Introduction to Bayesian Statistics

Fundamentals for applied research

Doug Leasure
Senior Researcher and Data Scientist
Demographic Science Unit
Dept. Population Health
University of Oxford

Learning outcomes

1. Understand the components of Bayes theorem:
   Priors, *likelihood*, *posteriors*
2. Familiarity with probability distributions and how to select among them
3. Experience coding Bayesian models and fitting them with MCMC using Stan software from R
4. Exposure to a few different types of models
   Linear regression, *generalised linear models*, *hierarchical models*
5. Practice with basic model diagnostics and cross-validation
6. Ability to develop and implement your own Bayesian models:
   Data model + process model + parameter model

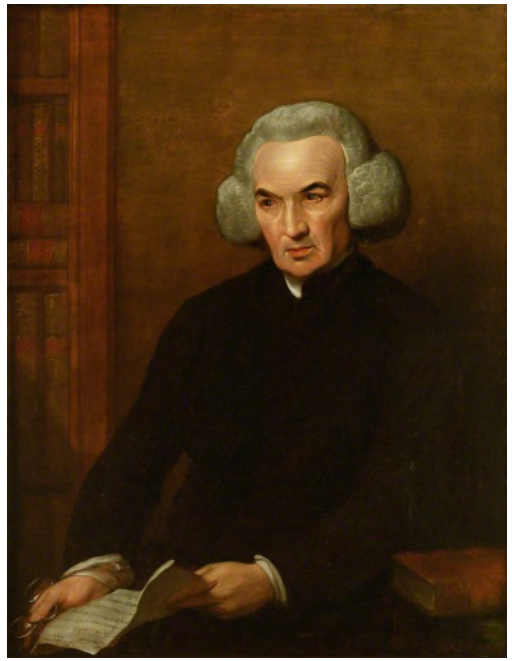
Reverend Thomas Bayes

c.1701-1761

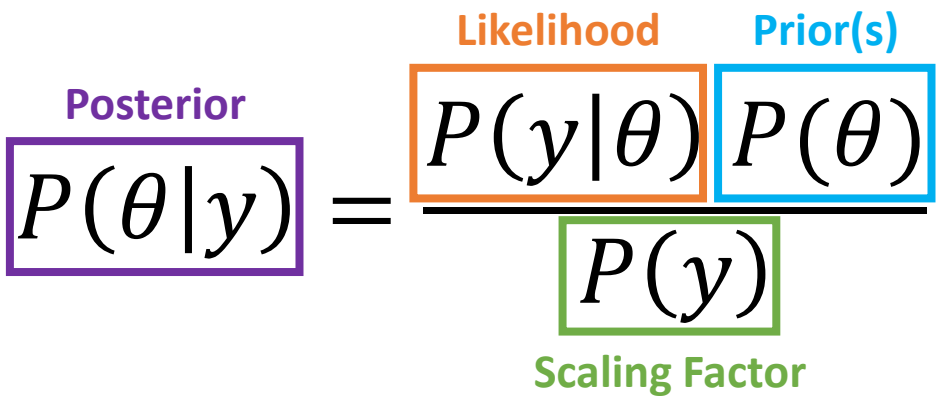
Bayes T. 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London.* 53:370-418.

Dr. Richard Price

1723-1791



Bayes' theorem



$y =$ data
$\theta =$ parameters (hypothesis)

Bayesian vs frequentist modes of inference

Frequentist

- How probable is this data set given the hypothesis?
- Likelihood $P(y|\theta)$
- Parameters fixed
- Data varies
- Confidence interval
- No prior information incorporated

Bayesian

- What is the probability of a hypothesis given the data?
- Posterior $P(\theta|y)$
- Parameters vary
- Data fixed
- Credible interval
- Incorporates prior information about parameters

What is a “deterministic” model?

$$N_{t+1} = N_t + B_t - D_t + I_t + E_t$$

The outcome on the left is always exactly the same for a given set of values on the right.

What is a “stochastic” model?

$$y_i = \alpha + \beta x_i + \epsilon_i$$

What does ϵ_i represent?

$$\epsilon_i \sim Normal(0, \sigma)$$

The outcome y_i has random variation for a given set of values for $\alpha, \beta, x_i,$ and σ.

A quick (but important) word on notation…

$$y_i = \alpha + \beta x_i + \epsilon_i$$

$$\epsilon_i \sim Normal(0, \sigma)$$

$$y_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta x_i$$

Bayes' theorem

Likelihood Prior(s)

Posterior

$$P(\theta|y) = \frac{P(y|\theta)\,P(\theta)}{P(y)}$$

$y =$ data
$\theta =$ parameters (hypothesis)

Likelihood

What is a likelihood, exactly?

$$P(y \mid \mu, \sigma)$$

$$y \sim Normal(\mu, \sigma)$$

Probability density function (pdf) for a normal distribution

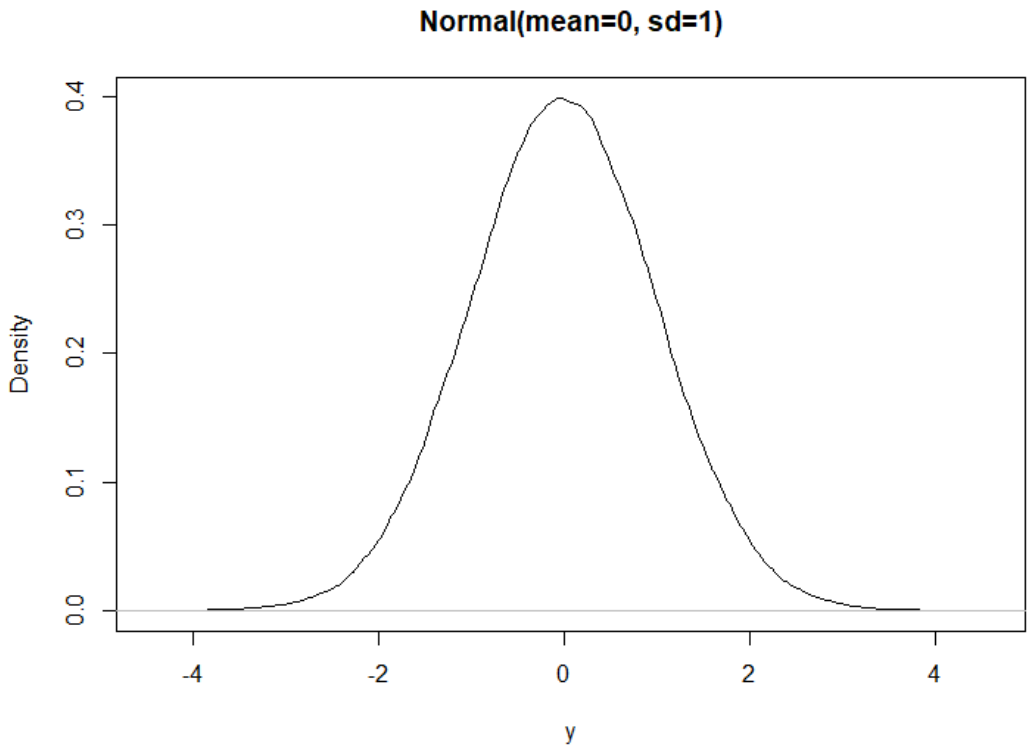
$$P(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)$$

What is a likelihood, exactly?

$$P(y \mid \mu = 0, \sigma = 1)$$

$$y \sim Normal(\mu = 0, \sigma = 1)$$

$$P(y \mid 0, 1) = \frac{1}{\sqrt{2\pi}1} \exp\left(-\frac{1}{2}\left(\frac{y - 0}{1}\right)^2\right)$$



What is a likelihood, exactly?

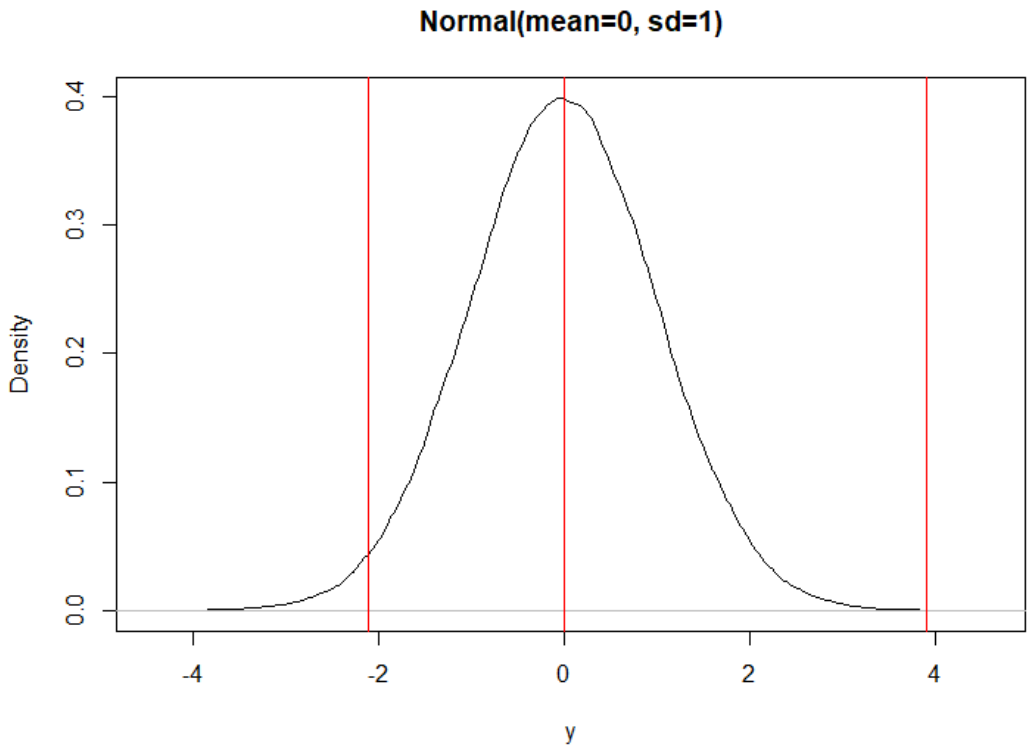
$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)$$

$$0.044 = P(y = -2.1 \mid \mu = 0, \sigma = 1)$$

$$0.399 = P(y = 0 \mid \mu = 0, \sigma = 1)$$

$$1.99e^{-4} = P(y = 3.9 \mid \mu = 0, \sigma = 1)$$

$$1.63e^{-34} = P(y = -12.4 \mid \mu = 0, \sigma = 1)$$



Note: These probability densities can be easily calculated in R…

$$dnorm(x = -12.4, mean = 0, sd = 1)$$

How do we calculate the likelihood for a full data set (rather than one data point)?

$$y = [-2.1, 0, 3.9, -12.4]$$

$$L(y \mid \mu, \sigma) = \prod_{i=1}^{I} P(y_i|0,1)$$

Log-likelihood

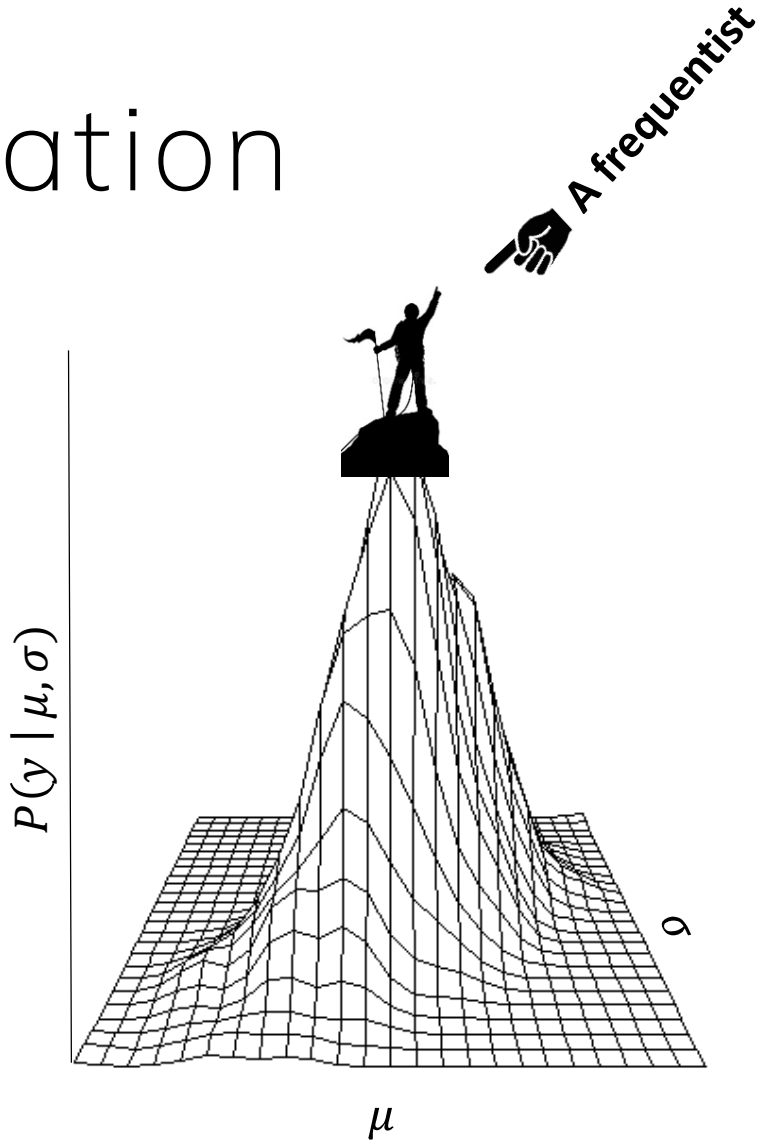
$$\sum_{i=1}^{I} \log\left(P(y_i|0,1)\right)$$

$$L(y \mid \mu, \sigma) = 0.044 \times 0.399 \times 1.99e^{-4} \times 1.63e^{-34} = 5.68e^{-40}$$

Maximum likelihood estimation

Repeat the process for many combinations of μ and σ.

Search parameter values until you find the maximum likelihood for your data set.



Bayes' theorem

Do we just plug in the maximum likelihood (i.e. a single number)?

$$\underbrace{P(\mu,\sigma|y)}_{\text{Posterior}} = \frac{\overbrace{P(y|\mu,\sigma)}^{\text{Likelihood}}\ \overbrace{P(\mu)P(\sigma)}^{\text{Priors}}}{P(y)}$$

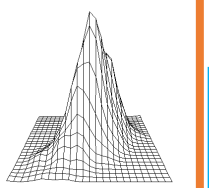
$y = \text{data}$

$\mu = \text{mean}$

$\sigma = \text{standard deviation}$

Bayes' theorem

Every term in Bayes theorem is a probability distribution.

$$\underbrace{P(\mu,\sigma|y)}_{\text{Posterior}} = \frac{\overbrace{\text{  }}^{\text{Likelihood}} \quad \overbrace{P(\mu)P(\sigma)}^{\text{Priors}}}{P(y)}$$

$y = \text{data}$

$\mu = \text{mean}$

$\sigma = \text{standard deviation}$

What does this likelihood tell us about our data and our hypothesis?

$$y_i \sim Normal(\mu, \sigma)$$

- ✓y is a continuous number.
- ✓It can be positive and negative.
- ✓We have I observations of y.
- ✓The most likely value of y is μ.
- ✓It is a symmetrical distribution, i.e. $mean(y) = median(y)$.
- ✓Variation around the mean is “bell-shaped” with a width defined by σ.
- ✓μ and σ are not dependent on each other or any other parameters or data.

Probability distributions

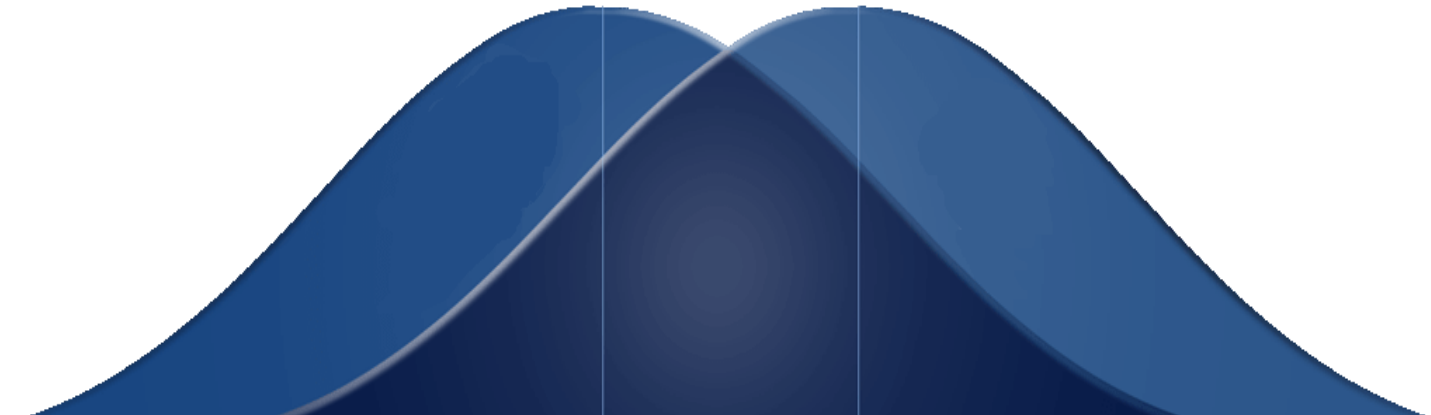
When designing likelihoods for data (i.e. dependent variable), we need to:

1. Choose an appropriate probability distribution for the data characteristics
2. Design a model for its parameters that reflects our hypothesis about the process that generated the data

Next up….

Probability distributions

A quick tour of common probability distributions…



Key considerations

- Continuous versus discrete
- Range of support
- Interpretability of parameters

Normal distribution

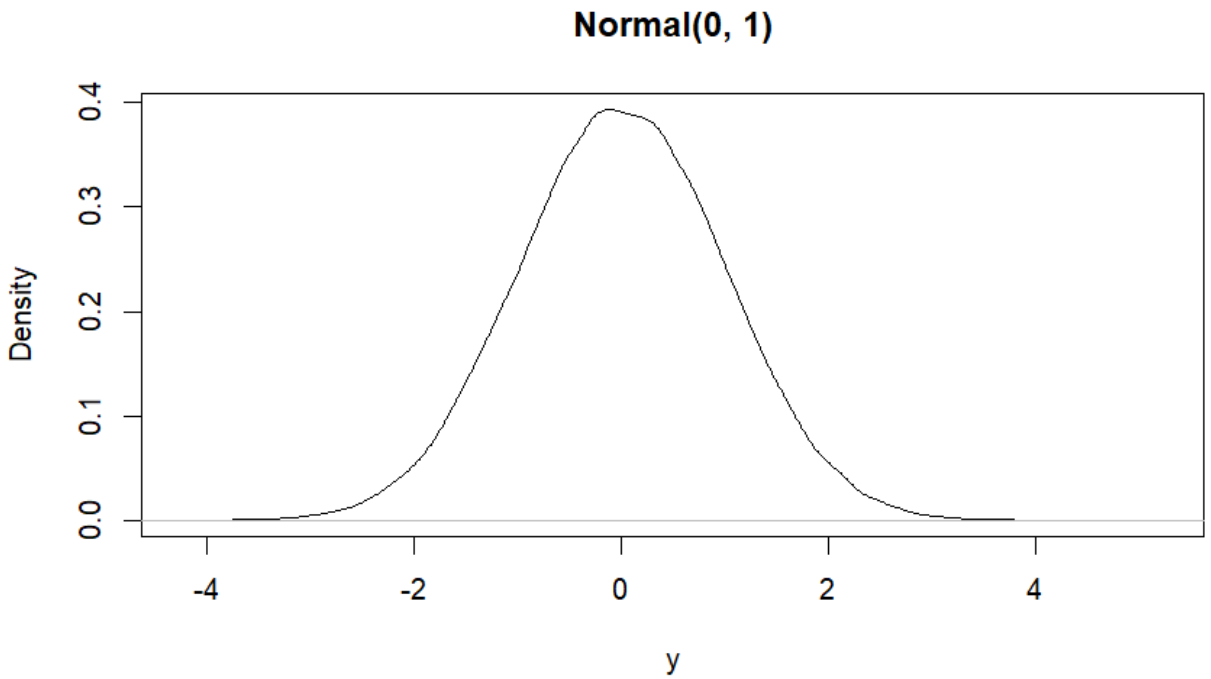
Data type: Continuous

Range of support: $[-\infty, \infty]$

Parameters:

$\mu\ (mean)$ $real[-\infty, \infty]$

$\sigma\ (sd)$ $real[0, \infty]$



Probability density function:

$$\text{Normal}(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)$$

```
y <- rnorm(n=1e5, mean=0, sd=1)
plot(density(y), main='Normal(0, 1)', xlab='y')
```

Uniform distribution

Data type: Continuous

Range of support: $[-\infty, \infty]$

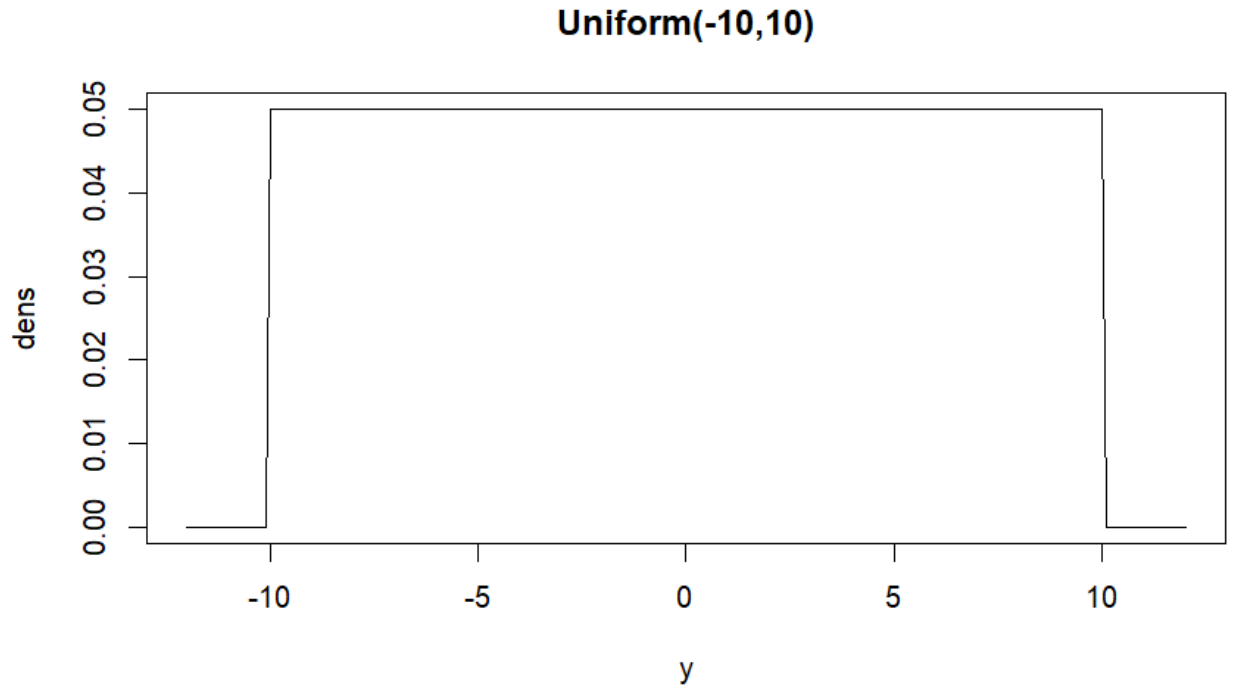
Parameters:

$\alpha\ (min)\ \ real[-\infty, \beta]$

$\beta\ (max)\ \ real[\alpha, \infty]$

Probability density function:

$$\text{Uniform}(y|\alpha, \beta) = \frac{1}{\beta - \alpha}$$




```
x <- seq(-12, 12, by=0.1)
dens <- dunif(x, min=-10, max=10)
plot(y=dens, x=x, main='Uniform(-10,10)', xlab='y', type='l')
```

LogNormal distribution

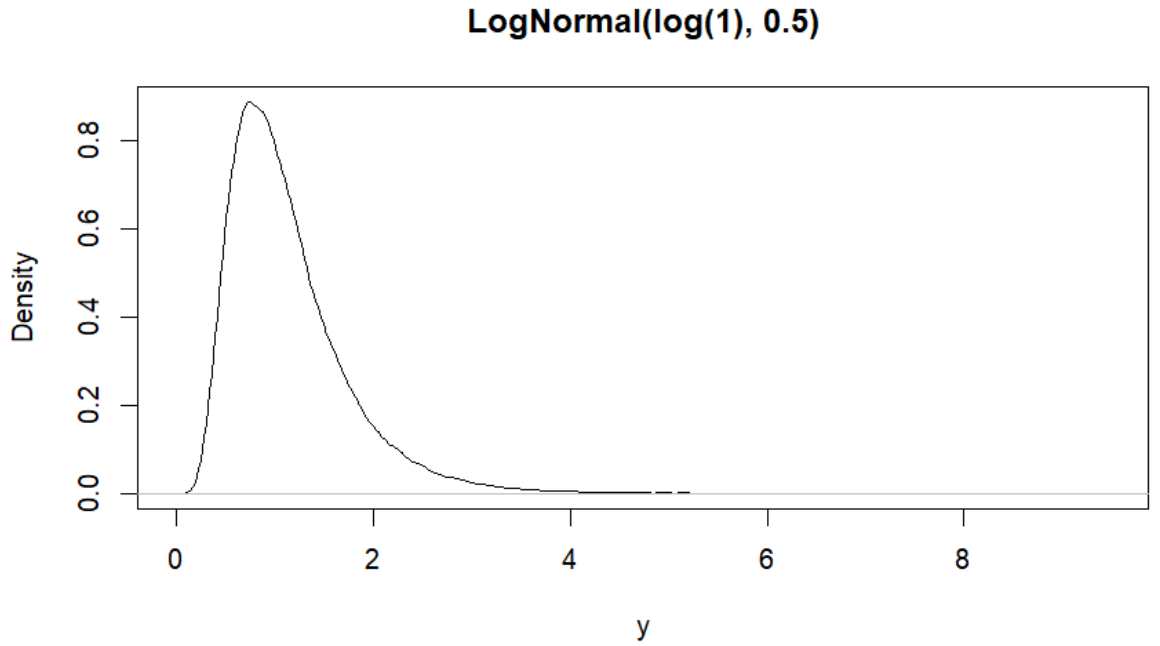
Data type: Continuous

Range of support: $[0, \infty]$

Parameters:

$\mu\ (meanlog)$ $\quad real[-\infty, \infty]$

$\sigma\ (sdlog)$ $\quad real[0, \infty]$



Probability density function:

$$\text{LogNormal}(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\,\frac{1}{y}\,\exp\left(-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2\right)$$

```
y <- rlnorm(n=1e5, meanlog=log(1), sd=0.5)
plot(density(y), main='LogNormal(log(1), 0.5)', xlab='y')
```

Gamma distribution

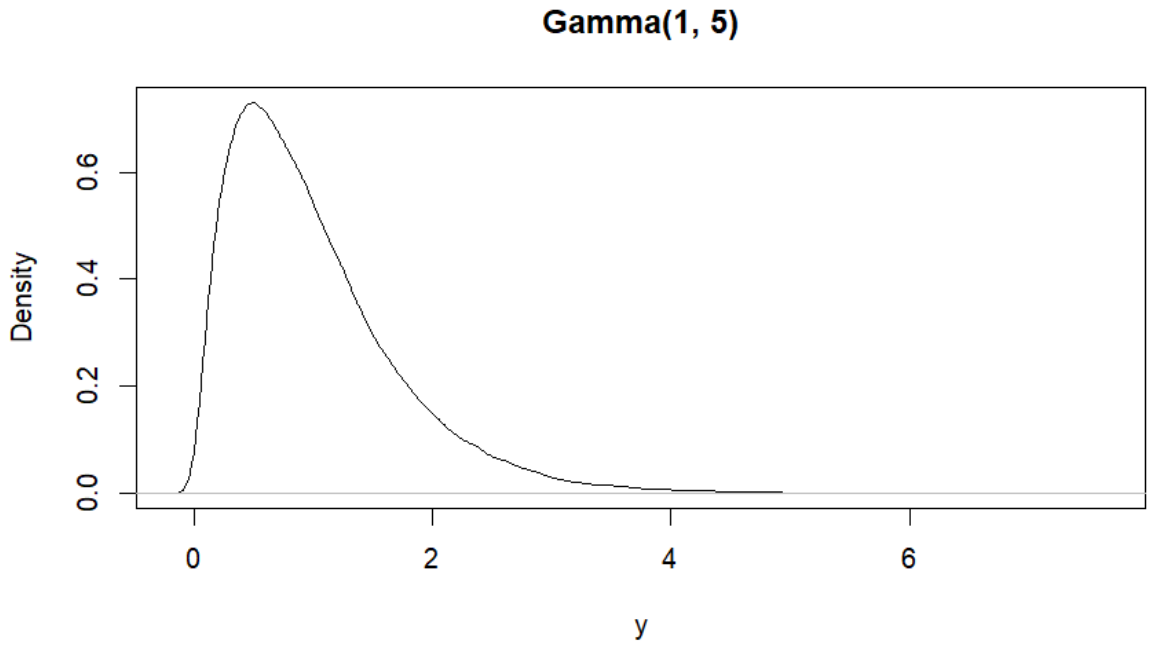
Data type: Continuous

Range of support: $[0, \infty]$

Parameters:

α $(shape)$ $real[0, \infty]$

β $(rate)$ $real[0, \infty]$



Probability density function:

$$\mathrm{Gamma}(y|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\, y^{\alpha-1} \exp(-\beta\, y)$$

```
y <- rgamma(n=1e5, shape=2, rate=2)
plot(density(y), main='Gamma(1, 5)', xlab='y')
```

Beta distribution

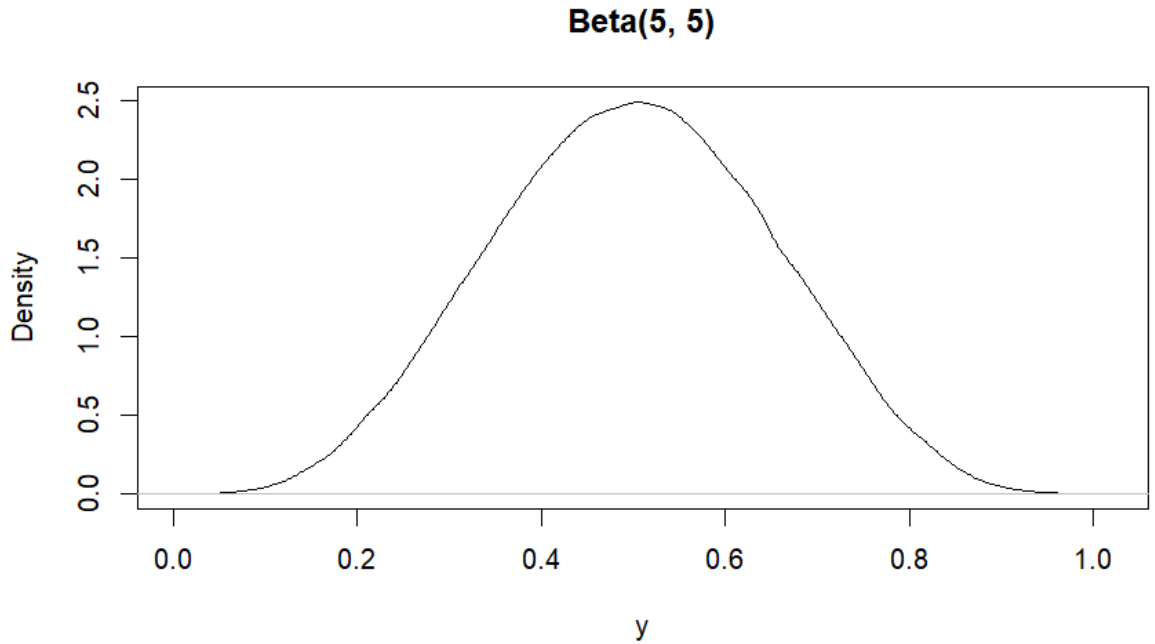
Data type: Continuous

Range of support: $[0, 1]$

Parameters:
$\alpha\ (shape1)\ real[0, \infty]$
$\beta\ (shape2)\ real[0, \infty]$

Probability density function:

$$\mathrm{Beta}(\theta|\alpha, \beta) = \frac{1}{\mathrm{B}(\alpha, \beta)}\, \theta^{\alpha-1}\, (1-\theta)^{\beta-1}$$




```
y <- rbeta(n=1e5, shape1=5, shape2=5)
plot(density(y), main='Beta(5, 5)', xlab='y')
```

Cauchy distribution

Data type: Continuous

Range of support: $[-\infty, \infty]$

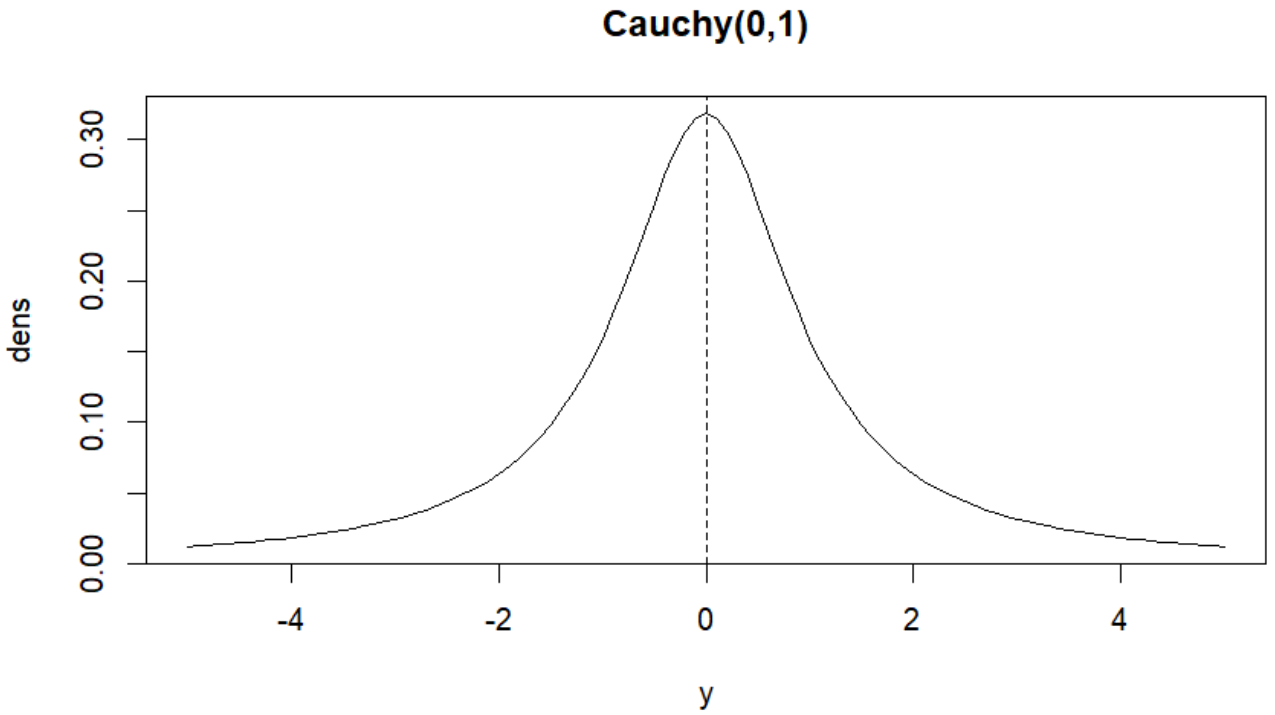
Parameters:

$\mu\ (location)$ $real[-\infty, \infty]$

$\sigma\ (scale)$ $real[0, \infty]$

Probability density function:

$$\text{Cauchy}(y|\mu,\sigma) = \frac{1}{\pi\sigma}\ \frac{1}{1+((y-\mu)/\sigma)^2}$$




```
x <- seq(-5, 5, by=0.1)
dens <- dcauchy(x)
plot(y=dens, x=x, main='Cauchy(0,1)', xlab='y', type='l')
abline(v=0, lty=2)
```

Poisson distribution

Data type: Discrete

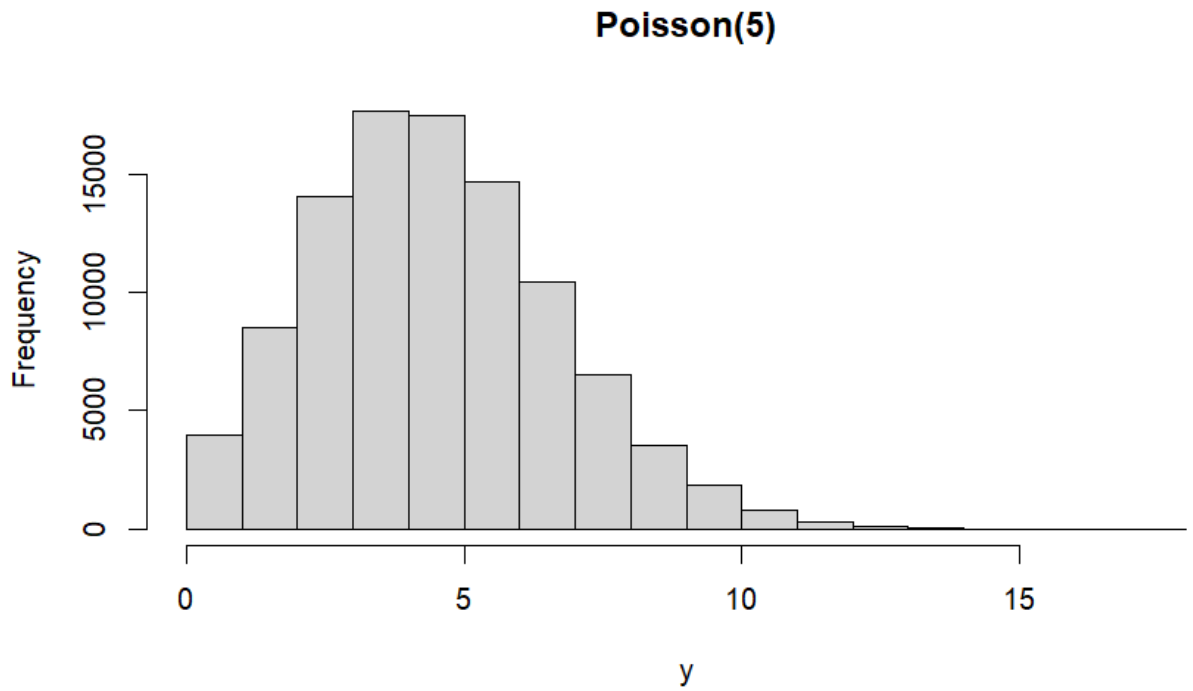
Range of support: $[0, \infty]$

Parameters:

$\lambda\ (lambda) \quad real[0, \infty]$

Probability density function:

$$\mathrm{Poisson}(n|\lambda) = \frac{1}{n!}\,\lambda^n\,\exp(-\lambda).$$




```
y <- rpois(1e5, lambda=5)
hist(y, main='Poisson(5)')
```

Binomial distribution

Data type: Discrete

Range of support: $[0, N]$

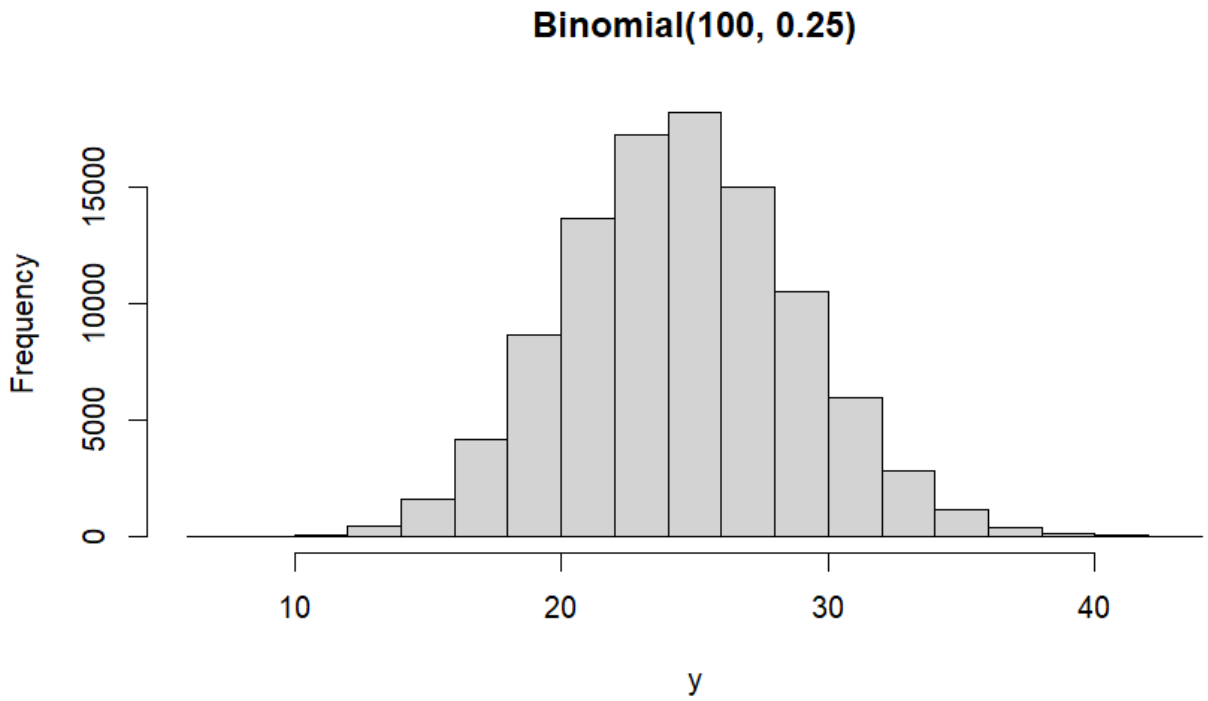
Parameters:

$N\ (size)$ $\quad int[0, \infty]$

$\theta\ (prob)$ $\quad real[0, 1]$

Probability density function:

$$\text{Binomial}(n \mid N, \theta) = \binom{N}{n} \theta^n (1-\theta)^{N-n}$$




```
y <- rbinom(1e5, size=100, prob=0.25)
hist(y, main='Binomial(100, 0.25)')
```

Bernoulli distribution

Data type: Discrete

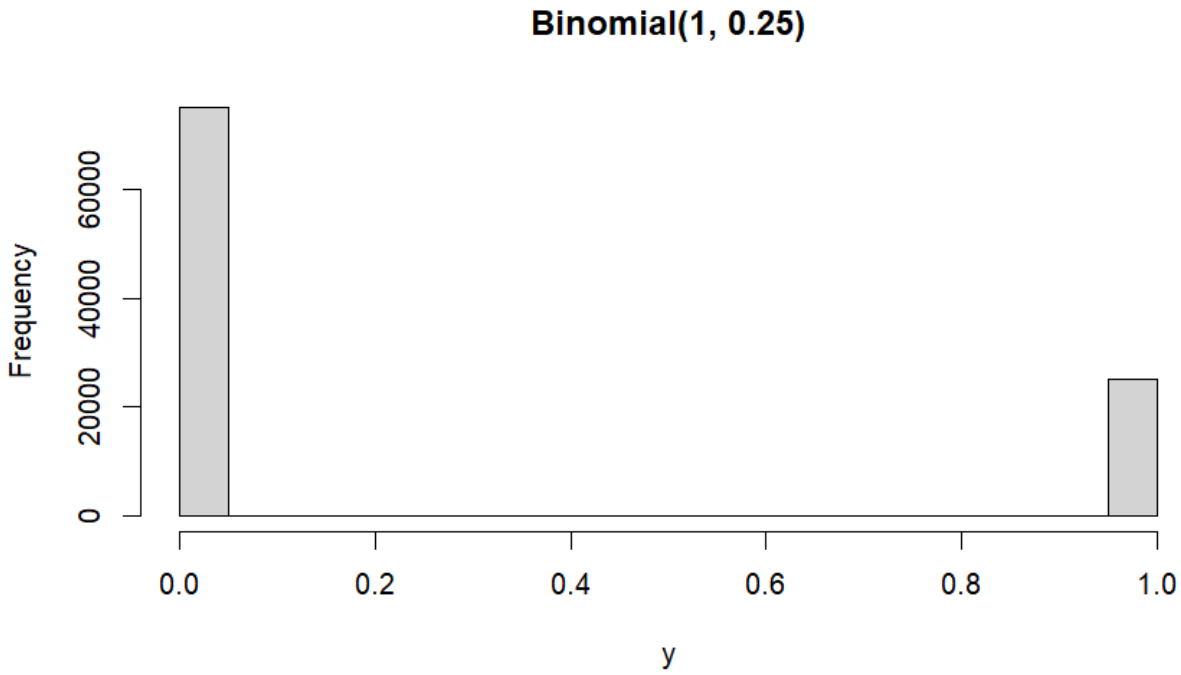
Range of support: $[0, 1]$

Parameters:

$\theta\ (prob) \quad real[0, 1]$

Probability density function:

$$\text{Bernoulli}(y \mid \theta) = \begin{cases} \theta & \text{if } y = 1, \text{ and} \\ 1 - \theta & \text{if } y = 0. \end{cases}$$




```
y <- rbinom(1e5, size=1, prob=0.25)
hist(y, main='Binomial(1, 0.25)')
```

```
> table(y)/1e5
y
      0       1
0.74987 0.25013
```

Multinomial distribution

Data type: Discrete

Range of support: $[0, N]$

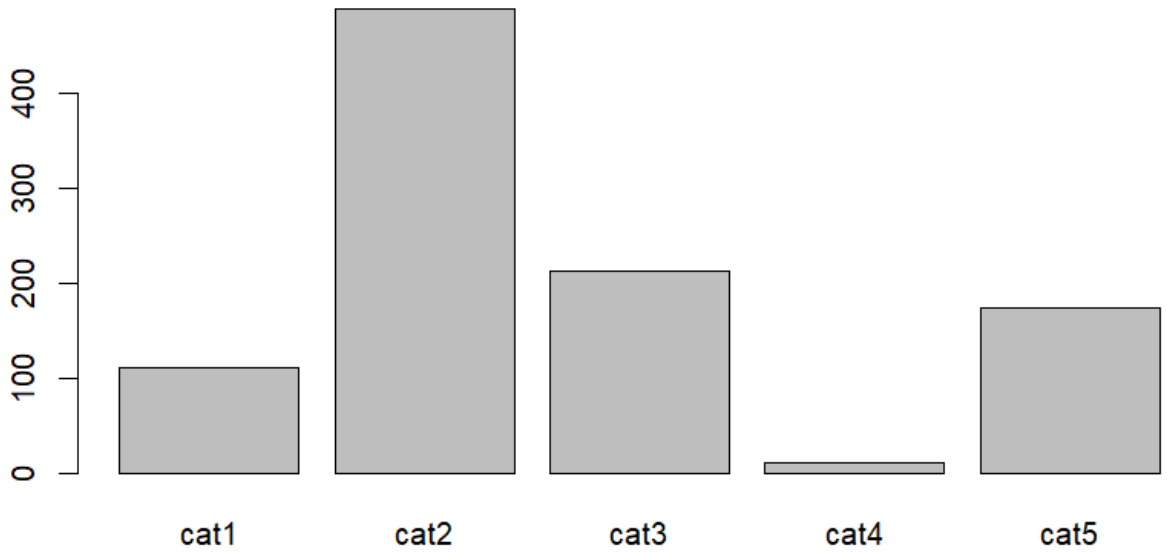
Parameters:

$N\ (size)$ $\quad int[0, \infty]$

$\theta_k\ (prob)$ $\quad real[0, 1]$

Probability density function:

$$\text{Multinomial}(y|\theta) = \binom{N}{y_1, \ldots, y_K} \prod_{k=1}^{K} \theta_k^{y_k},$$




```
y <- rmultinom(n=1, size=1e3, prob=c(0.1, 0.5, 0.2, 0.02, 0.18))
y <- as.vector(y)
names(y) <- paste0('cat', 1:5)
barplot(y)
```

Method of moments

- Moments: Statistics that describe a probability distribution (e.g. expected value, variance, skewness)
- Method of moments: Functions to calculate an estimate of the moments from the parameters of any distribution
- Example:

$$y_i \sim Gamma(\alpha, \beta)$$

$$\mu = \frac{\alpha}{\beta} \qquad \alpha = \frac{\mu^2}{\sigma^2}$$

$$\sigma^2 = \frac{\alpha}{\beta^2} \qquad \beta = \frac{\mu}{\sigma^2}$$

There are many more…

(see “Distributions Cheat Sheet” in the readings)

Priors

Bayes' theorem

$$\underbrace{P(\mu,\sigma|y)}_{\text{Posterior}} = \frac{\overbrace{P(y|\mu,\sigma)}^{\text{Likelihood}}\overbrace{P(\mu)P(\sigma)}^{\text{Priors}}}{P(y)}$$

$y =$ data
$\mu =$ mean
$\sigma =$ standard deviation

A prior probability …

- ✓contains our prior knowledge about a parameter in our model.
- ✓is a probability distribution.
- ✓is consistent with the numerical type of the parameter (continuous, discrete).
- ✓defines the “range of support” (i.e. range of possible values) for a parameter.
- ✓quantifies which parameter values are more likely than others.
- ✓can be informative, minimally informative, or uninformative (vague).

Probability distributions

When designing priors for parameters, we need to:

1. Choose an appropriate probability distribution for the parameter characteristics
2. Define values for its parameters that reflect our prior beliefs about the parameter

Let's design some priors together for this model…

$$y_i \sim Normal(\mu, \sigma)$$

$$\mu \sim Normal(0, 1)$$
$$\sigma \sim Normal(0, 1)$$

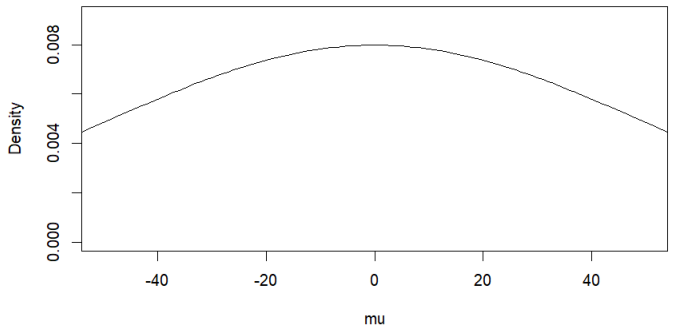
Are these priors reasonable?

… or might you have a few Qs first?

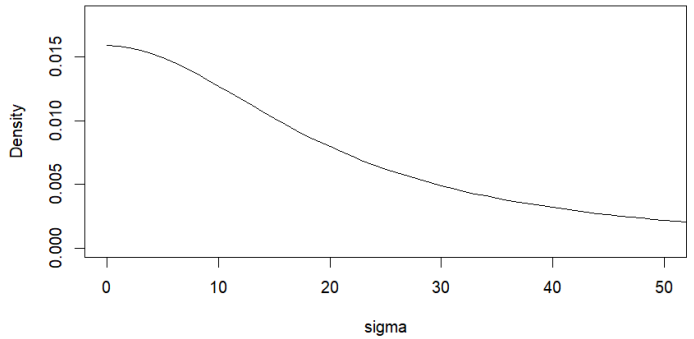
Let's design some priors together for this model…

$$y_i \sim Normal(\mu, \sigma)$$

y_i are temperatures measured at noon every day of 2023 in London.



$$\mu \sim Normal(0, 50)$$
$$\sigma \sim HalfCauchy(0, 20)$$



These are informative priors. What effect would that have on our posterior probabilities for μ and σ? What are potential justifications and alternatives?

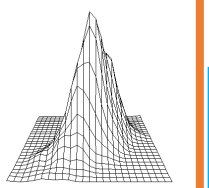
Bayes' theorem

Posterior

Likelihood

Priors

$$P(\mu, \sigma | y) = \frac{\text{Likelihood} \quad P(\mu)P(\sigma)}{P(y)}$$

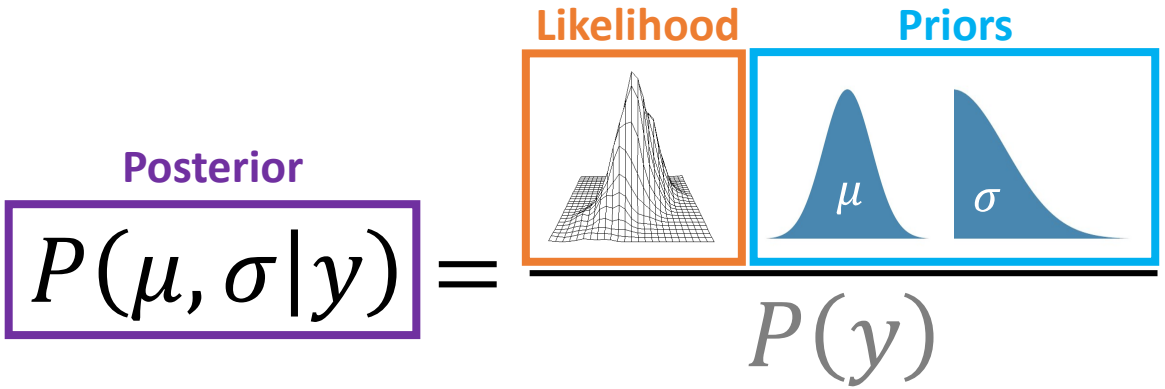


$y =$ data

$\mu =$ mean

$\sigma =$ standard deviation

Bayes' theorem



$y =$ data
$\mu =$ mean
$\sigma =$ standard deviation

Conjugate priors

- When the prior probability distribution has the same functional form as the likelihood, then the prior is “conjugate”
- Results in the posterior having the same distributional form as the prior (e.g. if the conjugate prior is Beta, then the posterior will be Beta)
- Facilitates an *analytical* solution to the posterior
- More efficient *numerical* solution to the posterior (i.e. MCMC) using Gibbs sampling

Conjugate priors

$$y_i \sim Normal(\mu, \sigma)$$

$$\frac{1}{\sigma^2} \sim Gamma(0.01, 0.01)$$

Conjugate priors

$$y_i \sim Poisson(\lambda)$$

$$\lambda \sim Gamma(0.01, 0.01)$$

Conjugate priors

$$y_i \sim Binomial(N, \theta)$$

$$\theta \sim Beta(1, 1)$$

Conjugate priors

$$y_k \sim Multinomial(N, \theta_k)$$

$$\theta_k \sim Dirichlet(1_k)$$

Conjugate priors

Do we need to use conjugate priors?

No, not really. But, it can improve computational efficiency.

Next up….

Posterior distribution & Markov Chain Monte Carlo (MCMC)

Bayes' theorem

Likelihood Priors

Posterior

$$P(\theta|y) = \frac{P(y|\theta)\,P(\theta)}{P(y)}$$

$y =$ data
$\theta =$ parameter

Bayes' theorem

Posterior

$$\boxed{P(\theta|y)} \propto P(y|\theta)\ P(\theta)$$

$y = \text{data}$
$\theta = \text{parameter}$

Markov chain Monte Carlo (MCMC)

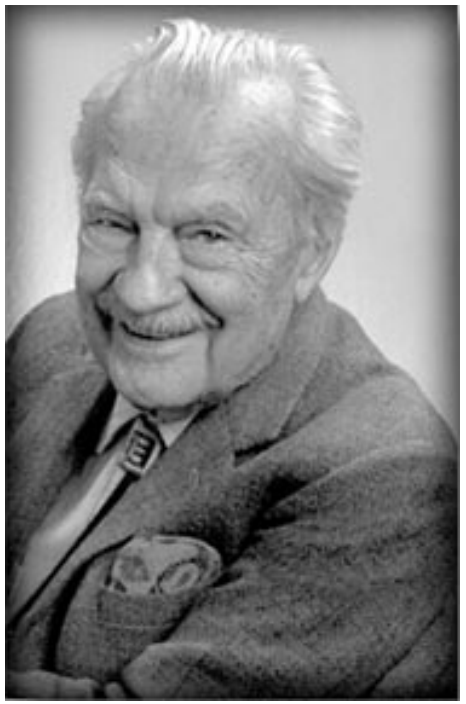
Andrei Markov
1856 - 1922

Monte Carlo Casino
Monte Carlo, Monaco

Stanislov Ulam
1909-1984

Markov chain Monte Carlo (MCMC)
Metropolis-Hastings Algorithm

Nicholas Metropolis
1915-1999



Wilfred K. Hastings
1930 - 2016

What is our goal with MCMC?

- The posterior distribution is *unknown*, but the likelihood and priors are *known.*
- We want to accumulate many random samples proportionate to their density in the posterior distribution
- MCMC generates these samples using the likelihood and the priors
- We can then use these samples to calculate statistics describing the distribution: mean, median, variance, credible intervals, etc.
- We can plot the posterior distribution using the MCMC samples the same way we plot the histogram using sample data.

Slide stolen from Mevin Hooten because I couldn't make one better.

How MCMC works

- MCMC produces a vector of parameter values θ_k sampled from the posterior distribution for K iterations of MCMC
- θ_k = value of parameter at iteration k
- θ' = value of parameter proposed at next iteration $k+1$ in the chain

We have a probability rule that allows us to compare the proposed value θ' with the current value θ_k. If θ' is more probable than θ_k, then it becomes θ_{k+1}. Otherwise, we keep $\theta_k = \theta_{k+1}$. The frequency of values of θ are proportional to the posterior density.

Slide stolen from Mevin Hooten because I couldn't make one better.

How MCMC works

k	1	2	3	4	K
Proposal θ'^k		θ'^2	θ'^3	θ'^4	. . . .
Rule		$P(\theta'^2) > P(\theta^1)$	$P(\theta^2) > P(\theta'^3)$	$P(\theta^3) > P(\theta'^4)$	. . . .
Chain (θ^k)	θ^1	$\theta^2 = \theta'^2$	$\theta^3 = \theta^2$	$\theta_4 = \theta_3$	

As the chain gets longer, the more probable values are retained more frequently than the less probable ones. This is what gives the proper shape to the simulated distribution—the frequency of values is proportionate to the probability density of the posterior distribution.

Slide stolen from Mevin Hooten because I couldn't make one better.

The wickedly clever idea behind MCMC

- We don't know the posterior distribution of θ, but we can *simulate* its density using a series of random draws accumulated in a Markov chain.
- The priors and the data arbitrate which draws are kept most frequently in the chain, thereby approximating the shape of the posterior distribution.

Slide stolen from Mevin Hooten because I couldn't make one better.

Metropolis: A flavor of MCMC

We start with an "arbitrary" value of $\theta = \theta_1$. We then draw a value θ' from an arbitrary, symmetric probability distribution (more about symmetric soon). We calculate a ratio:

$$P(\theta^{k+1} = \theta'^{k+1} \mid y) = \frac{\overbrace{P\left(y \mid \theta^{k+1} = \theta'^{k+1}\right)}^{\text{likelihood}} \overbrace{P\left(\theta^{k+1} = \theta'^{k+1}\right)}^{\text{prior}}}{\int_{\theta} P\left(y \mid \theta\right) P\left(\theta\right) d\theta}$$

$$P(\theta^{k+1} = \theta^{k} \mid y) = \frac{\overbrace{P\left(y \mid \theta^{k+1} = \theta^{k}\right)}^{\text{likelihood}} \overbrace{P\left(\theta^{k+1} = \theta^{k}\right)}^{\text{prior}}}{\int_{\theta} P\left(y \mid \theta\right) P\left(\theta\right) d\theta}$$

$$R = \frac{P(\theta^{k+1} = \theta'^{k+1} \mid y)}{P(\theta^{k+1} = \theta^{k} \mid y)}$$

Slide stolen from Mevin Hooten because I couldn't make one better.

The cunning bit:

$$P(\theta^{k+1} = \theta'^{k+1} \mid y) = \frac{\overbrace{P\left(y \mid \theta^{k+1} = \theta'^{k+1}\right)}^{\text{likelihood}} \overbrace{P\left(\theta^{k+1} = \theta'^{k+1}\right)}^{\text{prior}}}{\cancel{\int_{\theta} P(y \mid \theta) P(\theta) d\theta}}$$

$$P(\theta^{k+1} = \theta^{k} \mid y) = \frac{\overbrace{P\left(y \mid \theta^{k+1} = \theta^{k}\right)}^{\text{likelihood}} \overbrace{P\left(\theta^{k+1} = \theta^{k}\right)}^{\text{prior}}}{\cancel{\int_{\theta} P(y \mid \theta) P(\theta) d\theta}}$$

$$R = \frac{P(\theta^{k+1} = \theta'^{k+1} \mid y)}{P(\theta^{k+1} = \theta^{k} \mid y)}$$

Slide stolen from Mevin Hooten because I couldn't make one better.

When do we keep the proposal?

$$P_R = \min\left(1, \frac{\overbrace{P\left(y \mid \theta^{k+1} = \theta'^{k+1}\right)}^{\text{likelihood}} \overbrace{P\left(\theta^{k+1} = \theta'^{k+1}\right)}^{\text{prior}}}{\underbrace{P\left(y \mid \theta^{k+1} = \theta^{k}\right)}_{\text{likelihood}} \underbrace{P\left(\theta^{k+1} = \theta^{k}\right)}_{\text{prior}}}\right)$$

Keep θ'^{k+1} as the next value in the chain with probability P_R and keep θ^k with probability = 1-P_R.

Slide stolen from Mevin Hooten because I couldn't make one better.

Implementing Metropolis Algorithm

- Calculate R based on likelihoods and priors (i.e., the numerators of Bayes theorem)
- Draw a random number, U from uniform distribution 0,1
- If R > U, we keep θ'_k as the next value in the chain. Otherwise, we keep θ_k as the next value.

Slide stolen from Mevin Hooten because I couldn't make one better.

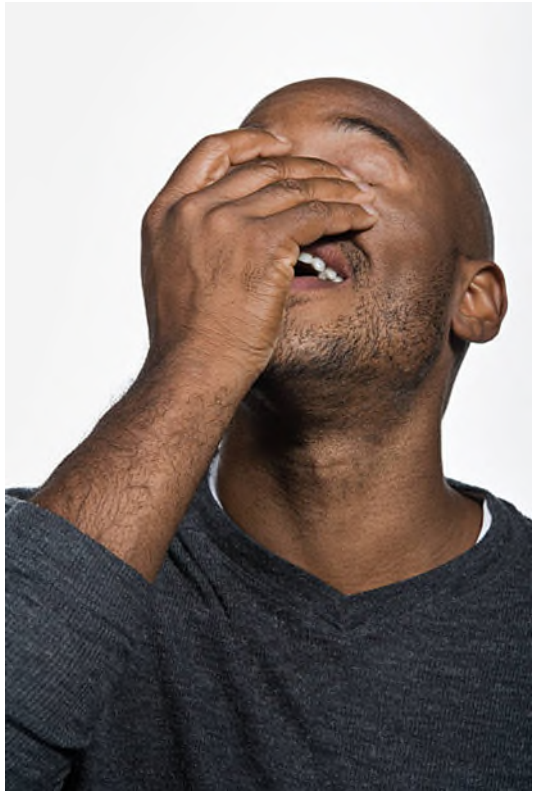
Are you ready to do this on your own?

MCMC mantra:

- This is easier done than said.
- This is easier done than said.
- This is easier done than said.
- This is easier done than said.
- This is easier done than said.
- This is easier done than said.
- This is easier done than said.

Slide stolen from Mevin Hooten because I couldn’t make one better.

Stan does everything we just talked about for you…



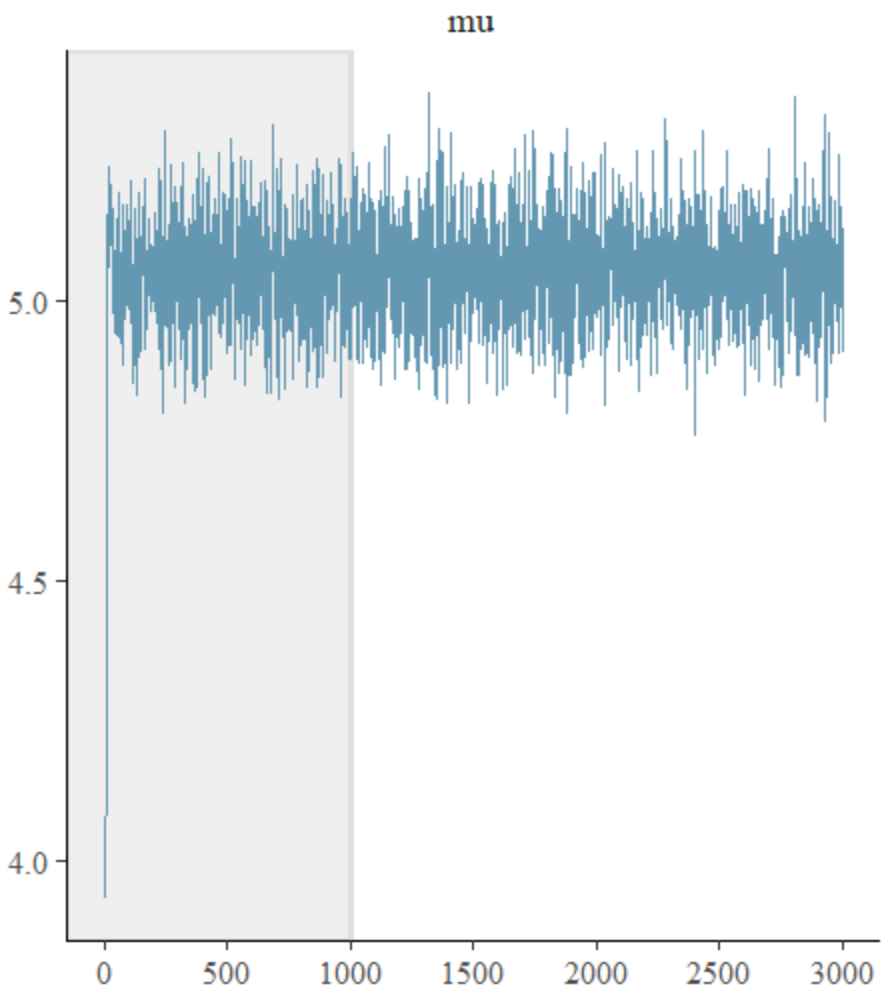
The basics

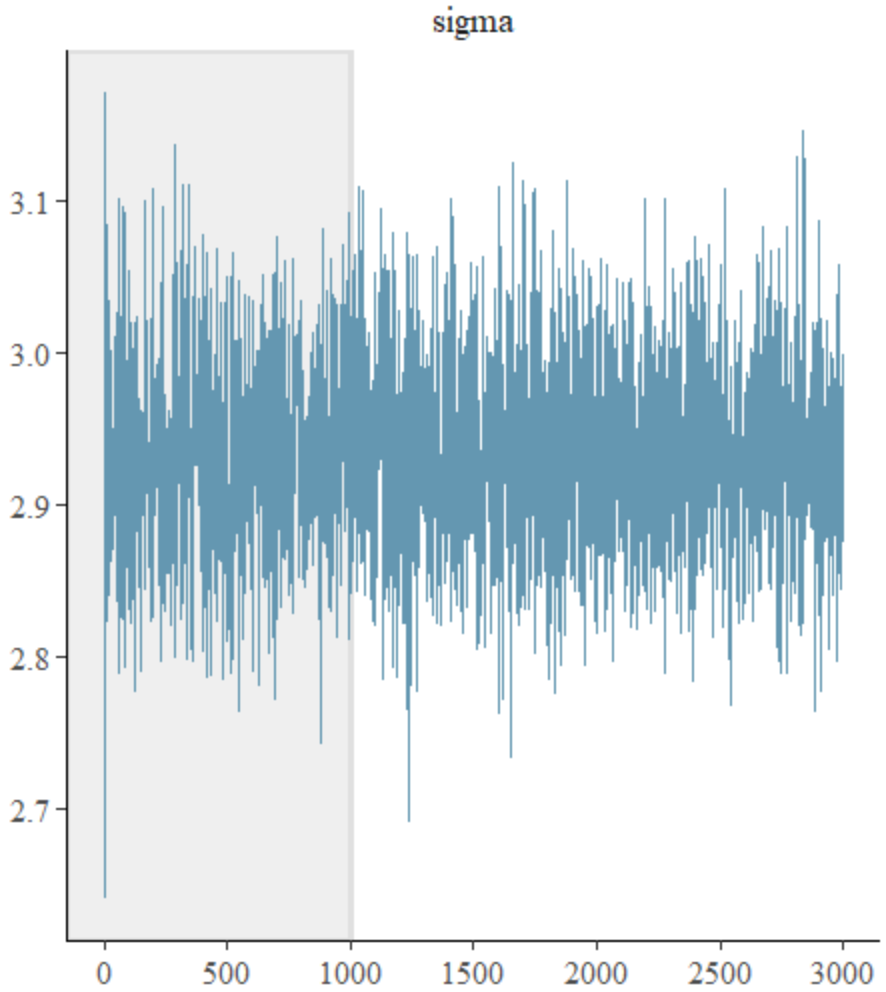
- Multiple MCMC chains
- Initial values for each parameter in each chain
- Burn-in period (a.k.a. warmup period) to be discarded
- Samples to keep for analysis
- Convergence of chains

A quick look…

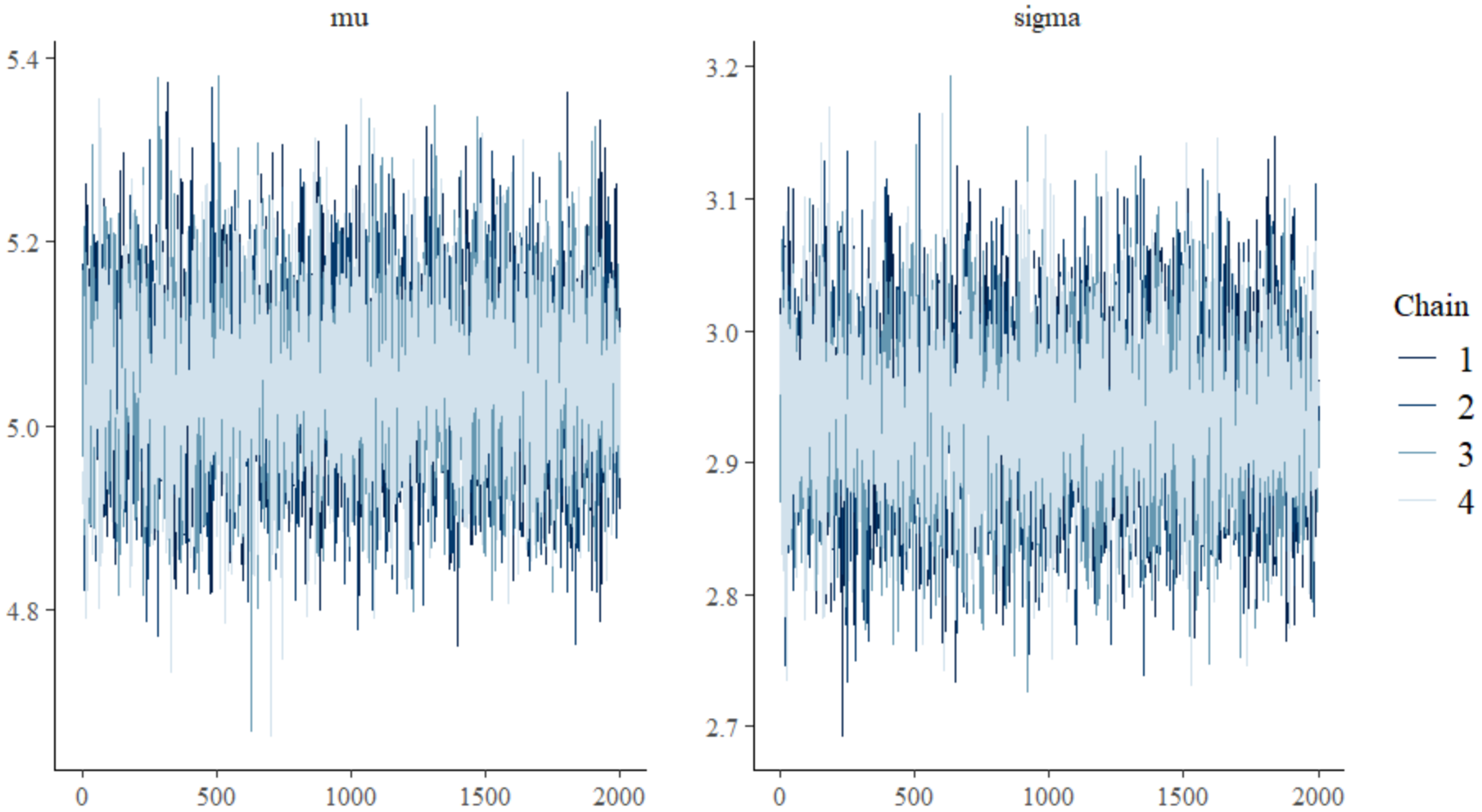
```
# run MCMC to sample from the posterior distribution
fit <- mod$sample(data = md,
                  parallel_chains = 4,
                  init = list(list(mu=3.8, sigma=1.1),
                              list(mu=-1.9, sigma=3.7),
                              list(mu=5.2, sigma=2.5),
                              list(mu=7.6, sigma=4.2)),
                  iter_sampling = 2000,
                  iter_warmup = 1000,
                  save_warmup = TRUE,
                  seed = md$seed)
```

A quick look…

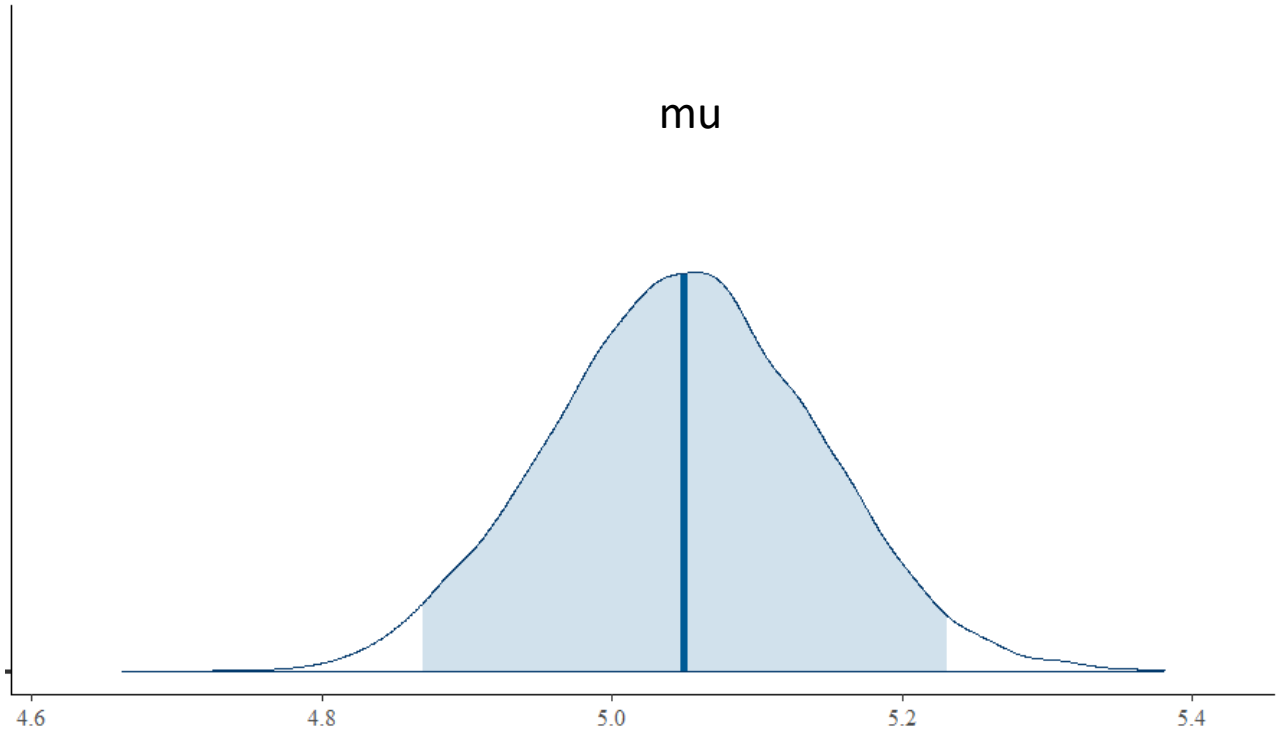

mu
5.0
4.5
4.0
0
500
1000
1500
2000
2500
3000

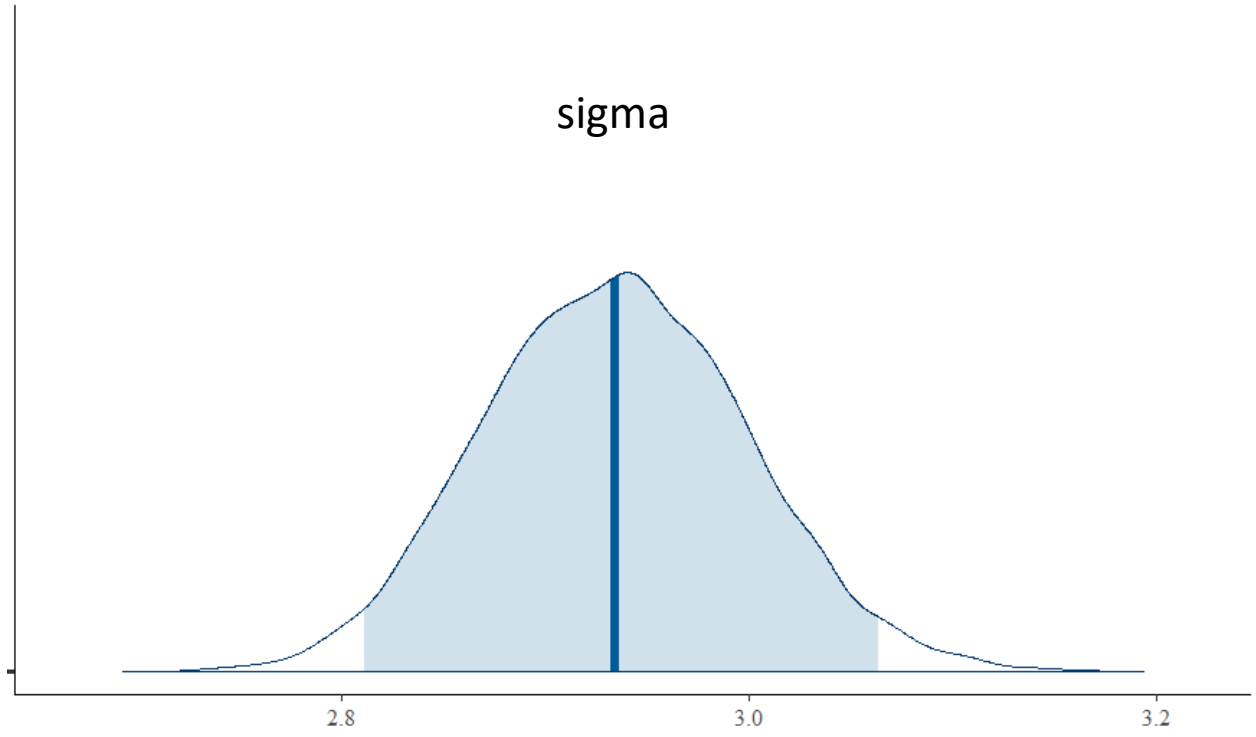

sigma
3.1
3.0
2.9
2.8
2.7
0
500
1000
1500
2000
2500
3000

A quick look…

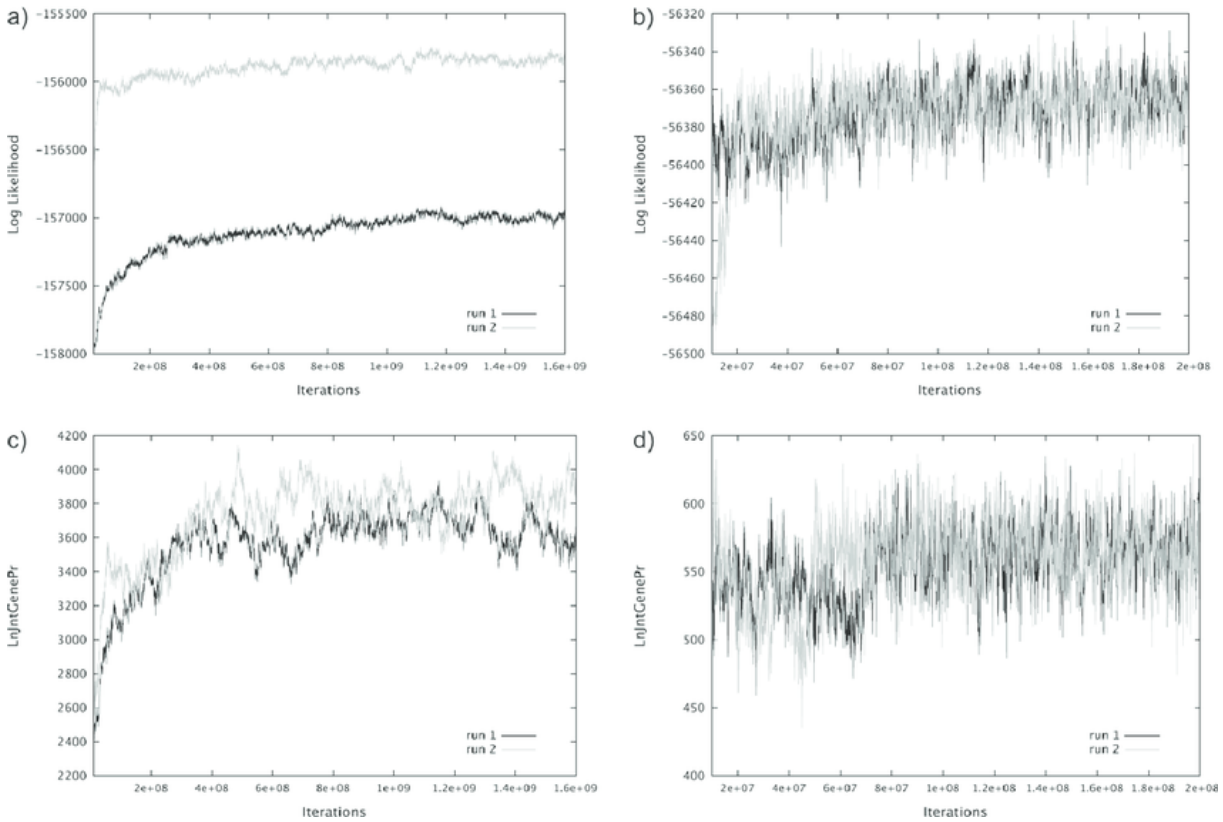

mu
5.4
5.2
5.0
4.8
0
500
1000
1500
2000
sigma
3.2
3.1
3.0
2.9
2.8
2.7
0
500
1000
1500
2000
Chain
1
2
3
4

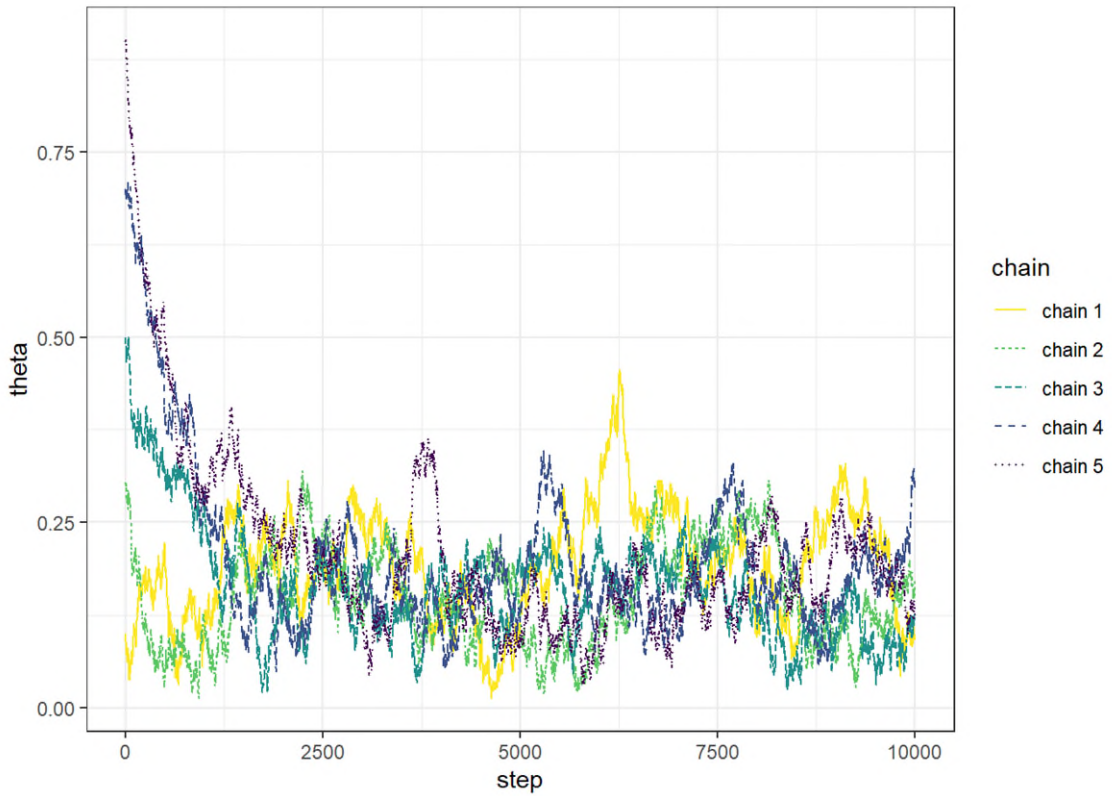
A quick look…





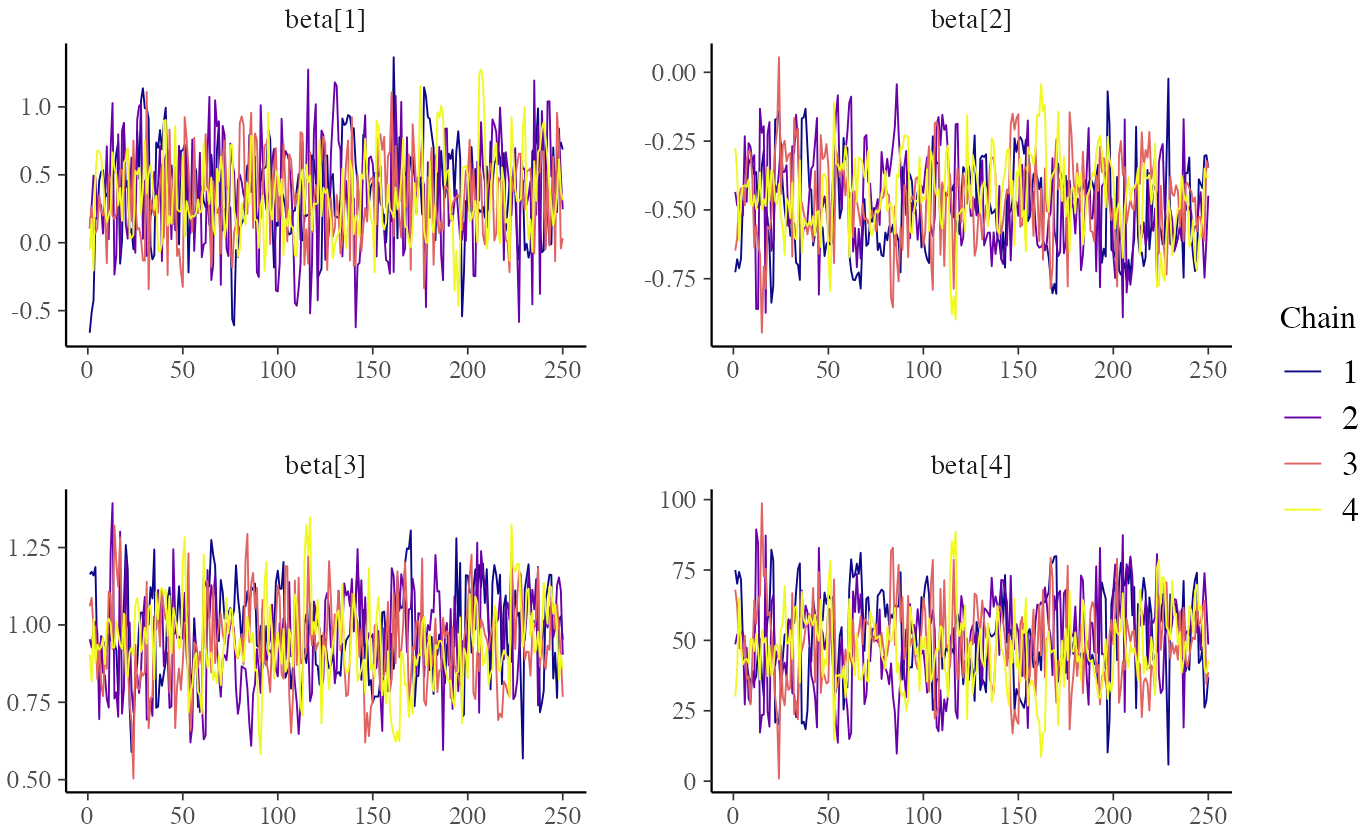
Convergence problems
Burn-in/warmup too short





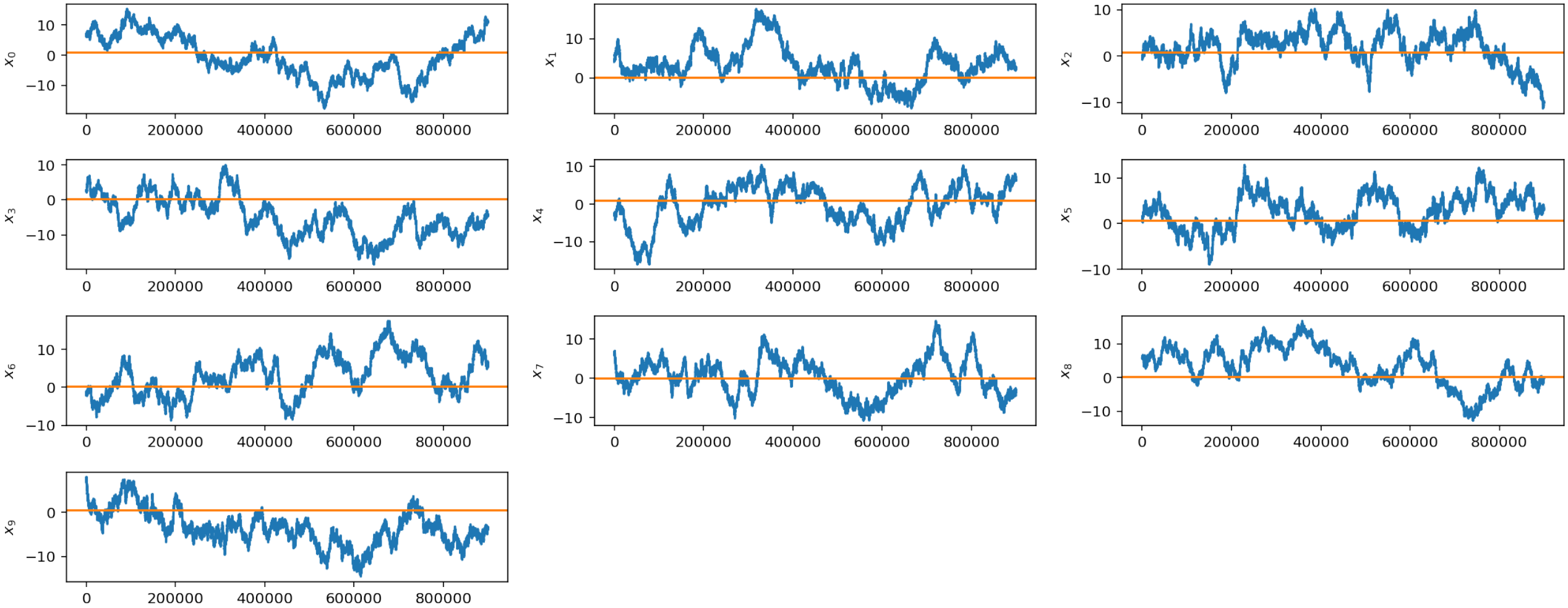
Convergence problems

Not enough samples (although getting close)

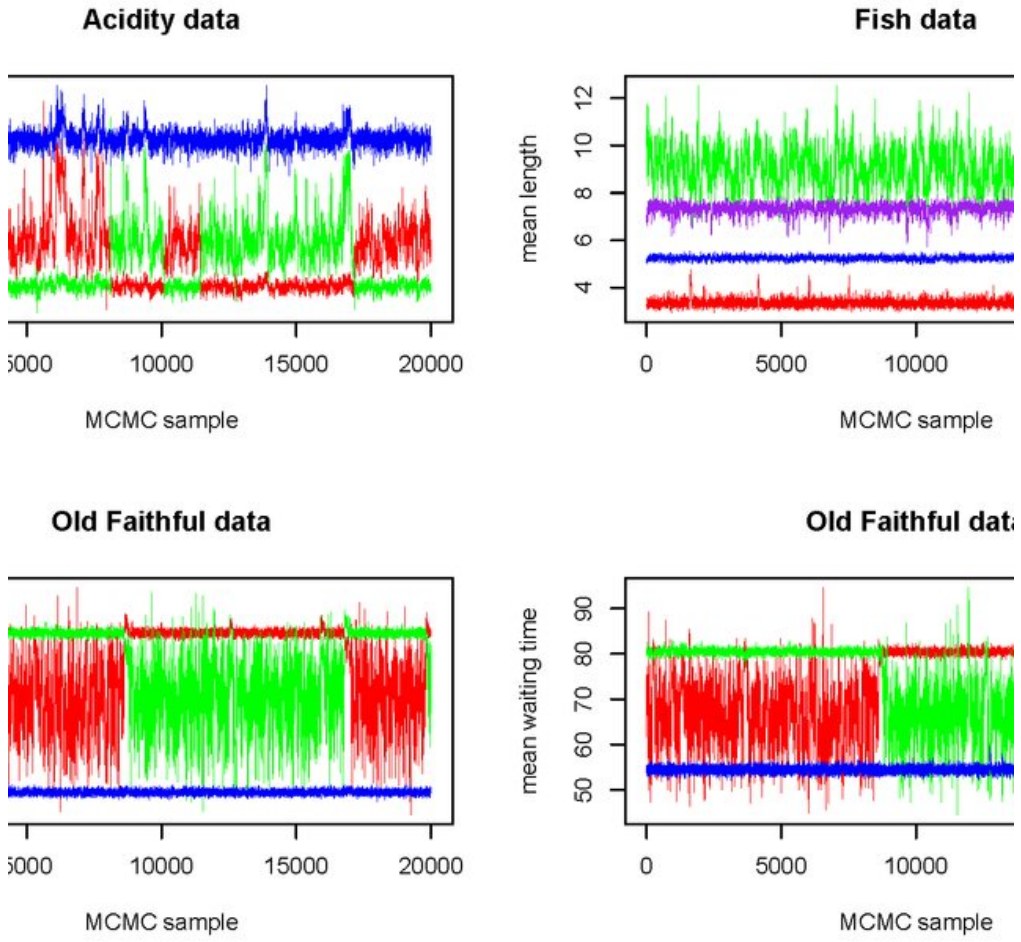


Convergence problems

Autocorrelation



Convergence problems
Non-identifiability

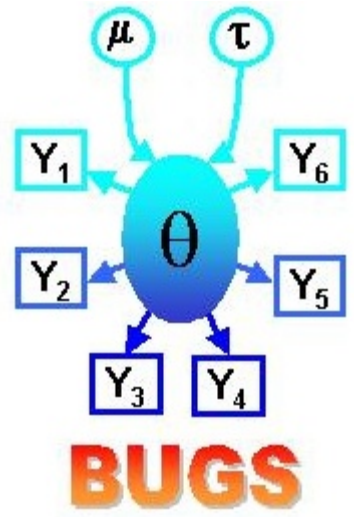


Quantitative diagnostics

- Rhat
- Divergences
- EBMI

Software

Program Bayesian models and run MCMC samplers



JAGS

Just Another Gibbs Sampler

Stan

Next up….

Designing your own model

Model design as a learning process

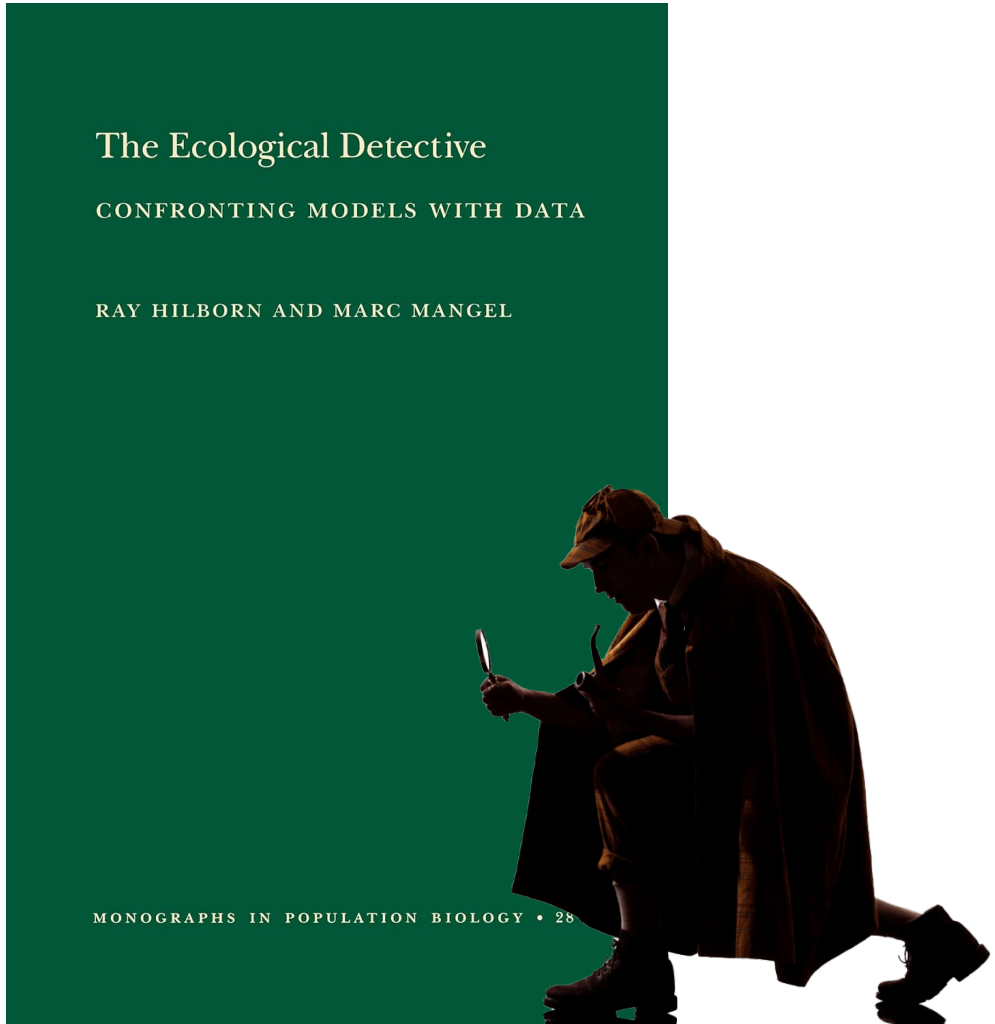
Formalise your mental picture (hypothesis) of the process that you are studying

Think about each distribution, parameter, and functional relationship

Design assumptions appropriate for the process and data.

Confront your model with data…

Learn, repeat.



Components of a model

1. Data model(s)

$$y_i \sim Normal(\mu_i, \sigma)$$

2. Process model(s)

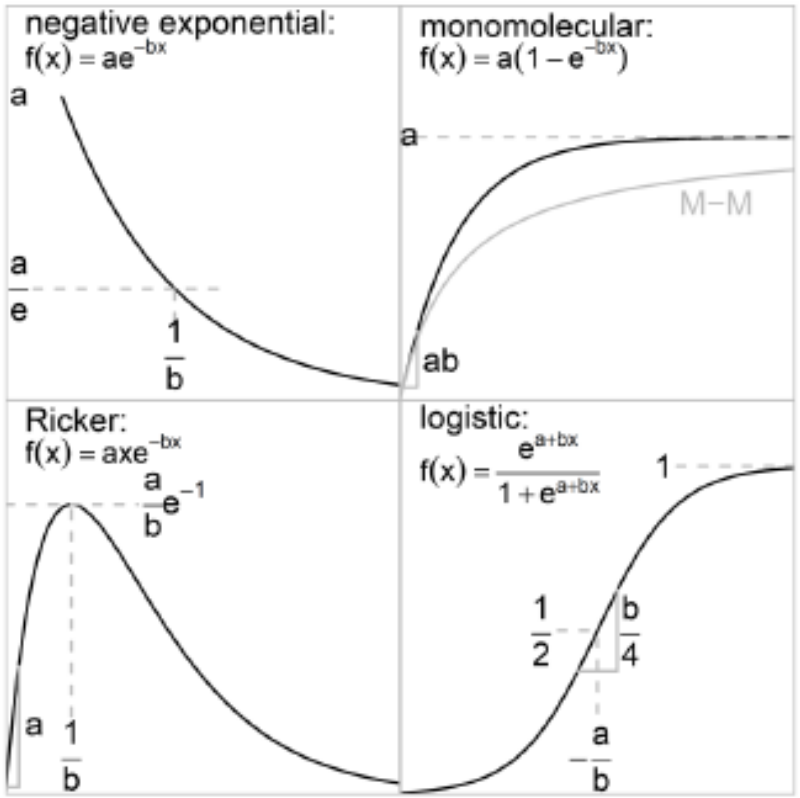
$$\mu_i = \alpha + \beta x_i$$

3. Parameter model(s)

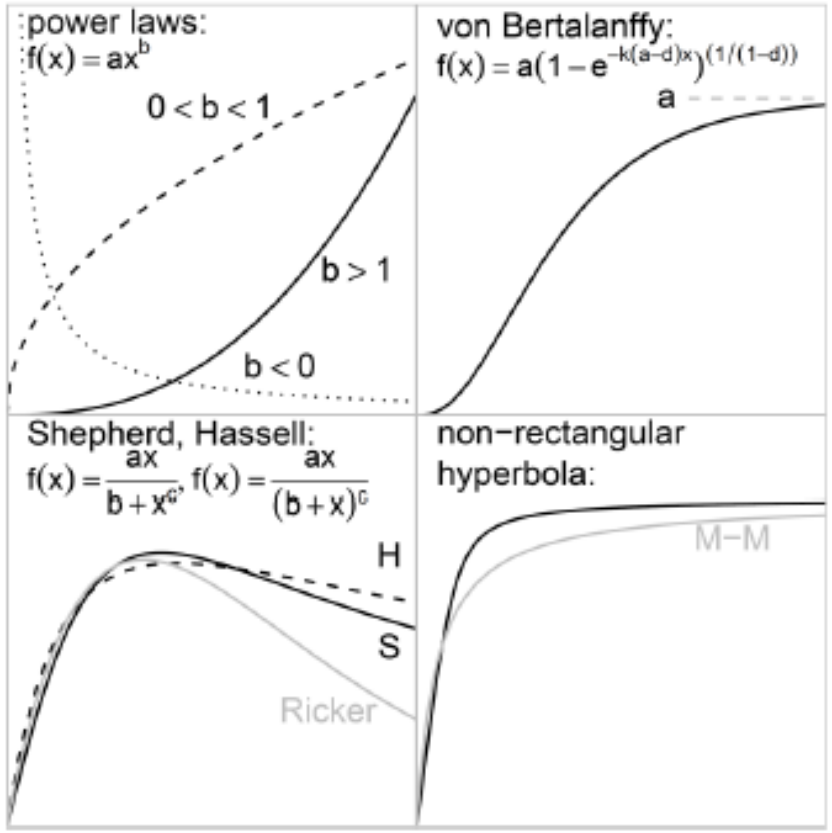
$$\alpha \sim Normal(0, 1)$$
$$\beta \sim Normal(0, 1)$$
$$\sigma \sim HalfCauchy(0, 1)$$

Process models
Deterministic functions

EXPONENTIAL-BASED

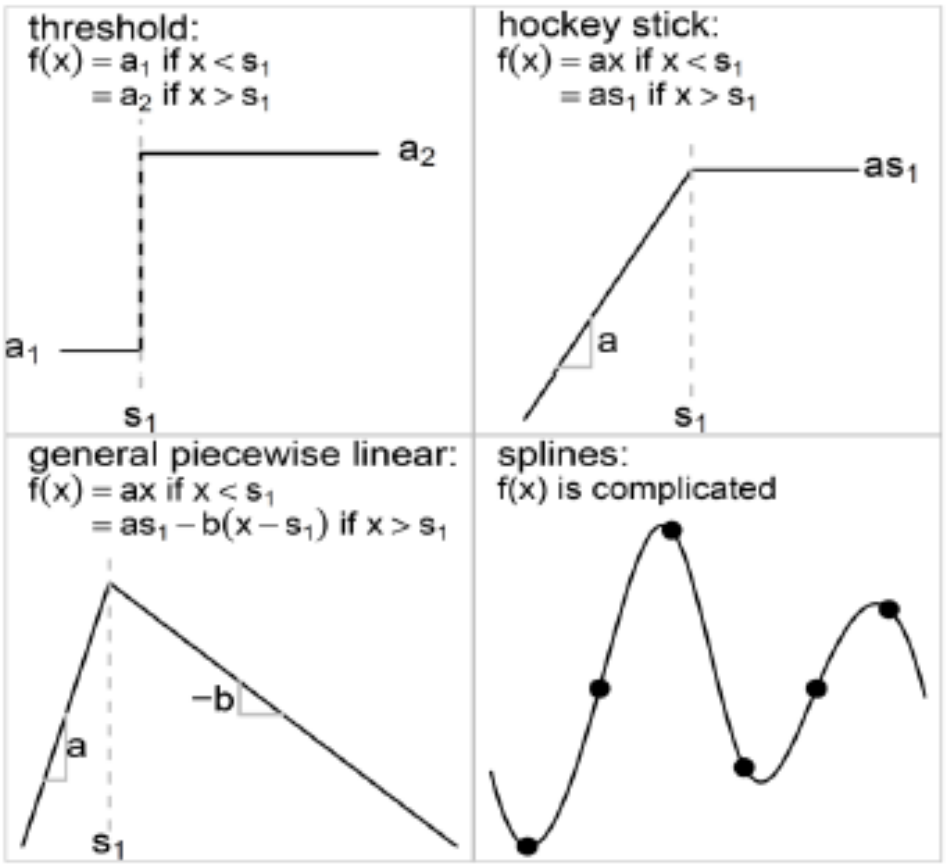


POWER-BASED

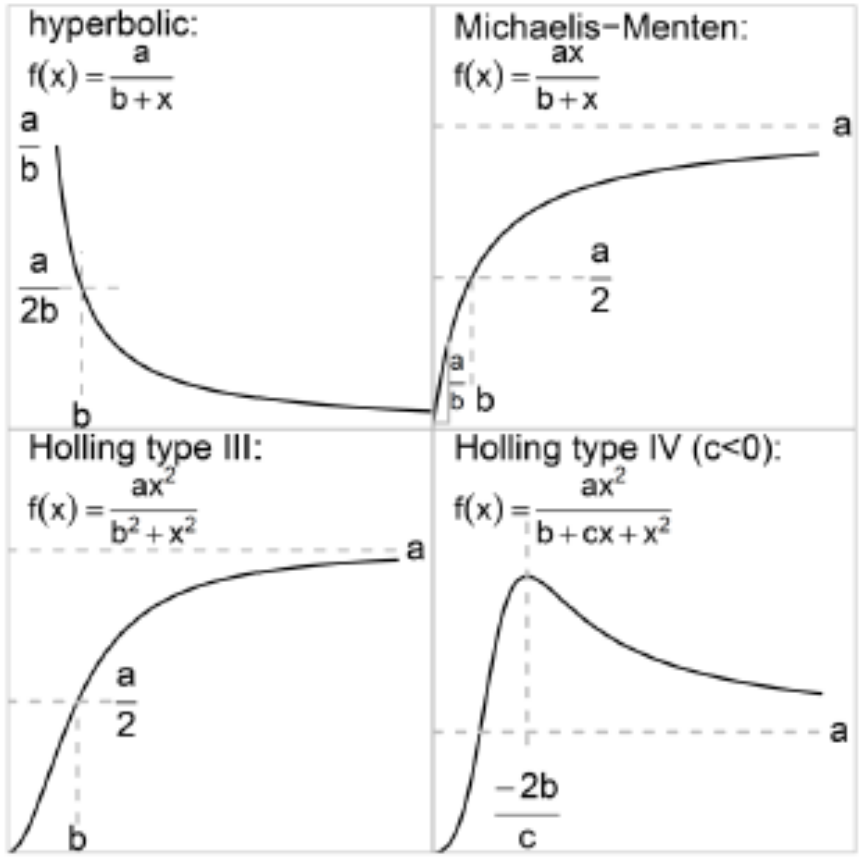


Process models
Deterministic functions

PIECEWISE POLYNOMIALS

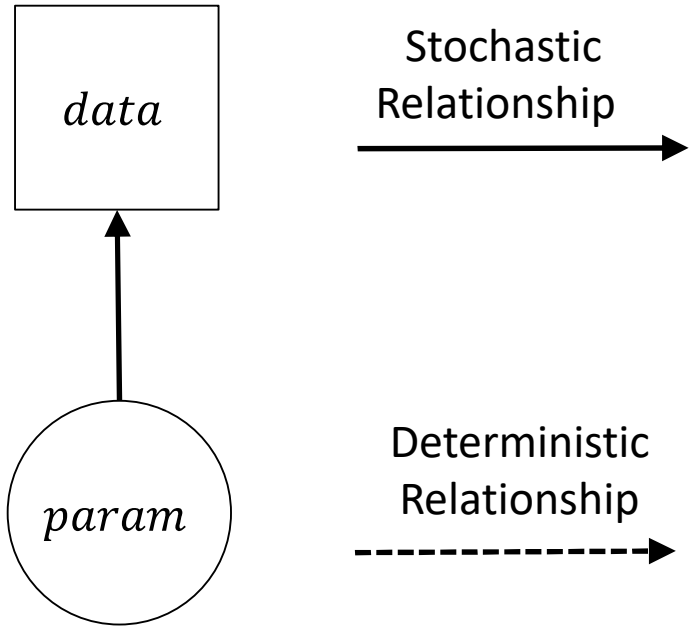


RATIONAL:



Directed Acyclic Graphs (DAGs)

- Graphical depiction of relationships between every node of your model
- Nodes = data or parameters
- Arrows point towards the node which is dependent on the other
- Arrow types indicate stochastic or deterministic relationships



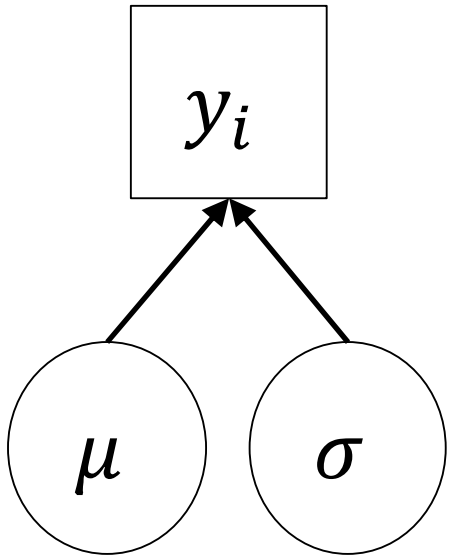
Directed Acyclic Graphs (DAGs)

$$y_i \sim Normal(\mu, \sigma)$$

Have we forgotten anything in this model?

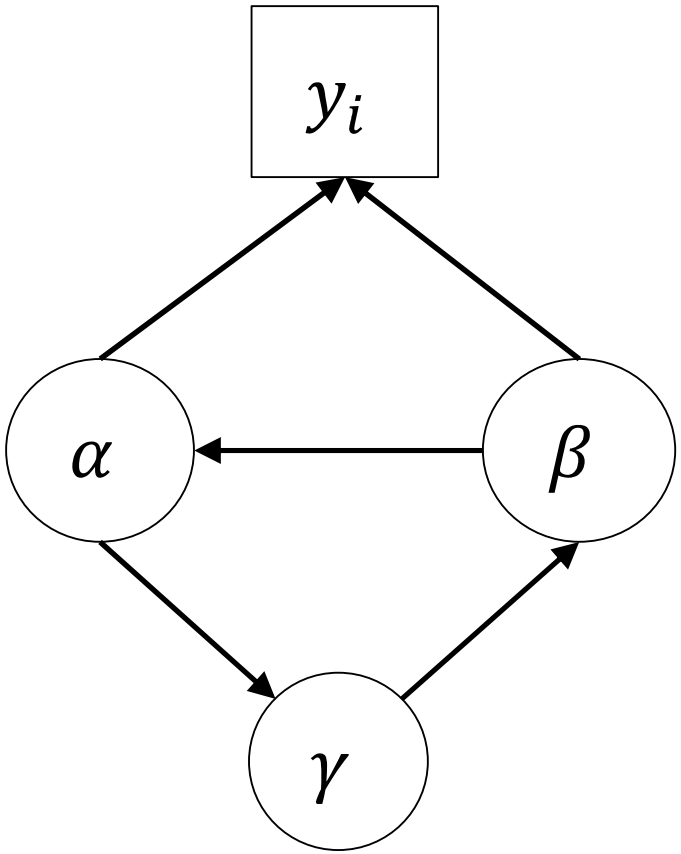
Priors!

$$\mu \sim Normal(0, 1)$$
$$\sigma \sim Uniform(0, 10)$$



Directed Acyclic Graphs (DAGs)

No loops (cycles) allowed

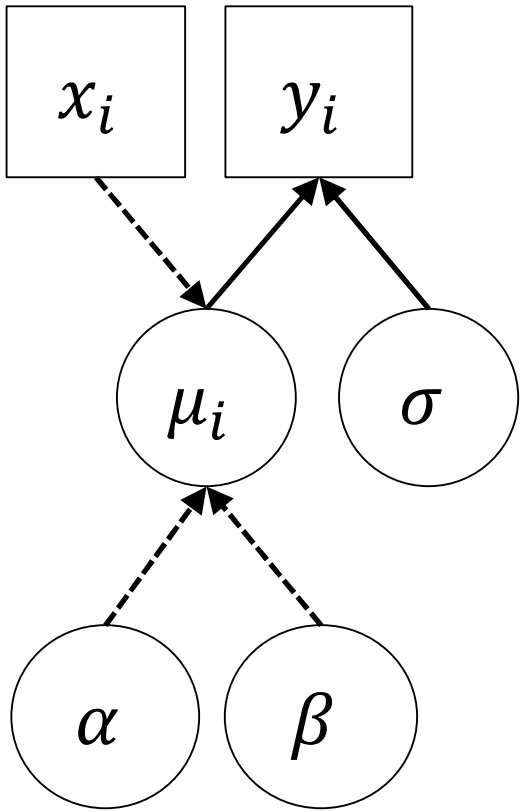


Directed Acyclic Graphs (DAGs)

$$y_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta x_i$$

$$\alpha \sim Normal(0, 1)$$
$$\beta \sim Normal(0, 1)$$
$$\sigma \sim HalfCauchy(0, 1)$$



Want to try to write a model from scratch together?

If we have time…

Recap

- Bayes theorem
- Likelihood
- Probability distributions
- Priors
- MCMC (and Posteriors)
- Components of a Bayesian model
- Directed Acyclic Graphs (DAGs)

Lab Activities

1. Run your first Bayesian model: Mean and variance
2. Explore likelihoods, priors, and posteriors
3. Gaussian linear regression