



UPPSALA
UNIVERSITET

• Introduction

Bayesian Statistics and Data Analysis

Lecture 3

Måns Magnusson

Department of Statistics, Uppsala University
Thanks to Aki Vehtari, Aalto University



UPPSALA
UNIVERSITET

● Introduction

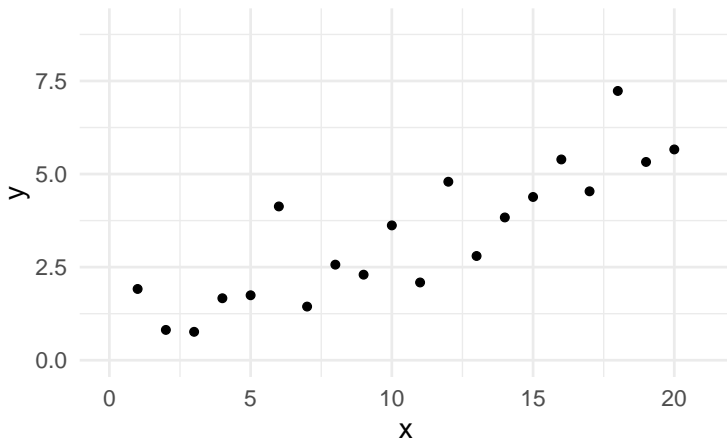
Section 1

Introduction

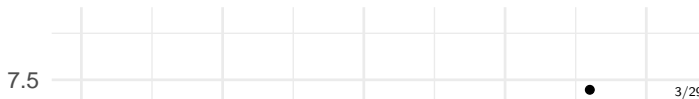


Example of uncertainty in modeling

Data



Posterior mean





- Introduction

- $\theta^{(s)}$ draws from $p(\theta | y)$ can be used
 - for visualization
 - to approximate expectations (integrals)

$$E_{p(\theta|y)}[\theta] = \int \theta p(\theta | y) \approx \frac{1}{S} \sum_{s=1}^S \theta^{(s)}$$



Marginalization

- Joint distribution of parameters

$$p(\theta_1, \theta_2 \mid y) \propto p(y \mid \theta_1, \theta_2)p(\theta_1, \theta_2)$$

- Marginalization

$$p(\theta_1 \mid y) = \int p(\theta_1, \theta_2 \mid y) d\theta_2$$

$p(\theta_1 \mid y)$ is a marginal distribution

- Monte Carlo approximation

$$p(\theta_1 \mid y) \approx \frac{1}{S} \sum_{s=1}^S p(\theta_1, \theta_2^{(s)} \mid y),$$

where $\theta_2^{(s)}$ are draws from $p(\theta_2 \mid y)$



- Introduction

- Marginalization over posterior distribution

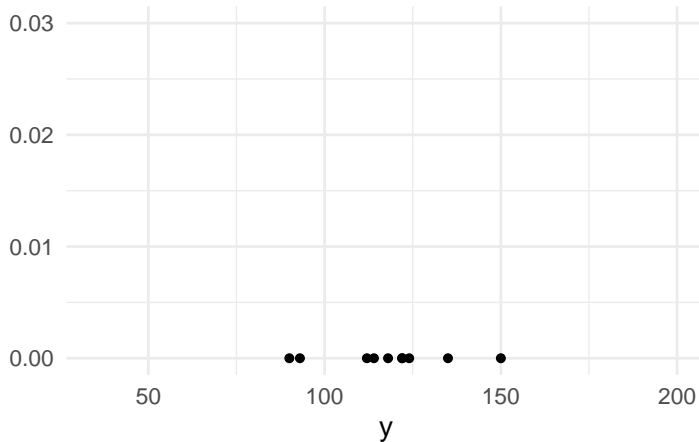
$$\begin{aligned} p(\tilde{y} | y) &= \int p(\tilde{y} | \theta) p(\theta | y) d\theta \\ &= \int p(\tilde{y}, \theta | y) d\theta \end{aligned}$$

$p(\tilde{y} | y)$ is a predictive distribution

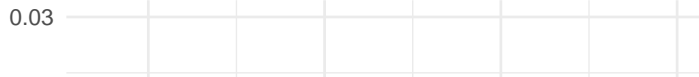


Gaussian example

Data



Gaussian fit with posterior mean

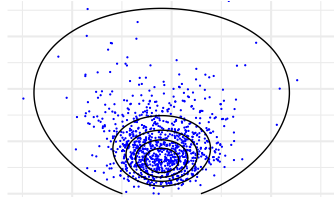




Joint posterior

$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

with $p(\mu, \sigma^2) \propto \sigma^{-2}$



$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-2} \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)$$

$$\begin{aligned} p(\mu, \sigma^2 \mid y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \\ &= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right]\right) \end{aligned}$$

$$\text{where } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2]\right)$$

$$\text{where } s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$



Gaussian - non-informative prior

$$\sum_{i=1}^n (y_i - \mu)^2$$

$$\sum_{i=1}^n (y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^n (y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

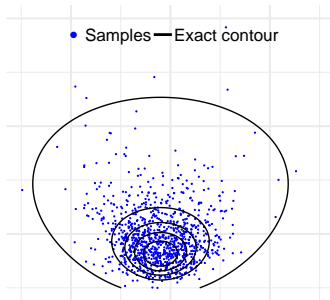
$$\sum_{i=1}^n (y_i^2 - 2y_i\bar{y} + \bar{y}^2) + \sum_{i=1}^n (\mu^2 - 2y_i\mu - \bar{y}^2 + 2y_i\bar{y})$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 + n(\mu^2 - 2\bar{y}\mu - \bar{y}^2 + 2\bar{y}\bar{y})$$

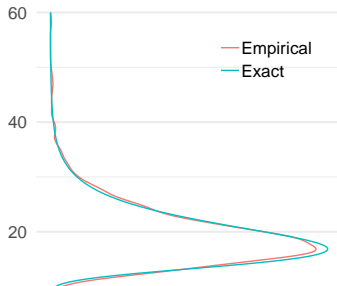
$$\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2$$



Joint posterior



Marginal of sigma





Marginal posterior $p(\sigma^2 | y)$ (easier for σ^2 than σ)

$$\begin{aligned} p(\sigma^2 | y) &\propto \int p(\mu, \sigma^2 | y) d\mu \\ &\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2]\right) d\mu \\ &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} (n-1)s^2\right) \\ &\quad \int \exp\left(-\frac{n}{2\sigma^2} (\bar{y} - \mu)^2\right) d\mu \\ &\quad \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} (y - \theta)^2\right) d\theta = \\ &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} (n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\ &\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \\ p(\sigma^2 | y) &= \text{Inv-}\chi^2(\sigma^2 | n-1, s^2) \end{aligned}$$



Gaussian - non-informative prior

Known mean

$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(n, \nu)$$

$$\text{where } \nu = \frac{1}{n} \sum_{i=1}^n (y_i - \theta)^2$$

Unknown mean

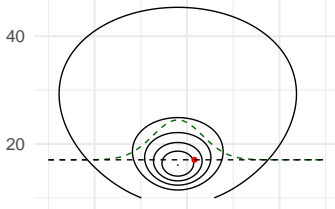
$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(n - 1, s^2)$$

$$\text{where } s^2 = \frac{1}{n - 1} \sum_{i=1}^n (y_i - \bar{y})^2$$

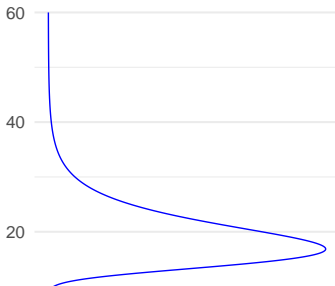


Joint posterior

60
- Exact contour plot - Cond. distribution of μ
Sample from joint post. — Sample from the marg.



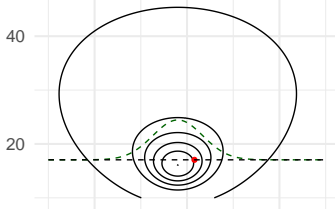
Marginal of sigma



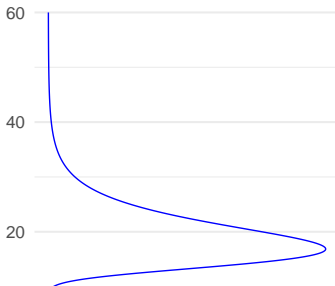


Joint posterior

60
- Exact contour plot - Cond. distribution of μ_1
Sample from joint post. — Sample from the marg.



Marginal of sigma





Marginal posterior $p(\mu \mid y)$

$$\begin{aligned} p(\mu \mid y) &= \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2 \\ &\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2]\right) d\sigma^2 \end{aligned}$$

Transformation

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{and} \quad z = \frac{A}{2\sigma^2}$$

$$p(\mu \mid y) \propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz$$

Recognize gamma integral $\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$

$$\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2}$$

$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

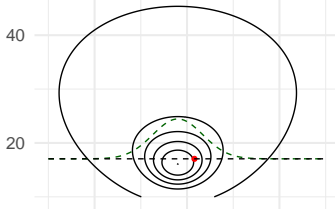
$$p(\mu \mid y) = t_{n-1}(\mu \mid \bar{y}, s^2/n) \quad \text{Student's } t$$



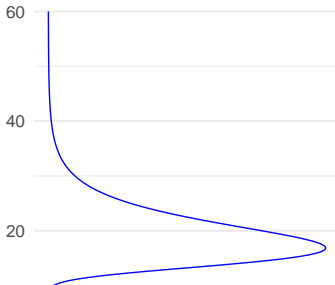
Joint posterior

60
40
20
0

- Exact contour plot - Cond. distribution of μ_1
Sample from joint post. — Sample from the marg.



Marginal of sigma





Posterior predictive distribution given known variance

$$\begin{aligned} p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\ &= \int \mathcal{N}(\tilde{y} \mid \mu, \sigma^2) \mathcal{N}(\mu \mid \bar{y}, \sigma^2/n) d\mu \\ &= \mathcal{N}(\tilde{y} \mid \bar{y}, (1 + \frac{1}{n})\sigma^2) \end{aligned}$$

this is up to scaling factor same as $p(\mu \mid \sigma^2, y)$

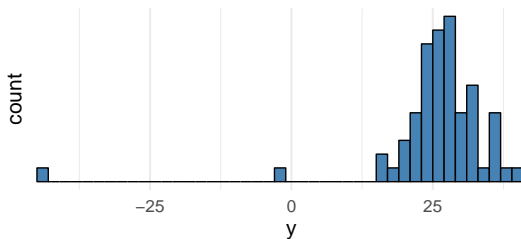
$$p(\tilde{y} \mid y) = t_{n-1}(\tilde{y} \mid \bar{y}, (1 + \frac{1}{n})s^2)$$



Simon Newcomb's light of speed experiment in 1882

Newcomb measured ($n = 66$) the time required for light to travel from his laboratory on the Potomac River to a mirror at the base of the Washington Monument and back, a total distance of 7422 meters.

Newcomb's measurements





Gaussian - conjugate prior

- Conjugate prior has to have a form $p(\sigma^2)p(\mu | \sigma^2)$ (see the chapter notes)
- Handy parameterization

$$\begin{aligned}\mu | \sigma^2 &\sim N(\mu_0, \sigma^2 / \kappa_0) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)\end{aligned}$$

which can be written as

$$p(\mu, \sigma^2) = \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2 / \kappa_0; \nu_0, \sigma_0^2)$$

- μ and σ^2 are a priori dependent
 - if σ^2 is large, then μ has wide prior



Gaussian - conjugate prior

Joint posterior (exercise 3.9)

$$p(\mu, \sigma^2 \mid y) = \text{N-Inv-}\chi^2(\mu_n, \sigma_n^2 / \kappa_n; \nu_n, \sigma_n^2)$$

where

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}$$

$$\kappa_n = \kappa_0 + n$$

$$\nu_n = \nu_0 + n$$

$$\nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n - 1) s^2 + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)^2$$



UPPSALA
UNIVERSITET

Multinomial model for categorical data

● Introduction

- Extension of binomial
- Observation model

$$p(y \mid \theta) \propto \prod_{j=1}^k \theta_j^{y_j},$$

- BDA3 p. 69–



- Introduction

- Observation model

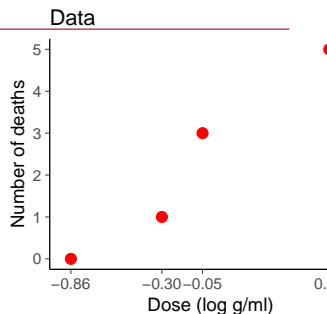
$$p(y \mid \mu, \Sigma) \propto |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} (y - \mu)^T \Sigma^{-1} (y - \mu) \right),$$

- BDA3 p. 72–
- New recommended LKJ-prior mentioned in Appendix A, see more in Stan manual



Bioassay

Dose, x_i (log g/ml)	Number of animals, n_i	Number of deaths, y_i
-0.86	5	0
-0.30	5	1
-0.05	5	3
0.73	5	5



Find out lethal dose 50% (LD50)

- used to classify how hazardous chemical is
- 1984 EEC directive has 4 levels (see the chapter notes)

Bayesian methods help to

- reduce the number of animals needed
- easy to make sequential experiment and stop as soon as desired accuracy is obtained

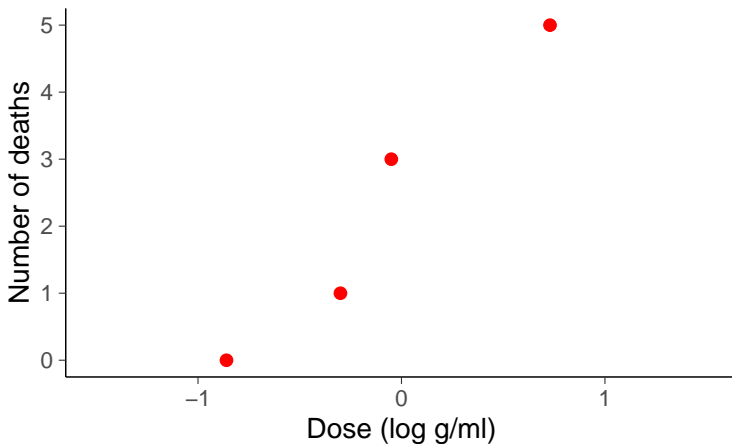


UPPSALA
UNIVERSITET

• Introduction

Bioassay

Data



Linear fit



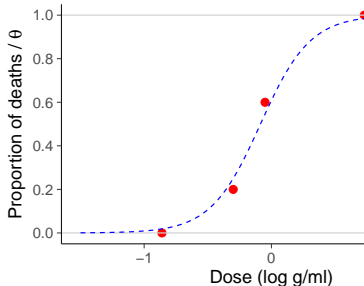


Bioassay

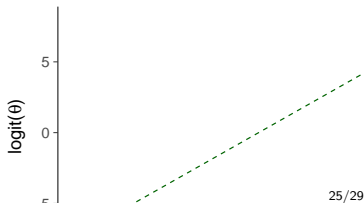
$$y_i \mid \theta_i \sim \text{Bin}(\theta_i, n_i)$$
$$\text{logit}(\theta_i) = \log\left(\frac{\theta_i}{1 - \theta_i}\right)$$
$$= \alpha + \beta x_i$$

$$\theta_i = \frac{1}{1 + \exp(-(\alpha + \beta x_i))}$$

Logistic regression fit



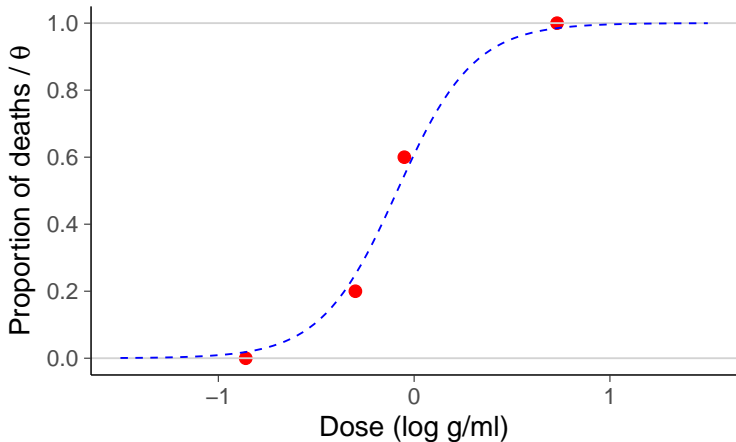
Logistic regression in latent space





Bioassay

Logistic regression fit



Posterior draws





Bioassay posterior

Binomial model

$$y_i \mid \theta_i \sim \text{Bin}(\theta_i, n_i)$$

Link function

$$\text{logit}(\theta_i) = \alpha + \beta x_i$$

Likelihood

$$p(y_i \mid \alpha, \beta, n_i, x_i) \propto \theta_i^{y_i} [1 - \theta_i]^{n_i - y_i}$$

$$p(y_i \mid \alpha, \beta, n_i, x_i) \propto [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i} [1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}$$

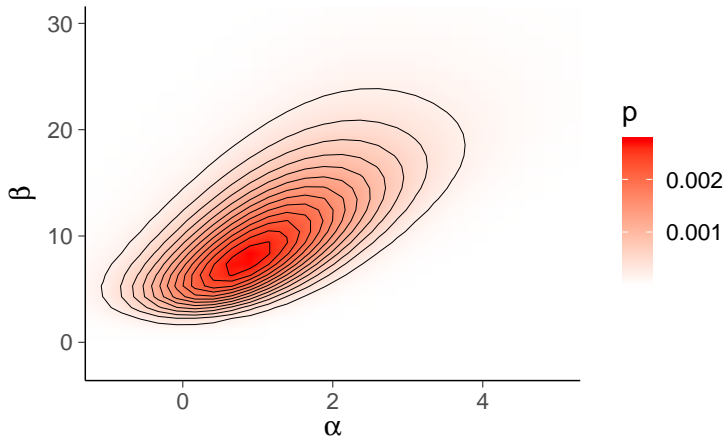
Posterior (with uniform prior on α, β)

$$p(\alpha, \beta \mid y, n, x) \propto p(\alpha, \beta) \prod_{i=1}^n p(y_i \mid \alpha, \beta, n_i, x_i)$$



Bioassay

Posterior density evaluated in a grid



Posterior density evaluated in a grid

30



Grid sampling

- Draws can be used to estimate expectations, for example

$$E[x_{LD50}] = E[-\alpha/\beta] \approx \frac{1}{S} \sum_{s=1}^S \frac{\alpha^{(s)}}{\beta^{(s)}}$$

- Instead of sampling, grid could be used to evaluate functions directly, for example

$$E[-\alpha/\beta] \approx \sum_{t=1}^T w_{\text{cell}}^{(t)} \frac{\alpha^{(t)}}{\beta^{(t)}},$$

where $w_{\text{cell}}^{(t)}$ is the normalized probability of a grid cell t , and $\alpha^{(t)}$ and $\beta^{(t)}$ are center locations of grid cells

- Grid sampling gets computationally too expensive in high dimensions