Workshop 7.2b: Introduction to Bayesian models

Murray Logan
February 7, 2017

Table of contents

1. Frequentist vs Bayesian

1.1. Frequentist

- P(D|H)
- long-run frequency
- simple analytical methods to solve roots
- conclusions pertain to data, not parameters or hypotheses
- compared to theoretical distribution when NULL is true
- probability of obtaining observed data or MORE EXTREME data

1.2. Frequentist

- P-value
  - probabulity of rejecting NULL
  - NOT a measure of the magnitude of an effect or degree of significance!
  - measure of whether the sample size is large enough

- 95% CI
  - NOT about the parameter it is about the interval
  - does not tell you the range of values likely to contain the true mean

1.3. Frequentist vs Bayesian

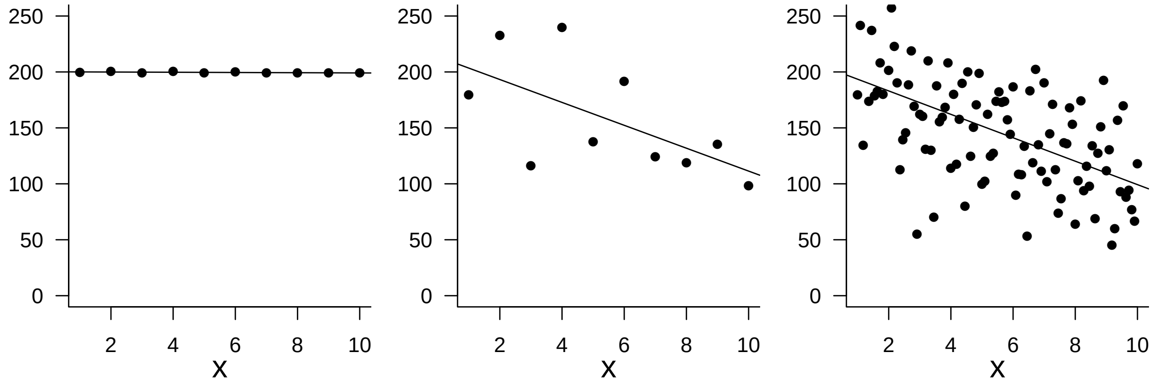
```
-------------------------------------------------
                Frequentist         Bayesian
--------------  ------------        ------------
Obs. data       One possible        Fixed, true

Parameters      Fixed, true         Random,
                                    distribution

Inferences      Data                Parameters

Probability     Long-run frequency  Degree of belief
                $P(D|H)$            $P(H|D)$
-------------------------------------------------
```

1.4. Frequentist vs Bayesian

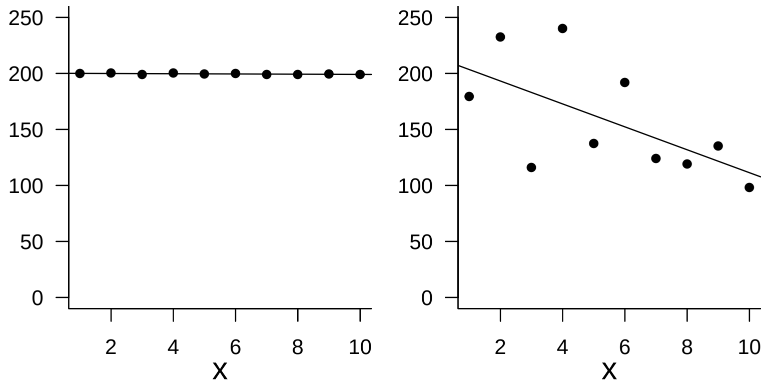


n: 10 Slope: -0.1022 t: -2.3252 p: 0.0485

n: 10 Slope: -10.2318 t: -2.2115 p: 0.0579

n: 100 Slope: -10.4713 t: -6.6457 p: 1.7101362 Œ 10-9

1.5. Frequentist vs Bayesian



	Population A	Population B
Percentage change	0.46	45.46
Prob. >5% decline	0	0.86

2. Bayesian Statistics

2.1. Bayesian

2.1.1. Bayes rule

$$P(H \mid D) = \frac{P(D \mid H) \times P(H)}{P(D)}$$

$$\text{posterior} \quad \textit{belief} \quad (\textit{probability}) = \frac{\textit{likelihood} \times \text{prior probability}}{\text{normalizing constant}}$$

2.2. Bayesian

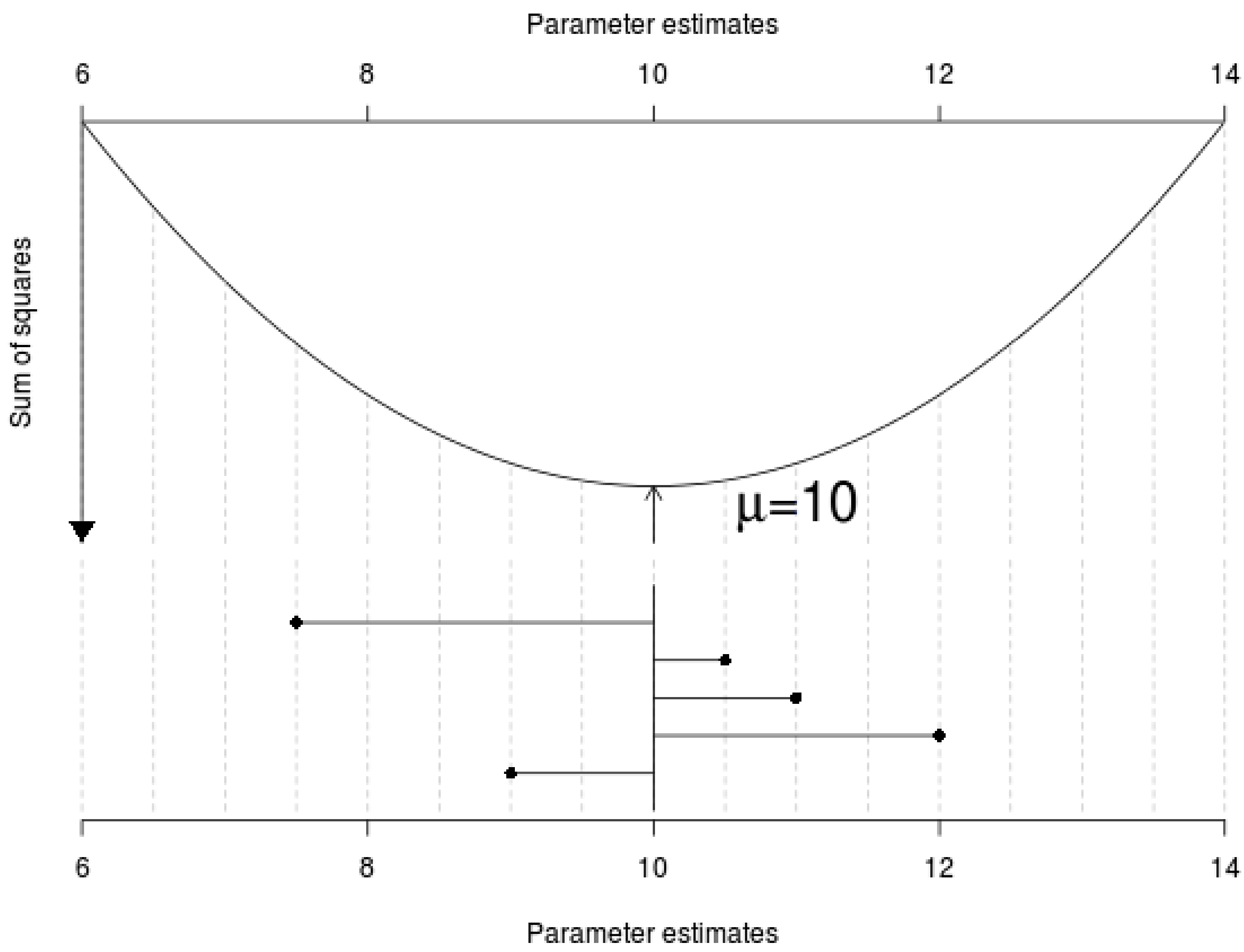
2.2.1. Bayes rule

$$P(H \mid D) = \frac{P(D \mid H) \times P(H)}{P(D)}$$

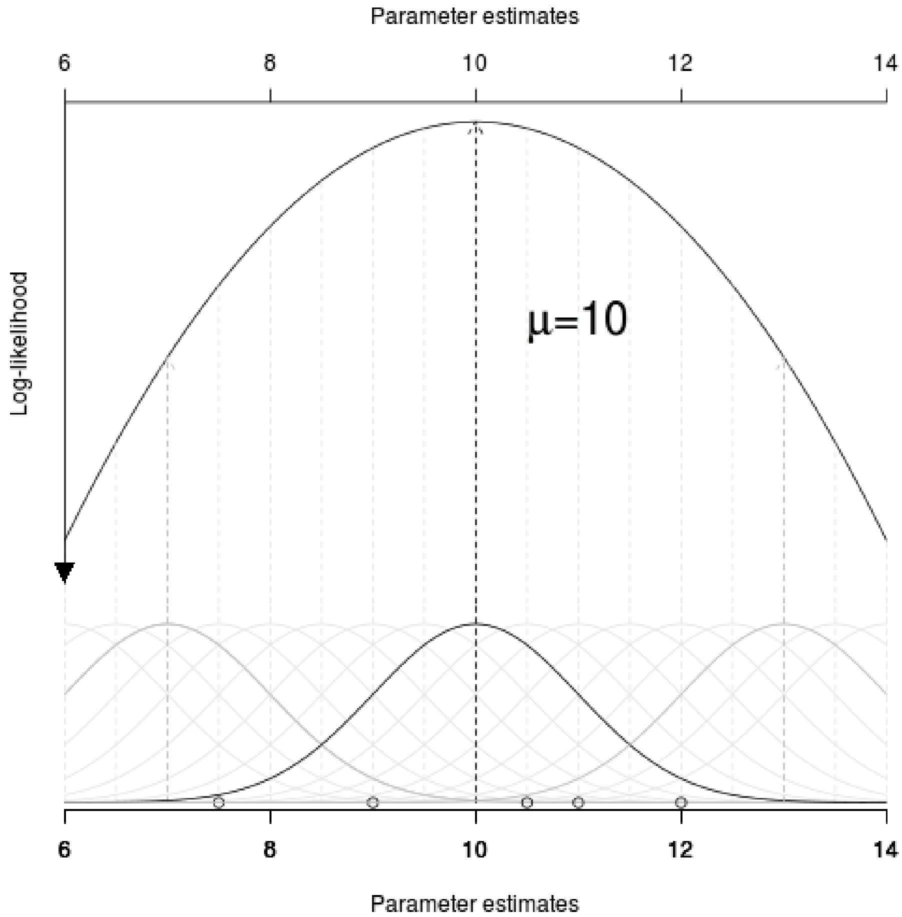
$$\begin{matrix}\text{posterior} \\ \textit{belief} \\ (\textit{probability})\end{matrix} = \frac{\textit{likelihood} \times \text{prior probability}}{\text{normalizing constant}}$$

The normalizing constant is required for probability - turn a frequency distribution into a probability distribution

2.3. Estimation: OLS



2.4. Estimation: Likelihood



$P(D \mid H)$

2.5. Bayesian

- conclusions pertain to hypotheses
- computationally robust (sample size,balance,collinearity)
- inferential flexibility - derive any number of inferences

2.6. Bayesian

- subjectivity?
- intractable

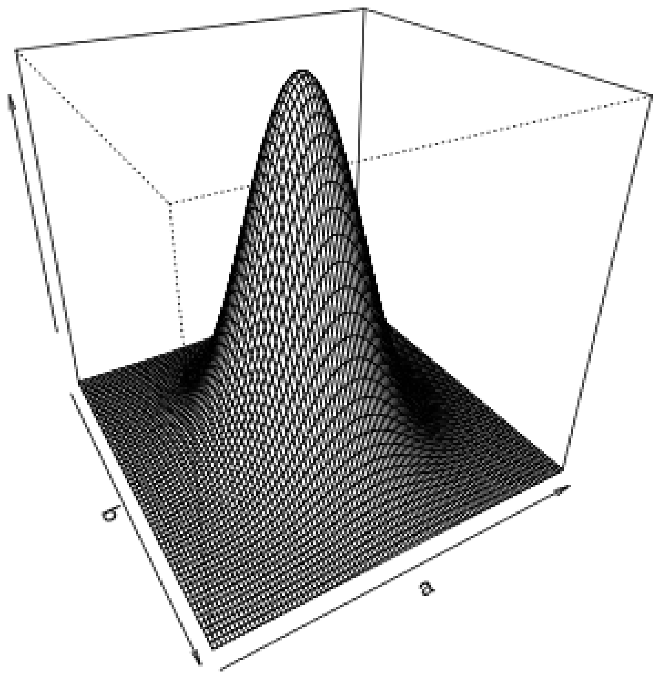
$$P(H \mid D) = \frac{P(D \mid H) \times P(H)}{P(D)}$$

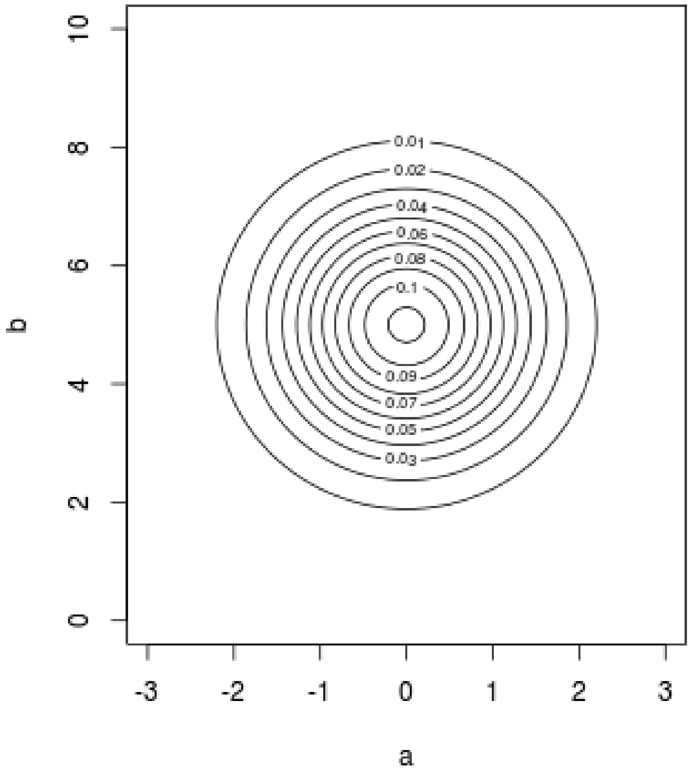
$P(D)$- probability of data from all possible hypotheses

2.7. MCMC sampling

Marchov Chain Monte Carlo sampling

- draw samples proportional to likelihood





two parameters α and β

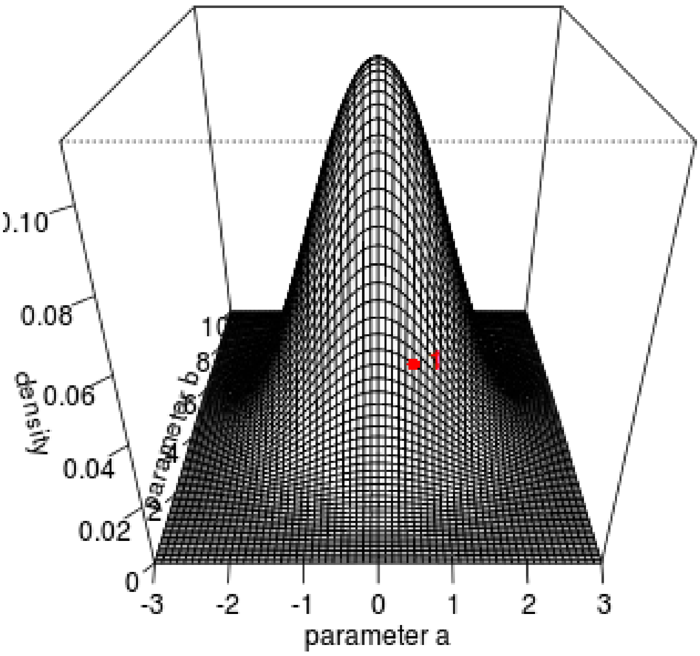
infinitely vague priors - posterior likelihood only

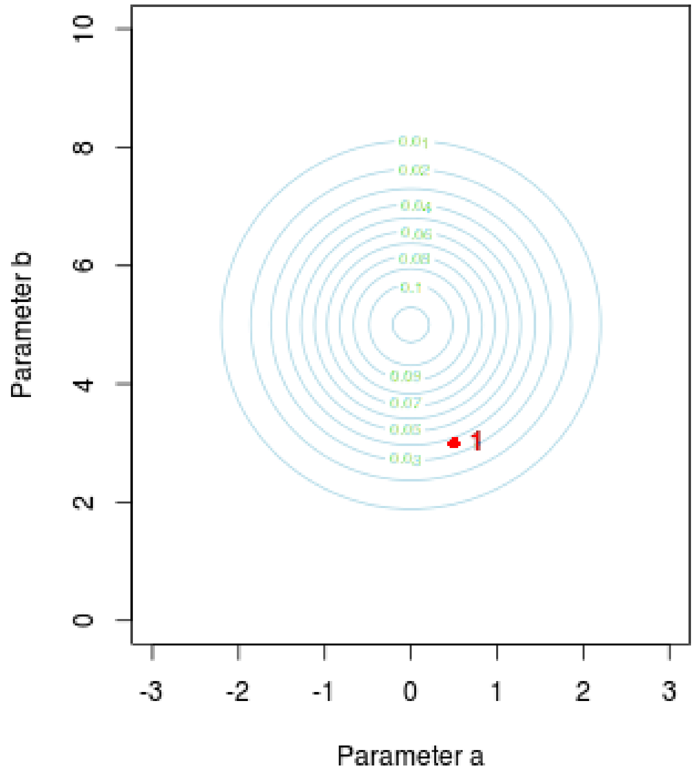
likelihood multivariate normal

2.8. MCMC sampling

Marchov Chain Monte Carlo sampling

- draw samples proportional to likelihood

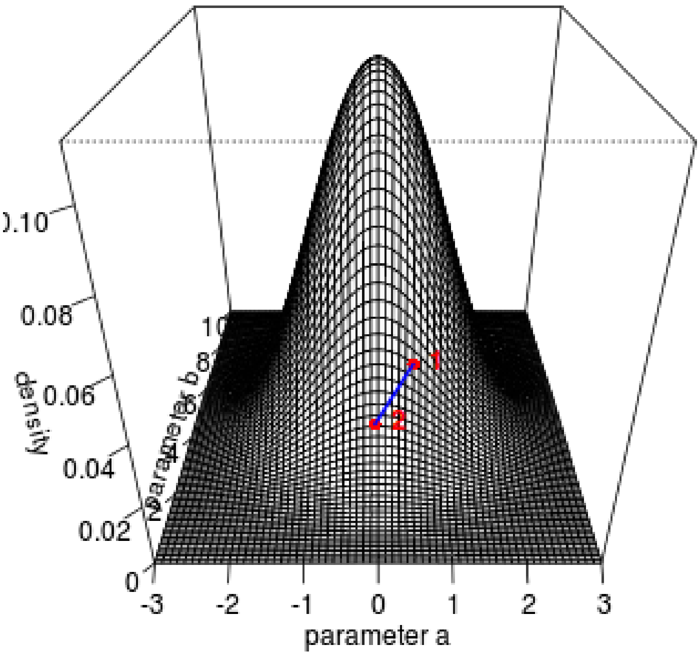


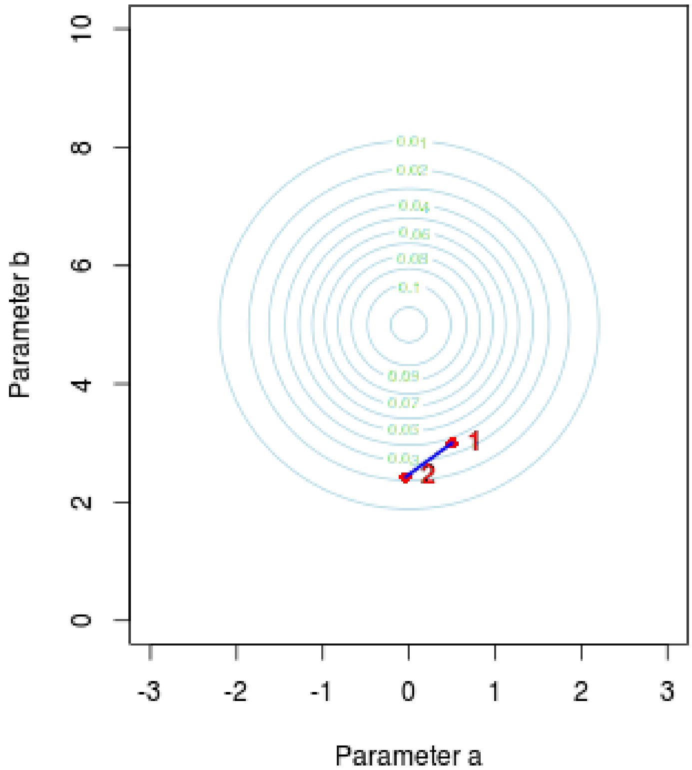


2.9. MCMC sampling

Marchov Chain Monte Carlo sampling

- draw samples proportional to likelihood

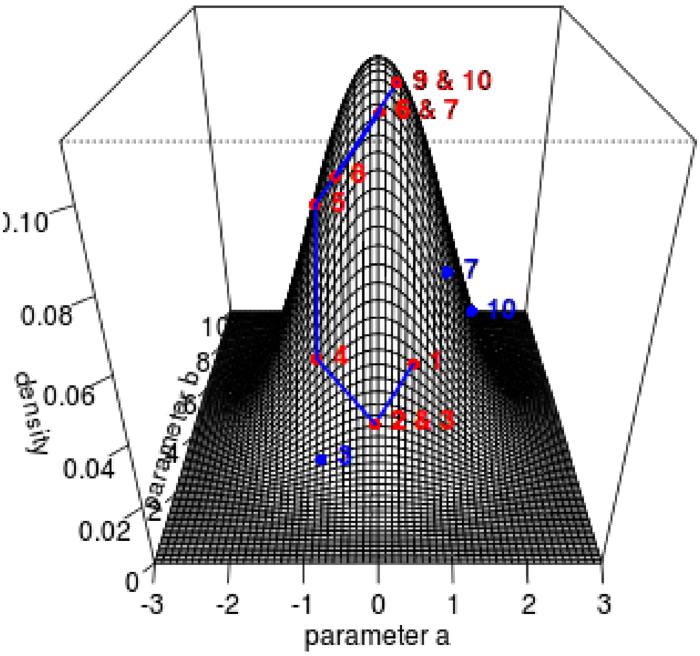


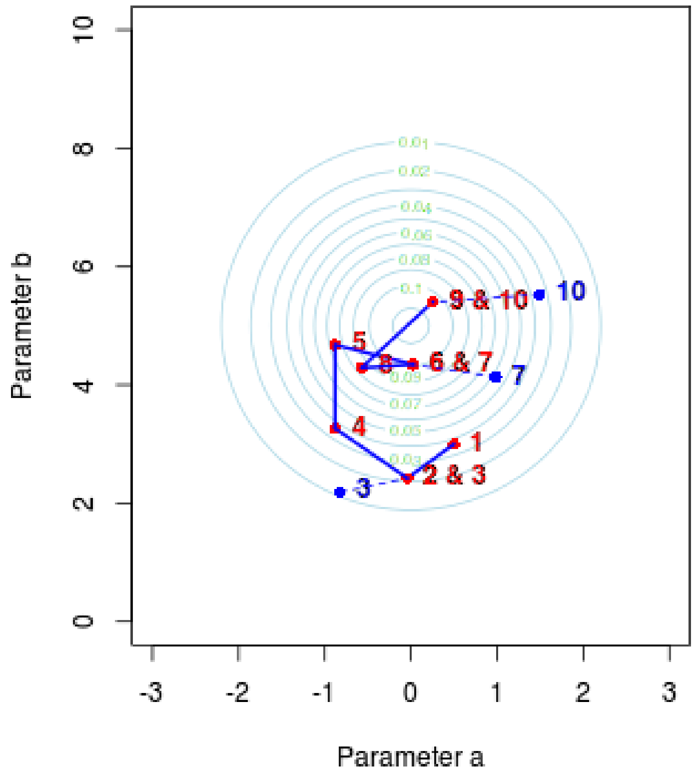


2.10. MCMC sampling

Marchov Chain Monte Carlo sampling

- chain of samples

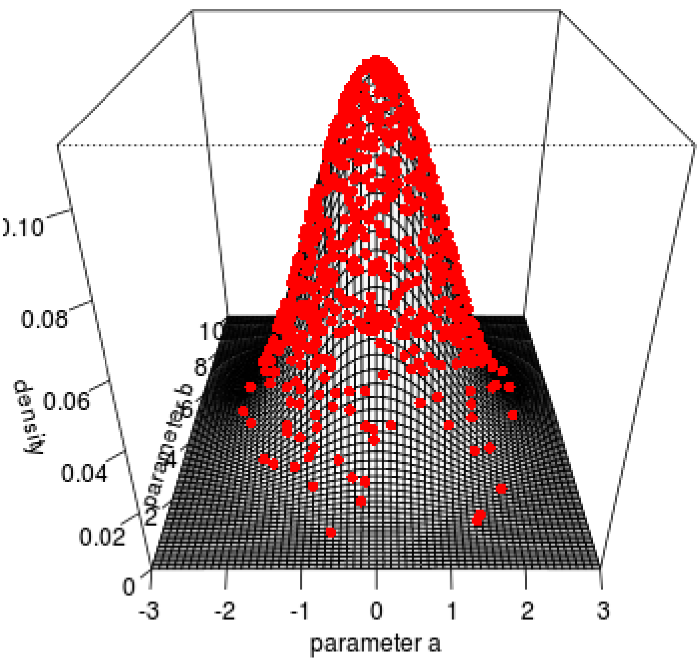


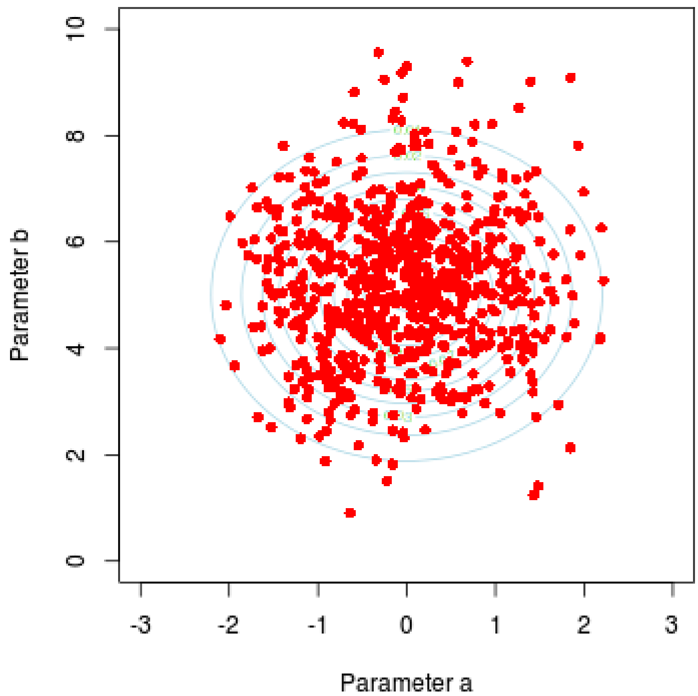


2.11. *MCMC sampling*

Marchov Chain Monte Carlo sampling

- 1000 samples

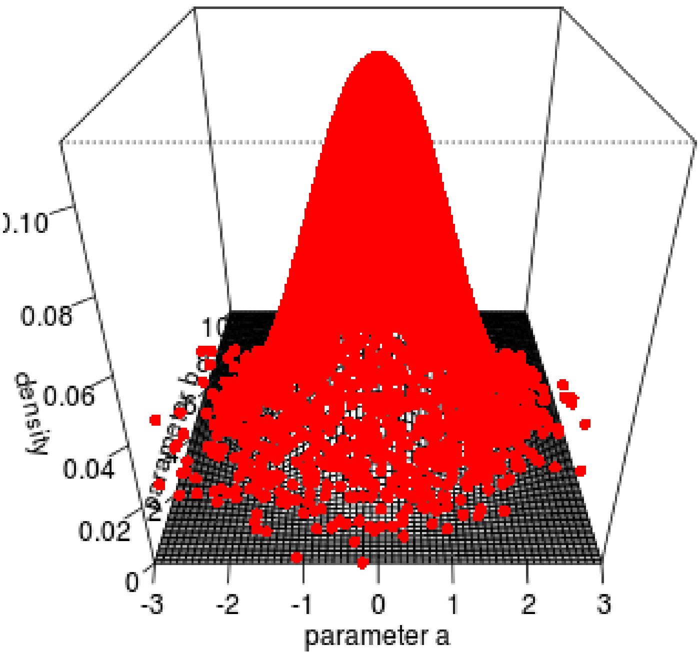


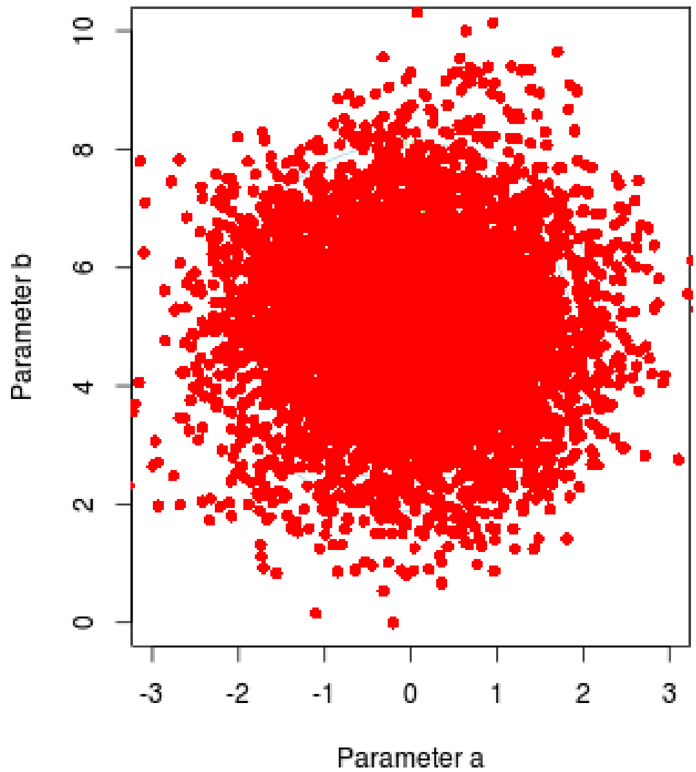


2.12. MCMC sampling

Marchov Chain Monte Carlo sampling

- 10,000 samples





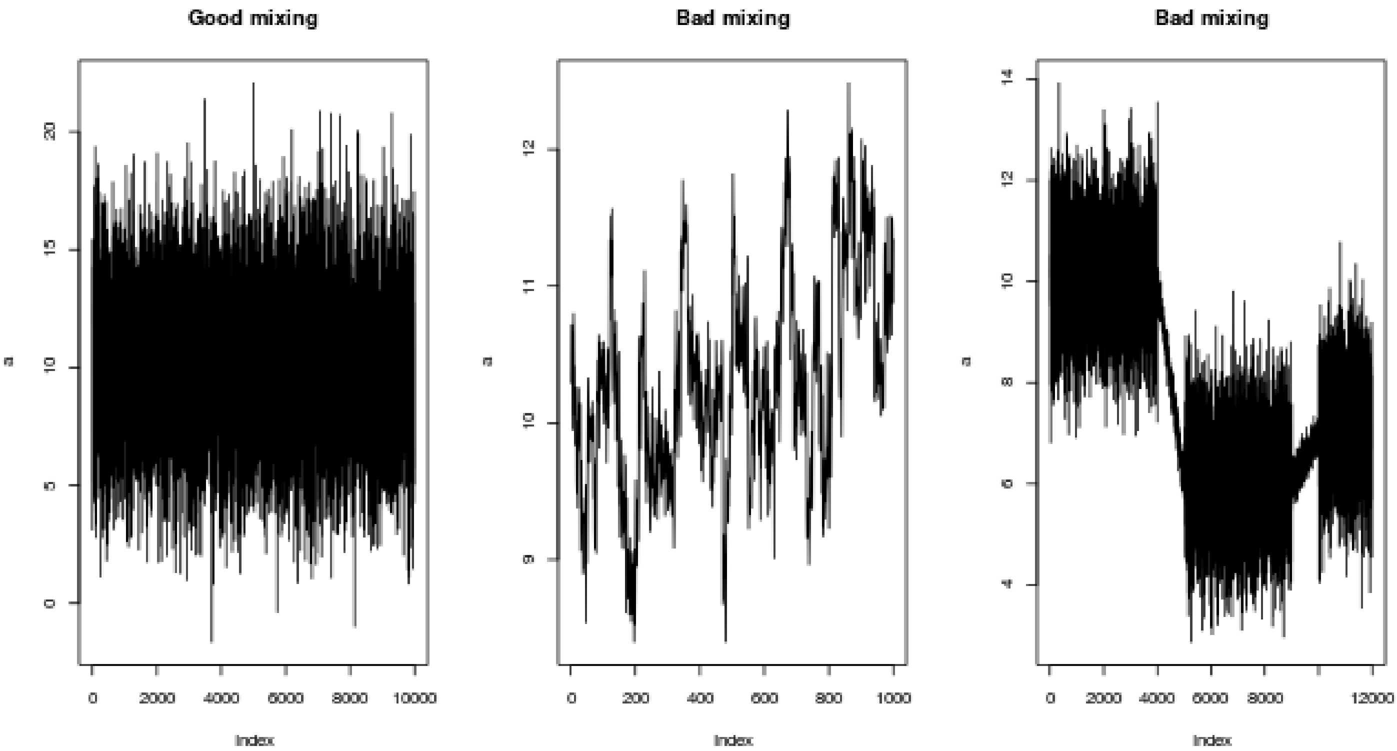
2.13. MCMC sampling

Marchov Chain Monte Carlo sampling

- Aim: samples reflect posterior frequency distribution
- samples used to construct posterior prob. dist.
- the sharper the multidimensional "features" - more samples
- chain should have traversed entire posterior
- inital location should not influence

2.14. MCMC diagnostics

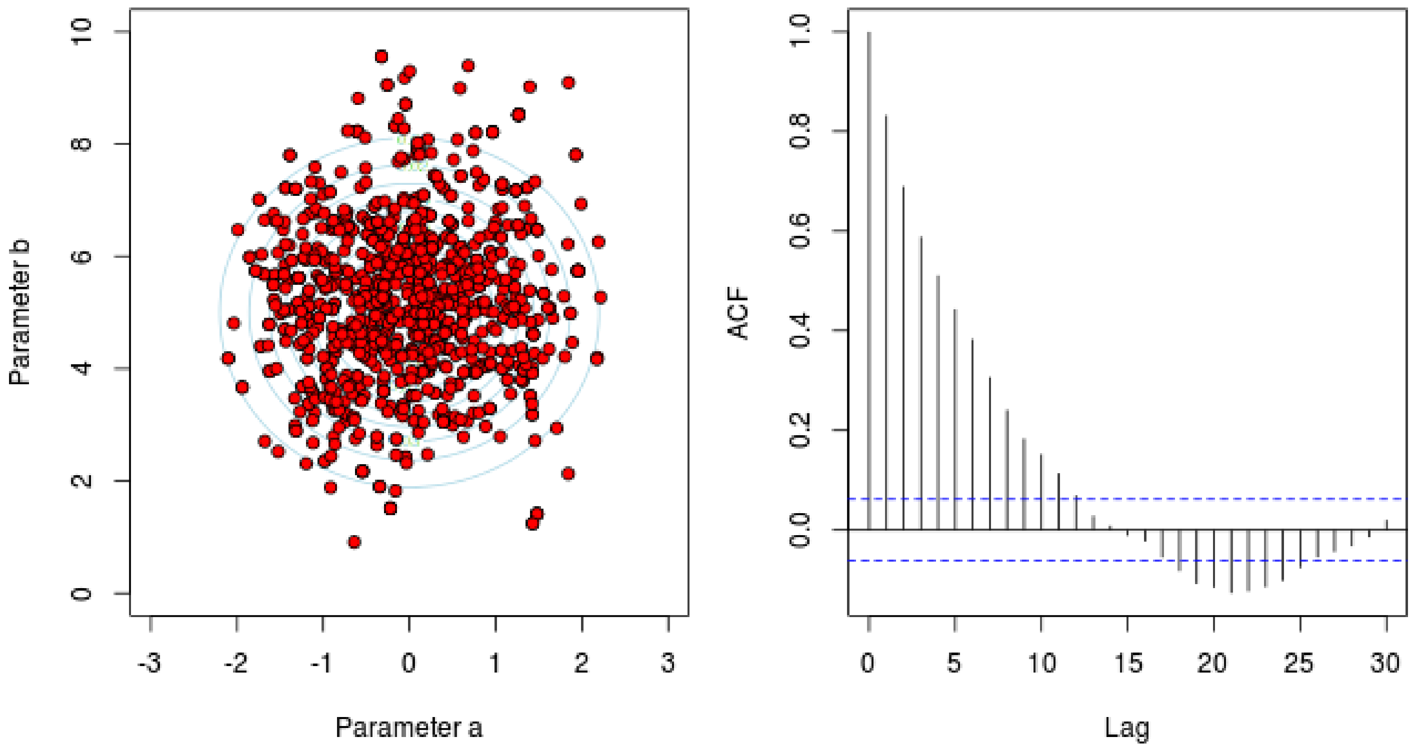
2.14.1. Trace plots



2.15. MCMC diagnostics

2.15.1. Autocorrelation

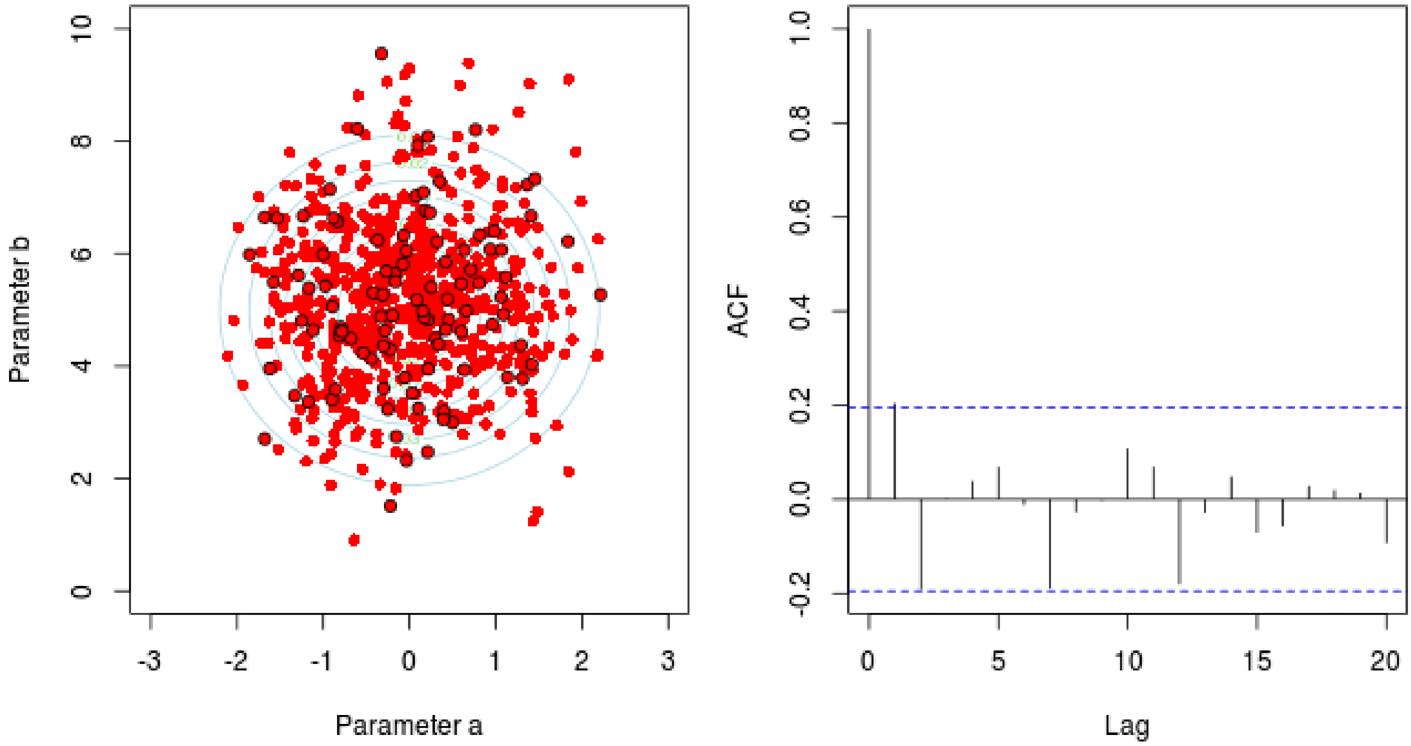
- Summary stats on non-independent values are biased
- Thinning factor = 1



2.16. MCMC diagnostics

2.16.1. Autocorrelation

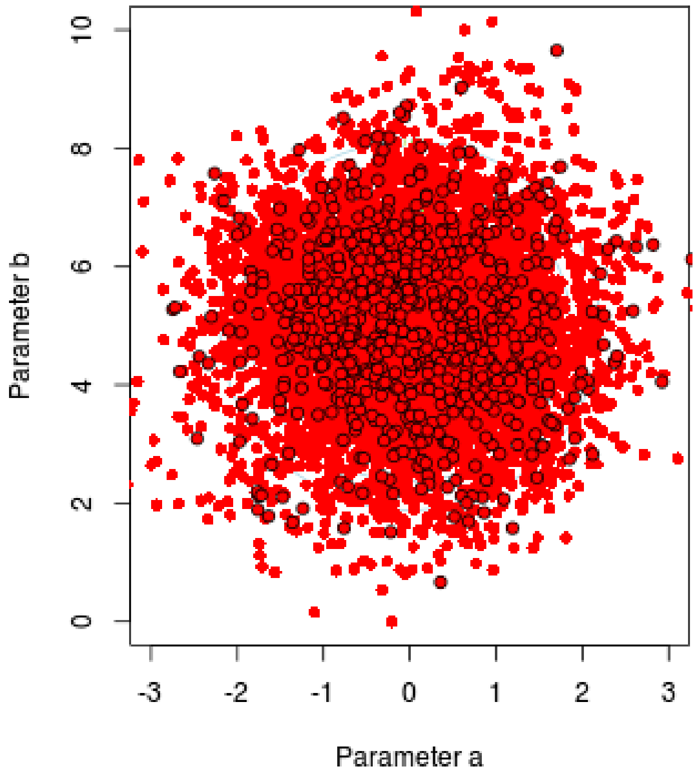
- Summary stats on non-independent values are biased
- Thinning factor = 10

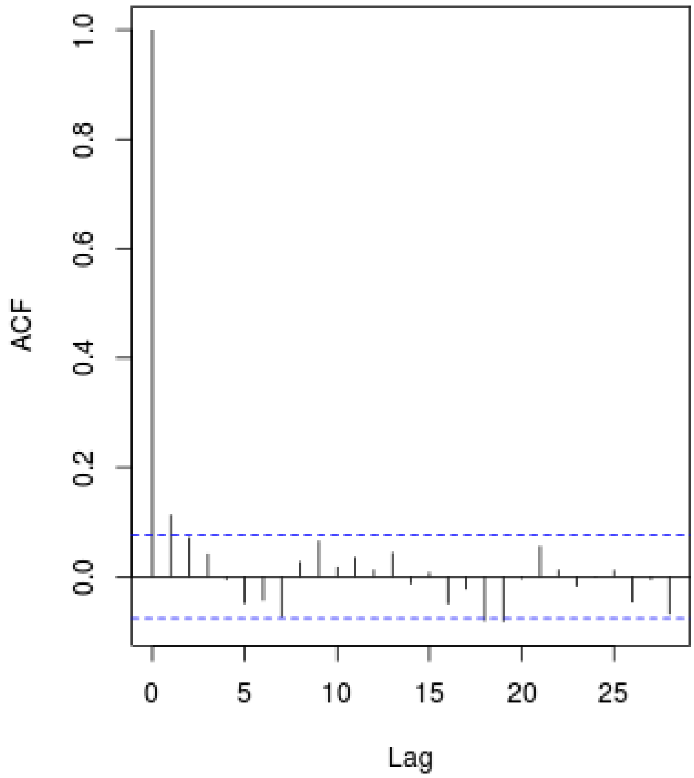


2.17. MCMC diagnostics

2.17.1. Autocorrelation

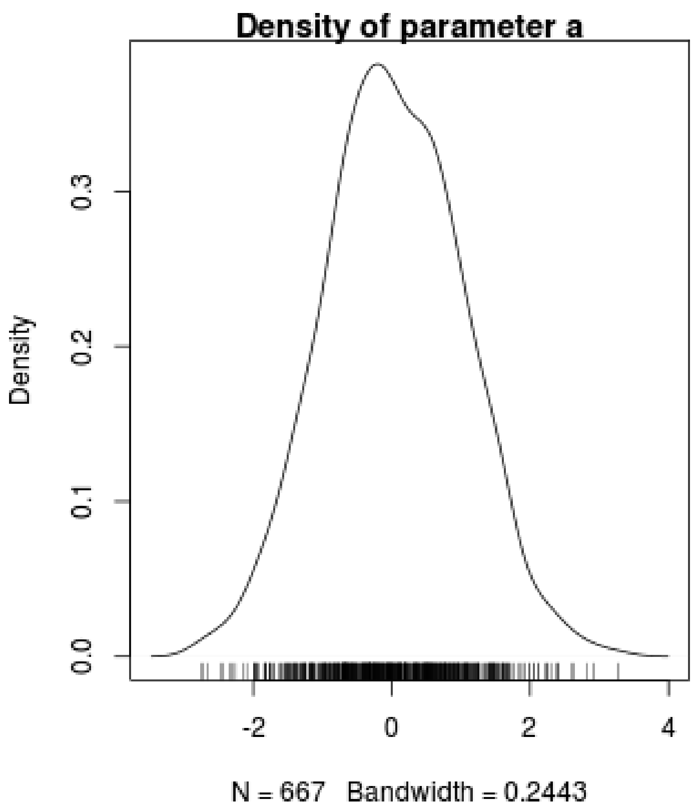
- Summary stats on non-independent values are biased
- Thinning factor = 10, n=10,000

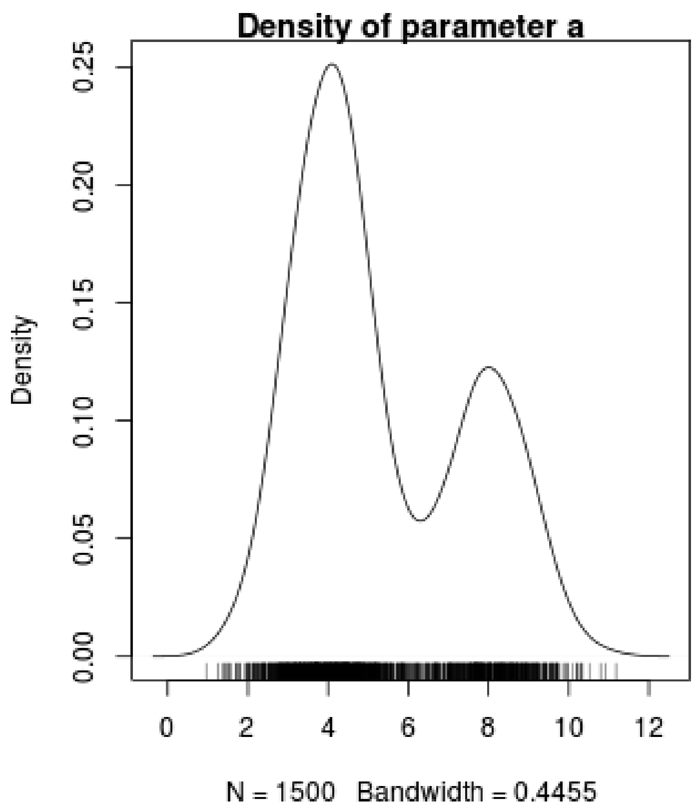




2.18. MCMC diagnostics

2.18.1. Plot of Distributions





2.19. Sampler types

Metropolis-Hastings

http://twiecki.github.io/blog/2014/01/02/visualizing-mcmc/

2.20. Sampler types

Gibbs

2.21. Sampler types

NUTS

2.22. Sampling

- thinning
- burning (warmup)
- chains

2.23. Bayesian software (for R)

- MCMCpack
- winbugs (R2winbugs)
- jags (R2jags)
- stan (rstan, brms)

2.24. BRMS

Extractor	Description
`residuals()`	Residuals
`fitted()`	Predicted values
`predict()`	Predict new responses
`coef()`	Extract model coefficients
`plot()`	Diagnostic plots
`stanplot(,type=)`	More diagnostic plots
`marginal_effects()`	Partial effects
`logLik()`	Extract log-likelihood
`LOO()` and `WAIC()`	Calculate WAIC and LOO
`influence.measures()`	Leverage, Cook's D
`summary()`	Model output
`stancode()`	Model passed to stan
`standata()`	Data list passed to stan

3. Worked Examples

3.1. Worked Examples

```
> fert <- read.csv('../data/fertilizer.csv', strip.white=T)
> fert
```

```
   FERTILIZER YIELD
1          25    84
2          50    80
3          75    90
4         100   154
5         125   148
6         150   169
7         175   206
8         200   244
9         225   212
10        250   248
```

```
> head(fert)
```

```
  FERTILIZER YIELD
1         25    84
2         50    80
3         75    90
4        100   154
5        125   148
6        150   169
```

```
> summary(fert)
```

```
   FERTILIZER         YIELD
 Min.   : 25.00   Min.   : 80.0
 1st Qu.: 81.25   1st Qu.:104.5
 Median :137.50   Median :161.5
 Mean   :137.50   Mean   :163.5
 3rd Qu.:193.75   3rd Qu.:210.5
 Max.   :250.00   Max.   :248.0
```

```
> str(fert)
```

```
'data.frame':   10 obs. of  2 variables:
 $ FERTILIZER: int  25 50 75 100 125 150 175 200 225 250
 $ YIELD     : int  84 80 90 154 148 169 206 244 212 248
```

3.2. Worked Examples

Question: is there a relationship between fertilizer concentration and grass yield?

Linear model:

Frequentist

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Bayesian

$$y_i \sim N(\eta_i, \sigma^2)$$
$$\eta_i = \beta_0 + \beta_1 x_i$$
$$\beta_0 \sim N(0, 1000)$$
$$\beta_1 \sim N(0, 1000)$$
$$\sigma^2 \sim cauchy(0, 4)$$