Workshop 7.2b: Introduction to Bayesian models

Murray Logan

07 Feb 2017

Section 1

Frequentist vs Bayesian

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

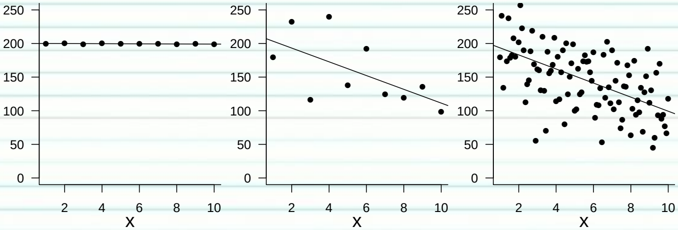
Frequentist

- 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

Frequentist vs Bayesian

	Frequentist	Bayesian
Obs. data	One possible	Fixed, true
Parameters	Fixed, true	Random, distribution
Inferences	Data	Parameters
Probability	Long-run frequency $P(D\|H)$	Degree of belief $P(H\|D)$

Frequentist vs Bayesian

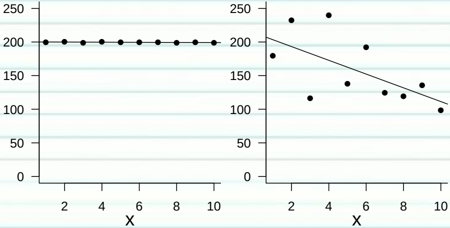


n: 10 Slope: -0.1022 t: -2.3252 p: 0.0485

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

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

Frequentist vs Bayesian



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

Section 2

Bayesian Statistics

Bayesian

BAYES RULE

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

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

Bayesian

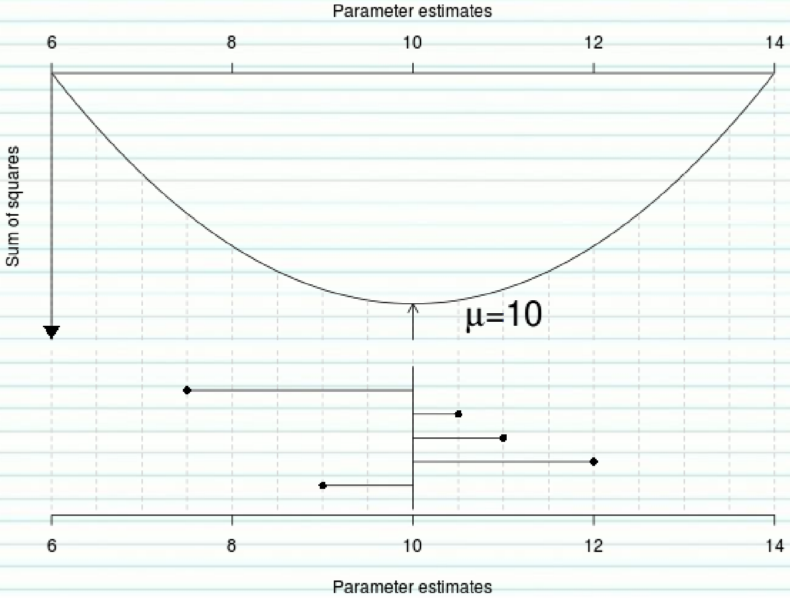
BAYES RULE

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

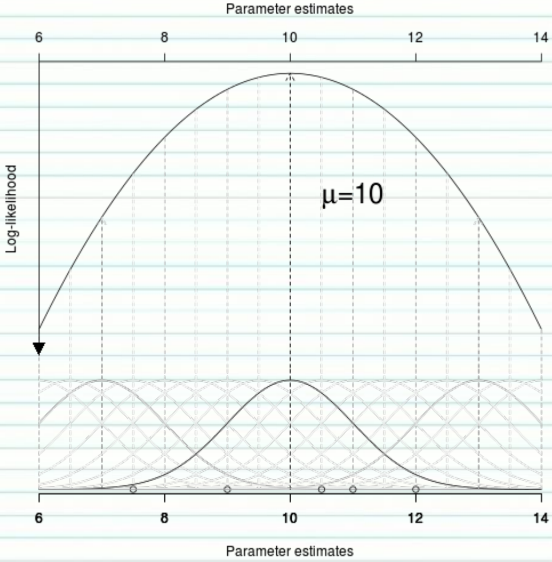
$$\text{posterior belief (probability)} = \frac{\text{likelihood} \times \text{prior probability}}{\text{normalizing constant}}$$

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

Estimation: OLS



Estimation: Likelihood



$P(D \mid H)$

Bayesian

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

Bayesian

- subjectivity?
- intractable

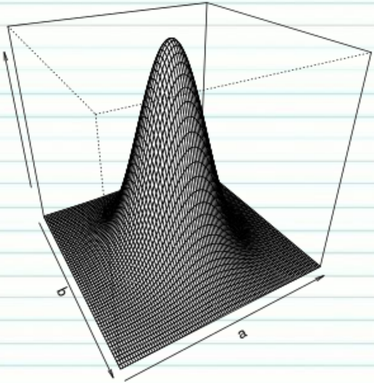
$$P(H \mid D) = \frac{P(D \mid H) \times P(H)}{P(D)}$$

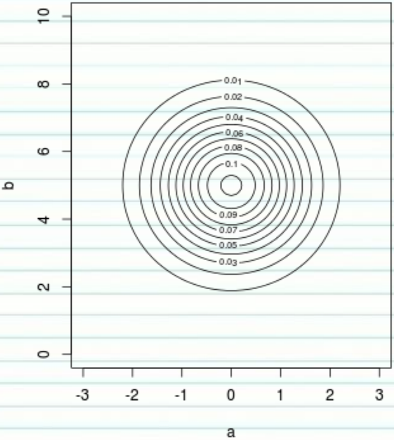
$P(D)$- probability of data from all possible hypotheses

MCMC sampling

Marchov Chain Monte Carlo sampling

- draw samples proportional to likelihood

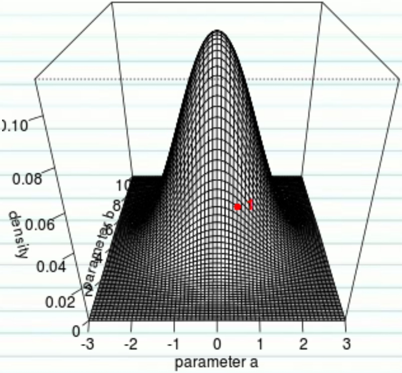


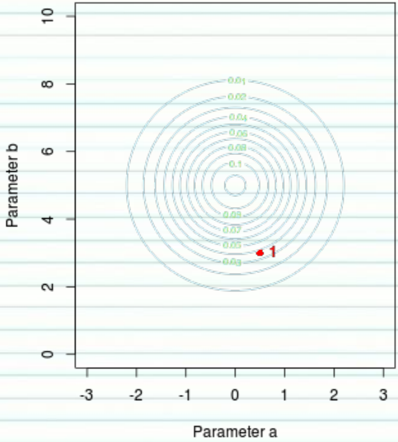


MCMC sampling

Marchov Chain Monte Carlo sampling

- draw samples proportional to likelihood

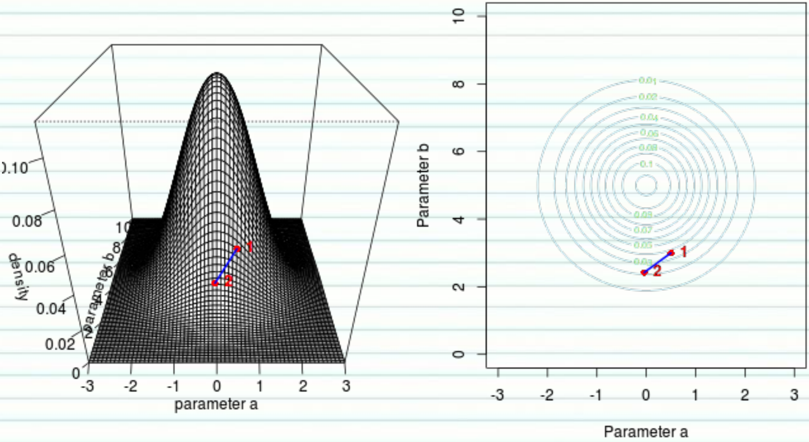




MCMC sampling

Marchov Chain Monte Carlo sampling

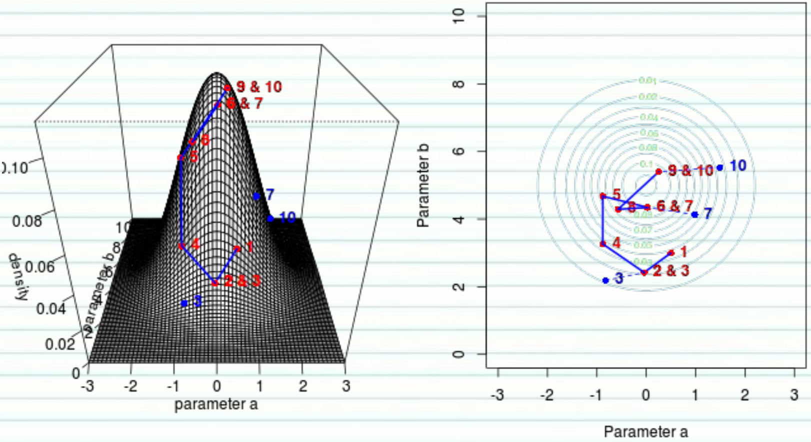
- draw samples proportional to likelihood



MCMC sampling

Marchov Chain Monte Carlo sampling

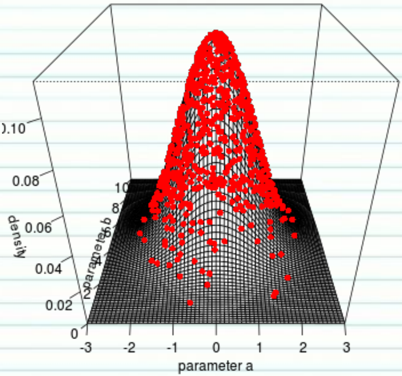
- chain of samples

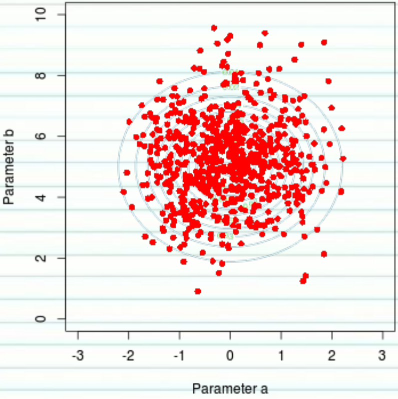


MCMC sampling

Marchov Chain Monte Carlo sampling

- 1000 samples

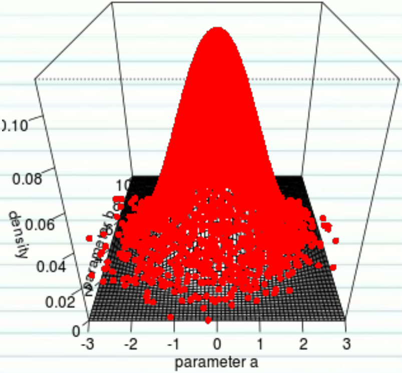


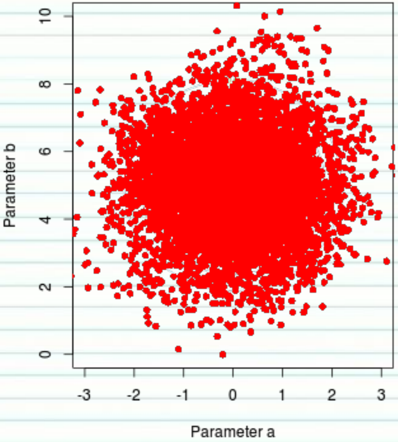


MCMC sampling

Marchov Chain Monte Carlo sampling

- 10,000 samples





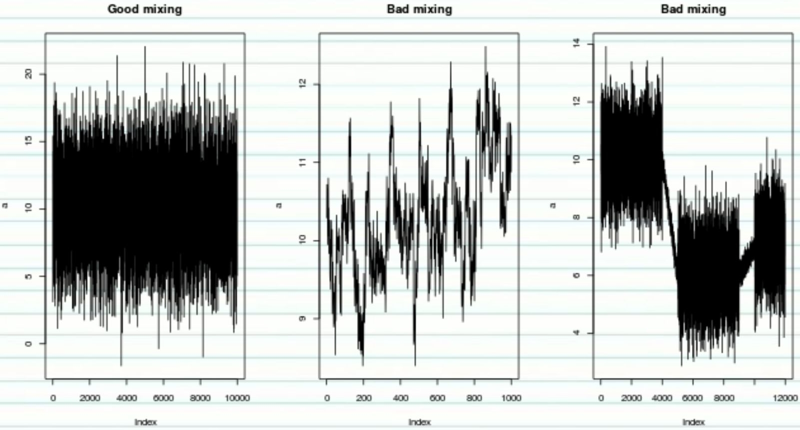
MCMC sampling

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

MCMC diagnostics

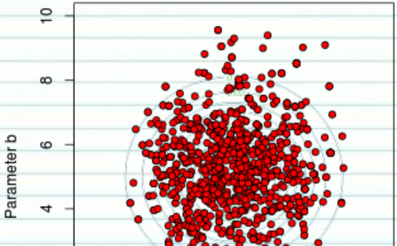
TRACE PLOTS



MCMC diagnostics

AUTOCORRELATION

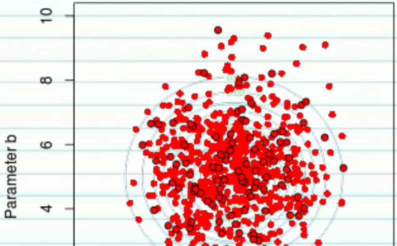
- Summary stats on non-independent values are biased
- Thinning factor = 1

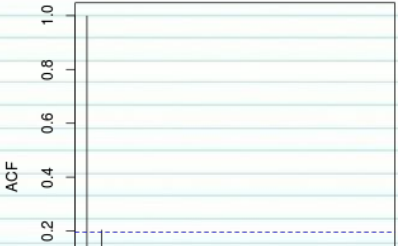


MCMC diagnostics

AUTOCORRELATION

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

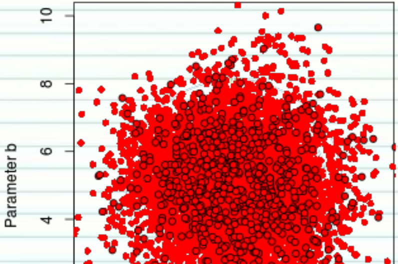


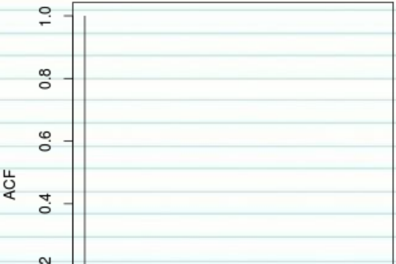


MCMC diagnostics

AUTOCORRELATION

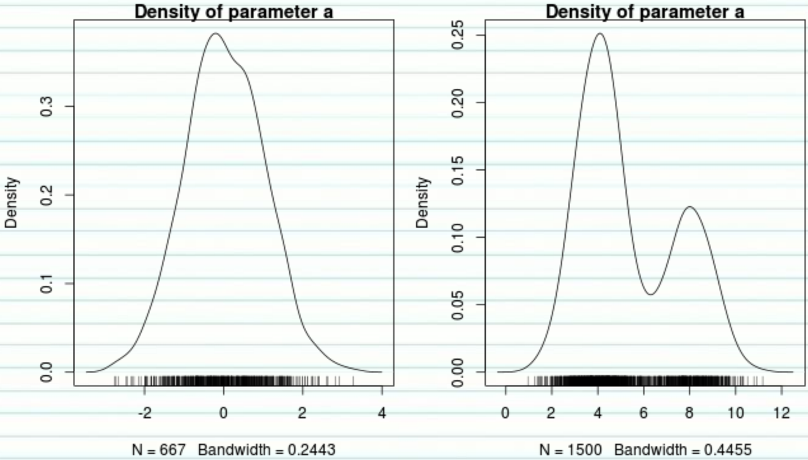
- Summary stats on non-independent values are biased
- Thinning factor = 10, n=10,000





MCMC diagnostics

PLOT OF DISTRIBUTIONS



Sampler types

Metropolis-Hastings

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

Sampler types

Gibbs

Sampler types

NUTS

Sampling

- thinning
- burning (warmup)
- chains

Bayesian software (for R)

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

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

Section 3

Worked Examples

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
```

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

$$\begin{aligned} y_i &\sim \mathtt{N}(\eta_i, \sigma^2) \\ \eta_i &= \beta_0 + \beta_1 x_i \\ \beta_0 &\sim \mathtt{N}(0, 1000) \\ \beta_1 &\sim \mathtt{N}(0, 1000) \\ \sigma^2 &\sim \mathtt{cauchy}(0, 4) \end{aligned}$$