# BAYESIAN BEST PRACTICES

## BAYESIAN STATISTICS FOR ECOLOGISTS

IGB 18. TO 26. NOVEMBER 2019

# FOUR STEPS TO AN ANALYSIS

1. Specify the joint posterior distribution of outcomes (i.e., response variables) and all unknowns/parameters
2. Draw from posterior distribution using MCMC
3. Evaluate model and revise if necessary (return to step 1)
4. Use posterior predictive distribution for inference

http://mc-stan.org/rstanarm/articles/rstanarm.html

# 1. JOINT POSTERIOR—LIKELIHOOD

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood
    - whenever possible use knowledge of the appropriate processes

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood
    - whenever possible use knowledge of the appropriate processes
    - transform parameters and sample in transformed space to improve behaviour

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood
    - whenever possible use knowledge of the appropriate processes
    - transform parameters and sample in transformed space to improve behaviour
        - use log(σ) instead of σ to avoid impossible negative variance in sampler (Stan does this automagically)

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood
  - whenever possible use knowledge of the appropriate processes
  - transform parameters and sample in transformed space to improve behaviour
    - use log(σ) instead of σ to avoid impossible negative variance in sampler (Stan does this automagically)
- Models should be specified to be **scale independent**

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood
  - whenever possible use knowledge of the appropriate processes
  - transform parameters and sample in transformed space to improve behaviour
    - use log(σ) instead of σ to avoid impossible negative variance in sampler
      (Stan does this automagically)
- Models should be specified to be **scale independent**
  - This is most easily accomplished by scaling your predictors to have mean=0, sd=1

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood
    - whenever possible use knowledge of the appropriate processes
    - transform parameters and sample in transformed space to improve behaviour
        - use log(σ) instead of σ to avoid impossible negative variance in sampler (Stan does this automagically)
- Models should be specified to be **scale independent**
    - This is most easily accomplished by scaling your predictors to have mean=0, sd=1
    - Be careful scaling outcomes (y variables)—it can affect your generative model.

# 1. JOINT POSTERIOR—LIKELIHOOD

- Specify a **generative model** (the distribution from which the observations are generated) as the likelihood
  - whenever possible use knowledge of the appropriate processes
  - transform parameters and sample in transformed space to improve behaviour
    - use log($\sigma$) instead of $\sigma$ to avoid impossible negative variance in sampler (Stan does this automagically)
- Models should be specified to be **scale independent**
  - This is most easily accomplished by scaling your predictors to have mean=0, sd=1
  - Be careful scaling outcomes (y variables)–it can affect your generative model.
  - Prefer to “scale” outcomes via a **link function**

# 1. JOINT POSTERIOR—PRIOR

# 1. JOINT POSTERIOR—PRIOR

- All remaining unknowns specified with a **prior** (or possibly a hierarchical generative model and hyperpriors)

# 1. JOINT POSTERIOR—PRIOR

- All remaining unknowns specified with a **prior** (or possibly a hierarchical generative model and hyperpriors)
- Prefer regularising priors to vague priors

# 1. JOINT POSTERIOR—PRIOR

- All remaining unknowns specified with a **prior** (or possibly a hierarchical generative model and hyperpriors)
- Prefer regularising priors to vague priors
    - Normal(0, 5) instead of Normal(0,500)

# 1. JOINT POSTERIOR—PRIOR

- All remaining unknowns specified with a **prior** (or possibly a hierarchical generative model and hyperpriors)
- Prefer regularising priors to vague priors
  - Normal(0, 5) instead of Normal(0,500)
- Avoid improper priors: Uniform(-Inf, Inf)

# 1. JOINT POSTERIOR—PRIOR

- All remaining unknowns specified with a **prior** (or possibly a hierarchical generative model and hyperpriors)
- Prefer regularising priors to vague priors
    - Normal(0, 5) instead of Normal(0,500)
- Avoid improper priors: Uniform(-Inf, Inf)
- Forget conjugacy unless you know what you are doing and why

# 1. JOINT POSTERIOR—PRIOR

# 1. JOINT POSTERIOR—PRIOR

- Avoid hard boundaries

# 1. JOINT POSTERIOR—PRIOR

- Avoid hard boundaries
    - ~~Uniform(0, 1000)~~

# 1. JOINT POSTERIOR—PRIOR

- Avoid hard boundaries
    - ~~Uniform(0, 1000)~~
    - Exponential(0.1)

# 1. JOINT POSTERIOR—PRIOR

- Avoid hard boundaries
  - ~~Uniform(0, 1000)~~
  - Exponential(0.1)
- With informative priors, be sure to specify reasonable initial values

# 1. JOINT POSTERIOR—PRIOR

- Avoid hard boundaries
    - ~~Uniform(0, 1000)~~
    - Exponential(0.1)
- With informative priors, be sure to specify reasonable initial values
- Begin inference with weaker priors, gradually strengthen once you know the model is working

# 1. JOINT POSTERIOR—PRIOR

- Avoid hard boundaries
    - ~~Uniform(0, 1000)~~
    - Exponential(0.1)
- With informative priors, be sure to specify reasonable initial values
- Begin inference with weaker priors, gradually strengthen once you know the model is working
- Specify priors for everything–avoid defaults

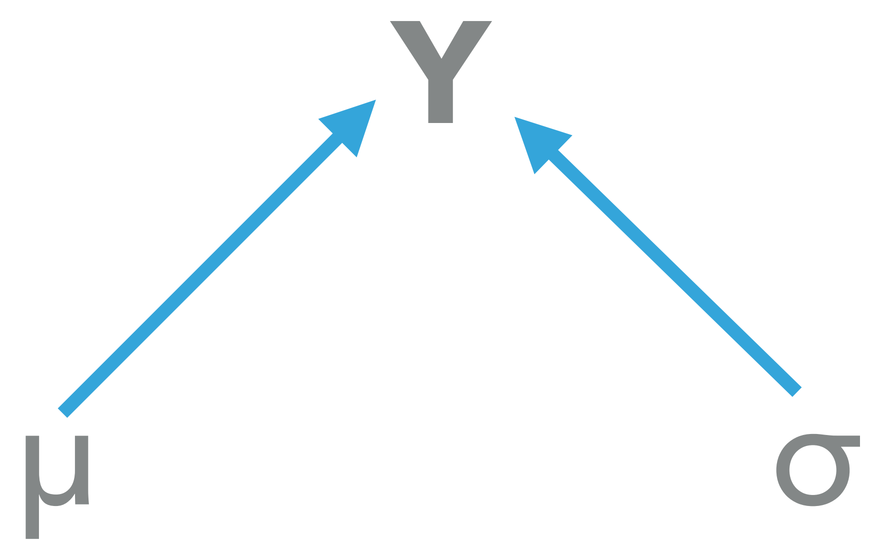
# 1. JOINT POSTERIOR—PRIOR

- Avoid hard boundaries
    - ~~Uniform(0, 1000)~~
    - Exponential(0.1)
- With informative priors, be sure to specify reasonable initial values
- Begin inference with weaker priors, gradually strengthen once you know the model is working
- Specify priors for everything–avoid defaults
- Can be useful to draw out your model as a **digraph** to make sure you don't miss anything

# DIGRAPH?

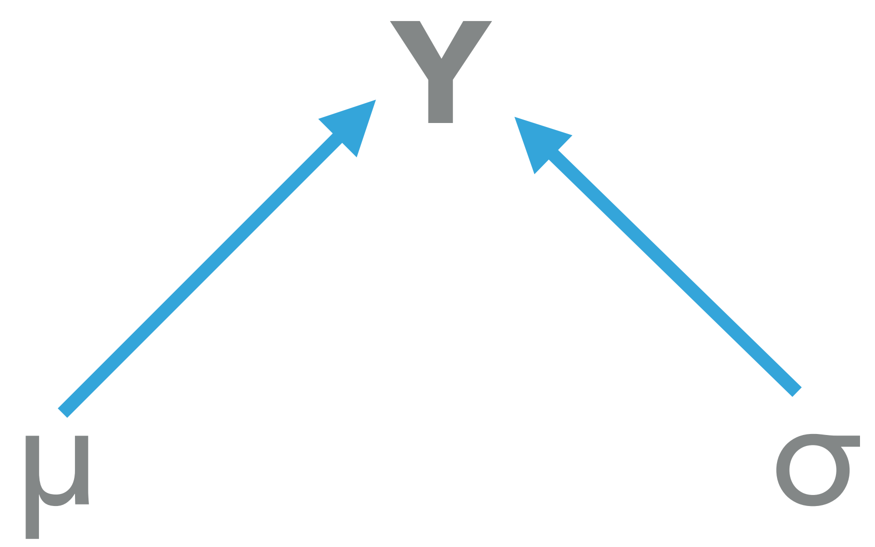
Y

# DIGRAPH?



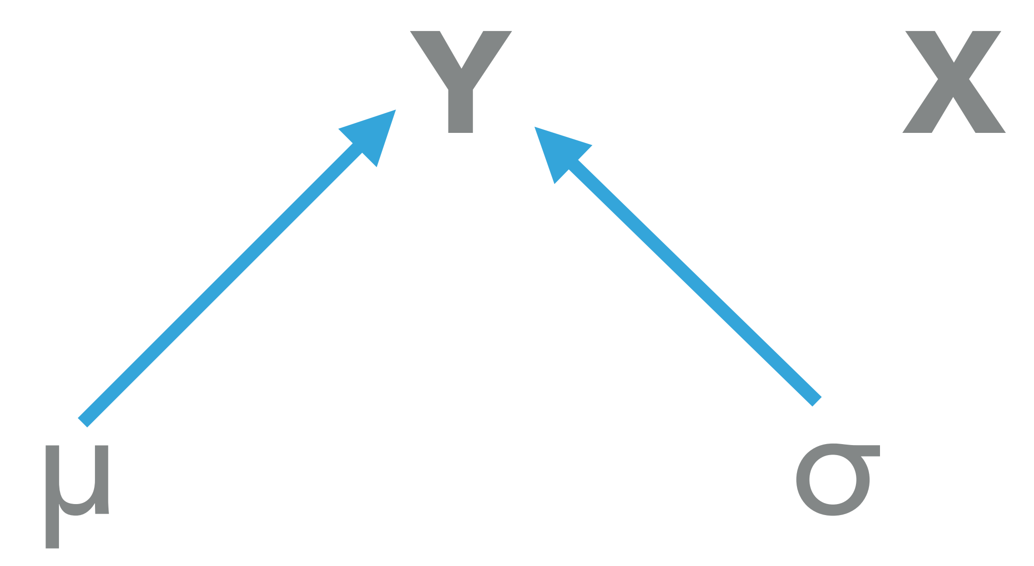
# DIGRAPH?

$Y \sim \text{Normal}(\mu, \sigma)$



# DIGRAPH?

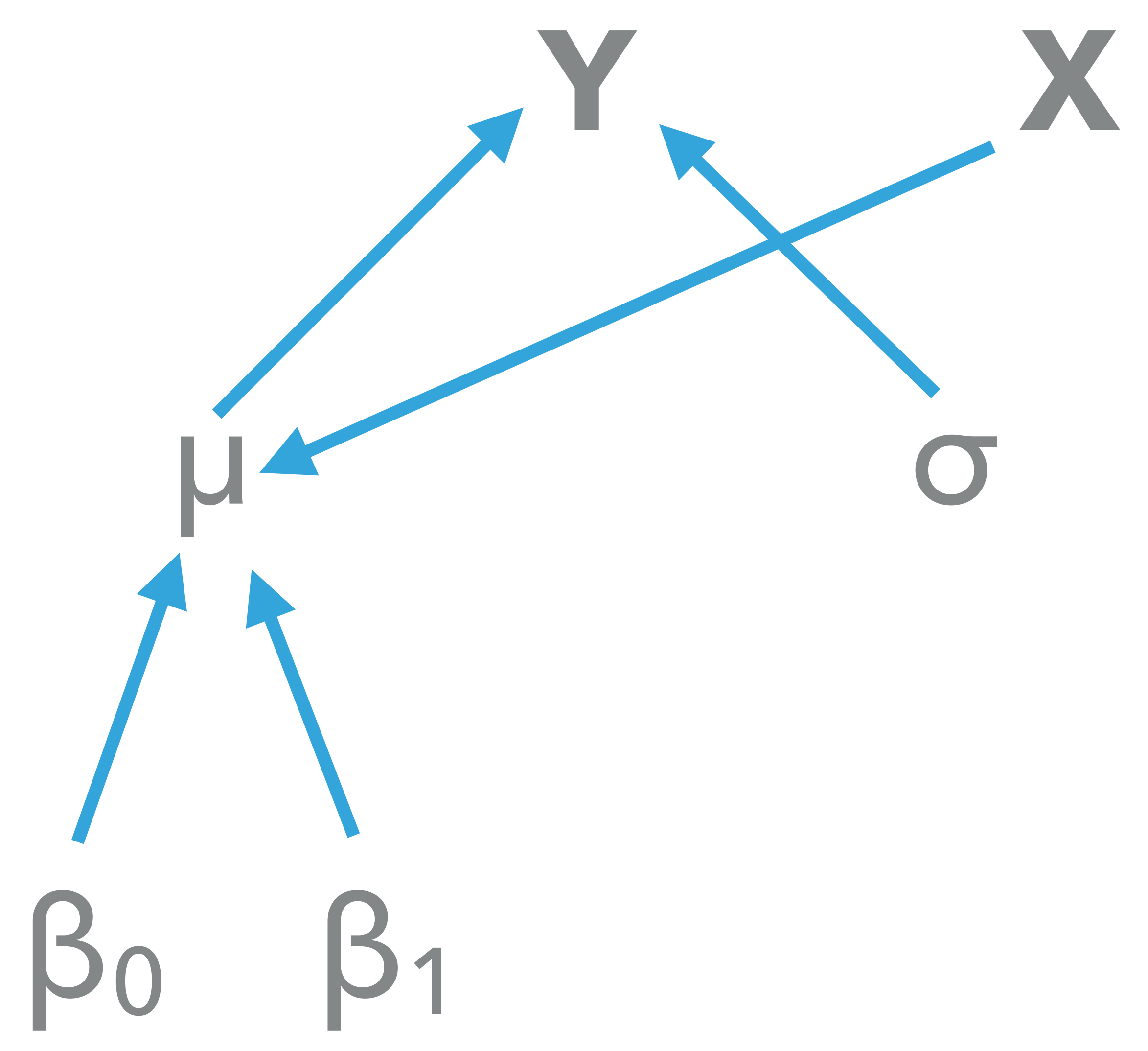
Y ~ Normal(μ, σ)



# DIGRAPH?

$Y \sim \text{Normal}(\mu, \sigma)$

$\mu = \beta_0 + \beta_1 X$



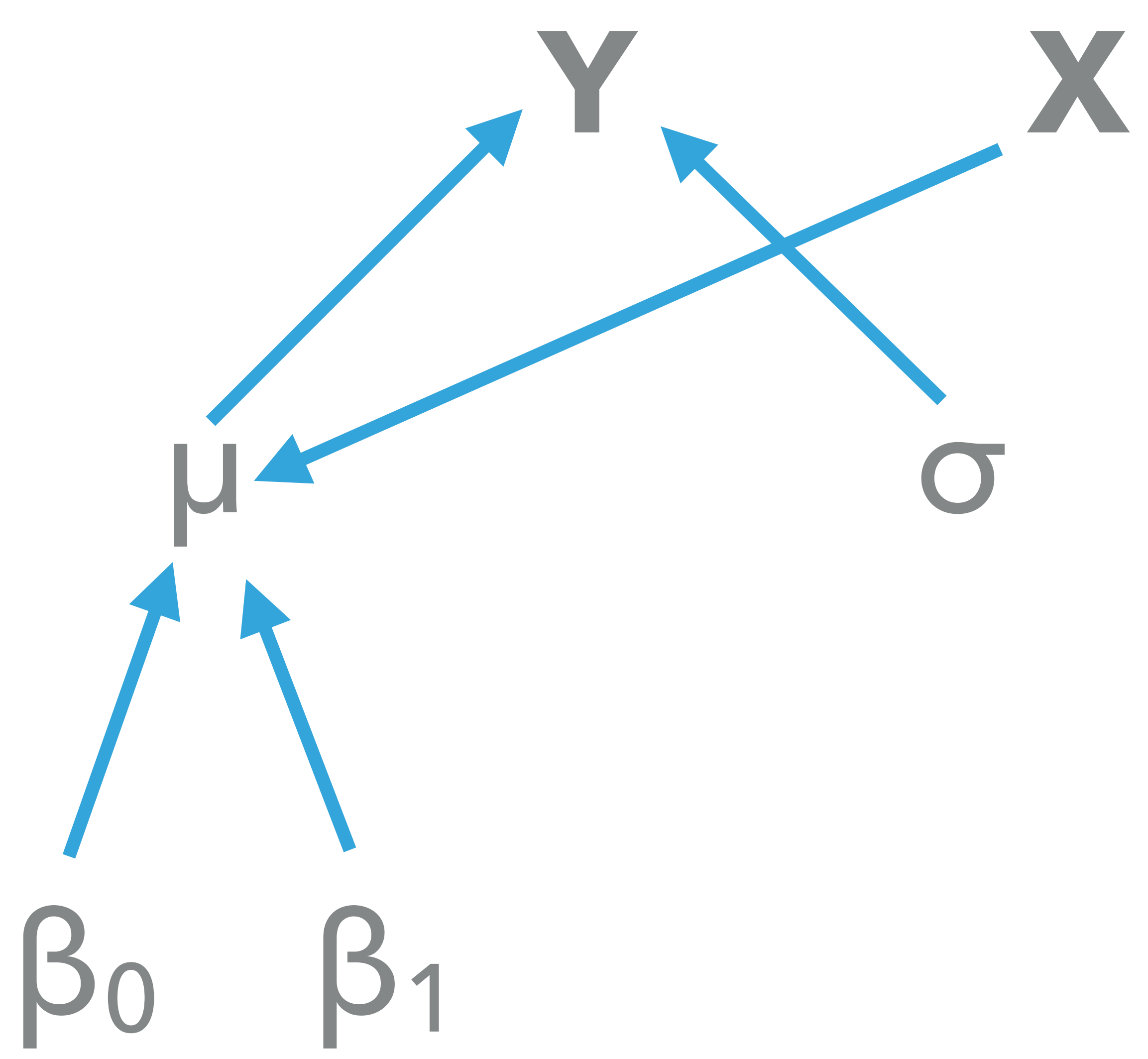
# DIGRAPH?

$Y \sim \text{Normal}(\mu, \sigma)$

$\mu = \beta_0 + \beta_1 X$

$\beta \sim \text{Normal}(0, 5)$

$\sigma \sim \text{Exponential}(0.1)$



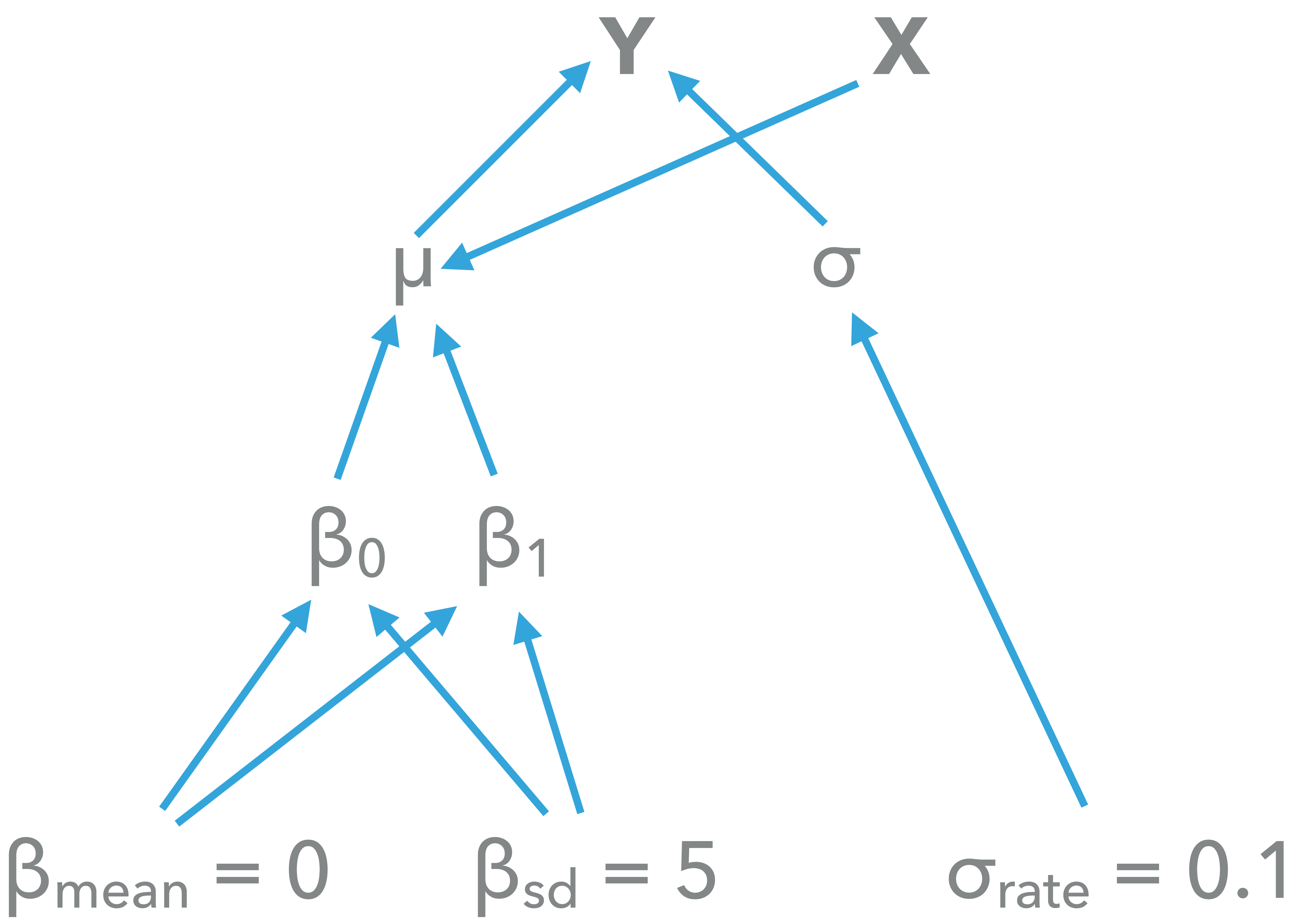
# DIGRAPH?

$Y \sim \text{Normal}(\mu, \sigma)$

$\mu = \beta_0 + \beta_1 X$

$\beta \sim \text{Normal}(0, 5)$

$\sigma \sim \text{Exponential}(0.1)$



# DIGRAPH?

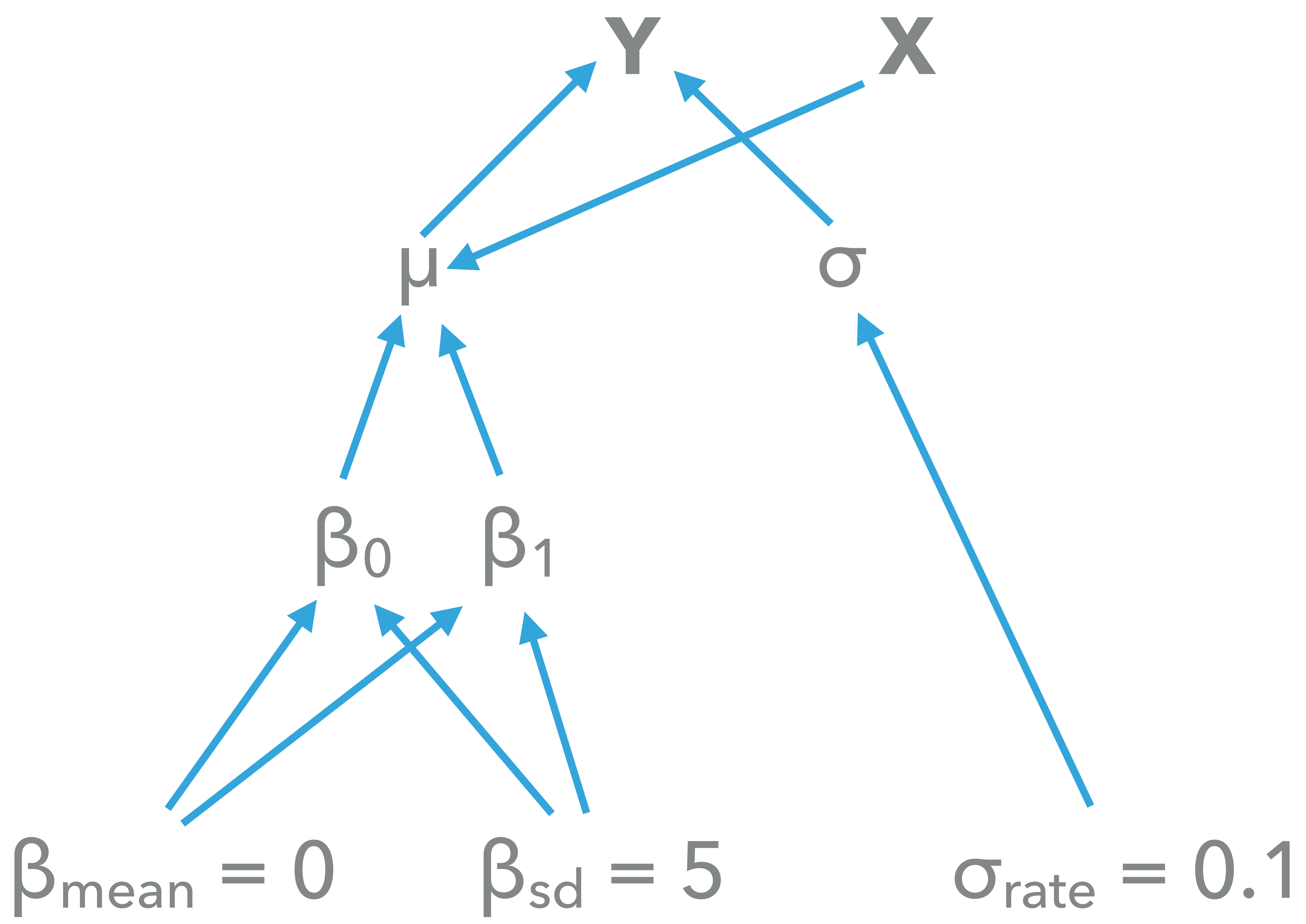
**Likelihood**

$Y \sim \text{Normal}(\mu, \sigma)$

$\mu = \beta_0 + \beta_1 X$

$\beta \sim \text{Normal}(0, 5)$

$\sigma \sim \text{Exponential}(0.1)$



# DIGRAPH?

**Likelihood**

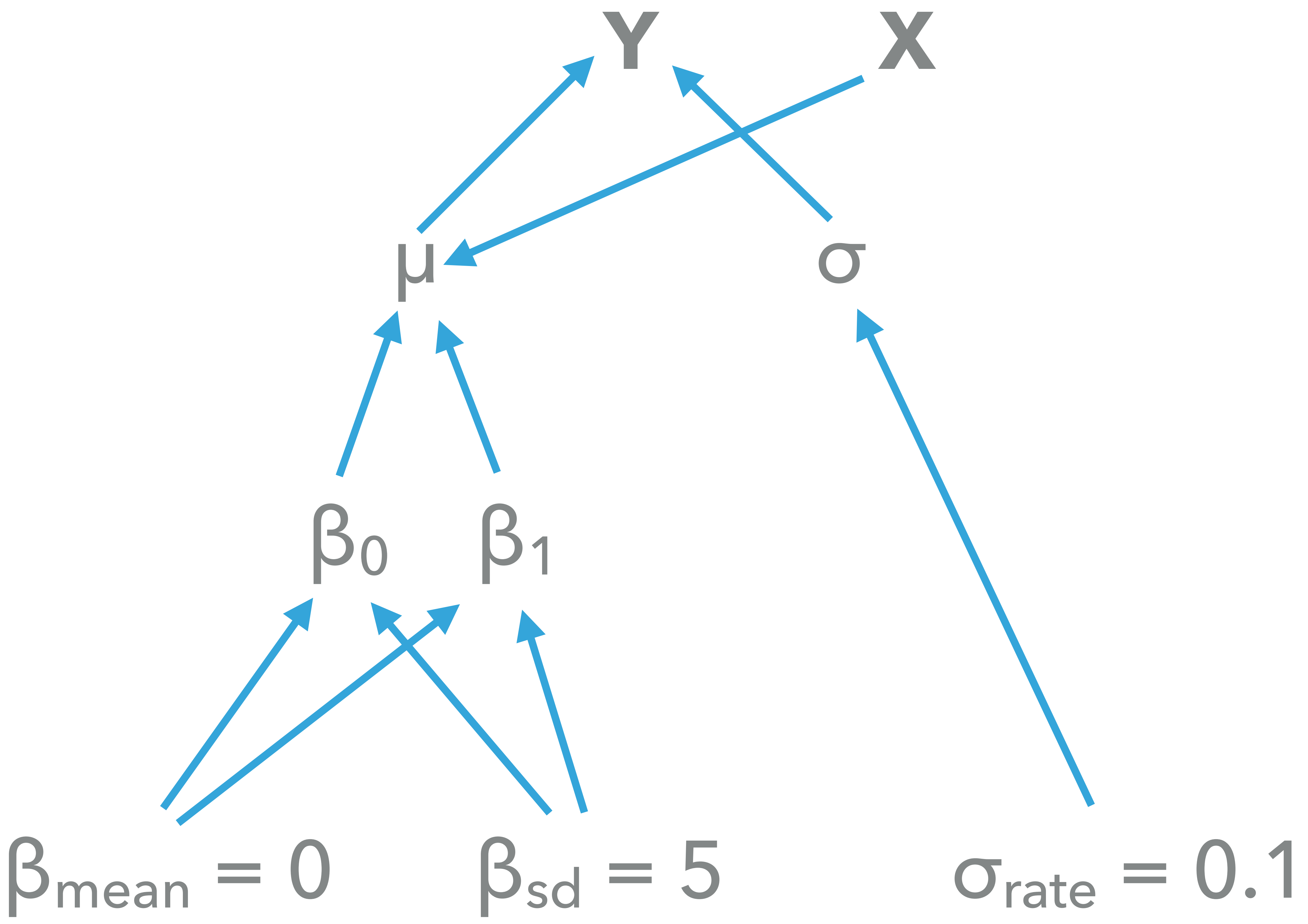
$Y \sim \text{Normal}(\mu, \sigma)$

$\mu = \beta_0 + \beta_1 X$

**Prior**

$\beta \sim \text{Normal}(0, 5)$

$\sigma \sim \text{Exponential}(0.1)$



# DIGRAPH?

**Likelihood**

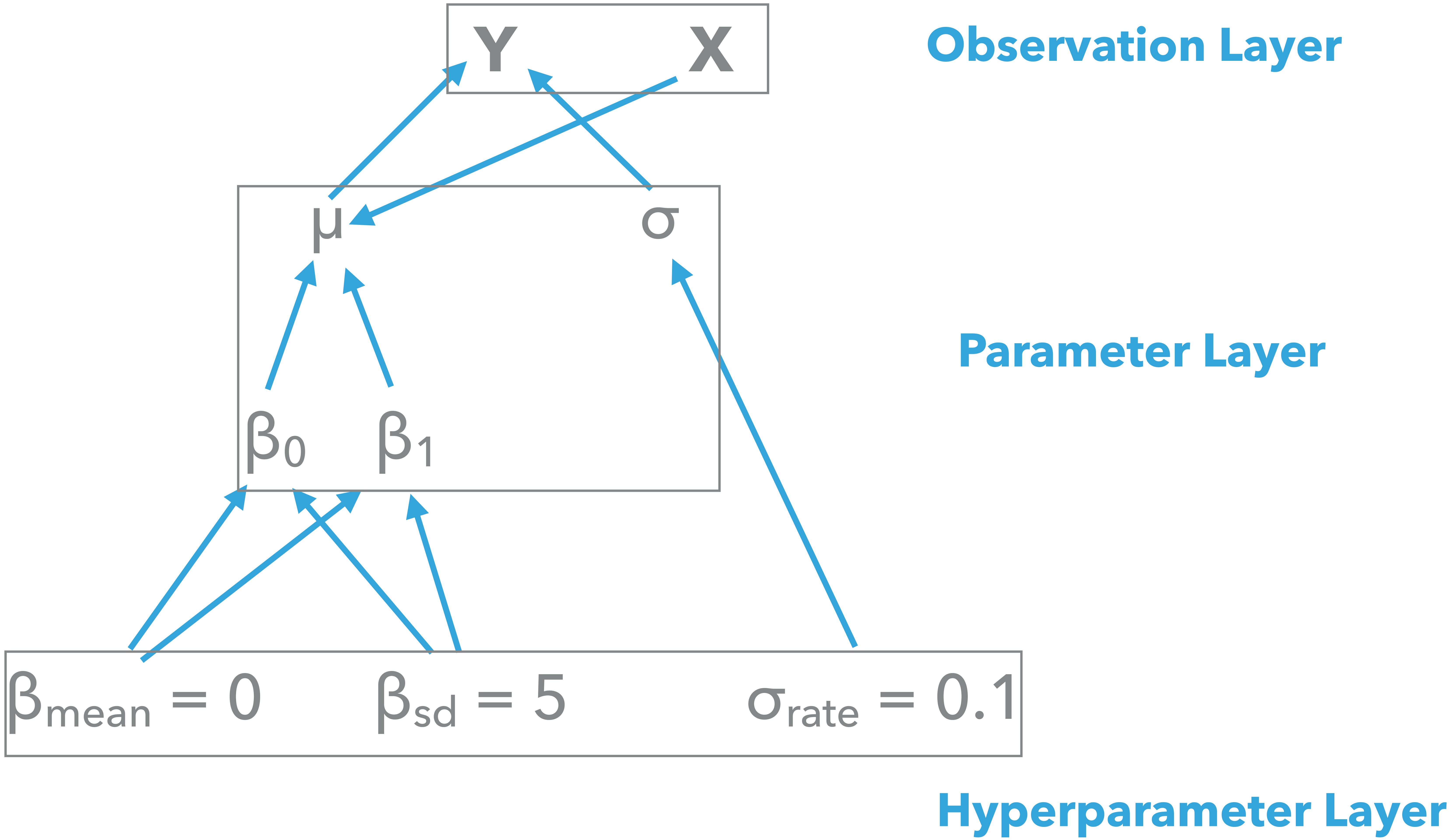
$Y \sim \text{Normal}(\mu, \sigma)$

$\mu = \beta_0 + \beta_1 X$

**Prior**

$\beta \sim \text{Normal}(0, 5)$

$\sigma \sim \text{Exponential}(0.1)$



# 2. SAMPLE FROM POSTERIOR

# 2. SAMPLE FROM POSTERIOR

- Start with a small run (10s-100s of iterations) to catch errors, make sure nothing surprising happens

# 2. SAMPLE FROM POSTERIOR

- Start with a small run (10s-100s of iterations) to catch errors, make sure nothing surprising happens
- Increase to a few thousand to view convergence

# 2. SAMPLE FROM POSTERIOR

- Start with a small run (10s-100s of iterations) to catch errors, make sure nothing surprising happens
- Increase to a few thousand to view convergence
- Select starting values and run until convergence (1.000s for Stan, 10.000s–100.000s or more for Metropolis)

# 3. EXAMINE MODEL—EFFICIENCY & CONVERGENCE

# 3. EXAMINE MODEL—EFFICIENCY & CONVERGENCE

- Positive-recurrent Markov chains have an asymptotic stationary distribution

# 3. EXAMINE MODEL—EFFICIENCY & CONVERGENCE

- Positive-recurrent Markov chains have an asymptotic stationary distribution
- Once a chain reaches this distribution, it will stay there

# 3. EXAMINE MODEL—EFFICIENCY & CONVERGENCE

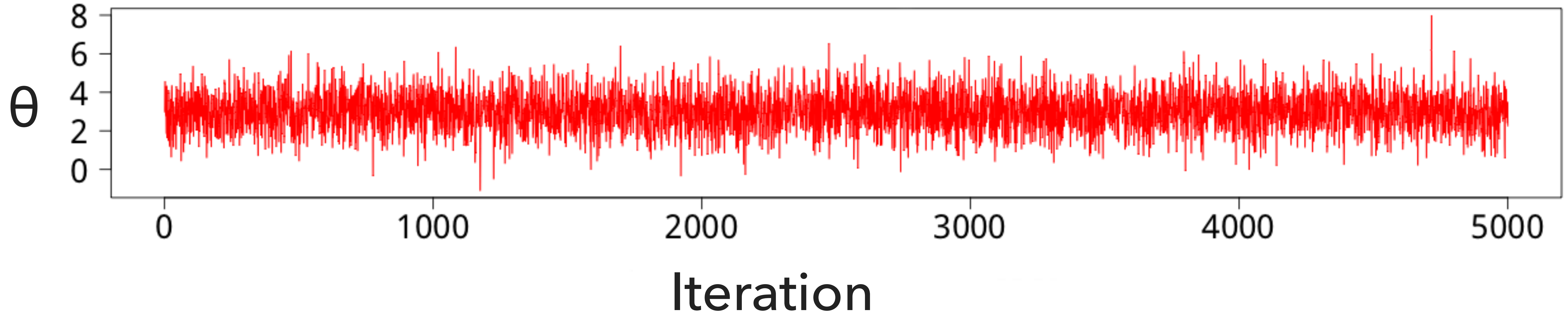
- Positive-recurrent Markov chains have an asymptotic stationary distribution
- Once a chain reaches this distribution, it will stay there
- For Bayesian MCMCs, this distribution approximates the **joint posterior**

# 3. EXAMINE MODEL—EFFICIENCY & CONVERGENCE

- Positive-recurrent Markov chains have an asymptotic stationary distribution
- Once a chain reaches this distribution, it will stay there
- For Bayesian MCMCs, this distribution approximates the **joint posterior**
- **Trace plots** are one way to visually examine this property

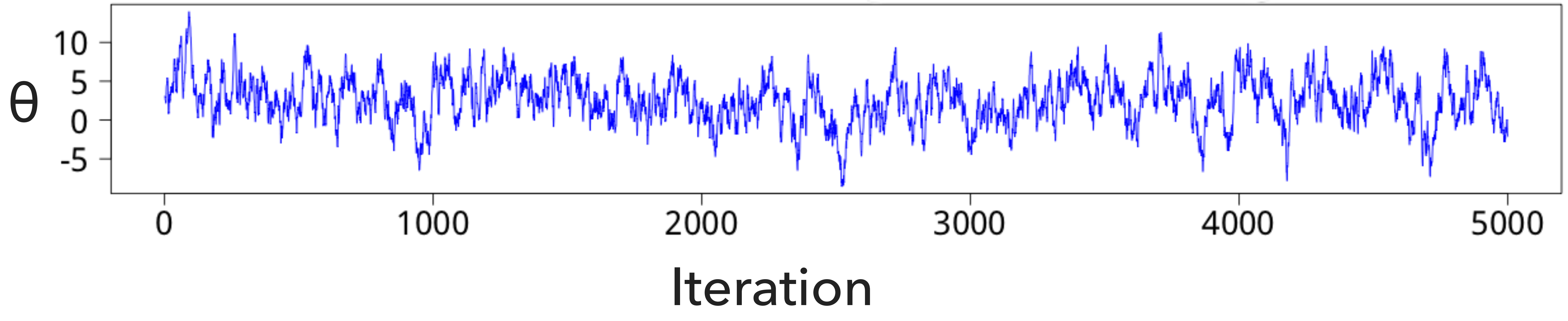
# 3. TRACE PLOTS

- Low autocorrelation
- Thorough coverage of range of parameter values



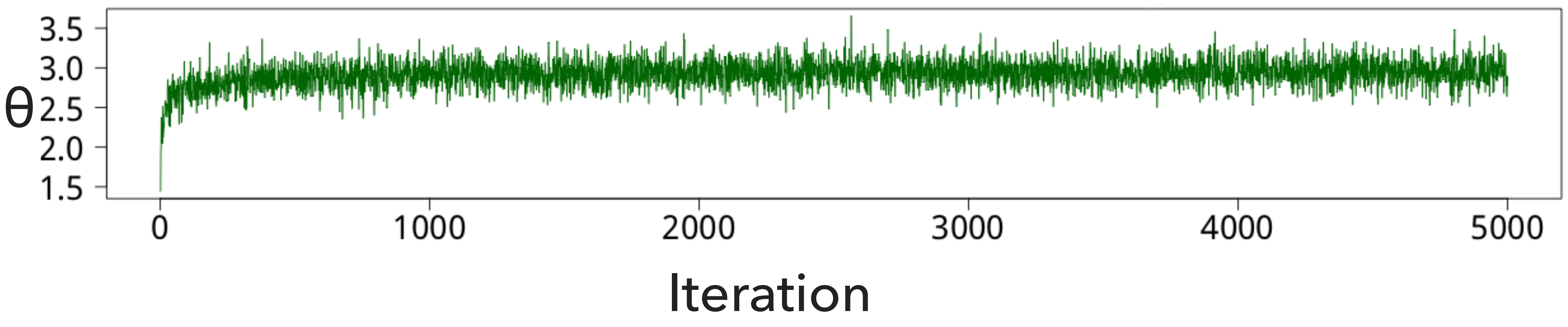
# 3. TRACE PLOTS

- High autocorrelation; can increase Metropolis step size
- Run longer
- Use thinning (not recommended)



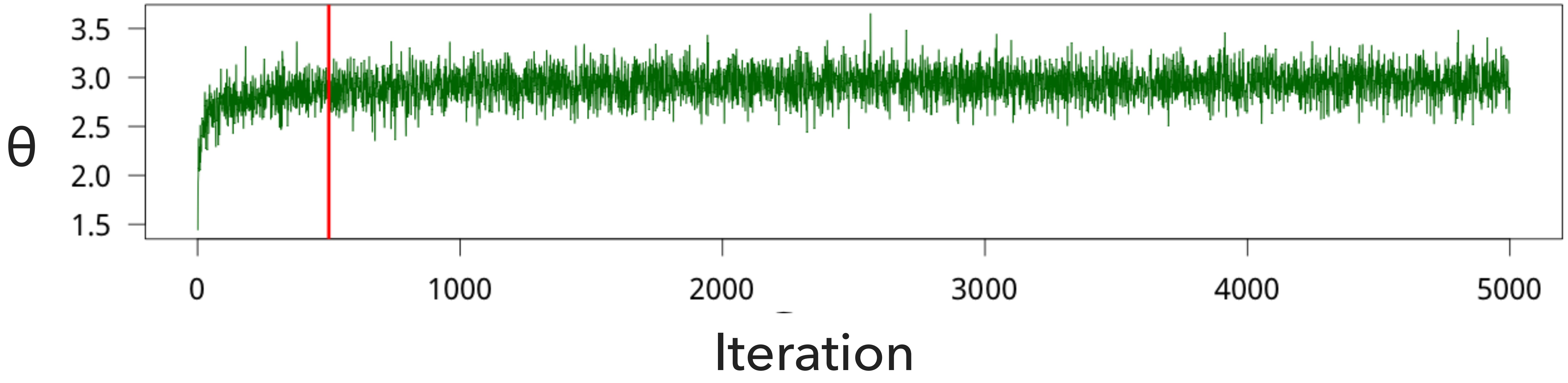
# 3. TRACE PLOTS

- Bad starting value
- Select new start
- Use burn-in (not the same as warm up!)



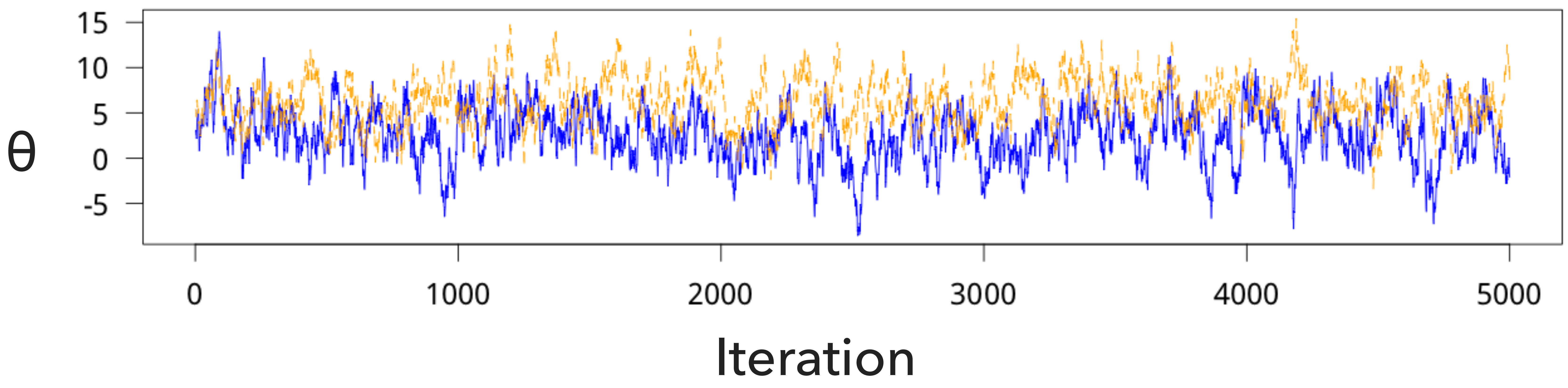
# 3. TRACE PLOTS

- A "burn-in" period selects a number of iterations to discard
- The idea is that the algorithm hasn't "forgotten" it's starting value
- Everything after burn-in approximates the stationary distribution



# 3. CONVERGENCE DIAGNOSTICS

- Using multiple chains allows us to compare within- and among-chain variance
- This is the **Gelman-Rubin statistic**; provided by Stan as r-hat
- Target value of 1.0. Less than 1.1 for all parameters is probably ok



# 3.5 EVALUATE AND COMPARE MODELS

# 3.5 EVALUATE AND COMPARE MODELS

- Before moving to inference, it is necessary to:

# 3.5 EVALUATE AND COMPARE MODELS

- Before moving to inference, it is necessary to:
    - select a model (or, to assign probabilities to competing models)

# 3.5 EVALUATE AND COMPARE MODELS

- Before moving to inference, it is necessary to:
  - select a model (or, to assign probabilities to competing models)
  - evaluate the fit of the model to data

# 3.5 EVALUATE AND COMPARE MODELS

- Before moving to inference, it is necessary to:
    - select a model (or, to assign probabilities to competing models)
    - evaluate the fit of the model to data
- Bayesian methods for model comparison range from fairly simple (DIC) to difficult (model averaging) to bewildering (RJ-MCMC) – will be covered later

# 3.5 EVALUATE AND COMPARE MODELS

- Before moving to inference, it is necessary to:
  - select a model (or, to assign probabilities to competing models)
  - evaluate the fit of the model to data
- Bayesian methods for model comparison range from fairly simple (DIC) to difficult (model averaging) to bewildering (RJ-MCMC) – will be covered later
- Model evaluation is a more complex topic – briefly on next week, in more detail if time allows

# 4. INFERENCE

# 4. INFERENCE

- There are multiple levels of inference

# 4. INFERENCE

- There are multiple levels of inference
- Examine posterior distributions of **parameters** (most informative for simple models)

# 4. INFERENCE

- There are multiple levels of inference
- Examine posterior distributions of **parameters** (most informative for simple models)
- **Posterior predictive distributions** – mean and standard error of outcomes

# 4. INFERENCE

- There are multiple levels of inference
- Examine posterior distributions of **parameters** (most informative for simple models)
- **Posterior predictive distributions** – mean and standard error of outcomes
- **Posterior predictive simulations** – generate new outcomes from model (incorporates all uncertainty, including process error)