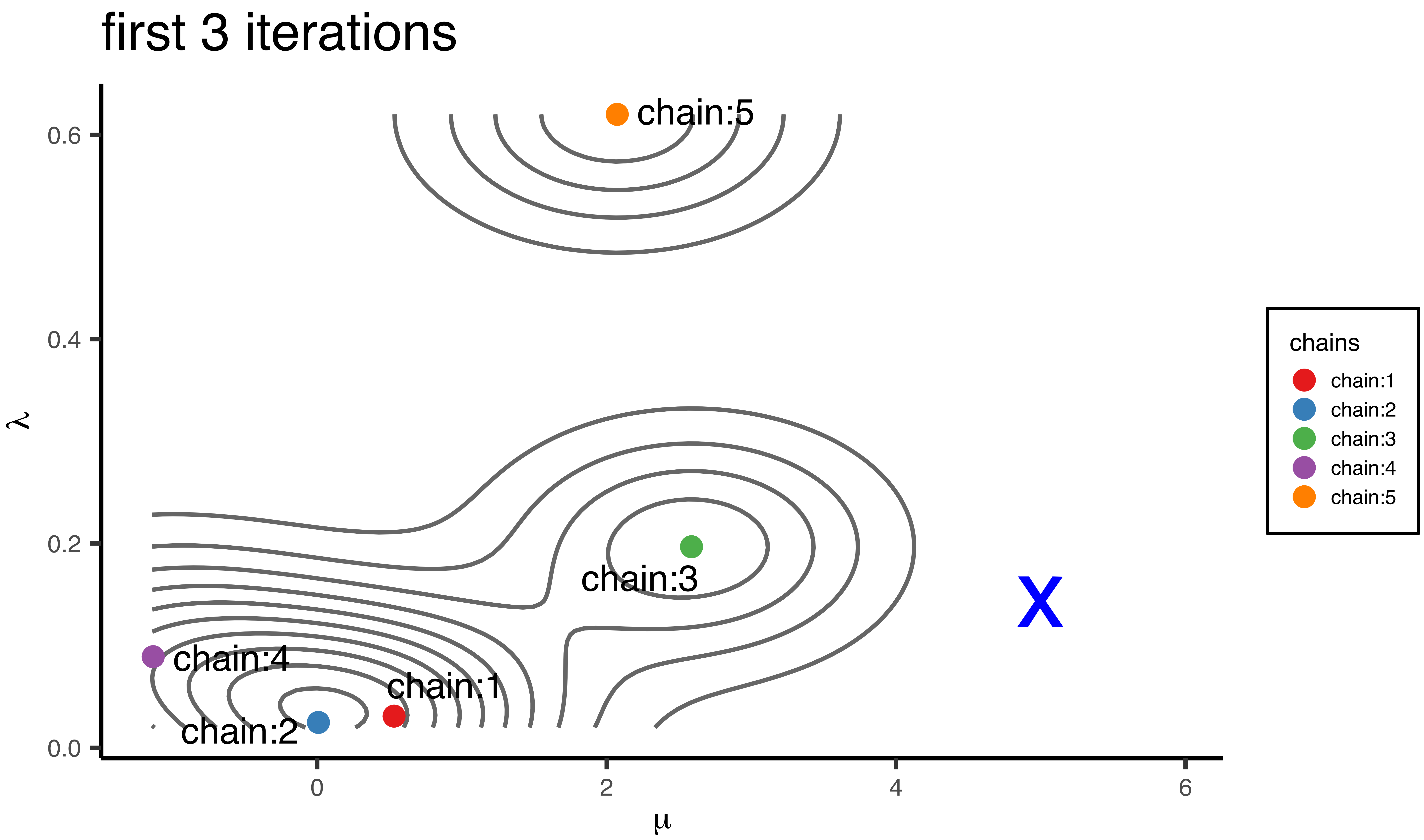
```r
dim(.mu)
```

[1] 1000 5 1

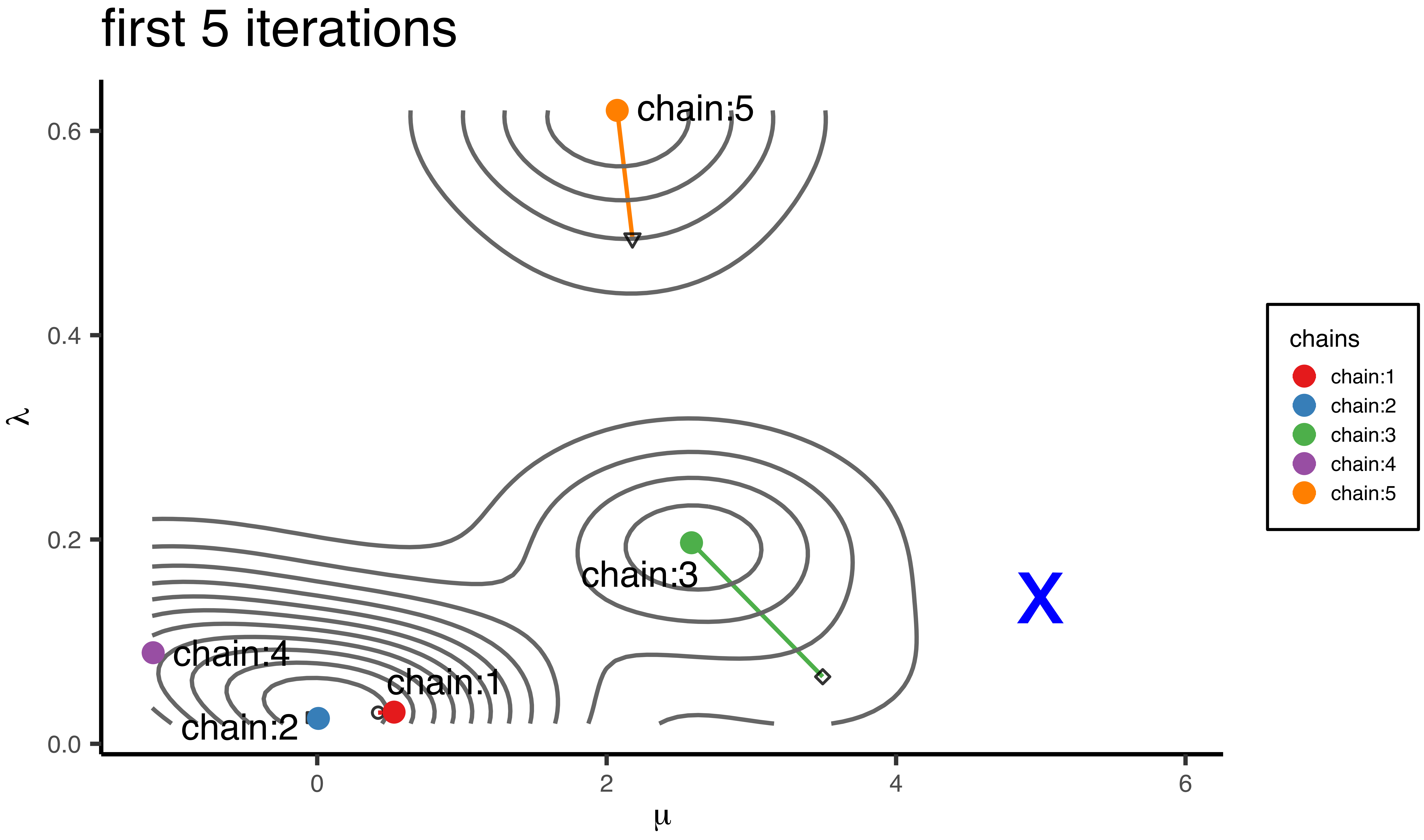
```r
dim(.lambda)
```

[1] 1000 5 1

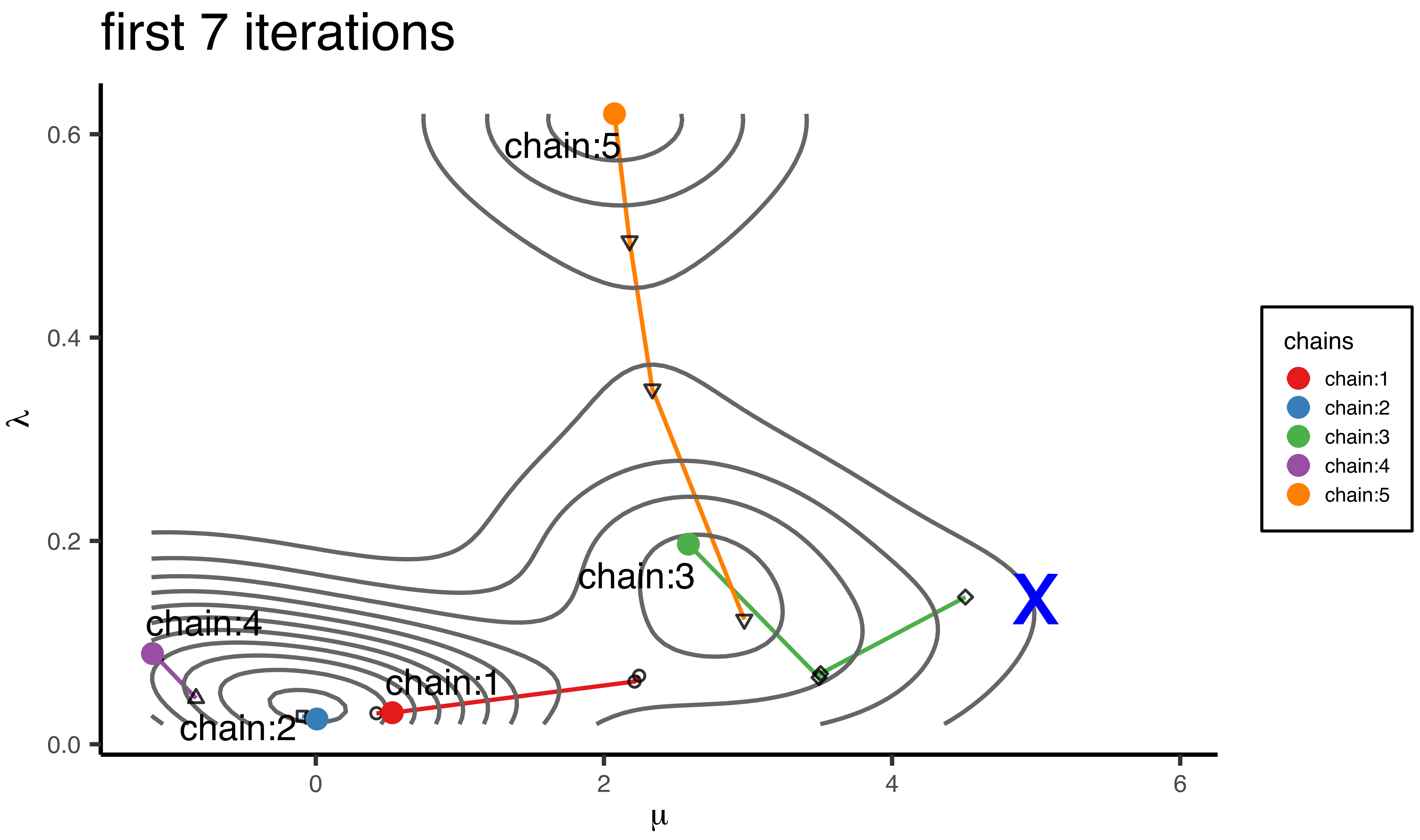
As a result of four independent MCMC runs



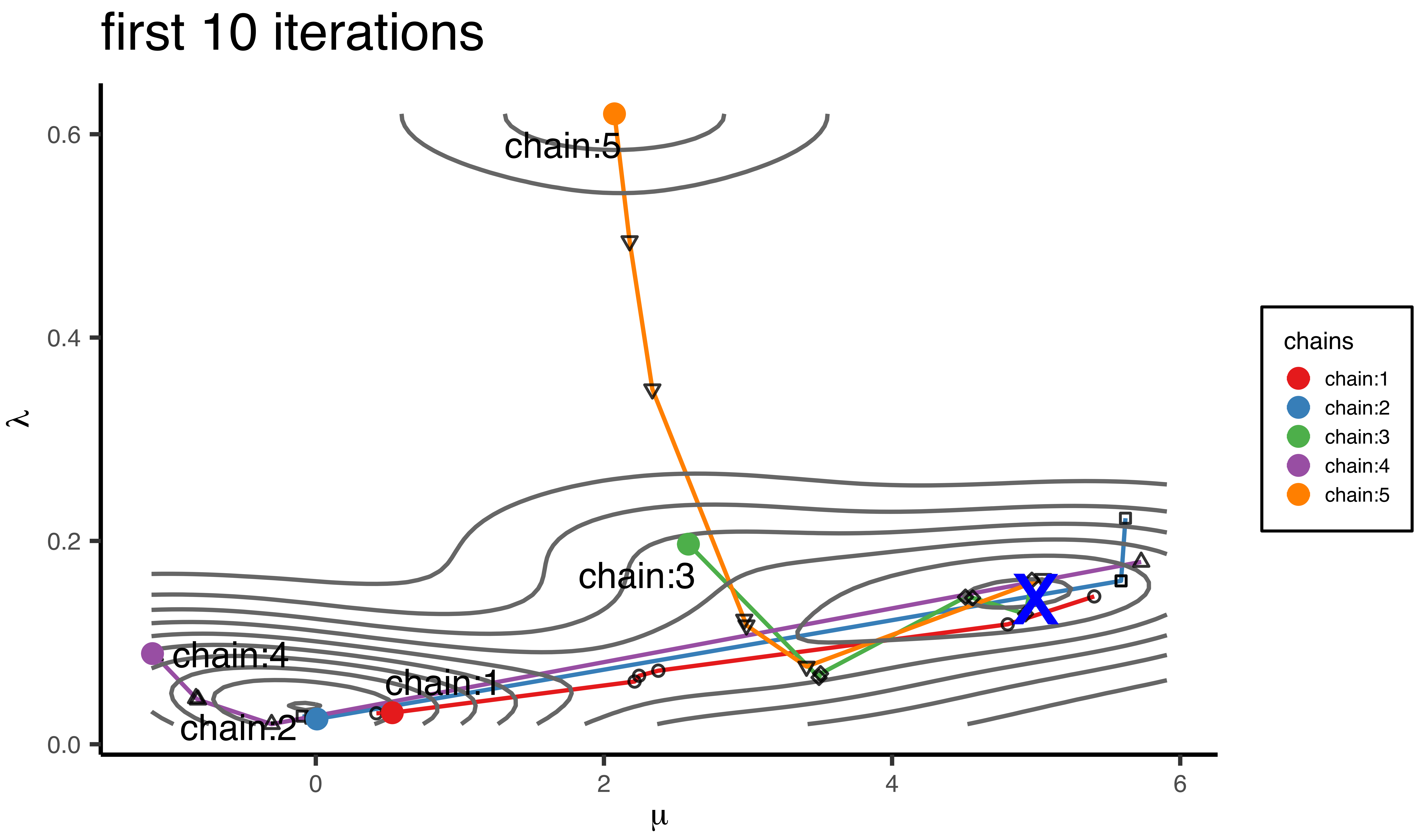
As a result of four independent MCMC runs



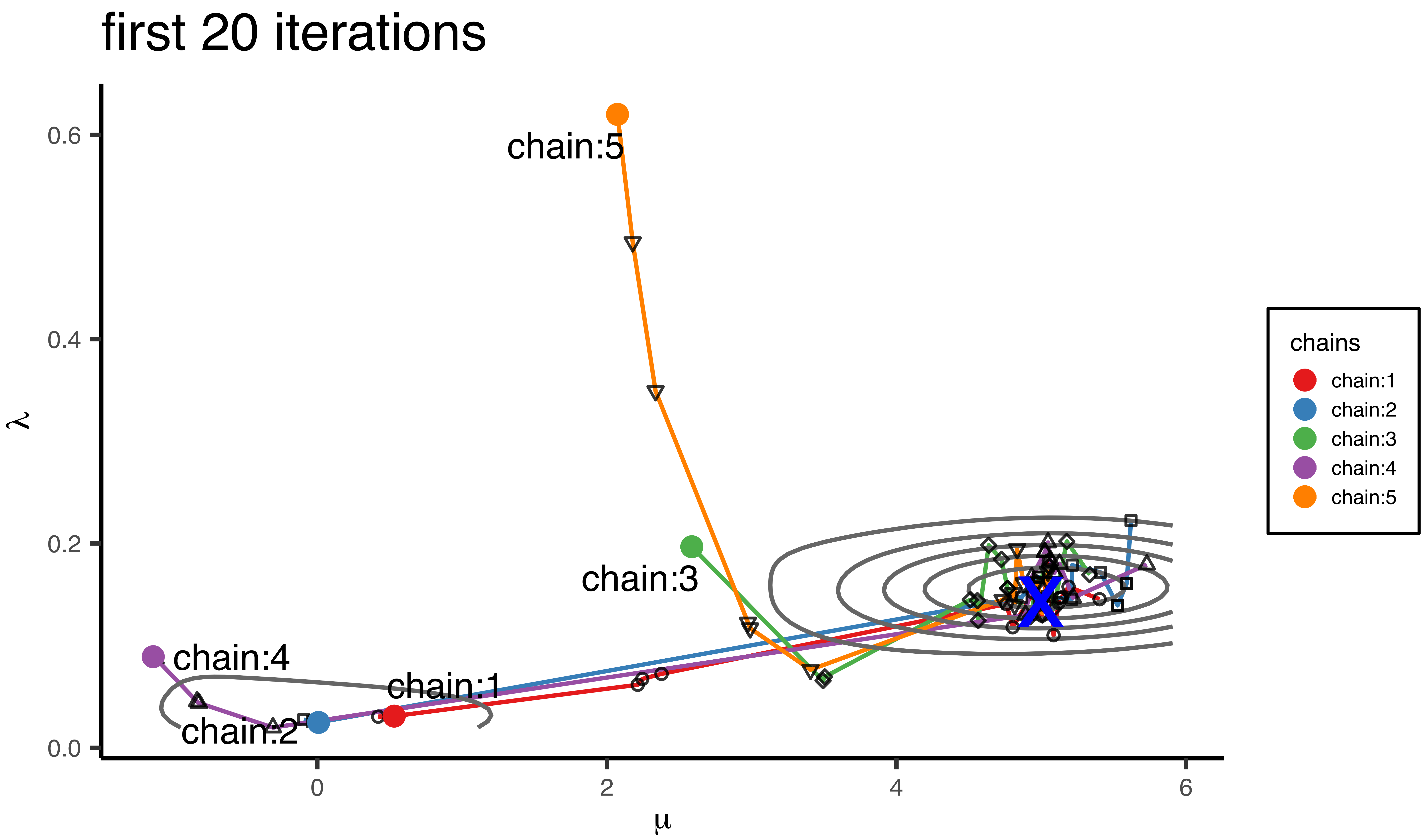
As a result of four independent MCMC runs



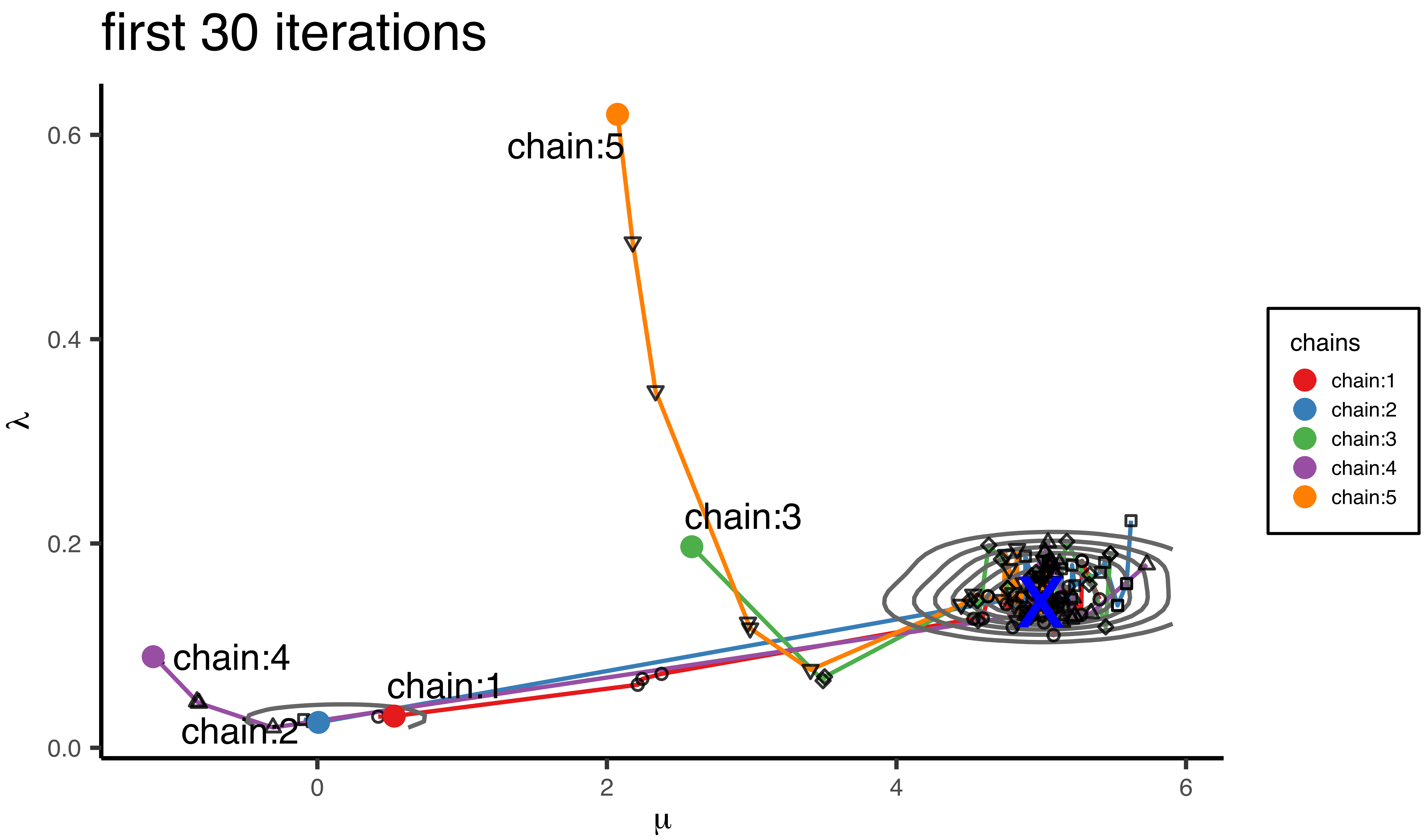
As a result of four independent MCMC runs



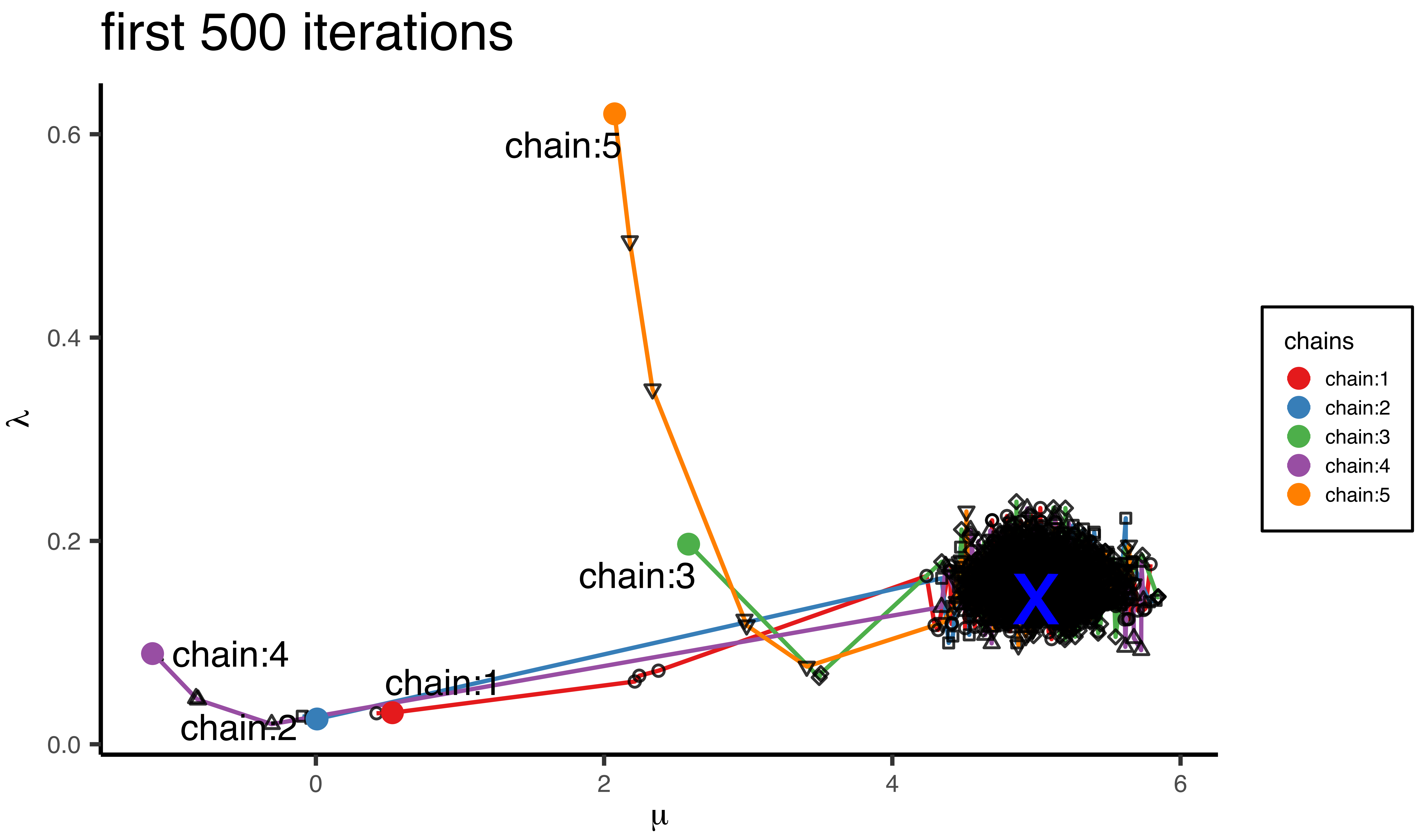
As a result of four independent MCMC runs



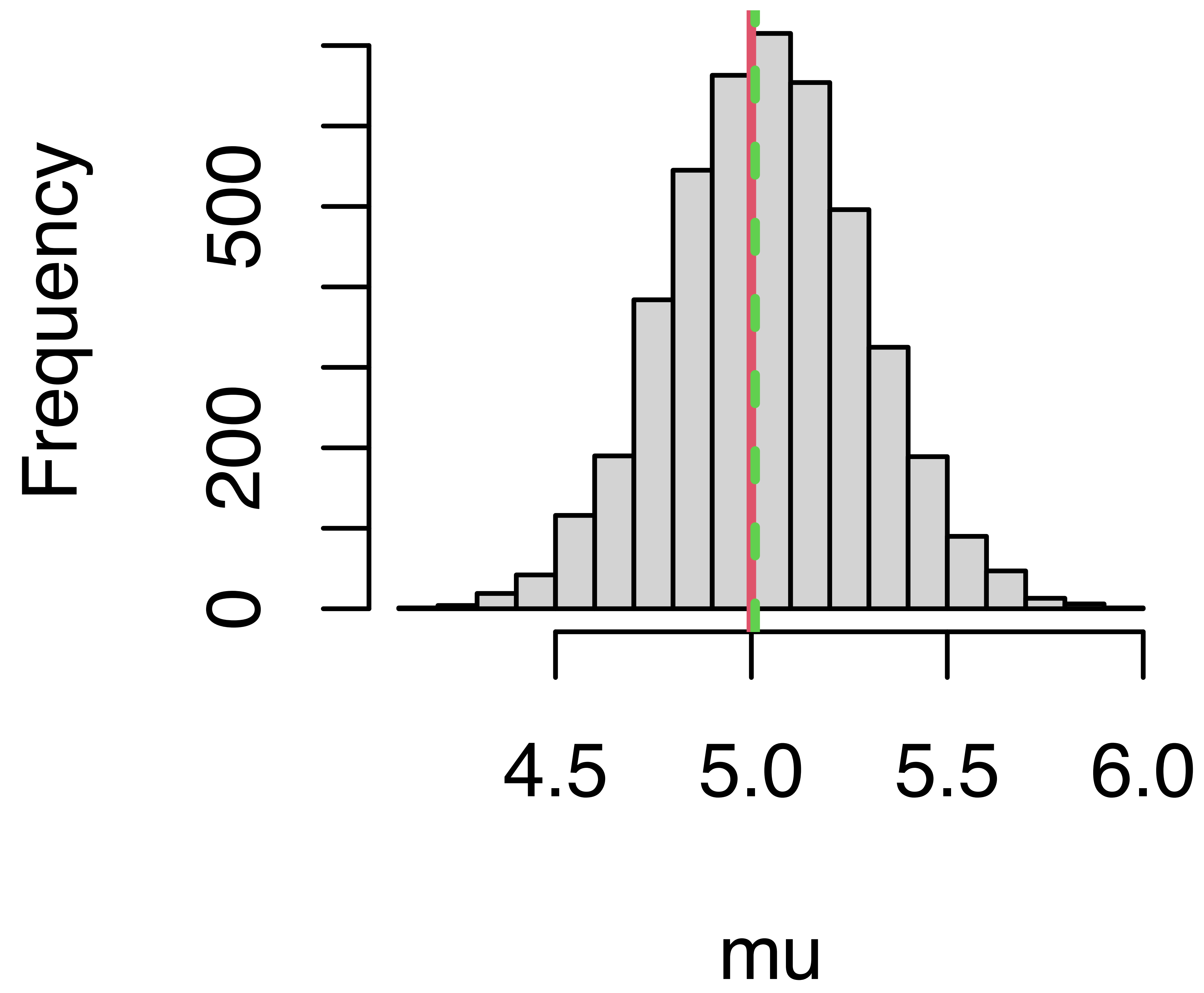
As a result of four independent MCMC runs

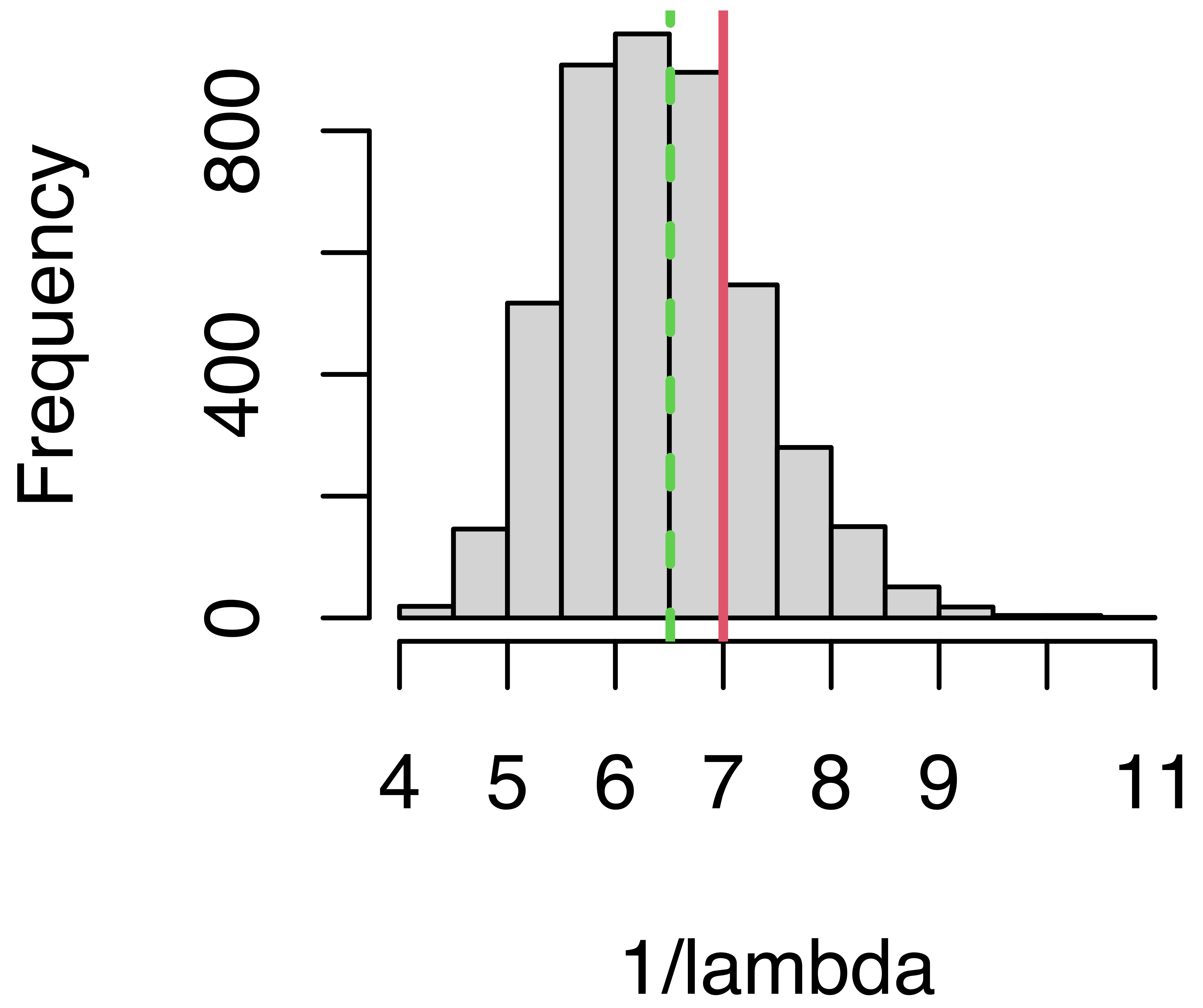


As a result of four independent MCMC runs



Posterior probability





Bayesian linear regression

Let's consider another toy example, Poisson regression

```
n <- 100        ## sample
p.tot <- 10     ## total no. of predictors
p <- 2          ## first `p` are true predictors
sigma <- 0.1    ## error StdDev

set.seed(1331)
X <- matrix(rnorm(n * p.tot), nrow=n, ncol=p.tot)
theta <- matrix(2 * rnorm(p), nrow=p, ncol=1)
y <- round(exp(X[, 1:p] %*% theta + rnorm(n) * sigma))
```

Can you draw the corresponding probabilistic graphical model?

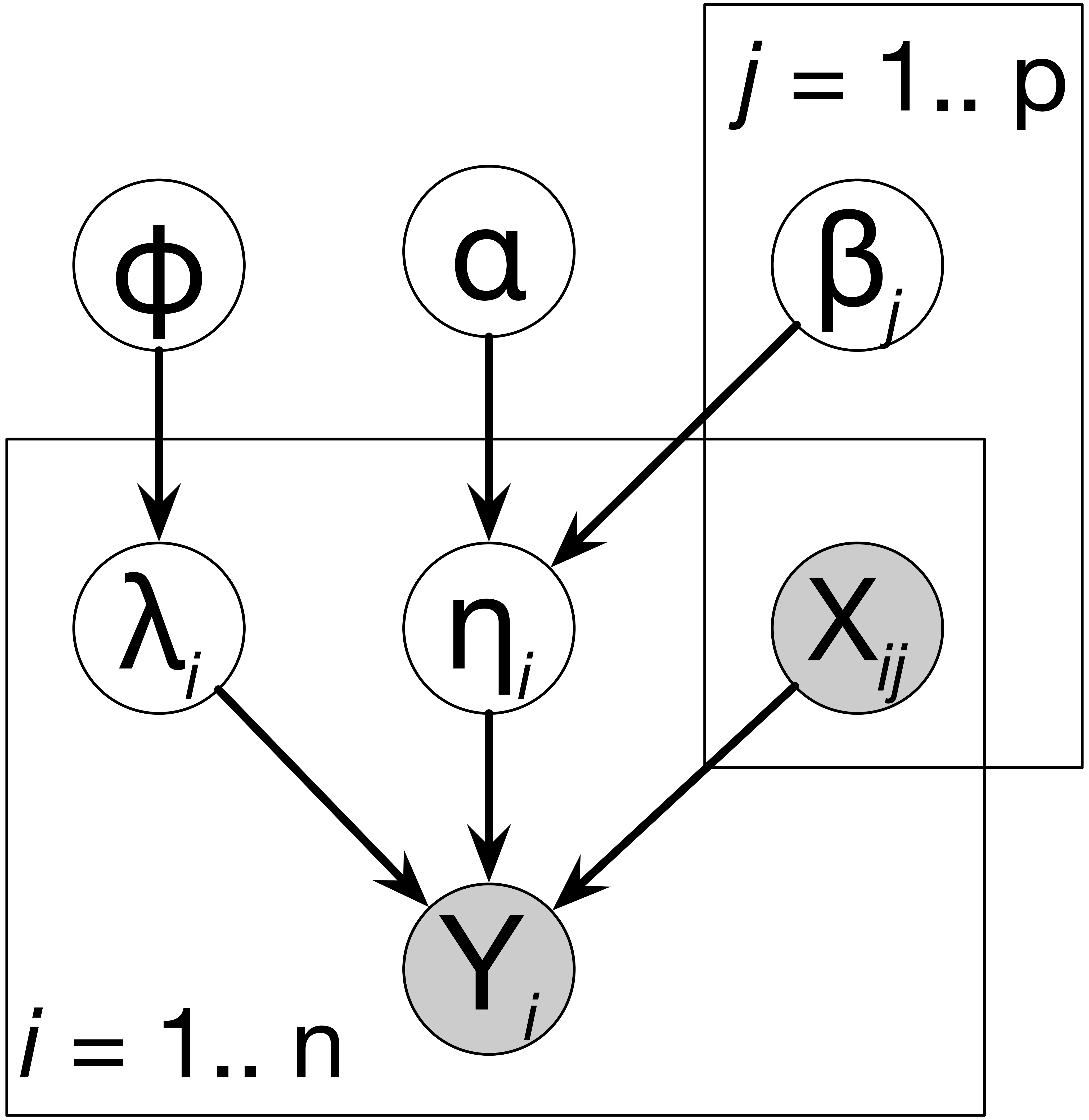
$$\alpha, \beta_j \sim \mathcal{N}(0, 1)$$

$$\phi \sim \text{Gamma}(1, 1)$$

$$\lambda_i | \phi \sim \text{Gamma}(\phi, 1)$$

$$\eta_i \leftarrow \mathbf{x}_i \boldsymbol{\beta} + \alpha$$

$$Y_i | \eta_i, \lambda_i \sim \text{Poisson}(\lambda_i \exp(\eta_i))$$



Let's write out a Poisson regression generative model

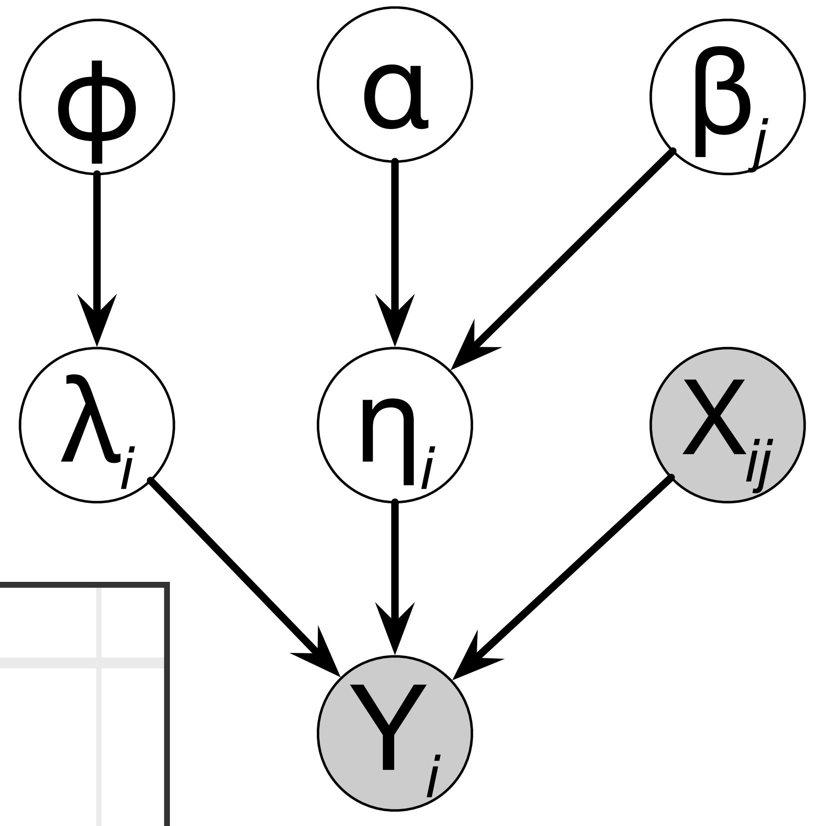
```
data {
  int n; int p;                                    // number of observations  & predictors
  matrix[n, p] X;                                  // design matrix
  int y[n];                                        // response vector
}
parameters {
  vector[p] beta;                                  // regression coefficients
  real alpha;                                      // intercept
  vector<lower=1e-8>[n] lambda;                    // overdispersion
  real phi;                                        // hyperparameter
}
transformed parameters {
  vector[n] eta;
  eta = X * beta + alpha;
}
model {
  alpha ~ normal(0, 1);  phi ~ gamma(1e4, 1e4);  //
  for (j in 1:p) { beta[j] ~ normal(0, 1); }     // coefficients
  for (i in 1:n) {                               //
    lambda[i] ~ gamma(phi, 1);                   // random overdispersion
    y[i] ~ poisson(lambda[i] * exp(eta[i]));     // poisson rate
  }
}
```

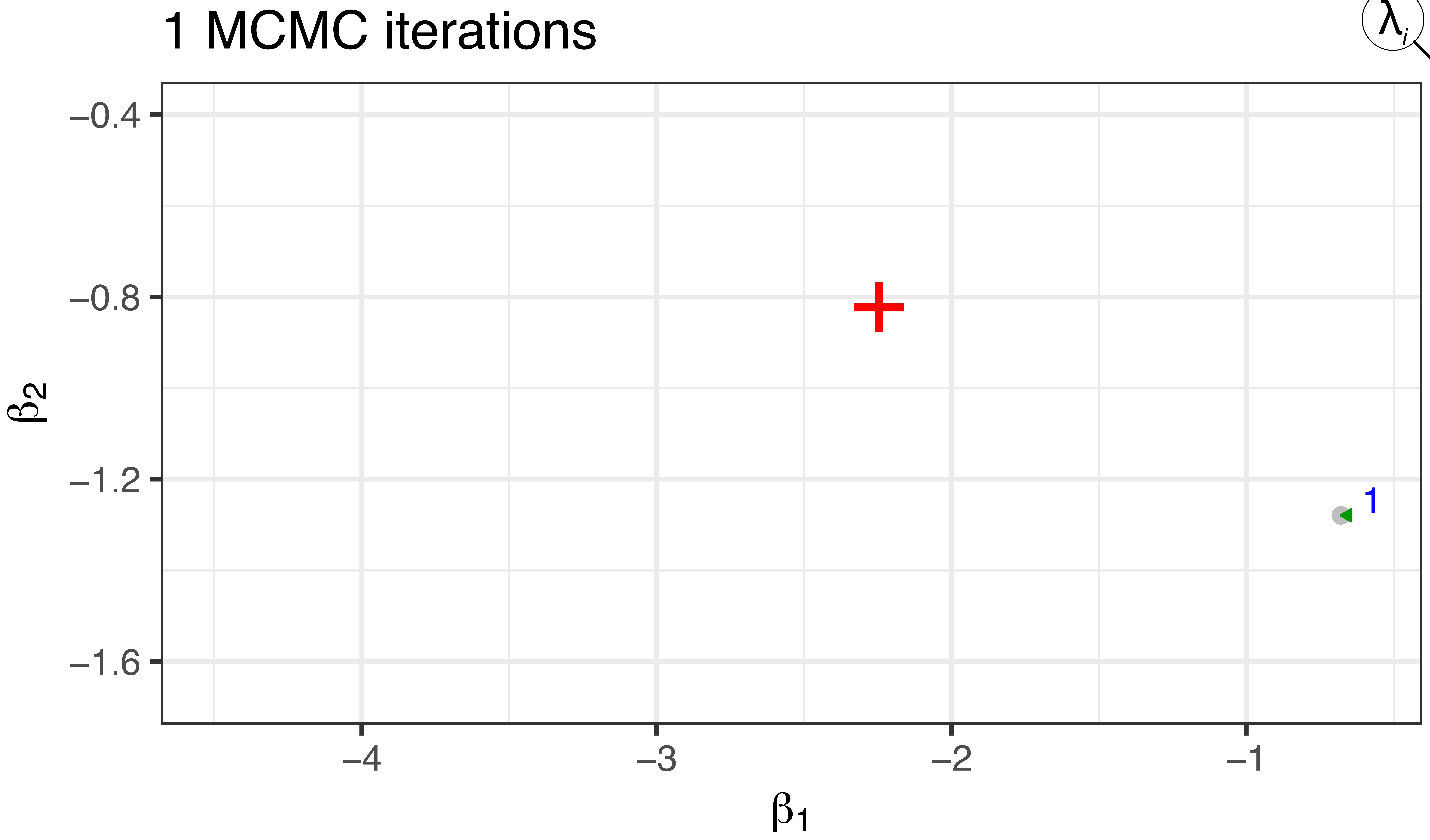
```
.code <- "
data {
  int n; int p;                                   // number of observations  & predictors
  matrix[n, p] X;                                 // design matrix
  int y[n];                                       // response vector
}
parameters {
  vector[p] beta;                                 // regression coefficients
  real alpha;                                     // intercept
  vector<lower=1e-8>[n] lambda;                   // overdispersion
  real phi;                                       // hyperparameter
}
transformed parameters {
  vector[n] eta;
  eta = X * beta + alpha;
}
model {
  alpha ~ normal(0, 1);   phi ~ gamma(1e4, 1e4); //
  for (j in 1:p) { beta[j] ~ normal(0, 1); }     //
  for (i in 1:n) {                                //
    lambda[i] ~ gamma(phi, 1);                    // random overdispersion
    y[i] ~ poisson(lambda[i] * exp(eta[i]));      // poisson rate
  }
}
"
```

Run stan to sample the parameters of interest

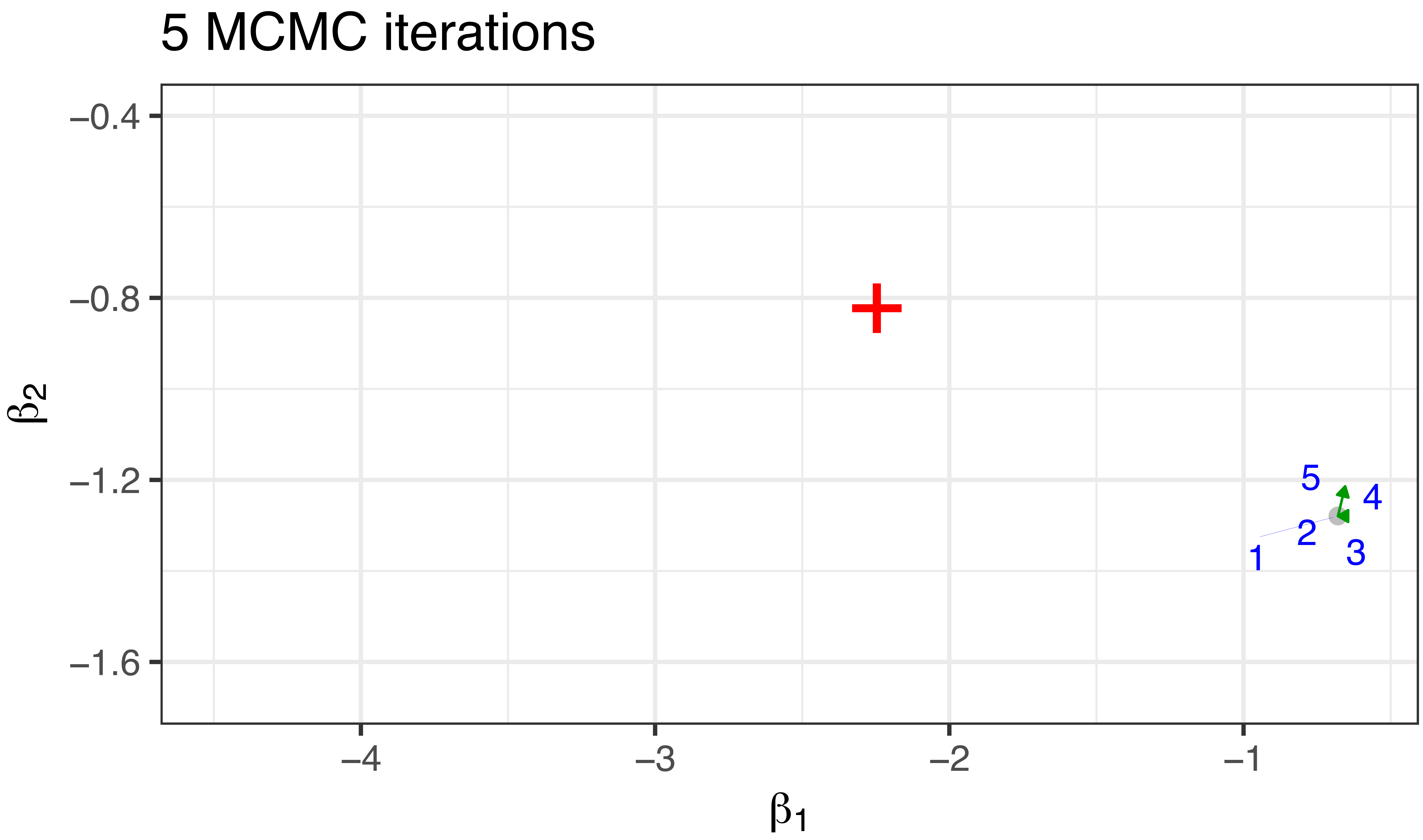
```
if.needed("example_stan_regression.rds", {        ## don't re-reun

    lm.fit <- stan(model_code=.code,                  ## model code
                   data = list(n=nrow(X),             ## samples
                               p=ncol(X),             ## predictors
                               y=y[,1],               ## response
                               X=X),                  ## design matrix
                   pars = c("beta","alpha"),          ## parameters
                   iter=1000,                         ## MCMC
                   chains=5)                          ## chains

    saveRDS(lm.fit, "example_stan_regression.rds")
})

lm.fit <- readRDS("example_stan_regression.rds") ## save the results

.beta <- extract(lm.fit, pars="beta", inc_warmup=TRUE, permuted=FALSE)
```

Let's take a look at truly non-zero coefficients

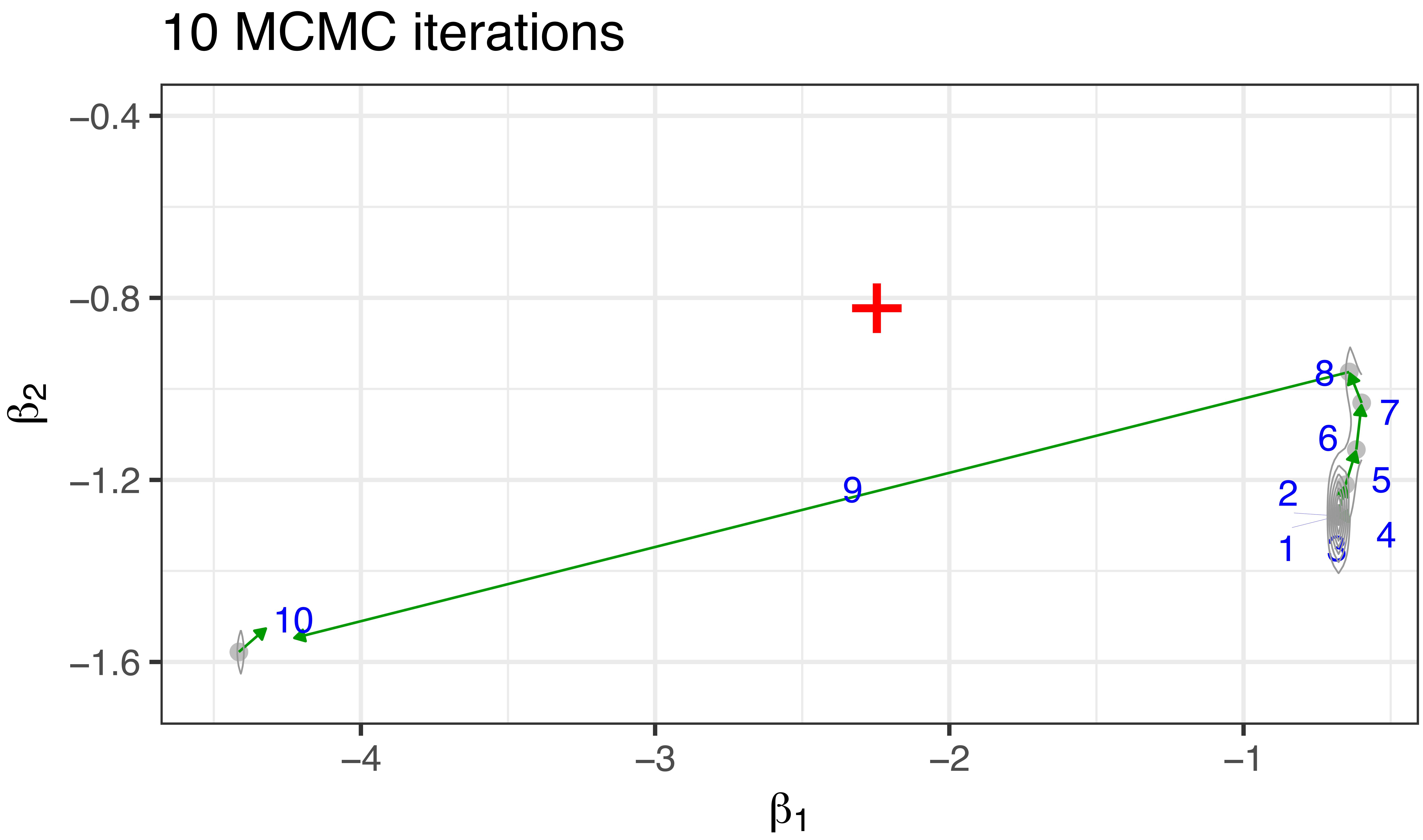




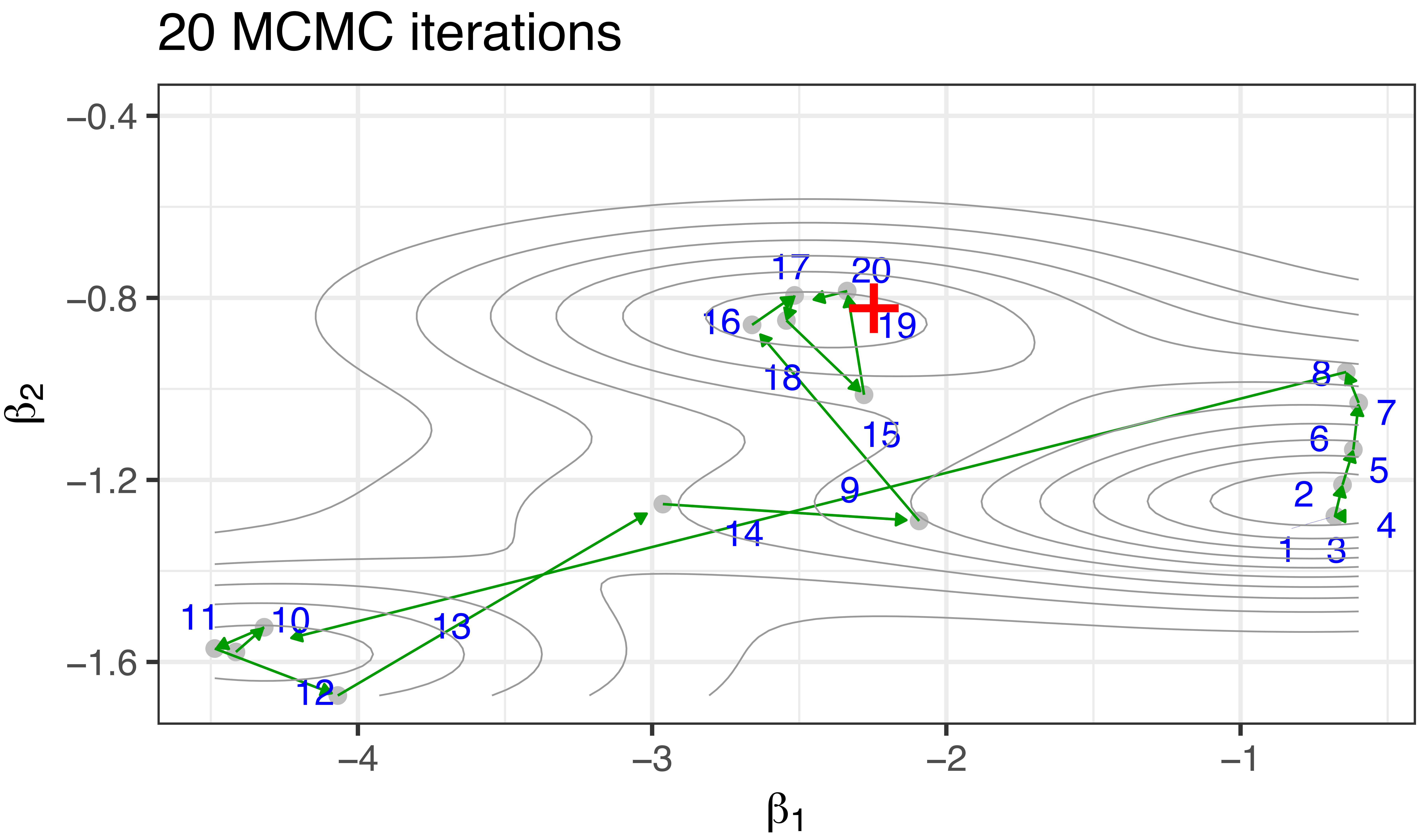
Let's take a look at truly non-zero coefficients



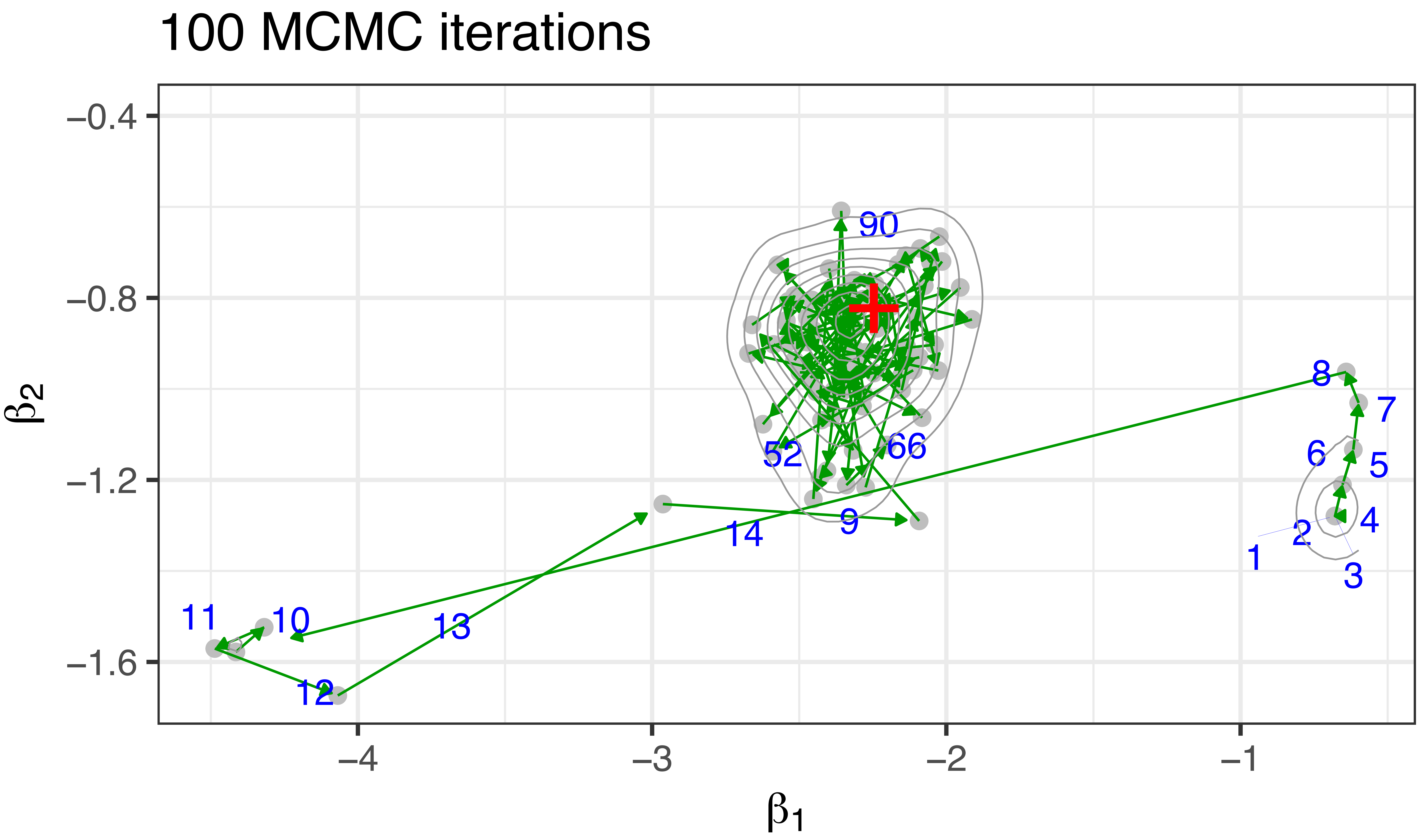
Let's take a look at truly non-zero coefficients



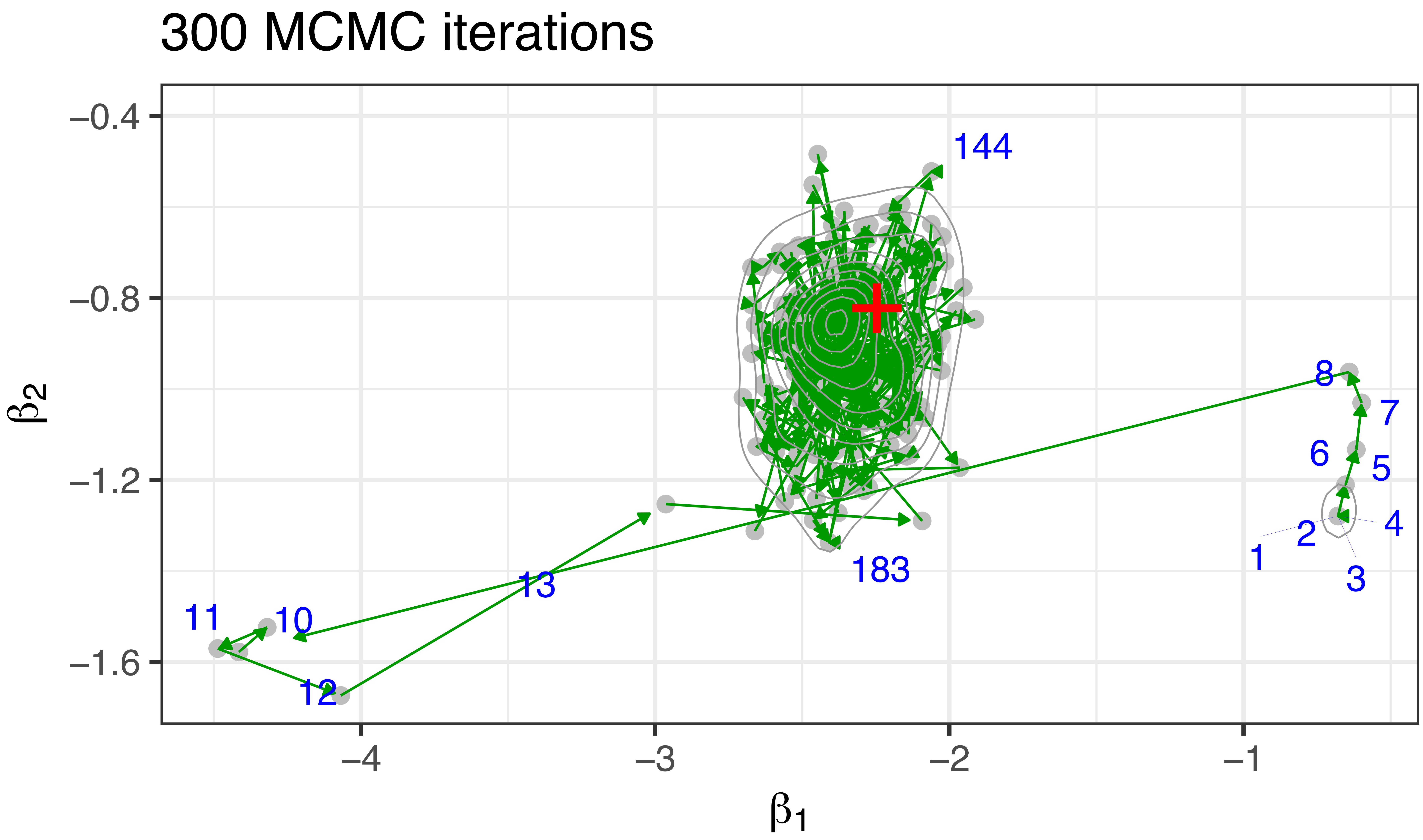
Let's take a look at truly non-zero coefficients



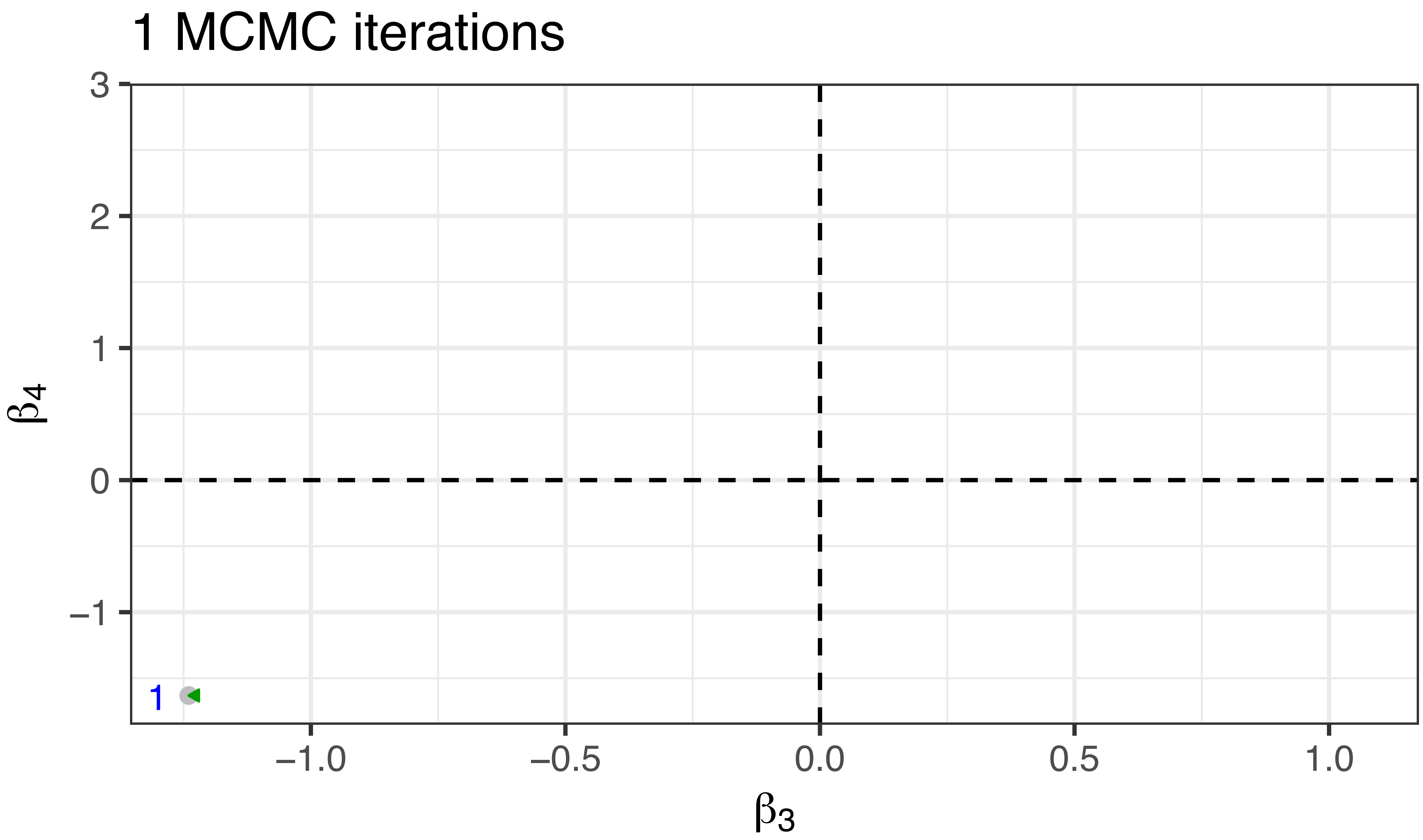
Let's take a look at truly non-zero coefficients



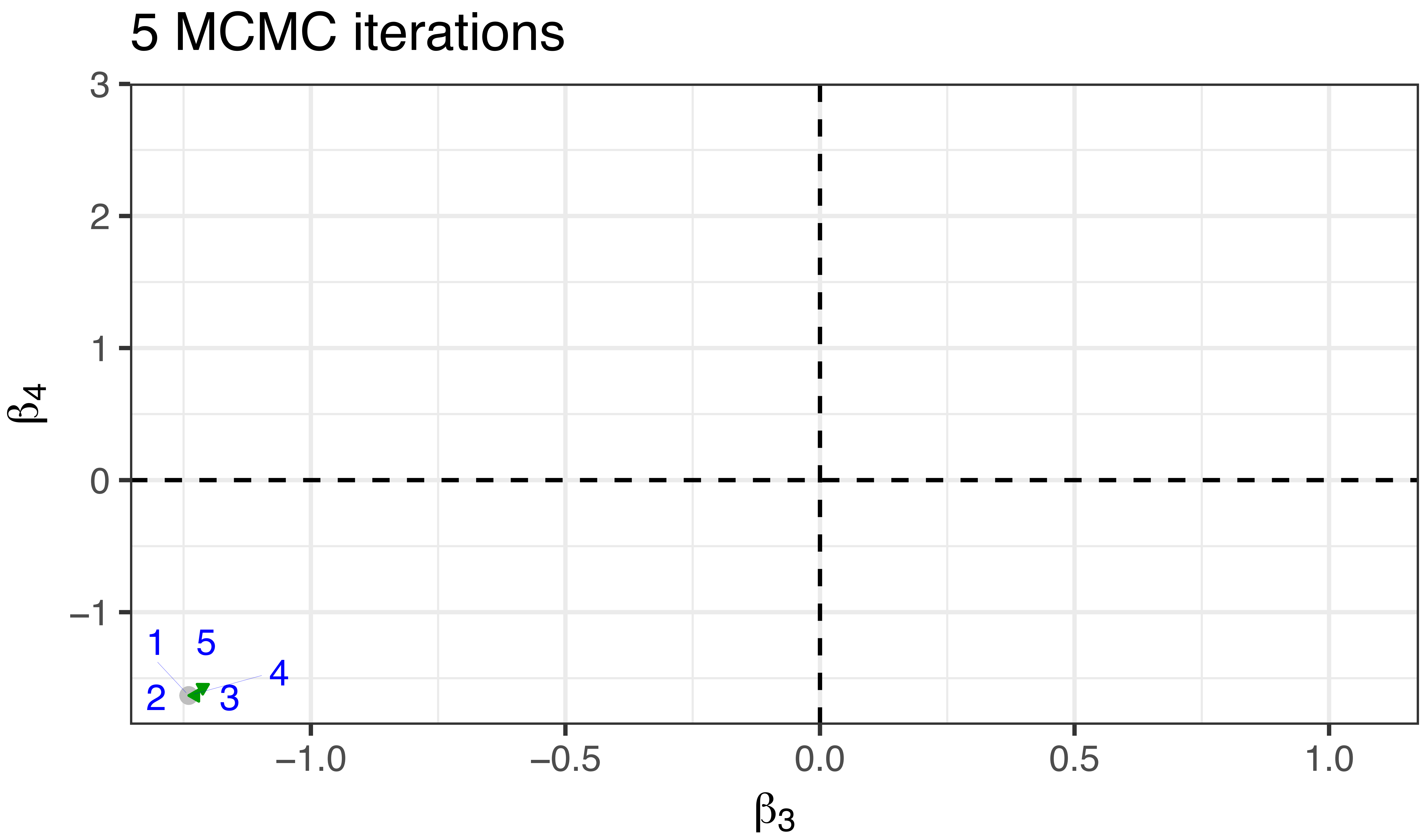
Let's take a look at truly non-zero coefficients



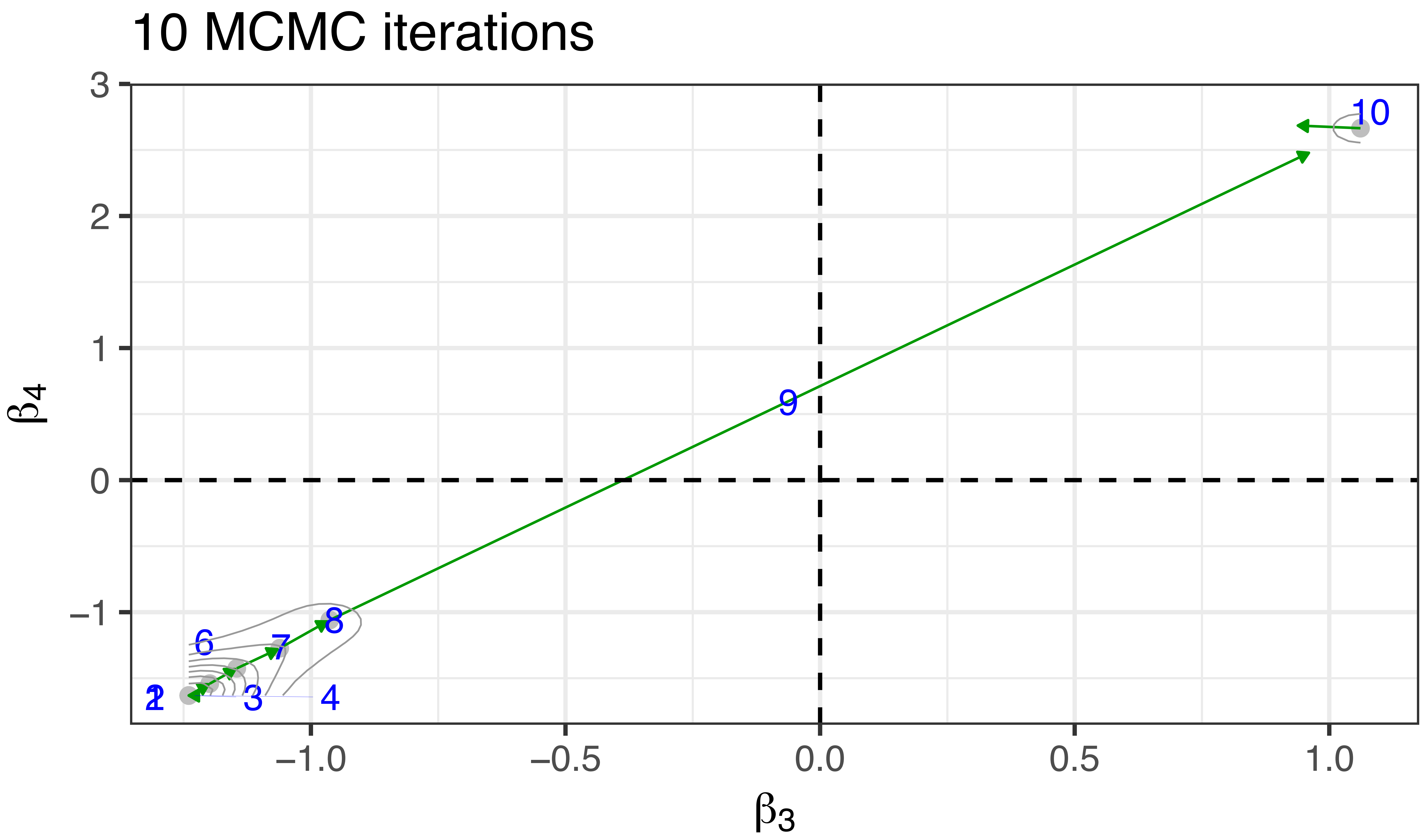
Let's take a look at other "null" coefficients



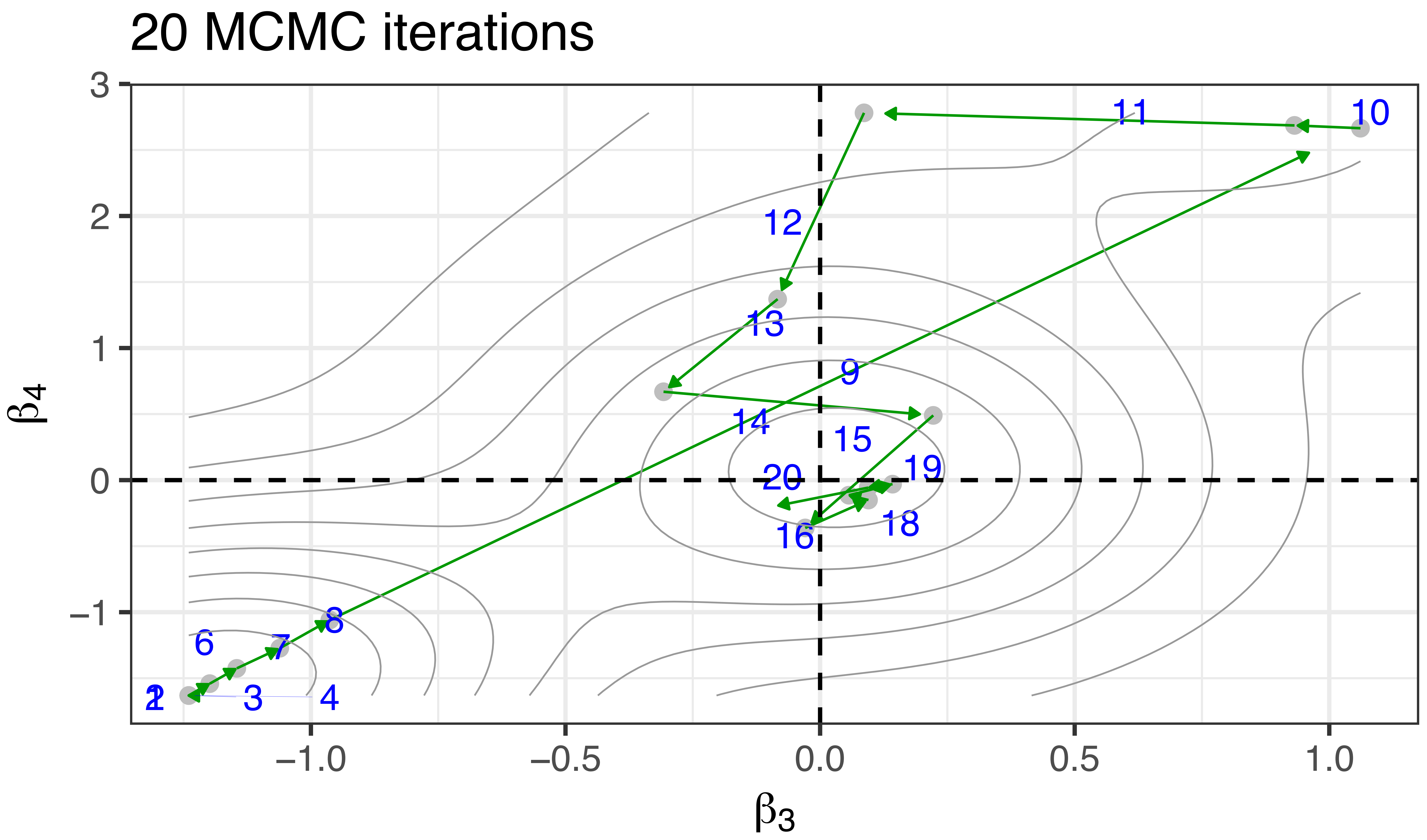
Let's take a look at other "null" coefficients



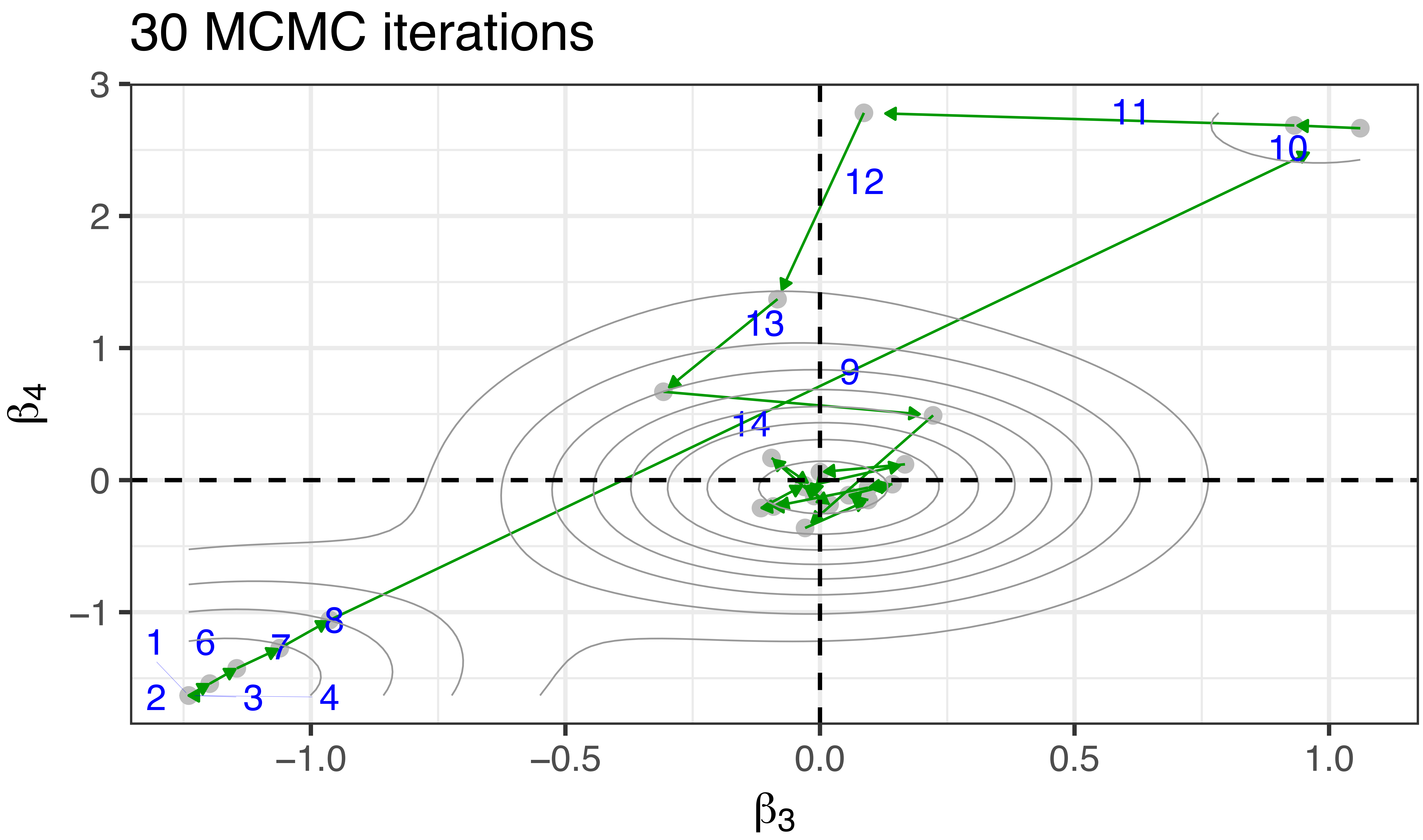
Let's take a look at other "null" coefficients



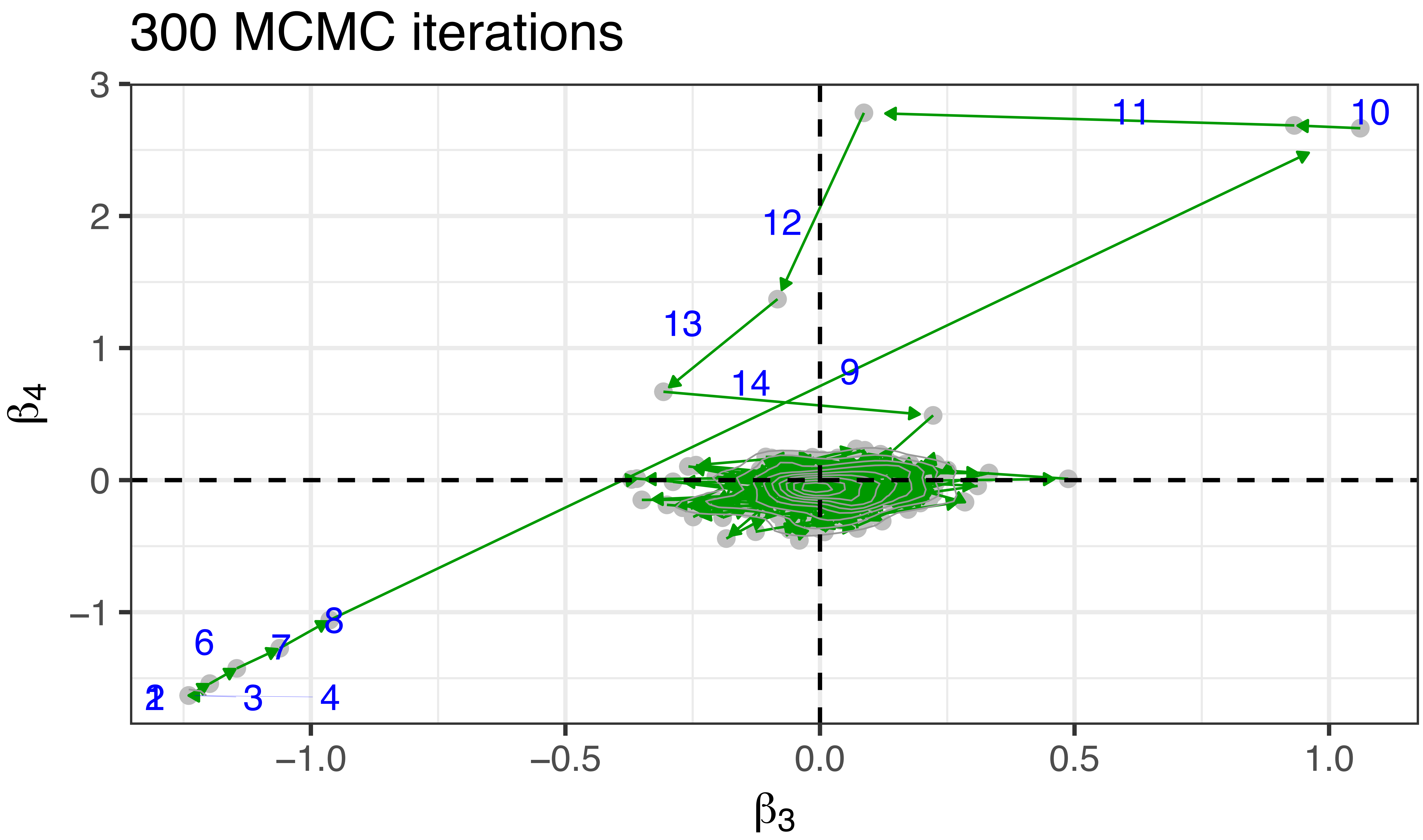
Let's take a look at other "null" coefficients



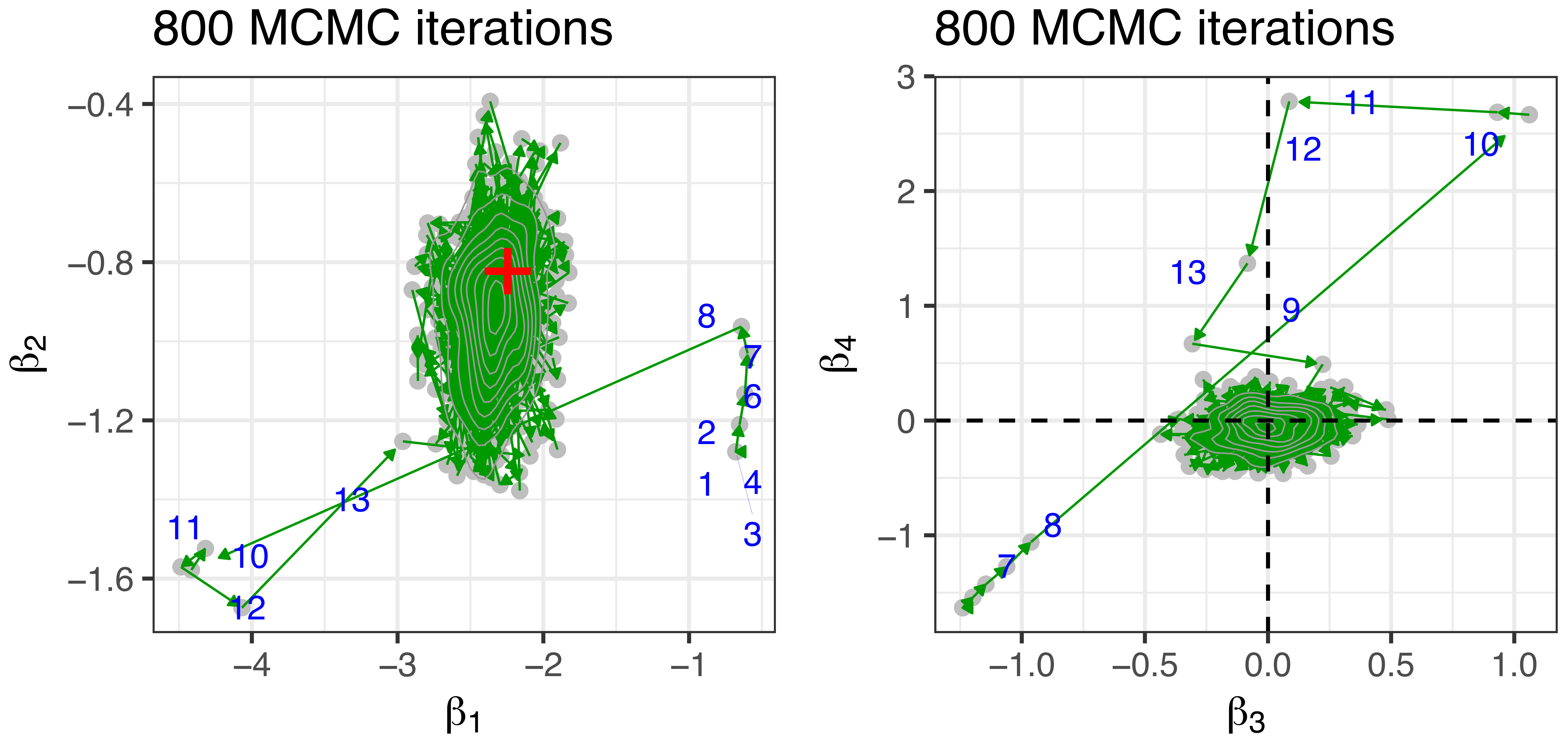
Let's take a look at other "null" coefficients



Let's take a look at other "null" coefficients



Compare true non-zero vs. zero coefficients



Today's lecture: Bayesian, PGM, Causality

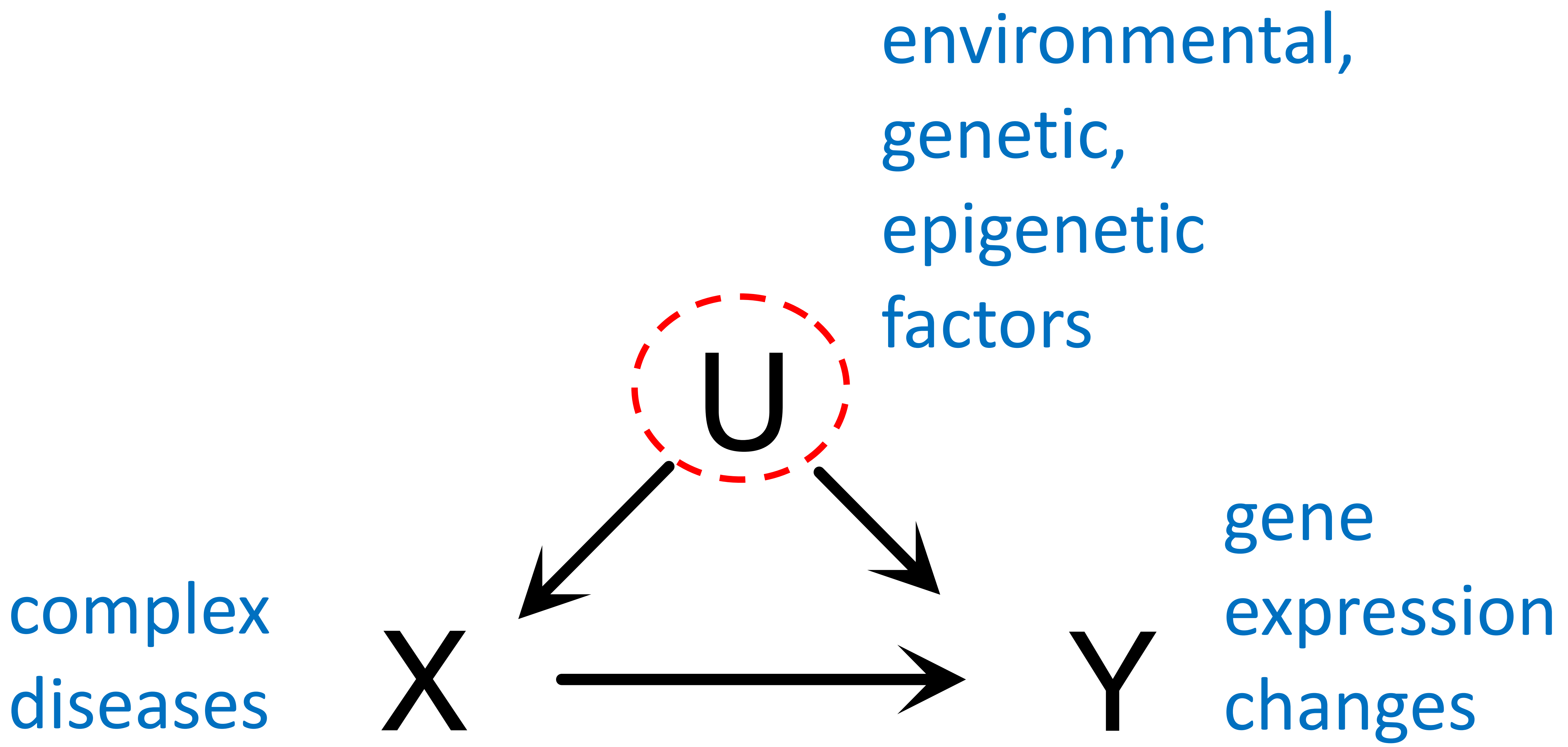
- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

Causal inference came back!

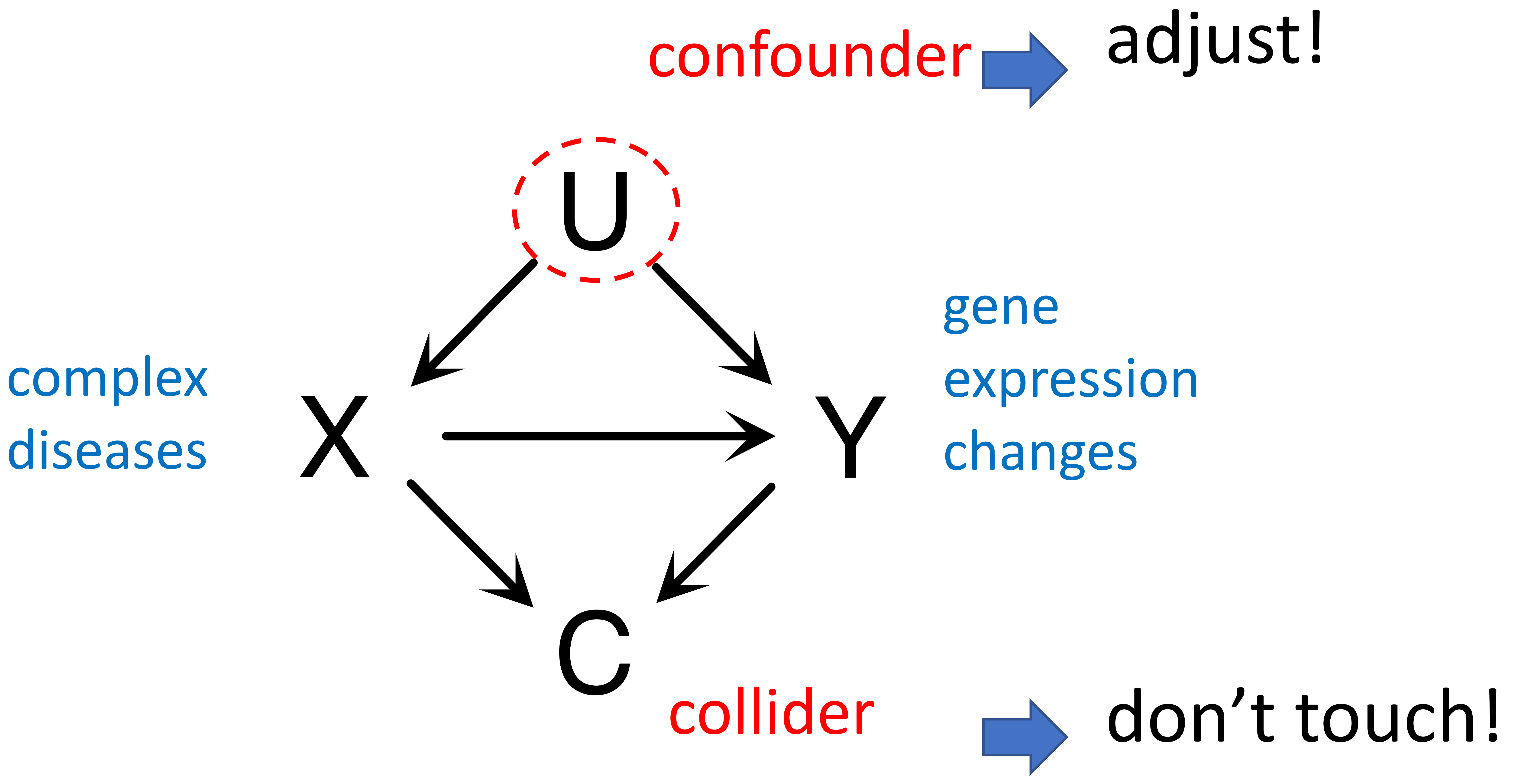
https://unsplash.com/photos/Q5qYZWlko9A

In the experimental design part of Lecture #3, we wanted to discuss causality.

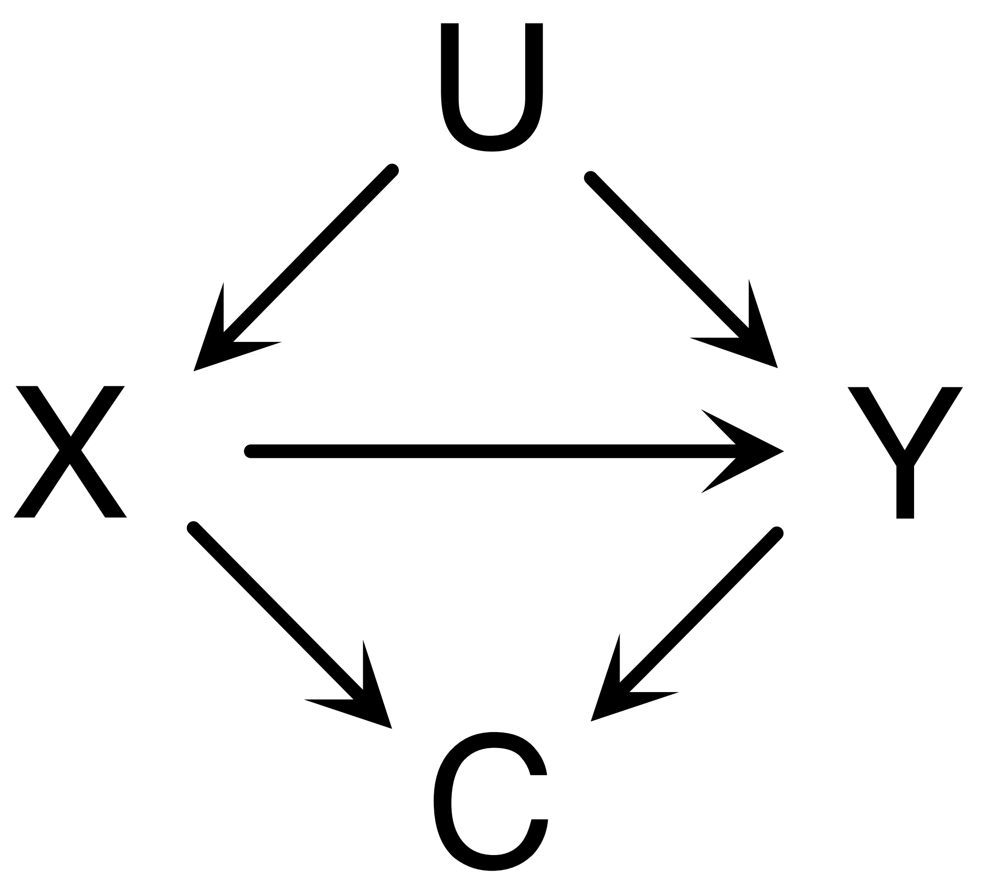
What if we have very little knowledge of confounding effects?



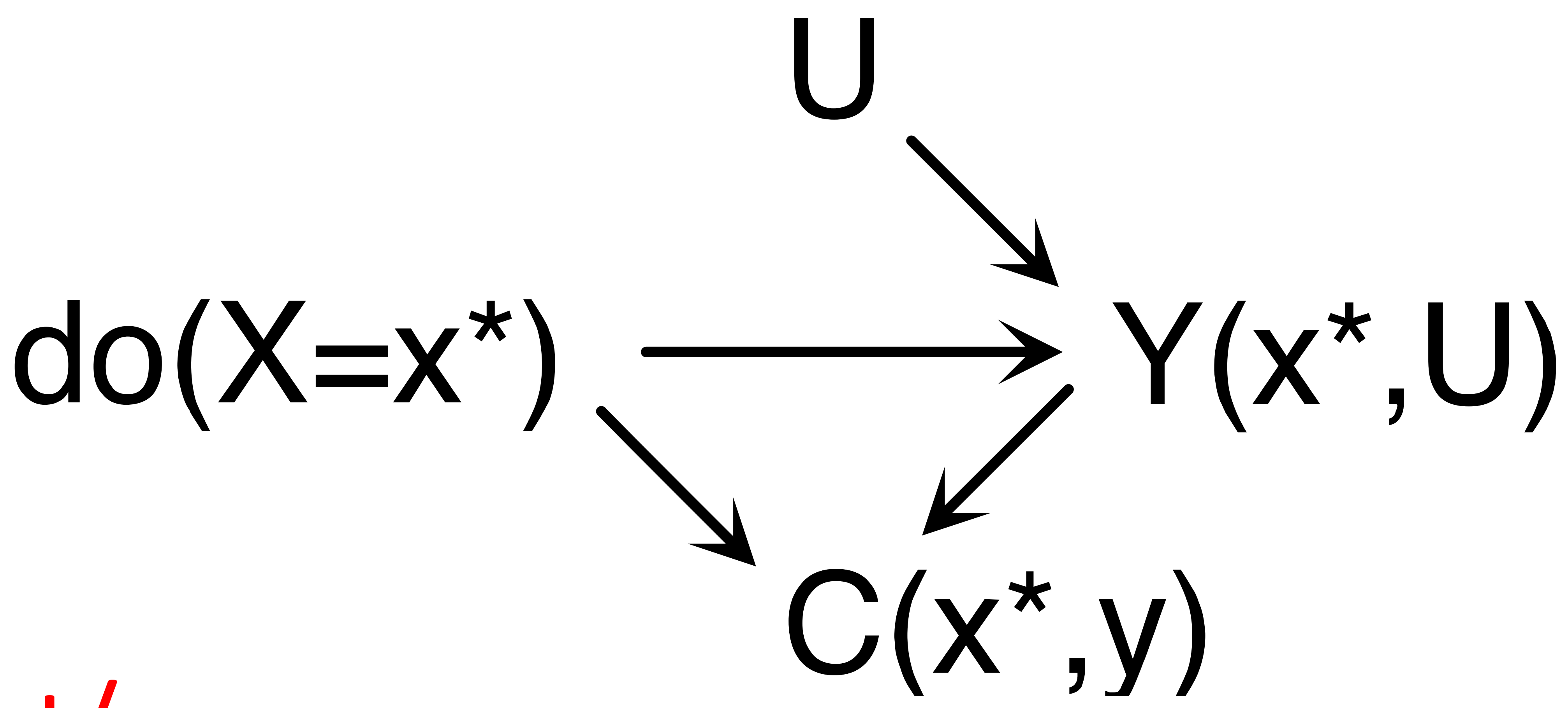
What if we don't know how to distinguish confounder vs. collider?



Randomized Control Trial: the gold standard experiment for causal discovery



What if we randomly assign X=x*



experiment/ intervention to change the underlying causal structure

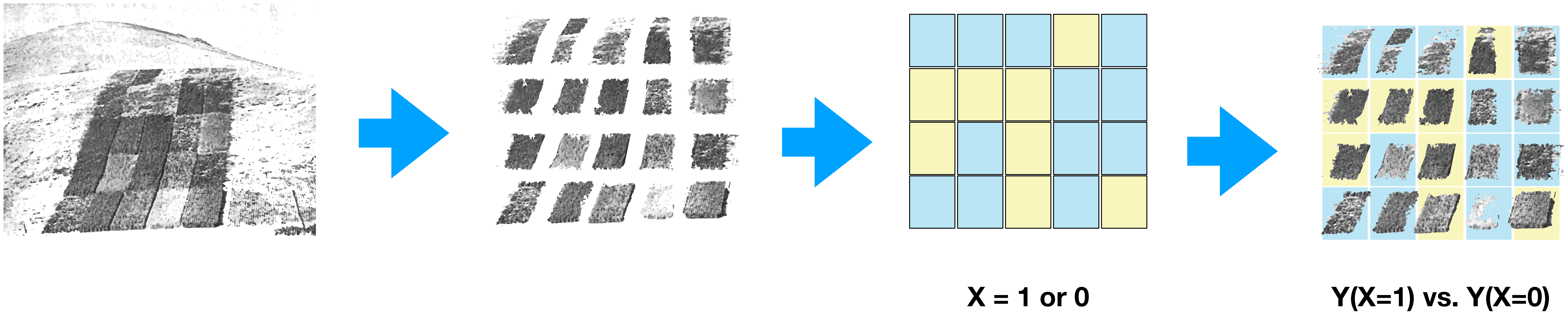
The arrangement of Field Experiments (RA Fisher, 1922-1926)

- Randomly select plots to treat manure (X=1) or not (X=0).
- Measure the yield of crops (Y).

ROTHAMSTED RESEARCH

Langkjær-Bain, “Where the seeds of modern statistics were sown” (2018)

The first Randomized Control Trial experiments



Today’s lecture: Bayesian, PGM, Causality

- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

Again, the R codes for the examples are all available online

https://github.com/STAT540-UBC/lectures/blob/main/lect13-causality_bayesian/causality.Rmd

A graphical language for causal inference (used throughout this lecture)

Probabilistic graphical models care about dependency

Causal relationship

$X \longrightarrow Y \quad \Leftrightarrow \quad$ "X causes Y" $\quad \Rightarrow \quad P(Y|X)$

$\overset{?}{\Leftarrow}$

Very frequently not true

Causal path & reachability

$X \longrightarrow Y \longrightarrow Z$

The effect of X can influence Z (flow)

Conditioning/adjustment

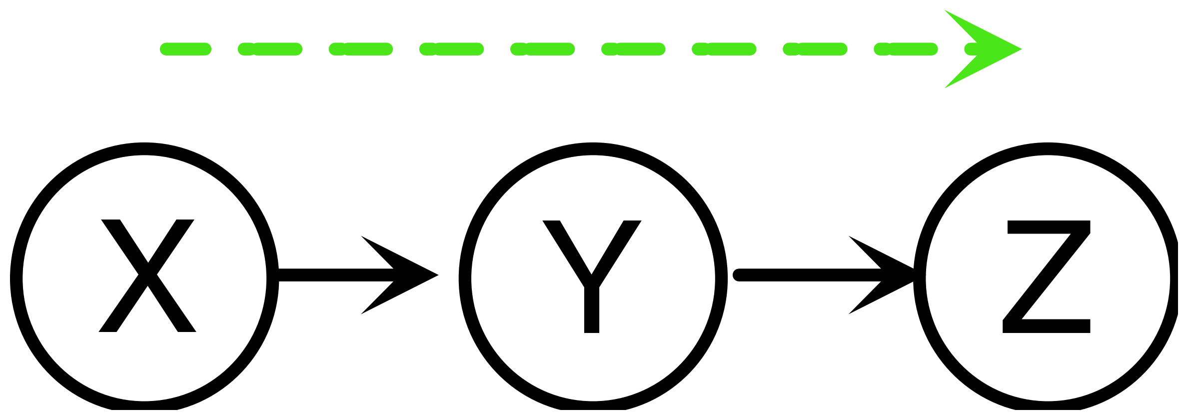
$X \rightarrow Y \rightarrow Z$

Open or no circle: not conditioned

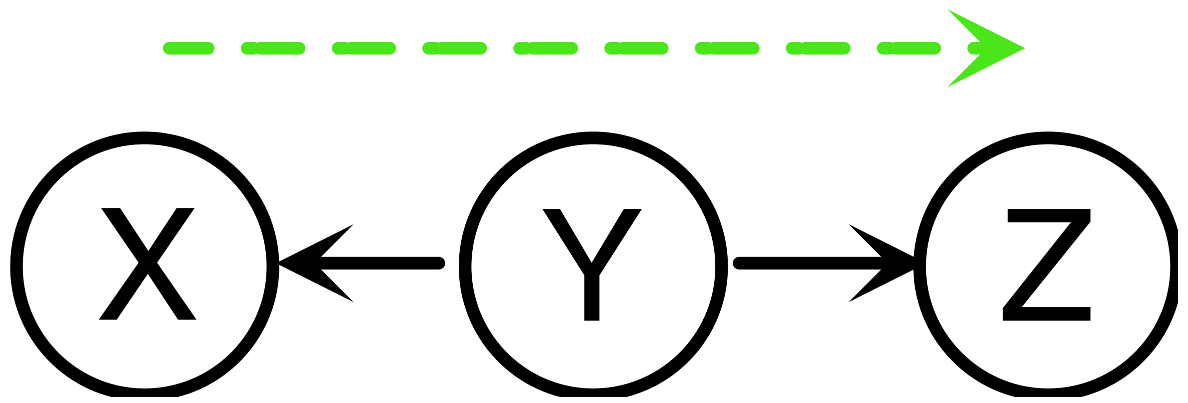
$X \rightarrow Y \rightarrow Z$

Closed circle: conditioned (e.g., setting/given Y = y*)

d-separation: testing conditional independency (flow vs. no flow)

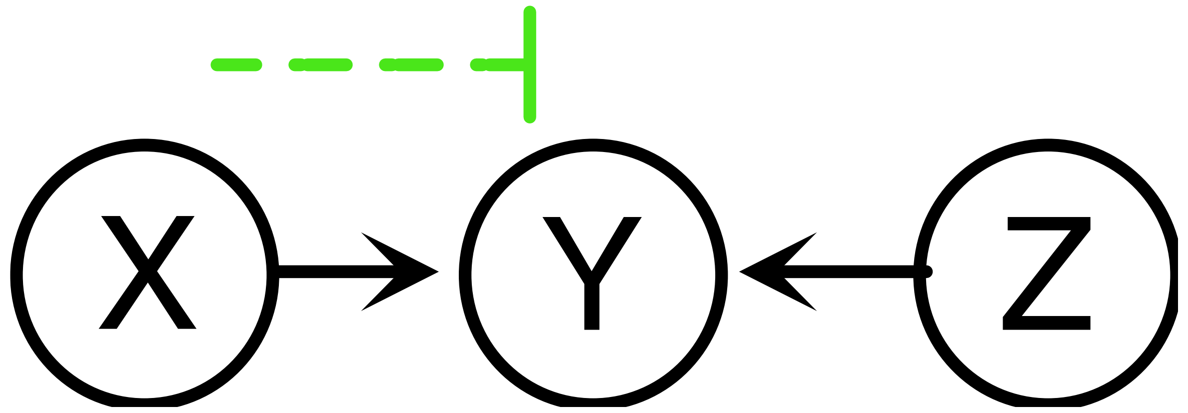


- “The outcome of Z depends on the effect of X”
- Because the outcome of Z depends on the outcome of Y
- And the outcome of Y depends on X



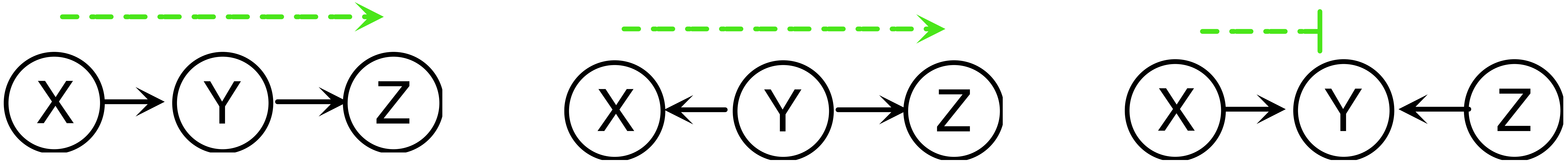
- “The outcome of Z depends on the effect of Y”
- Since “the outcome of X depends on Y”
- The outcomes of X and Z are dependent on each other

Collider (v-structure)

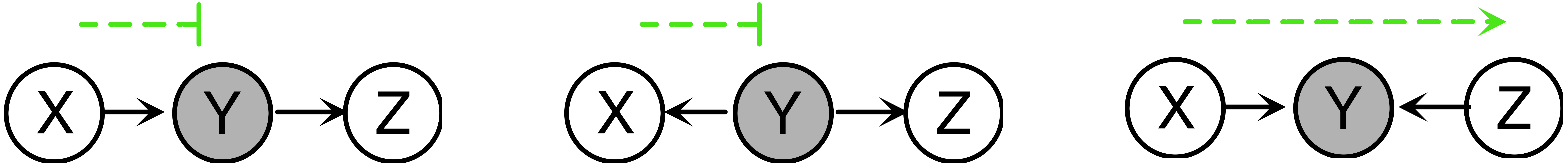


- The outcome of Z doesn’t depend on Y or X
- (Neither X nor Y causes Z)

d-separation: testing conditional independency (flow vs. no flow)



"Explain away"

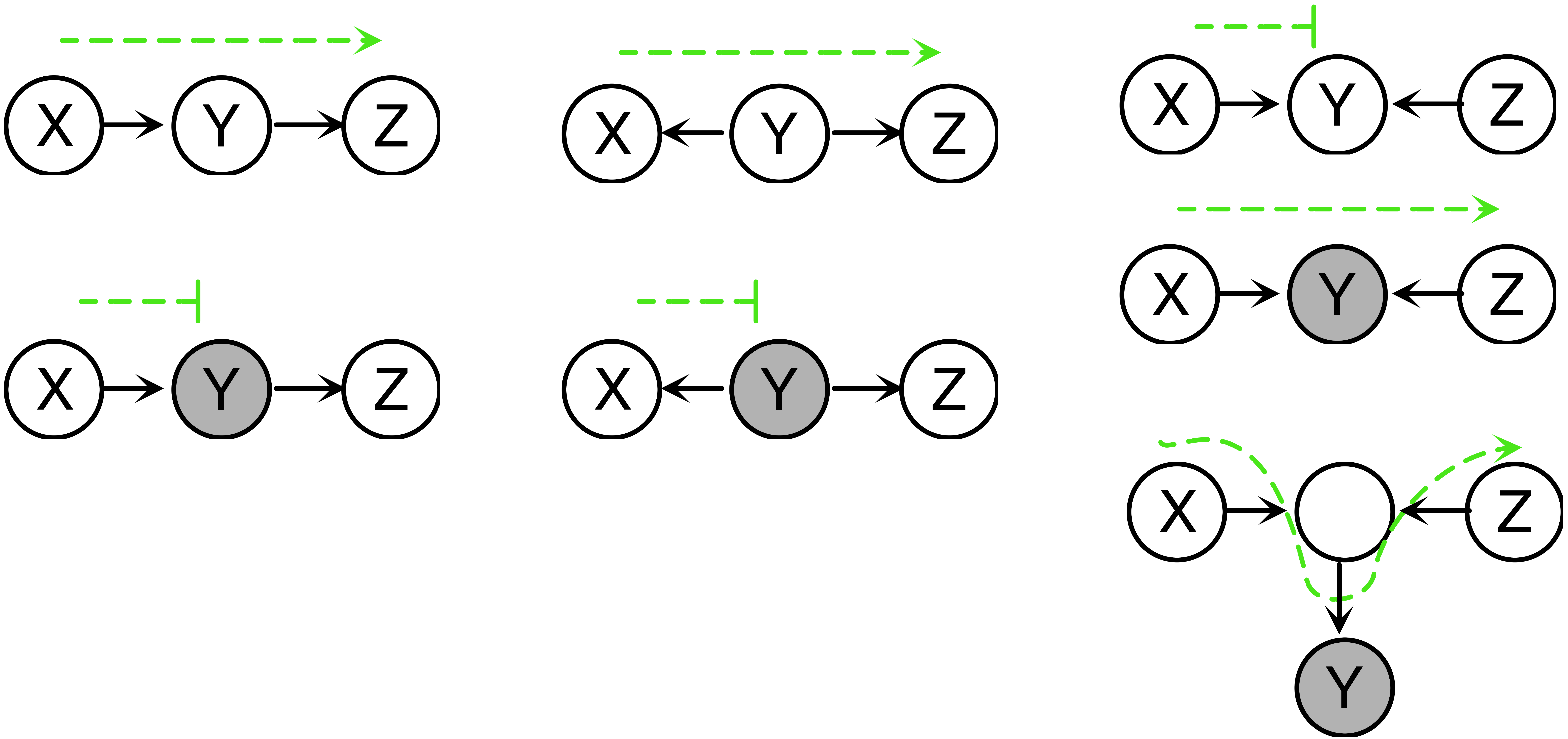


- Given Y=y*, Z doesn't depend on X

- Given Y=y*, Z only depends on Y

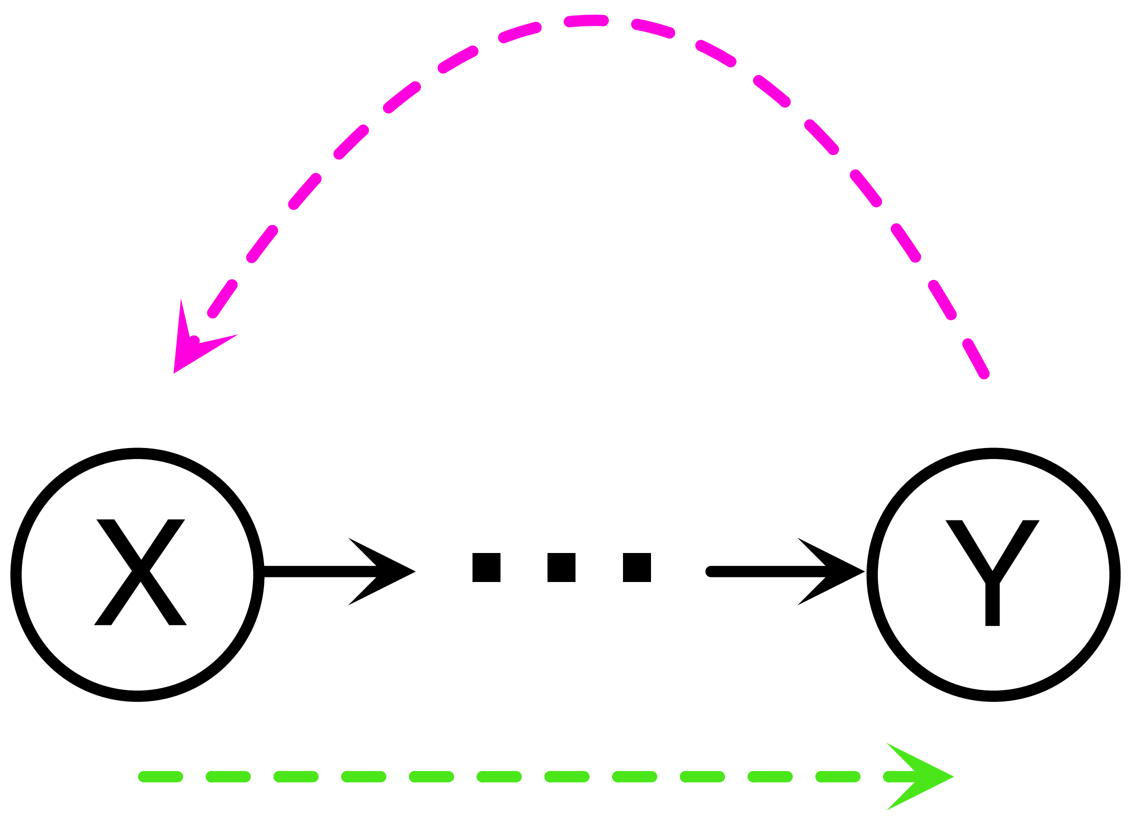
- Either X or Z can explain {Y = y*}
- If {X = x*} can explain {Y = y*}, Z doesn't explain Y
- vice versa

d-separation: testing conditional independency (flow vs. no flow)



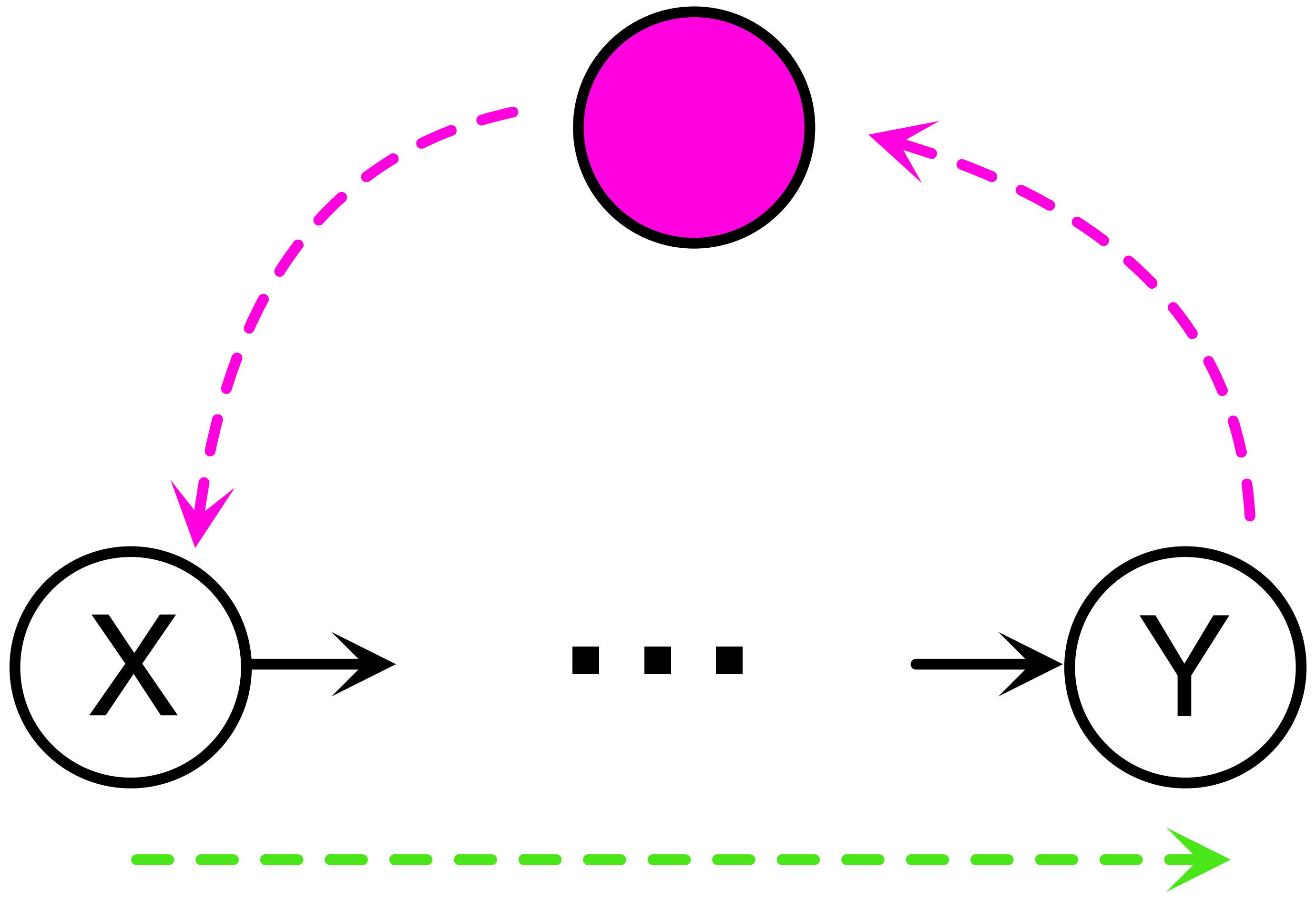
"Backdoor" exercise

Question: Which nodes should be "conditioned" and/or "adjusted" to block a reverse path from Y to X?

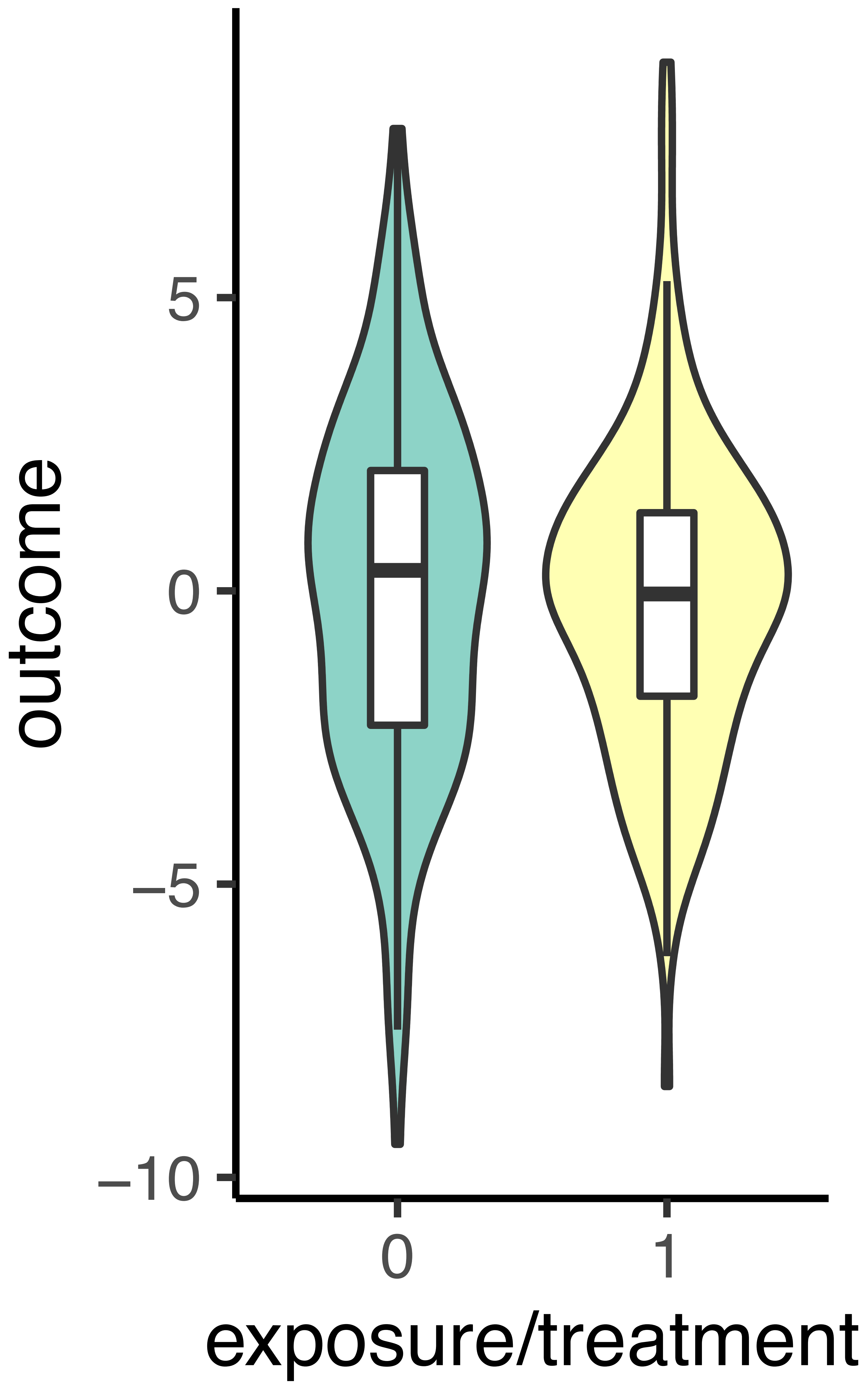


"Backdoor" exercise

Question: Which nodes should be "conditioned" and/or "adjusted" to block a reverse path from Y to X?



A working example: confounder adjustment in case-control study

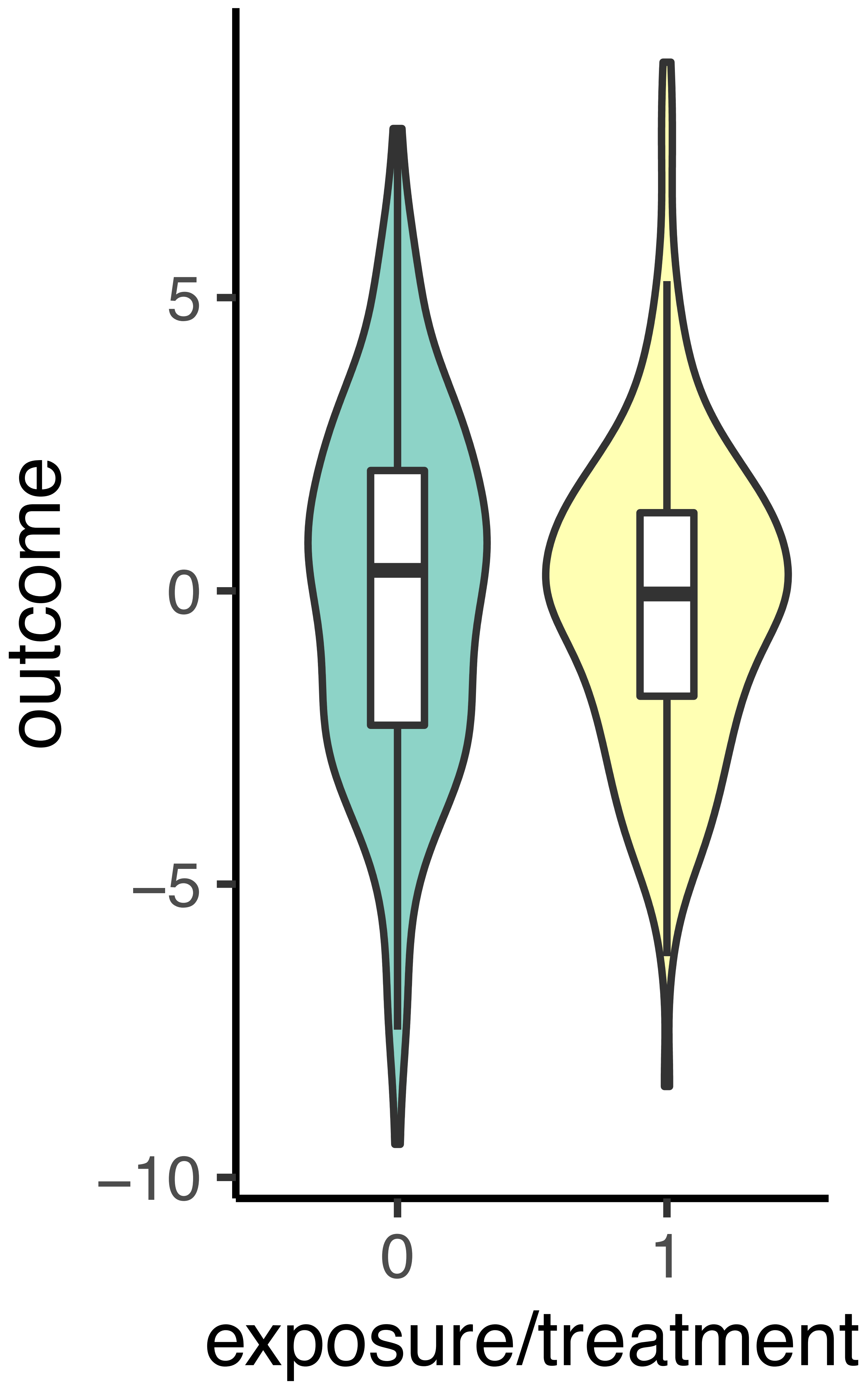


Key question: $X \overset{?}{\to} Y$

X: exposure variable

Y: outcome variable

A working example: confounder adjustment in case-control study



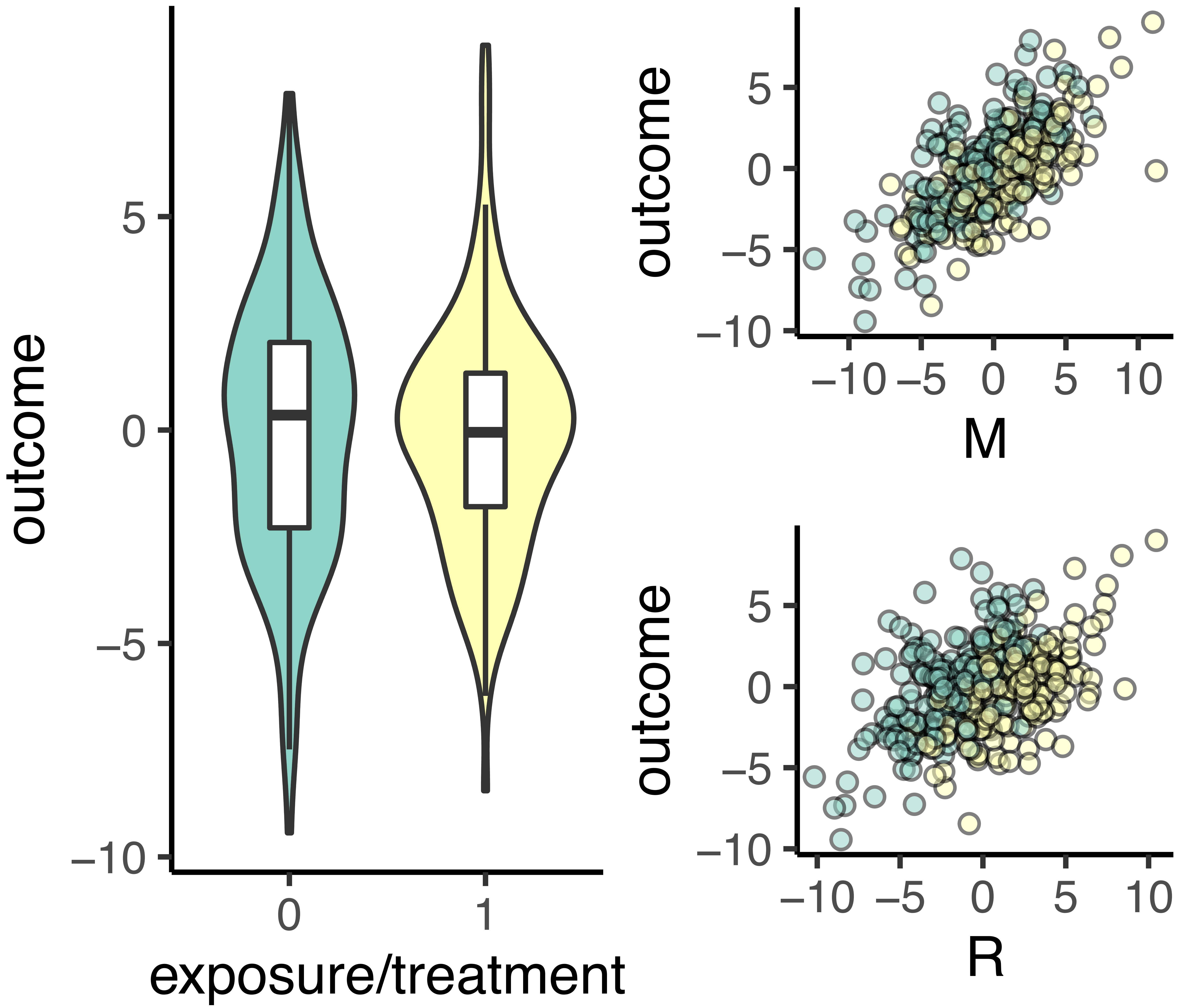
Key question: $X \overset{?}{\rightarrow} Y$

$\mathbf{X}$: exposure variable

$\mathbf{Y}$: outcome variable

$\mathbf{C_1}$ and $\mathbf{C_2}$: covariates

A working example: confounder adjustment in case-control study



Key question: $X \overset{?}{\to} Y$

X: exposure variable

Y: outcome variable

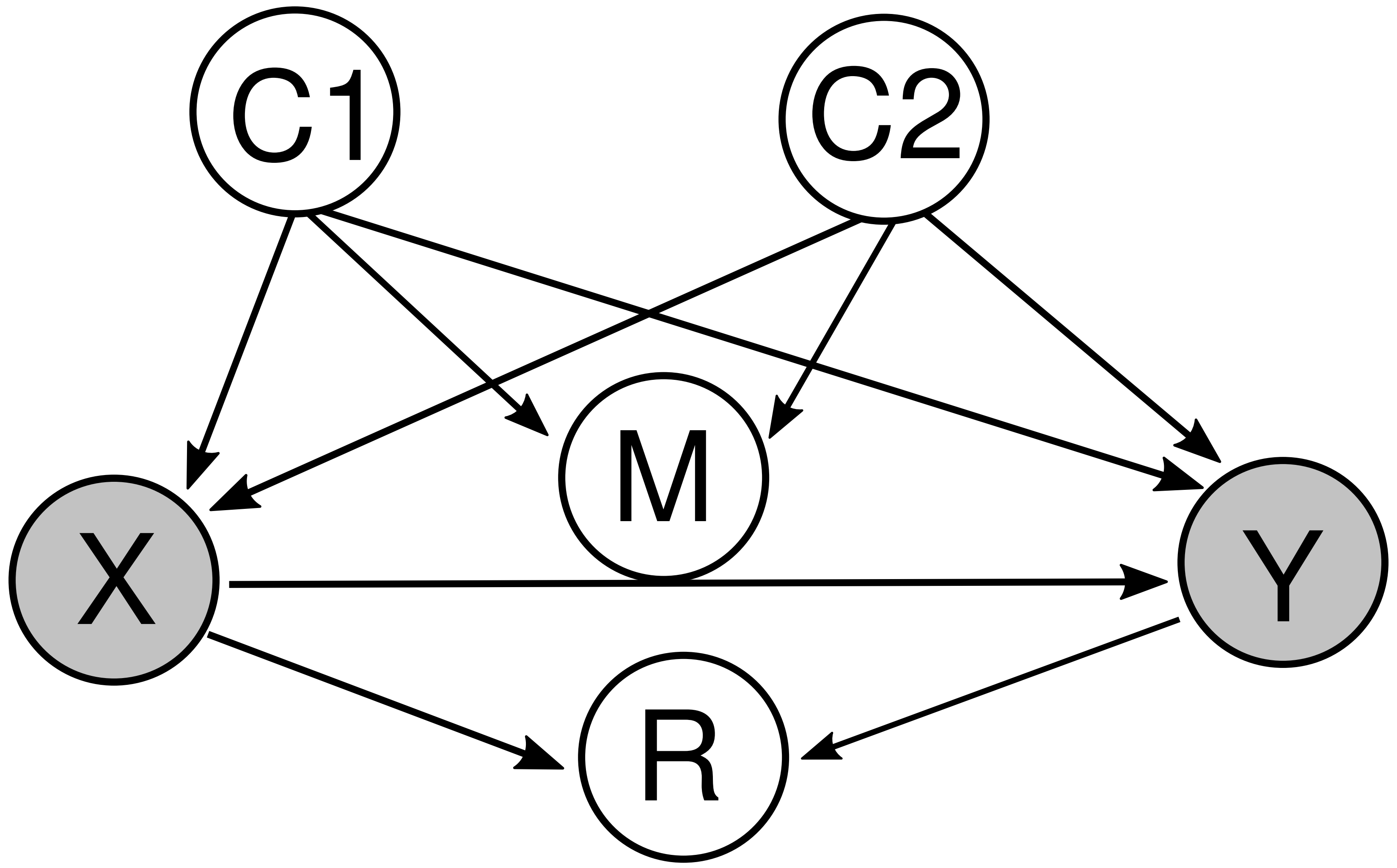
$\mathbf{C_1}$ and $\mathbf{C_2}$: covariates

M: other covariate

R: other covariate

Causal inference with a graphical model

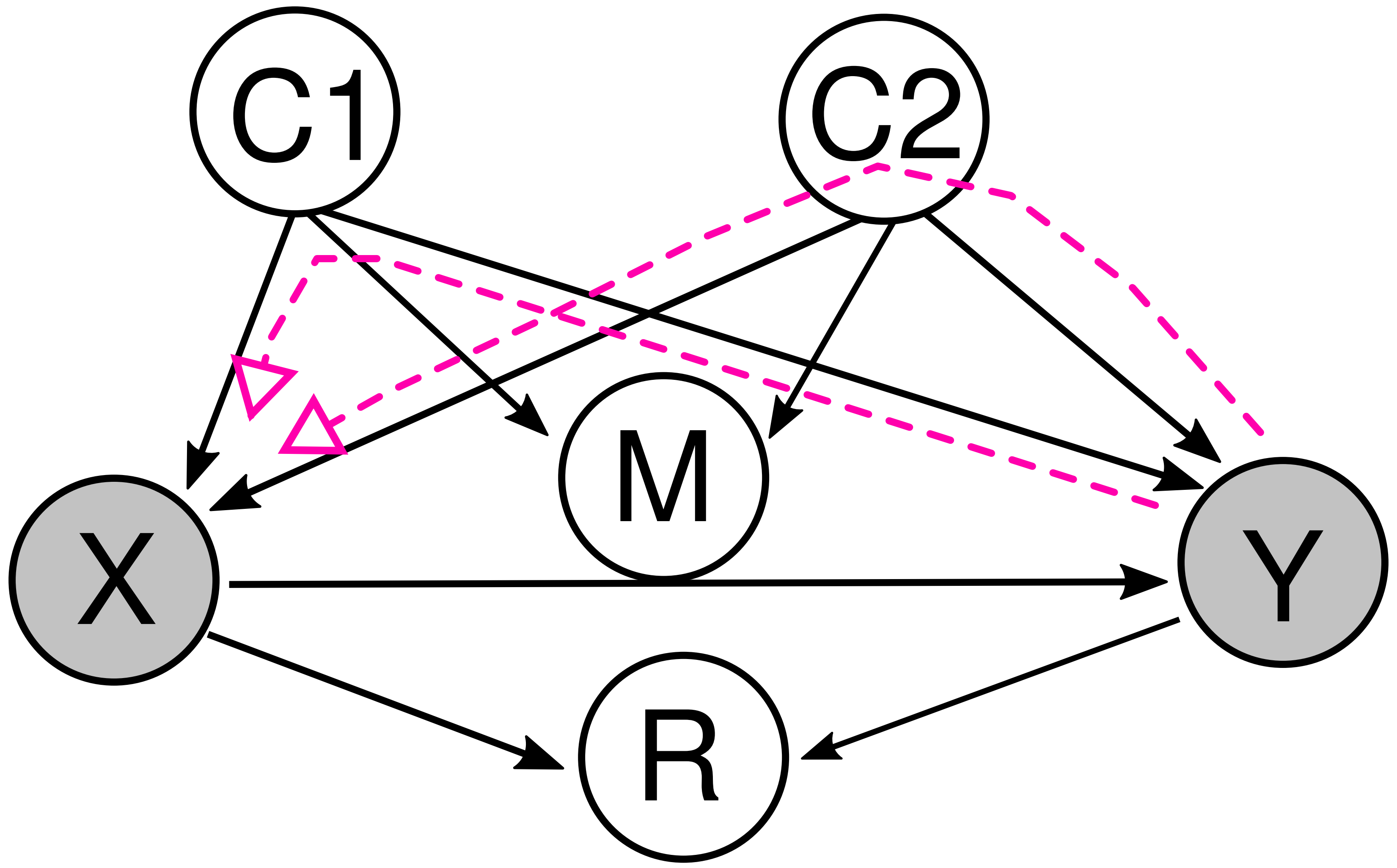
1. Build a causal structural model



What are potential backdoors?

Causal inference with a graphical model

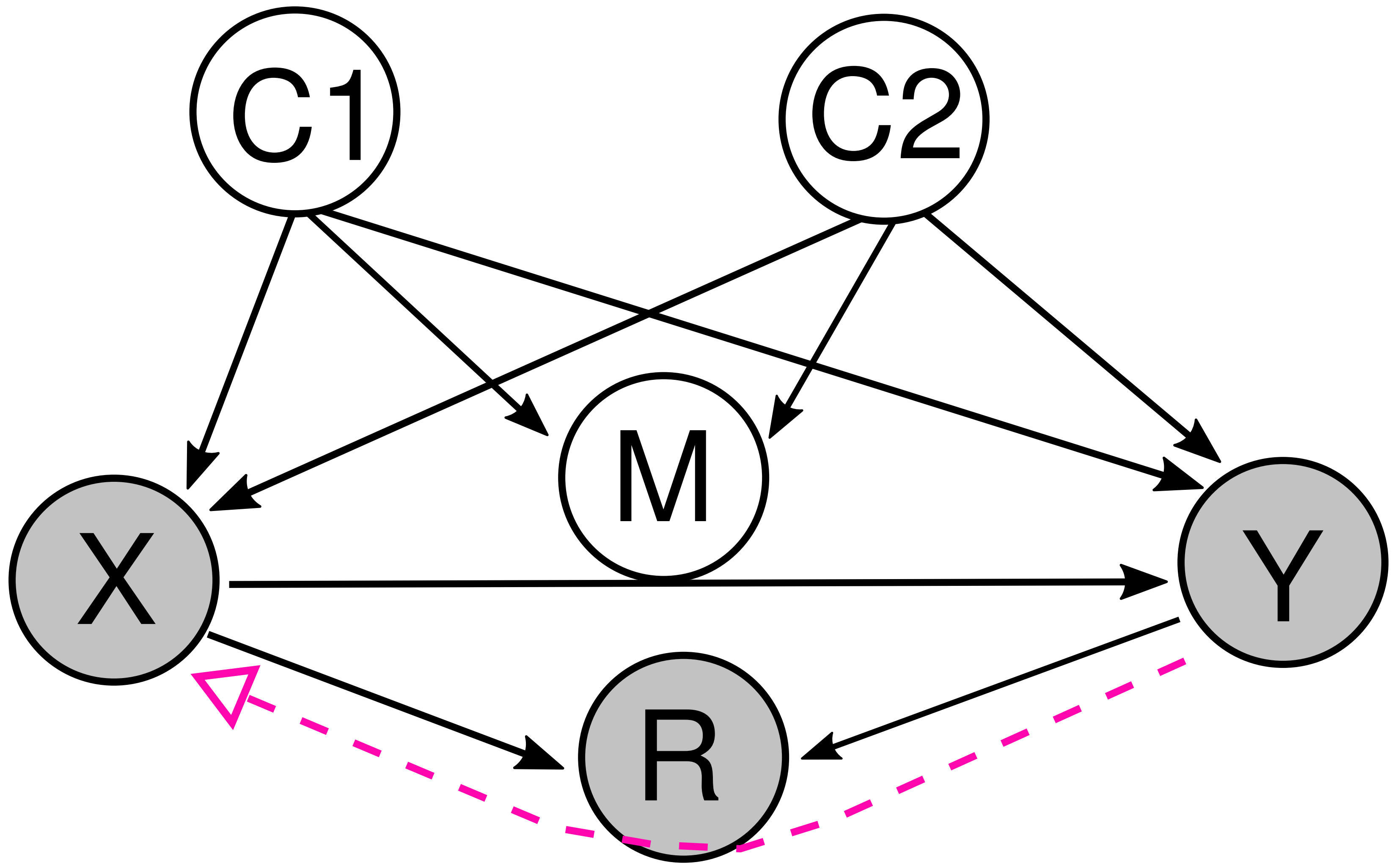
1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \rightarrow X$, *while opening* $X \rightarrow Y$)



How do we close them?

Causal inference with a graphical model

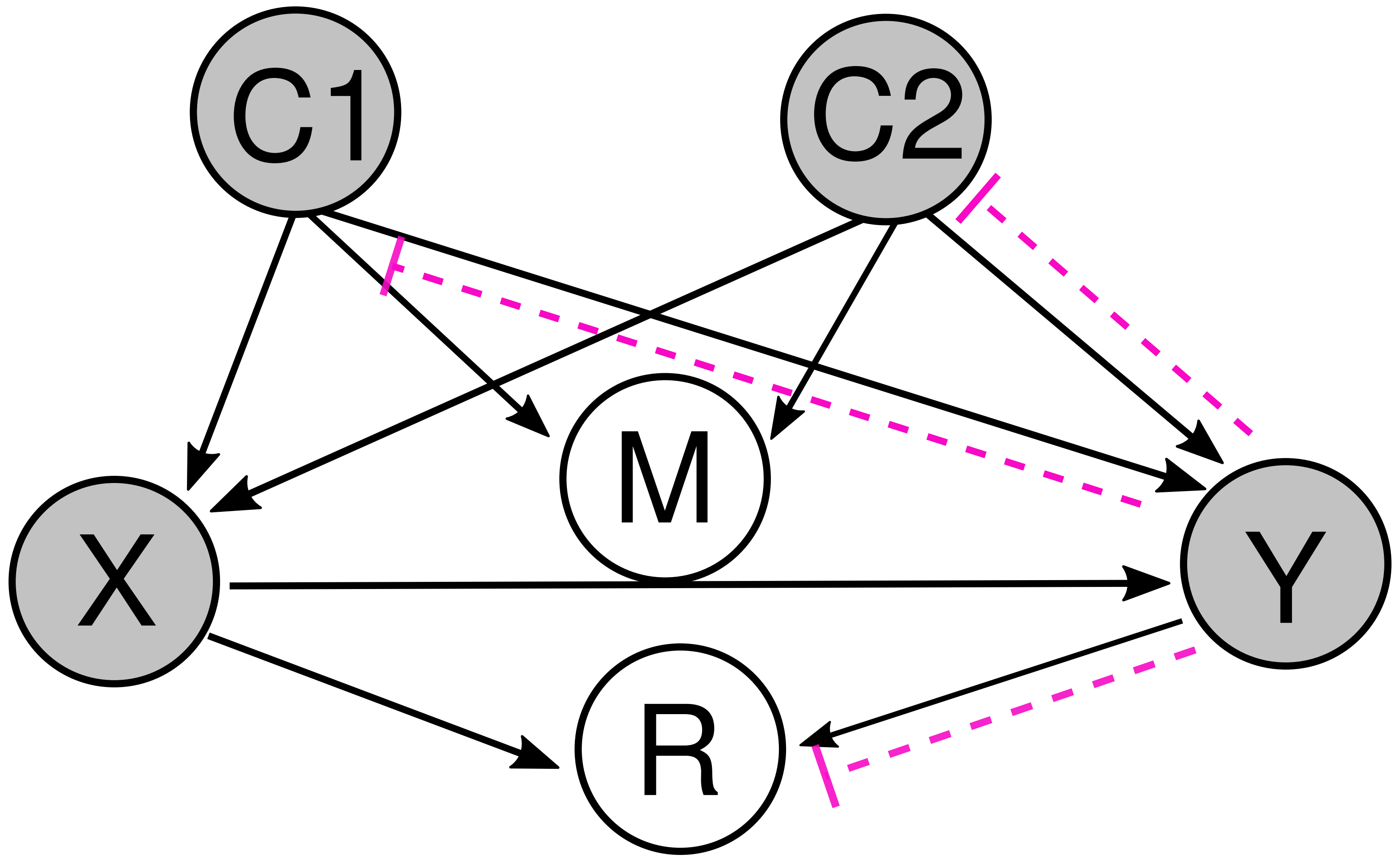
1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \rightarrow X$, *while opening* $X \rightarrow Y$)



What about this?

Causal inference with a graphical model

1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \to X$, *while opening* $X \to Y$)



Is this enough?

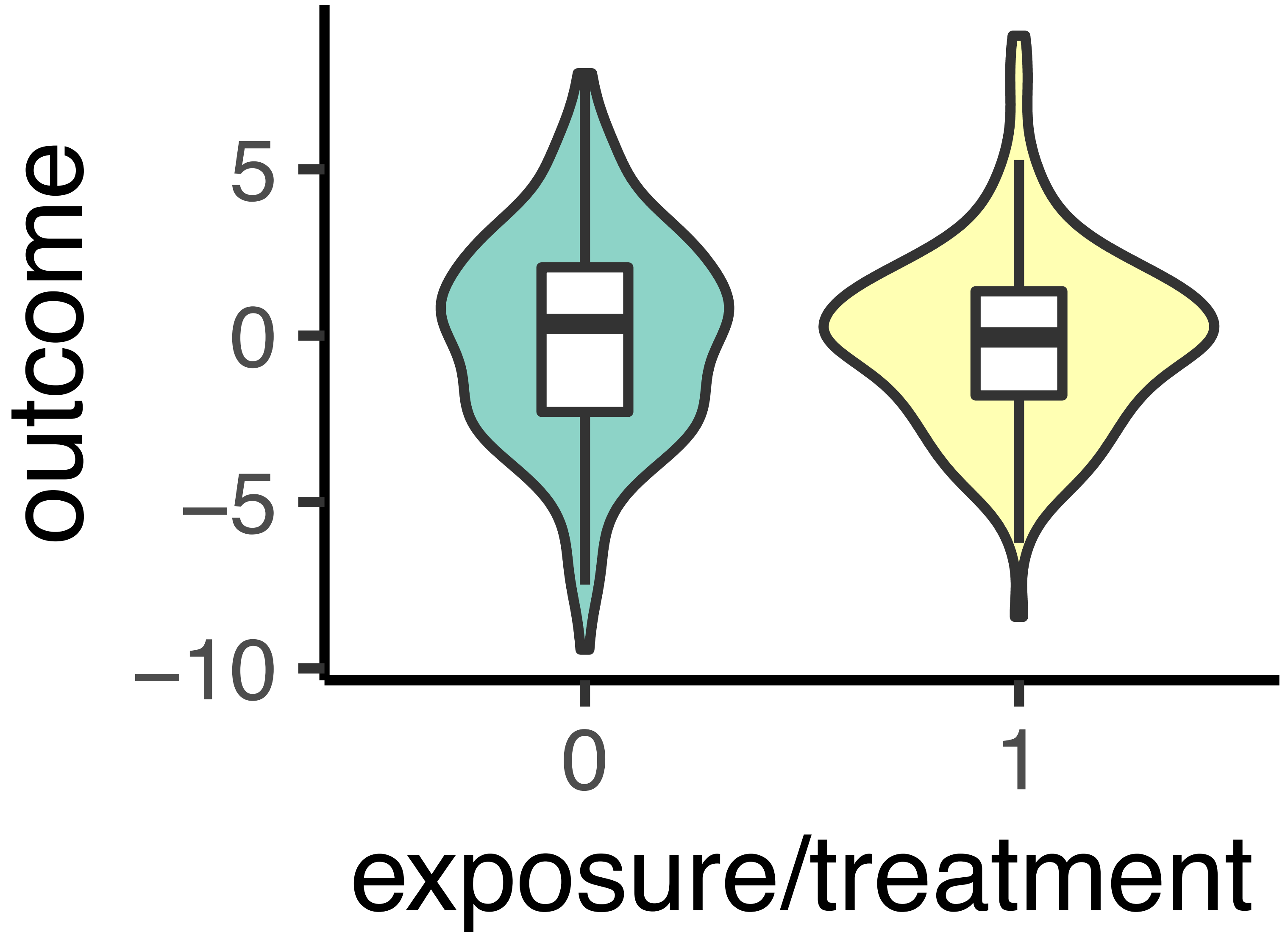
Causal inference with a graphical model

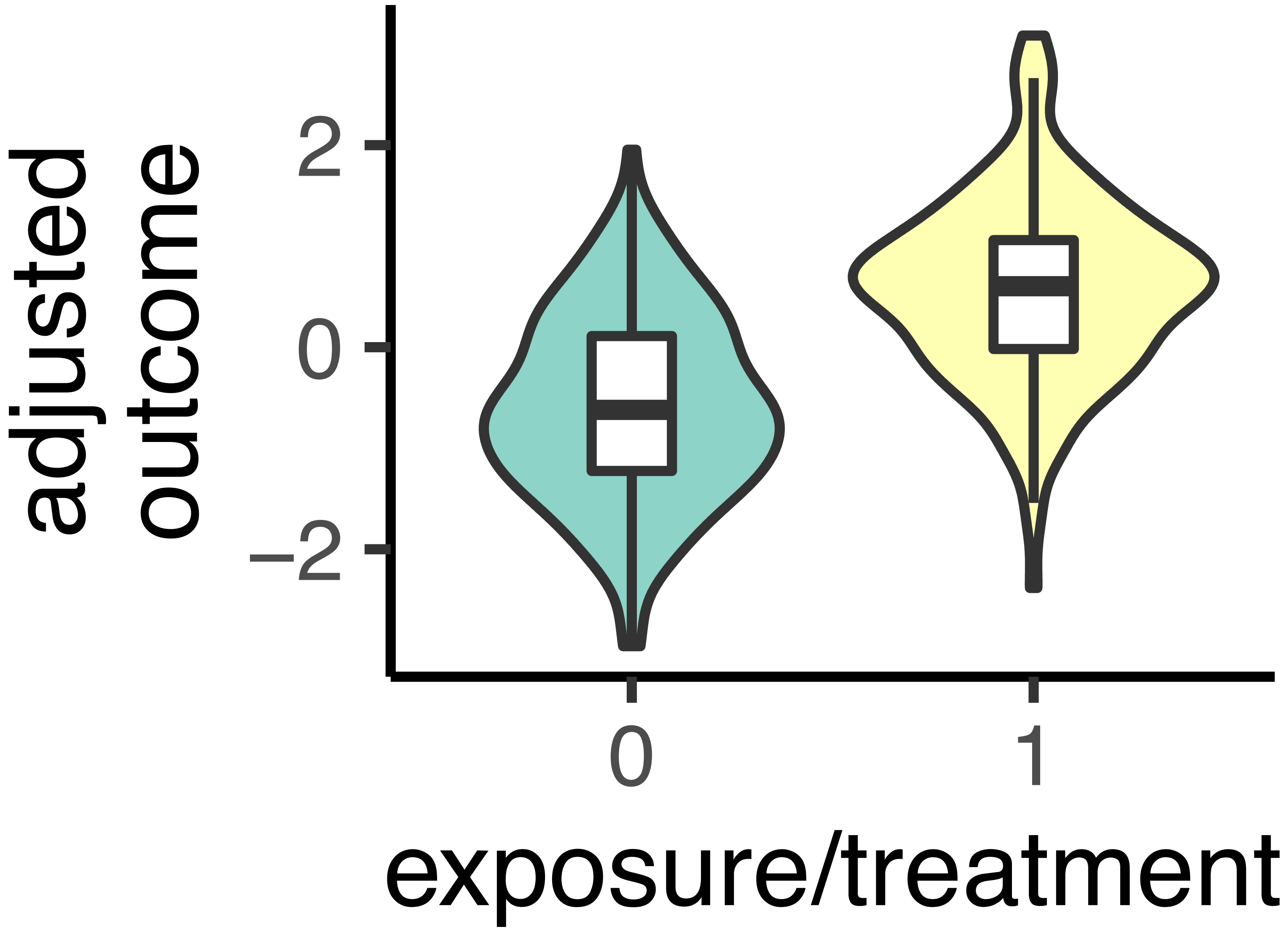
1. Build a causal structural model
2. Identify "back-door" paths/variables (*closing* $Y \to X$, *while opening* $X \to Y$)
3. Adjust "back-door" variables
4. Estimate causal effects

$$Y \leftarrow Y - \sum_{k=1}^{2} C_k \hat{\beta}_k$$

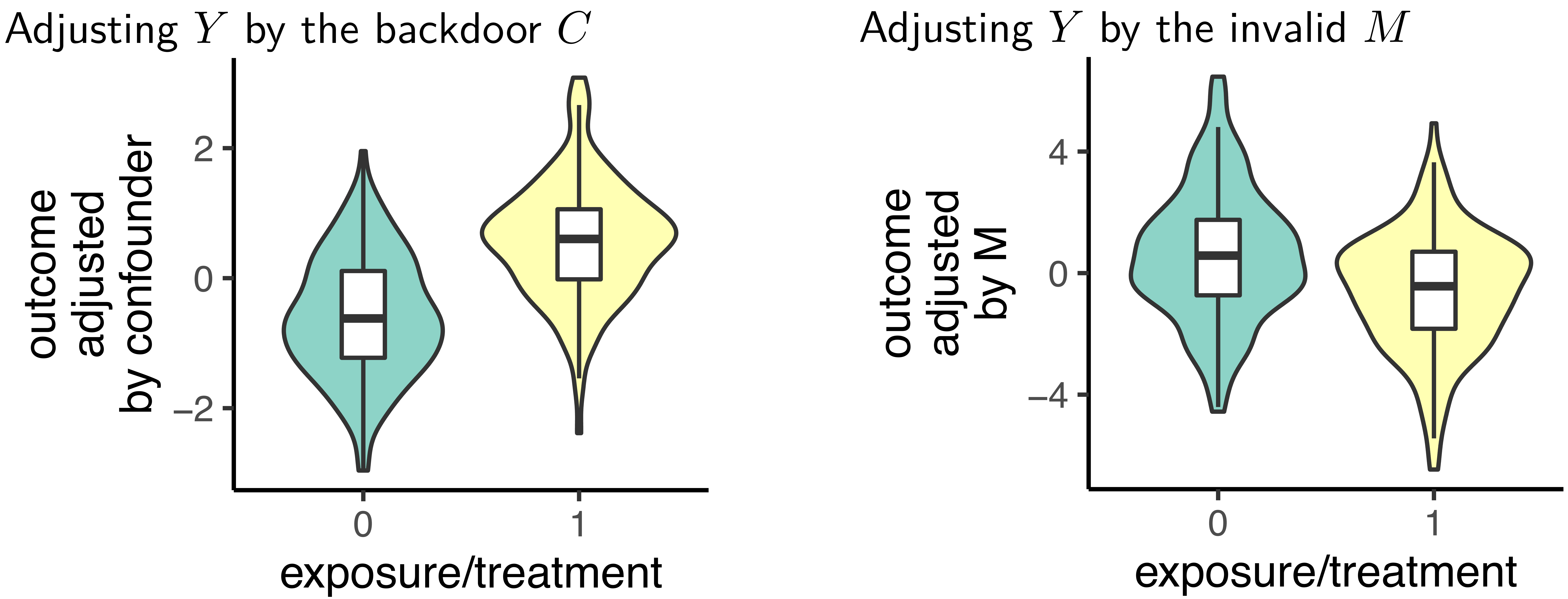
which approximates

$$p(Y|X) = \int_C p(Y|X,C)p(C)dC$$



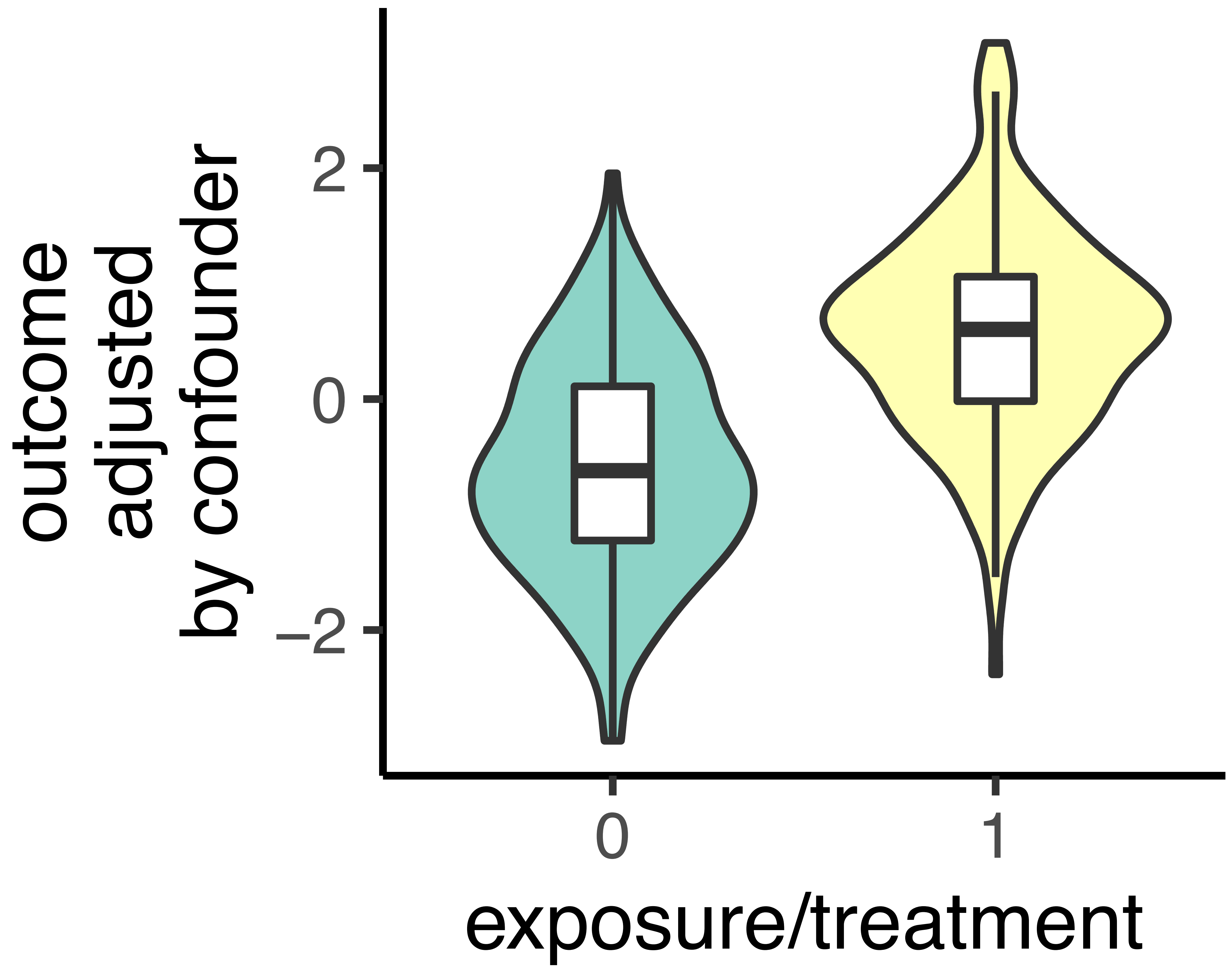


What would happen if we close a wrong "backdoor" variable?

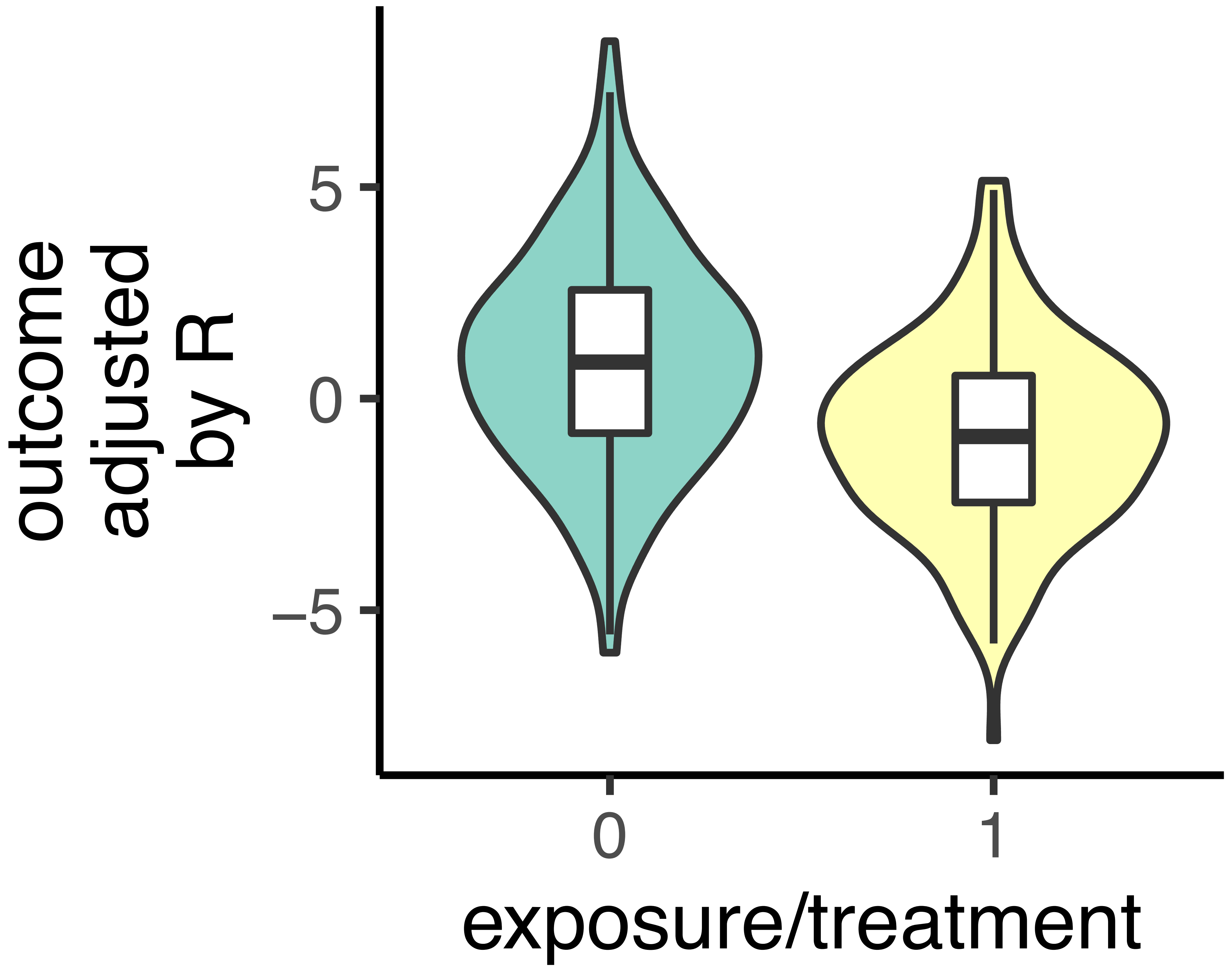


What would happen if we close a wrong "backdoor" variable?

Adjusting Y by the backdoor C



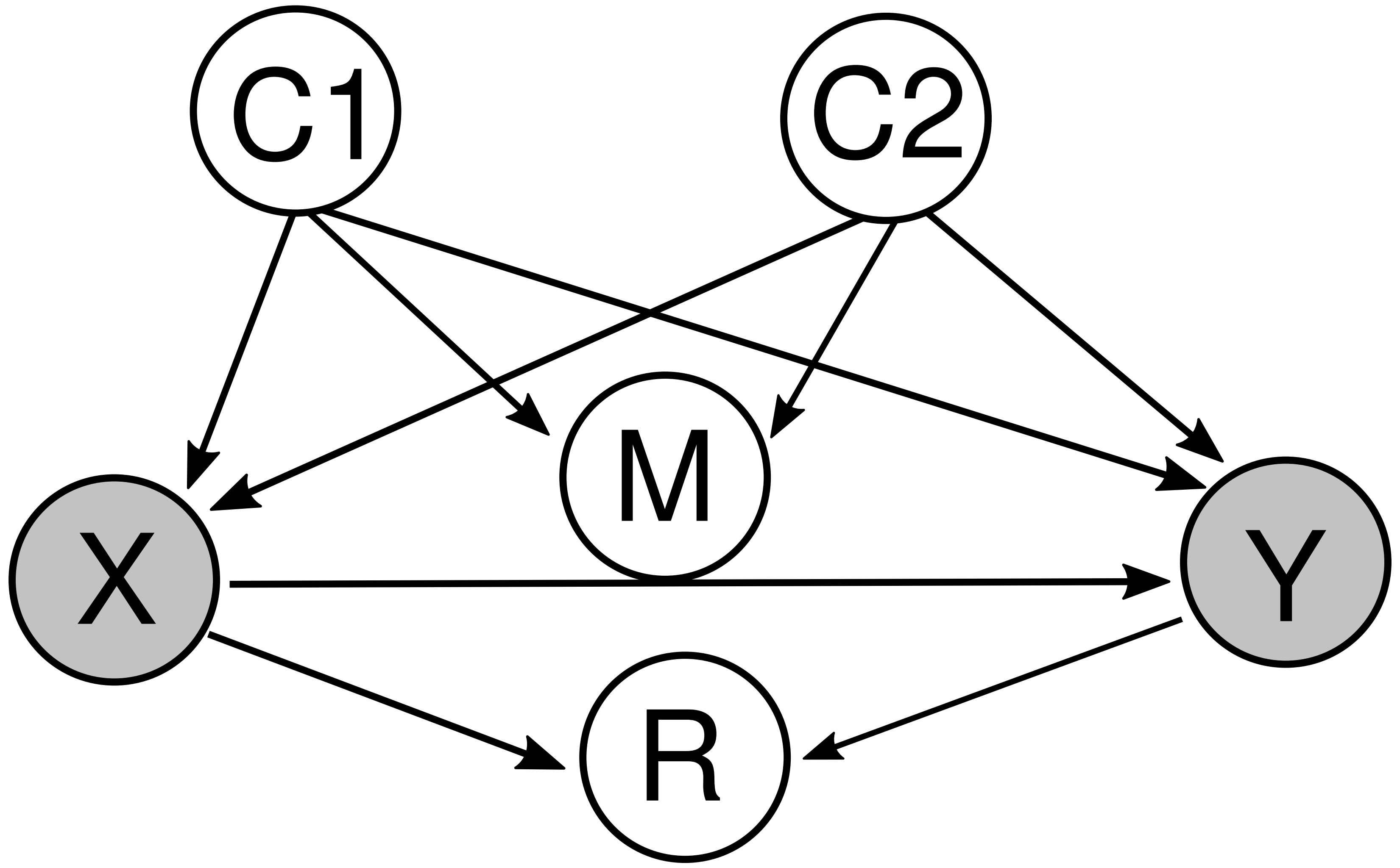
Adjusting Y by the collider R



Today's lecture: Bayesian, PGM, Causality

- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

The same example with causal relationship from X to Y

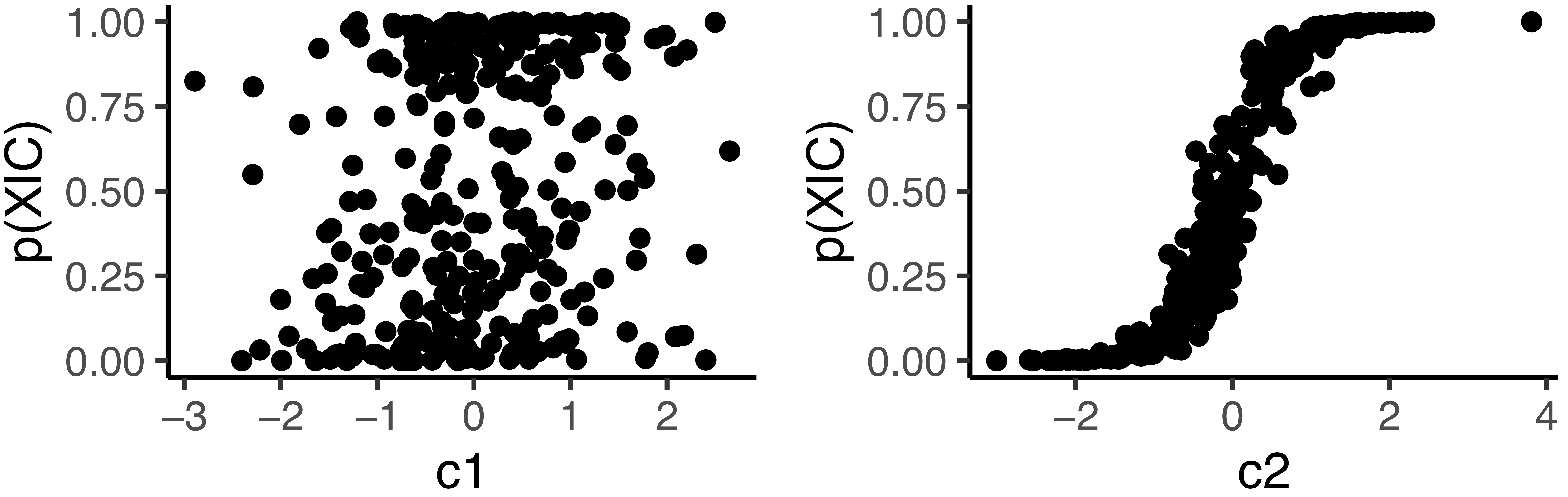


Inverse Propensity Weighting

What is the model for X given confoudners (the backdoor variables)?

Propensity $=$ probability of assignment $X = 1$:

$$p(X|C_1, C_2) \approx \frac{1}{1 + \exp(-\beta_0 - \beta_1 C_1 - \beta_2 C_2)}$$



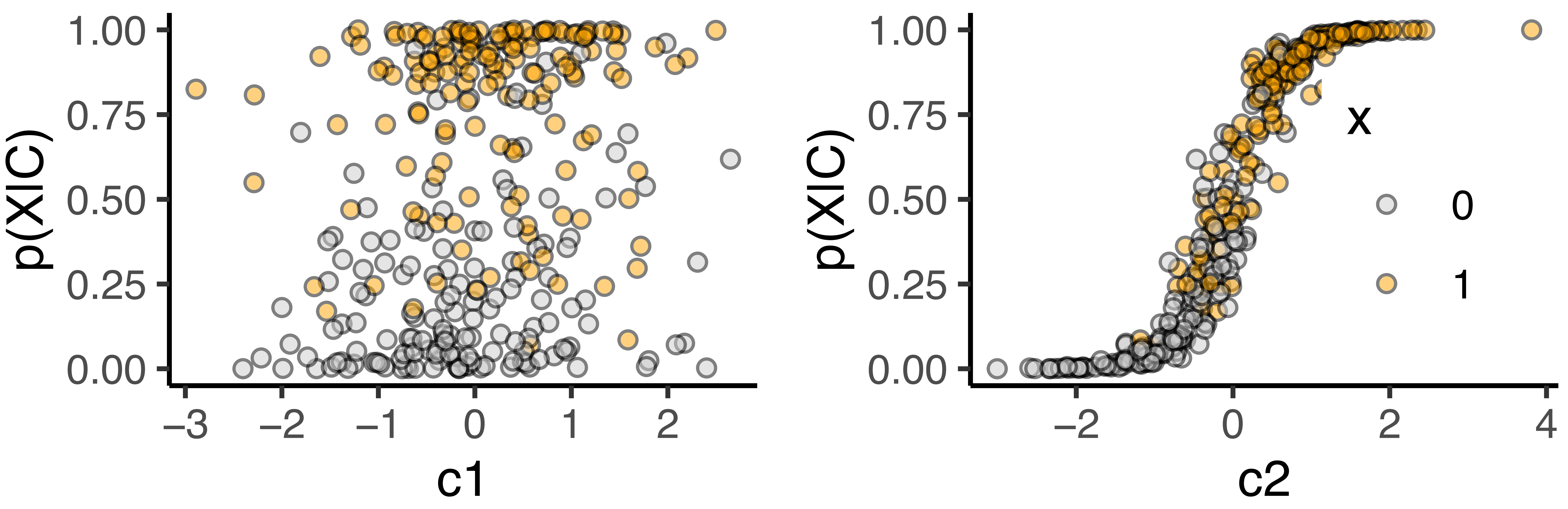
Horvitz and Thomson, *JASA*, 1952

Inverse Propensity Weighting

What is the model for X given confoudners (the backdoor variables)?

Propensity $=$ probability of assignment $X = 1$:

$$p(X|C_1, C_2) \approx \frac{1}{1 + \exp(-\beta_0 - \beta_1 C_1 - \beta_2 C_2)}$$



Horvitz and Thomson, *JASA*, 1952

Intuition behind IPW

What if we have assigned $X = 1$ unconfounded by C?

Intuition behind IPW

What if we have assigned $X = 1$ unconfounded by C?

In other words, what if samples could be drawn more from the underrepresented group?

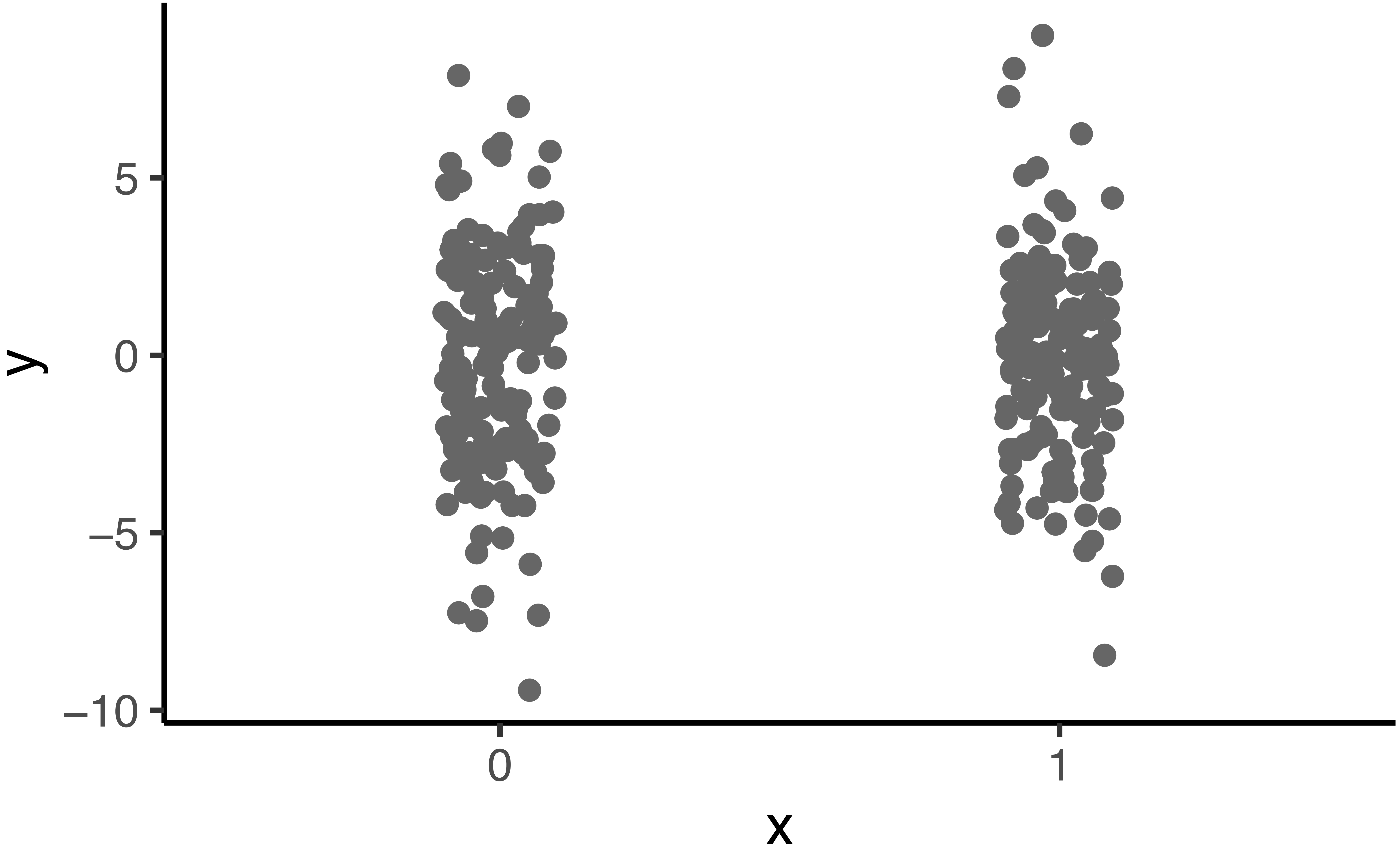
Intuition behind IPW

What if we have assigned $X = 1$ unconfounded by C?

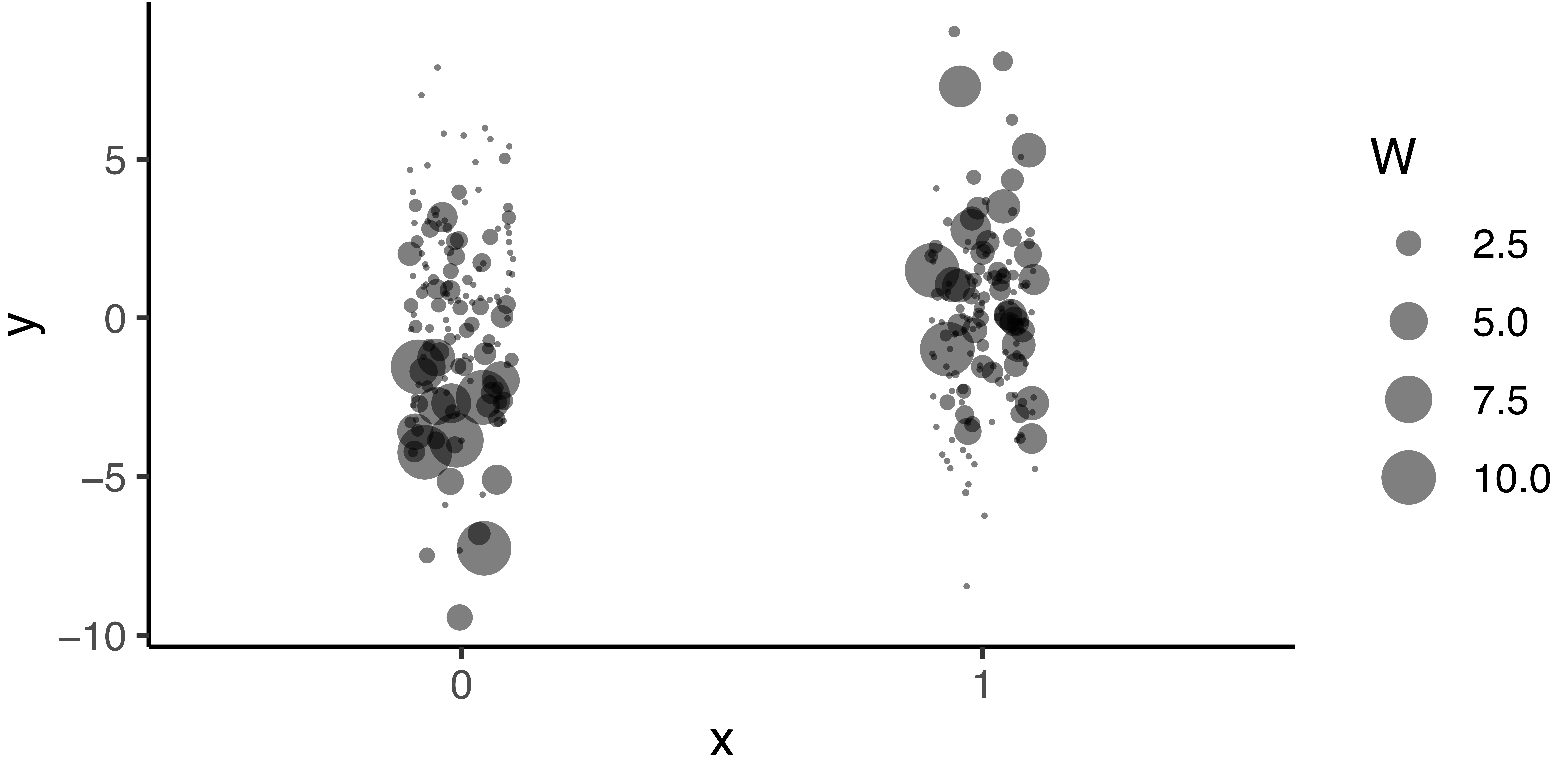
In other words, what if samples could be drawn more from the underrepresented group?

Likewise, what if samples could be dropped in the overrepresented group?

Take samples inversely proportional to propensity

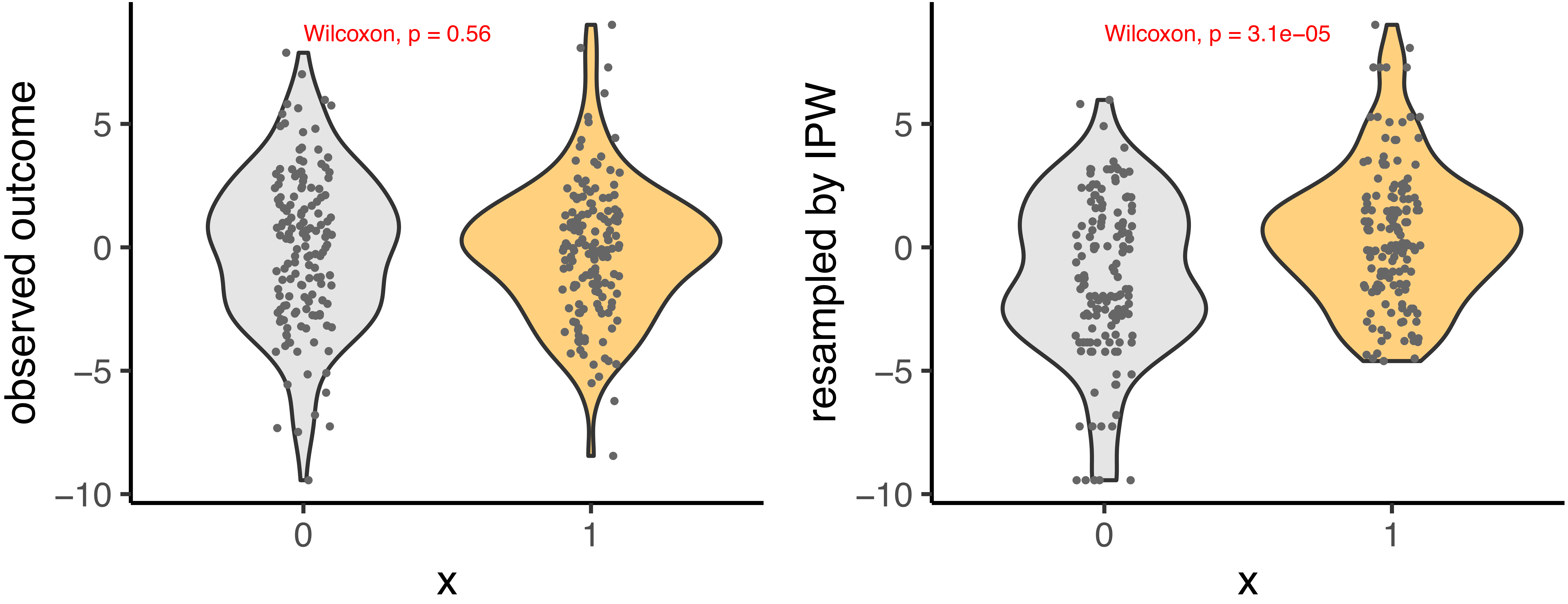


Take samples inversely proportional to propensity



$$W_i \propto \begin{cases} 1/p(X_i = 1|C_i) & X_i = 1 \\ 1/p(X_i = 0|C_i) & X_i = 0 \end{cases}$$

Take samples inversely proportional to propensity

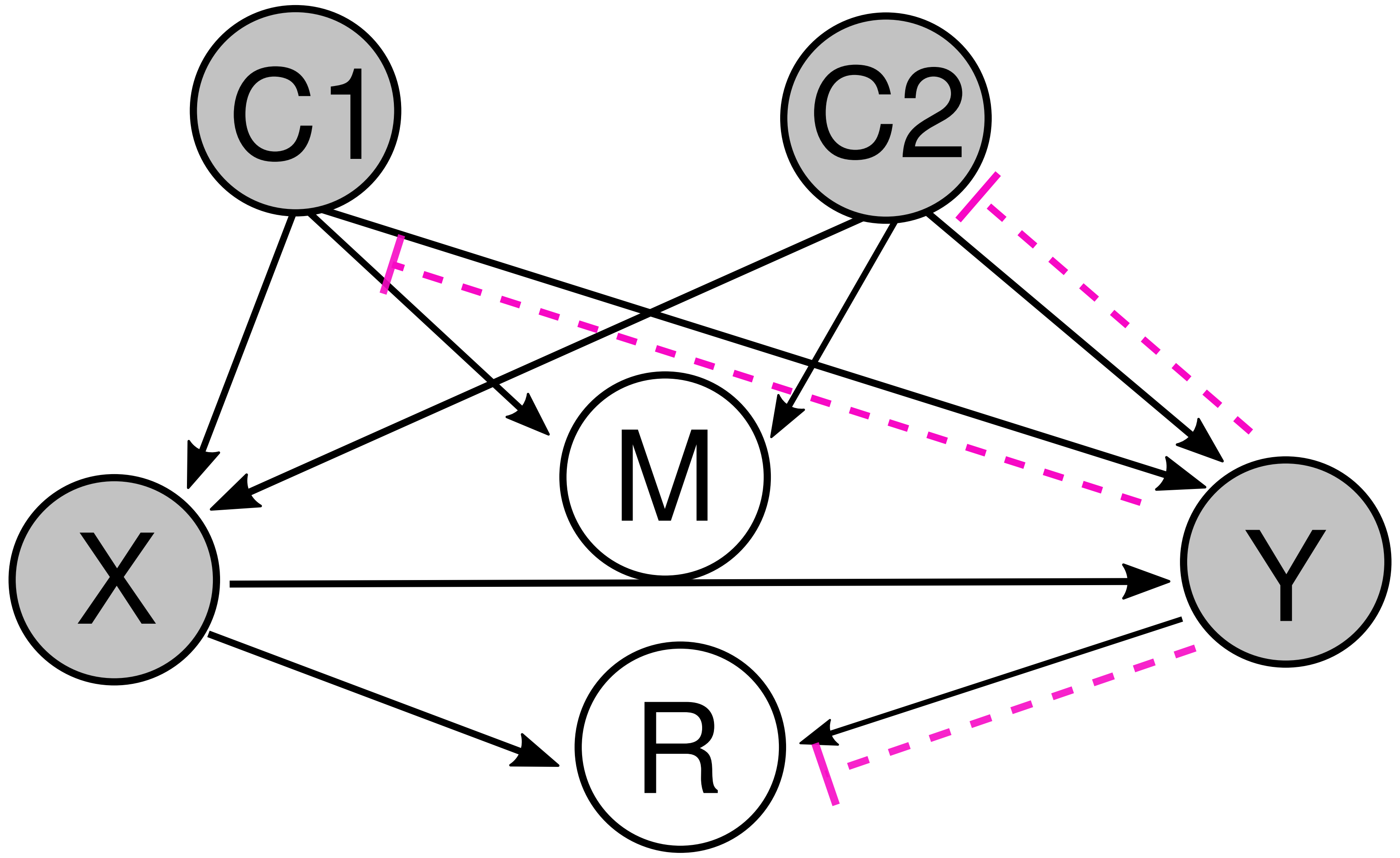


$$W_i \propto \begin{cases} 1/p(X_i = 1|C_i) & X_i = 1 \\ 1/p(X_i = 0|C_i) & X_i = 0 \end{cases}$$

Today’s lecture: Bayesian, PGM, Causality

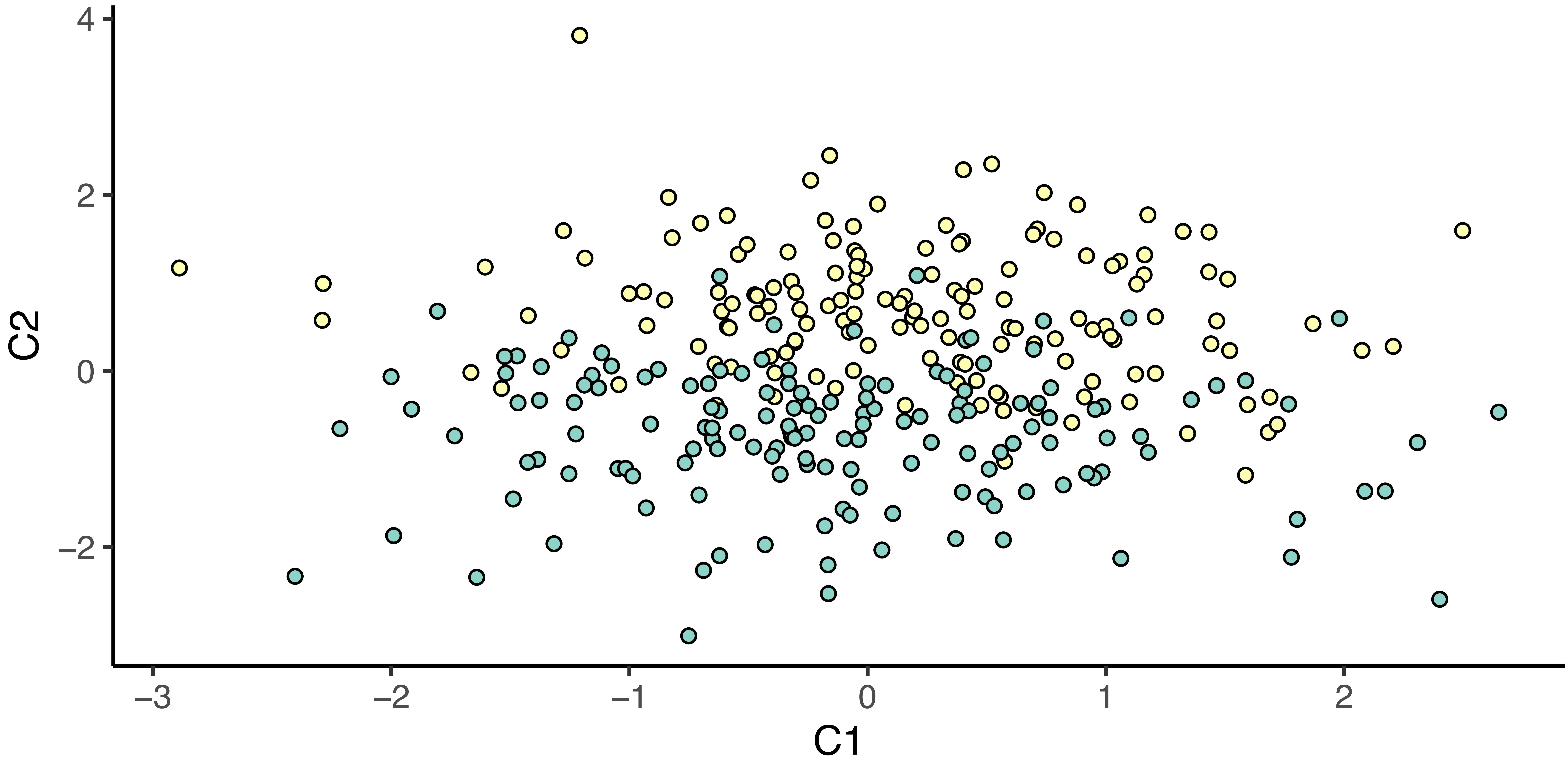
- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

Estimating potential outcome by matching

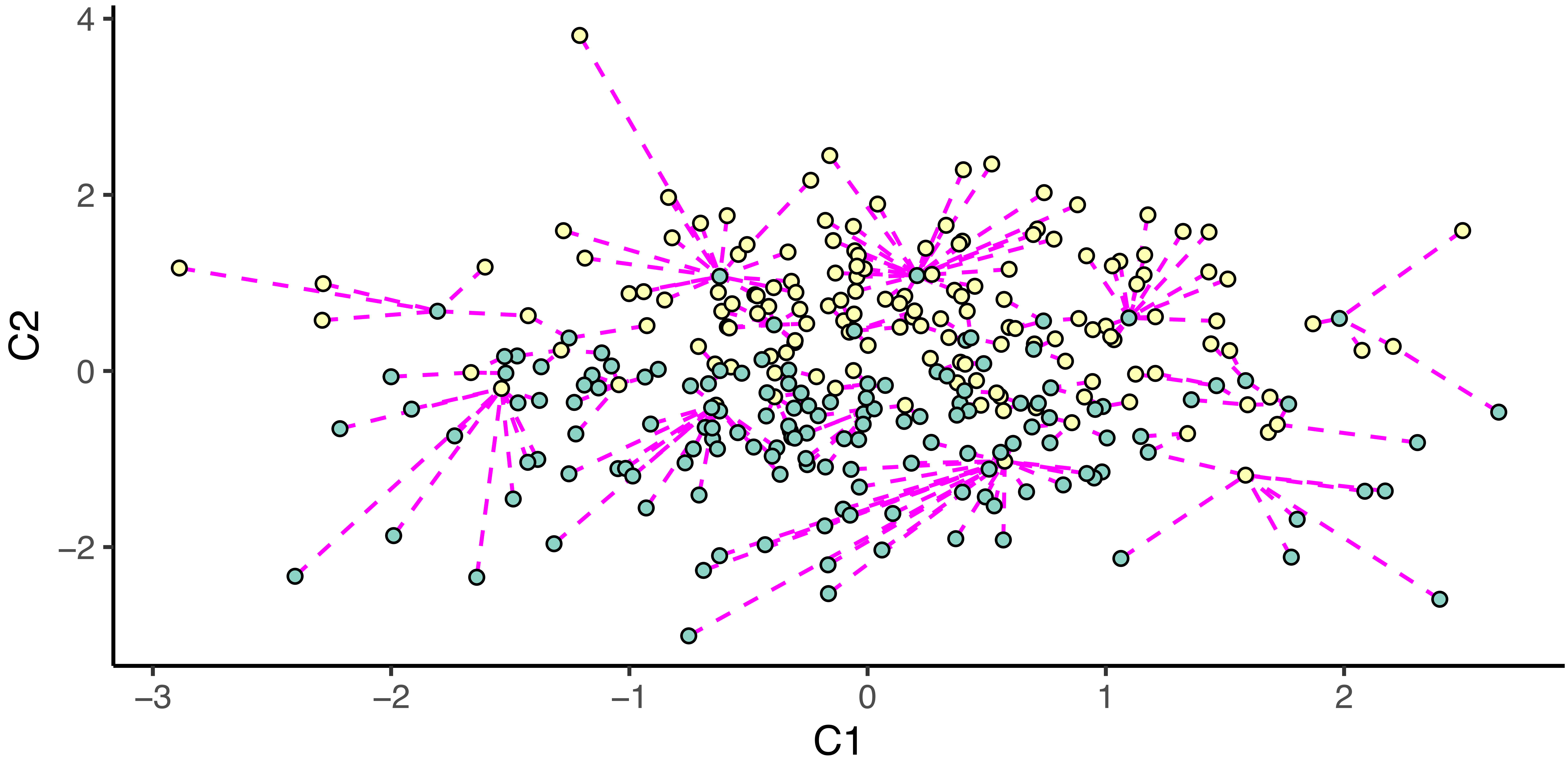


- Estimtae $\mathbb{E}\left[Y_i^{(0)} | C_{i1}, C_{i2}\right]$ for $X_i = 1$ to compare with $\mathbb{E}[Y_i | X_i = 1]$
- Estimtae $\mathbb{E}\left[Y_i^{(1)} | C_{i1}, C_{i2}\right]$ for $X_i = 0$ to compare with $\mathbb{E}[Y_i | X_i = 0]$

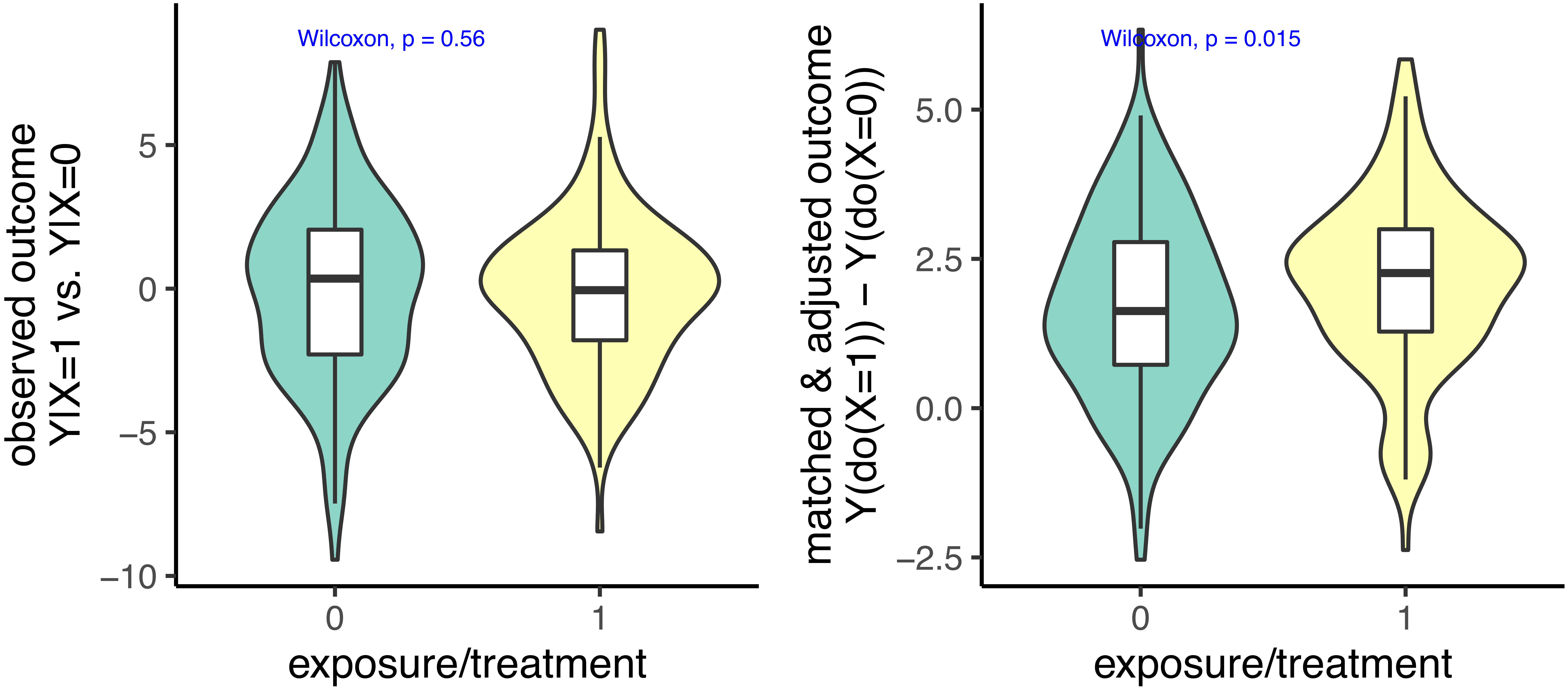
Estimating potential outcome by matching



Estimating potential outcome by matching



Estimating potential outcome by matching



Today’s lecture: Bayesian, PGM, Causality

- **Bayesian Inference**
  - Why is it worth knowing about Bayesian inference?
  - Graphical language in probabilistic modelling
  - Examples of (practical) Bayesian inference
- **Causal inference**
  - Observation vs. Experimentation
  - Identification of unwanted bias/variance
  - More general causal inference approaches

Bayesian Additive Regression Tree (BART) approach

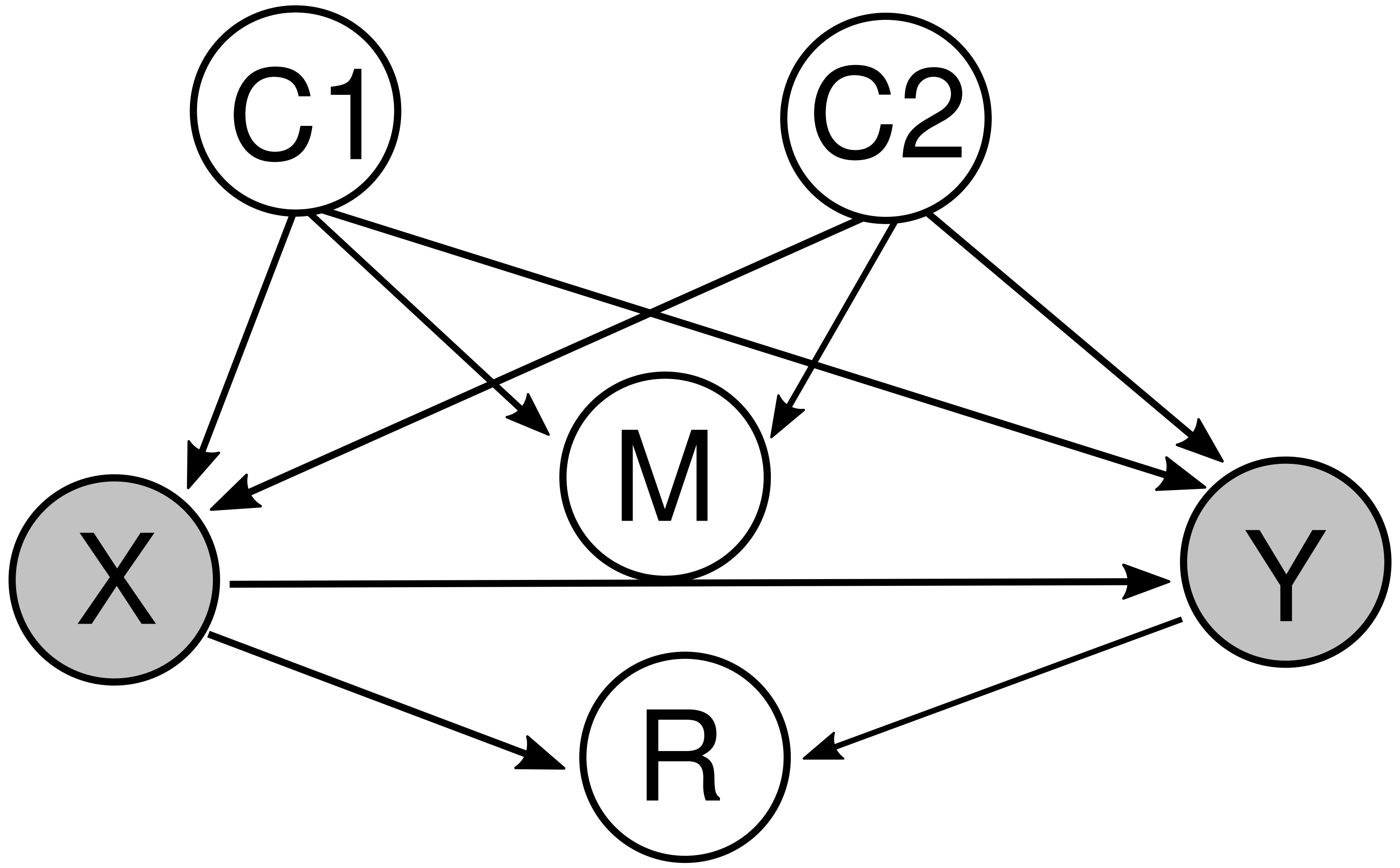
- G-formula $\rightarrow$ outcome regression models
- Regression model for $Y \sim S$ for each $X = 1$ and $X = 0$ using BART

$$\mathbb{E}\left[Y^{(1)}\right] = \int_S \mathbb{E}[Y|X = 1, S]dS$$
$$\mathbb{E}\left[Y^{(0)}\right] = \int_S \mathbb{E}[Y|X = 0, S]dS$$

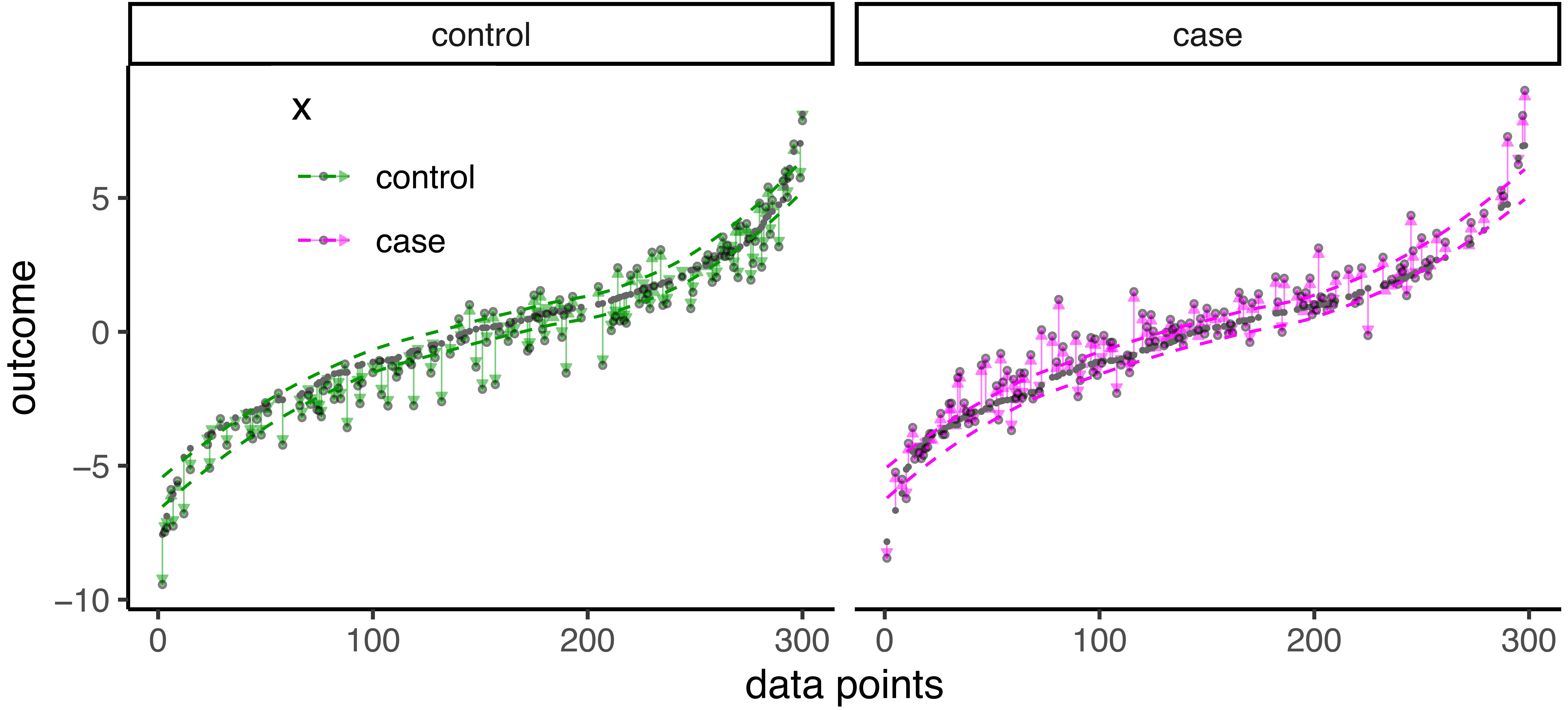
- Estimate causal effect: $\mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]$

Hill, *Bayesian Nonparametric Modeling for Causal Inference* (2011)

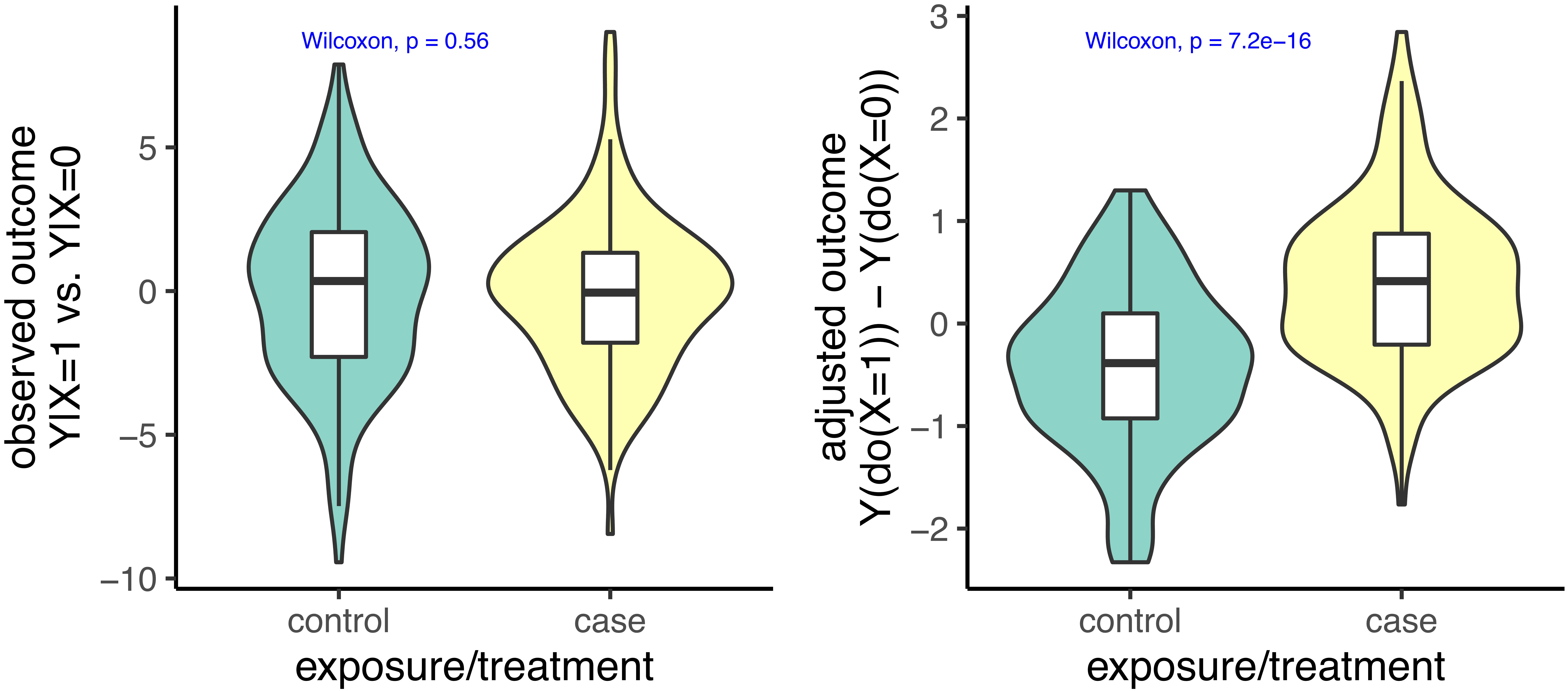
The same example with causal relationship from X to Y



BART: a regression model to impute potential outcomes



BART: a regression model to impute potential outcomes



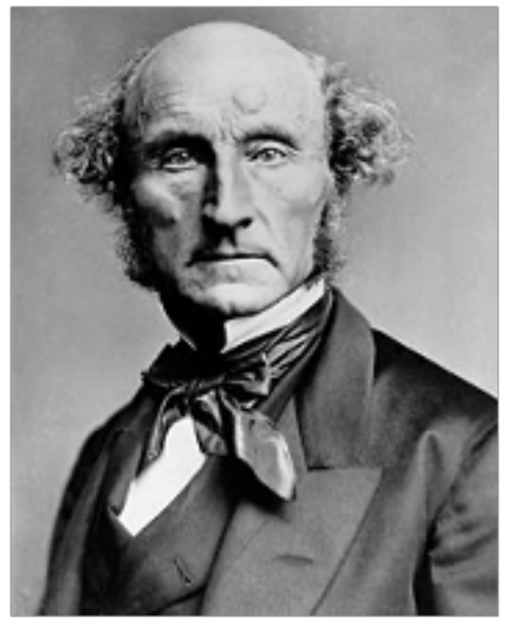
Today’s lecture: Bayesian, PGM, Causality

- **Bayesian Inference**
    - Why is it worth knowing about Bayesian inference?
    - Graphical language in probabilistic modelling
    - Examples of (practical) Bayesian inference
- **Causal inference**
    - Observation vs. Experimentation
    - Identification of unwanted bias/variance
    - More general causal inference approaches

How can we learn "true" scientific mechanisms from data collected in observational studies?

A causal model!

Munafo & Davey Smith, "Repeating Experiments is not enough" Nature (2018)



John Stuart Mill

SECOND CANON.

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.

JS Mill, A system of Logic (1843)

Peter Lipton

Contrastive Explanation & causal triangulation, Philosophy of Science (1991)

George Davey Smith