# Statistics 568 Bayesian Analysis

## Multiparameter Models

Ruobin Gong

Department of Statistics
Rutgers University

Week 2 - 01/28/21

# Noninformative and weakly informative priors

Sometimes, You may want Your prior distribution to play a minimal role in the posterior distribution. *Let the data speak for themselves*.

- "Vague", "flat", or "diffuse" prior;
- **Reference prior**;

Perhaps **weakly informative priors** should be preferred. Or, use hierarchical modeling and sensitivity analysis.

# Proper vs improper prior distributions

$p(\theta|y) \sim N(y, \sigma^2)$

In the Normal location model

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

a flat prior for $\theta$, i.e. $p(\theta) \propto 1$, is *improper*.

### Definition

A prior density is proper if it does not depend on data and integrates to 1.

- Non-informative priors can often be improper.
- Improper prior distributions can (but not always) lead to proper posterior distributions. **Avoid improper posterior distributions at all cost**!*

*Unless it is Your intention to specify a merely finitely additive probability model.

Dawid, Stone & Zidek (1973) J.R.SSB. Marginalization Paradox

**Jeffreys' invariance principle.** Suppose $\phi = \phi(\theta)$ is a one-to-one transformation of the parameter. A noninformative prior $p(\theta)$ should expresses equivalent probability assignment for the transformed $\phi$. That is, we want a prior that's invariant to parametrization:

$$p(\phi) = p(\theta) \left| \frac{d\theta}{d\phi} \right| .$$

# Reference priors

This means that one should set

$$p(\theta) \propto |J(\theta)|^{1/2}$$

where

$$J(\theta) = E\left(\left[\frac{d \log p(y \mid \theta)}{d\theta}\right]^2 \mid \theta\right) = -E\left(\frac{d^2 \log p(y \mid \theta)}{d\theta^2} \mid \theta\right)$$

is the Fisher information for $\theta$. *Why?*

$$J(\phi) = -E\left(\frac{\partial^2 \log p(y \mid \phi)}{\partial \phi^2}\right)$$

$$= -E\left(\frac{\partial^2 \log p(y \mid \theta = \phi^{-1}(\phi))}{\partial \theta^2}\left|\frac{\partial \theta}{\partial \phi}\right|^2\right) = J(\theta)\left|\frac{\partial \theta}{\partial \phi}\right|^2$$

# Reference priors

**Example.** For Binomial model $y \mid \theta \sim Bin\,(n, \theta)$, Jeffreys' prior is

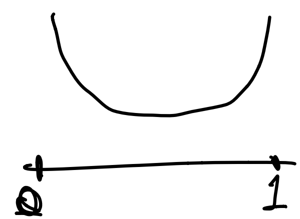
$$\theta \sim Beta(1/2, 1/2).$$

$$\log p(y \mid \theta) \propto y \log \theta + (n-y) \log(1-\theta)$$

$$-\frac{d^2}{d\theta^2} \log p(y \mid \theta) \propto \frac{y}{\theta^2} + \frac{n-y}{(1-\theta)^2}$$

$$J(\theta) \propto \frac{n}{\theta(1-\theta)}$$

$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$$



# Reference priors

**Note.**

- Jeffreys' prior may be improper;
- Extension to multiparameter models may be controversial;
- Jeffreys' prior is entirely determined by the sampling model. Is it still a prior in the truest sense?

# Multiparameter models: marginalization

Model parameter is multi-dimensional: $\theta = (\theta_1, \theta_2)$, in which $\theta_1$ is of interest, and $\theta_2$ is a **nuisance parameter**.

The **marginal posterior** distribution for $\theta_1$:

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

# Multiparameter models: marginalization

The posterior predictive distribution is in a sense a marginal posterior distribution:

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

# Location-scale Normal model, noninformative prior

Sampling model:

$$y \mid \mu, \sigma^2 \sim N(\mu, \sigma^2).$$

A noninformative prior that is uniform on $(\mu, \log \sigma)$:

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

$$\Downarrow$$

*Is this prior proper?*

$$\begin{cases} p(\mu) \propto 1 \\ p(\sigma^2) \propto \frac{1}{\sigma^2} \end{cases}$$

$a = \sigma^2$

$p(a) \propto \frac{1}{a}$

## Location-scale Normal model, noninformative prior

**Joint posterior**:

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

where

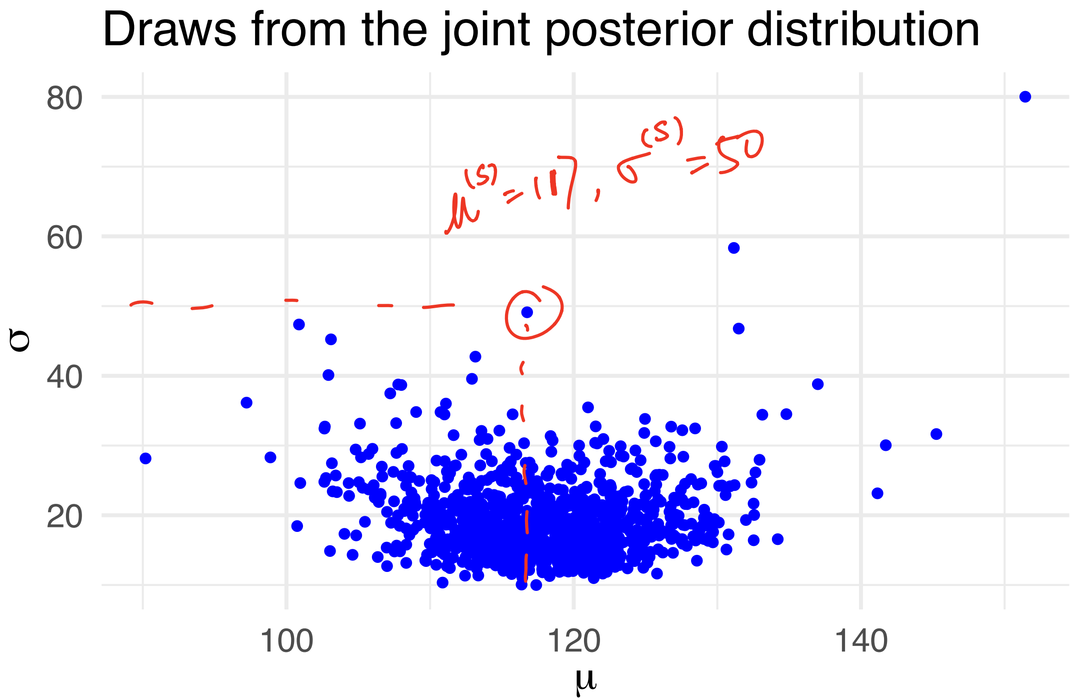
$$
\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2
$$

are the sample mean and sample variance.

# Location-scale Normal model, noninformative prior

Joint posterior draws

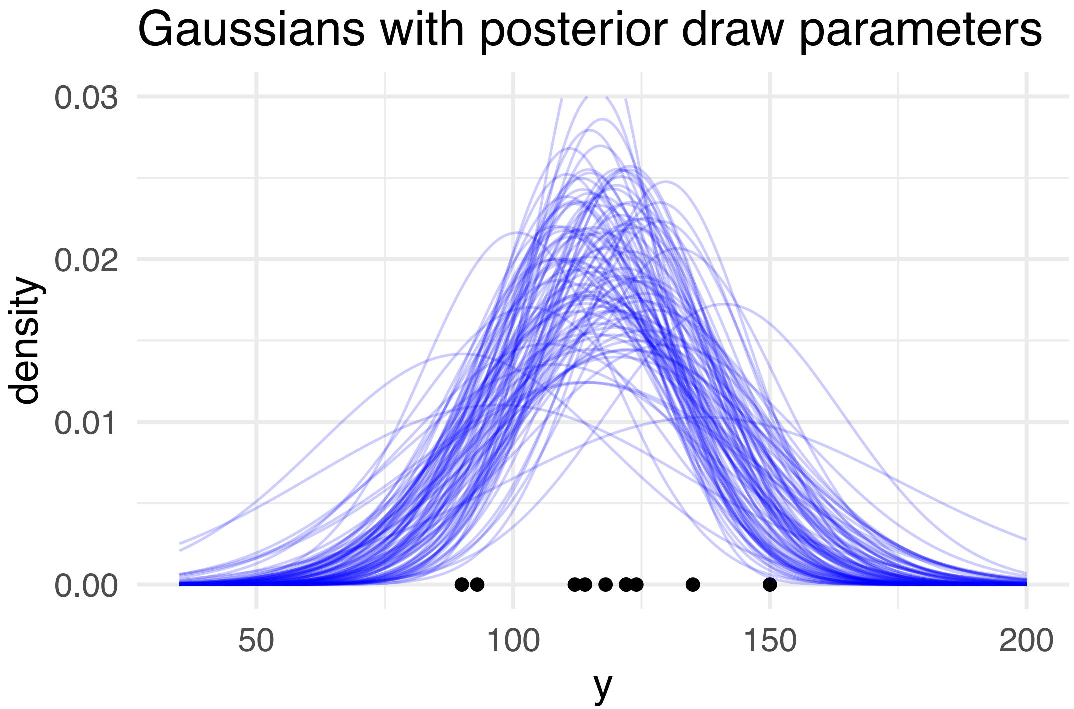
$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$



# Location-scale Normal model, noninformative prior

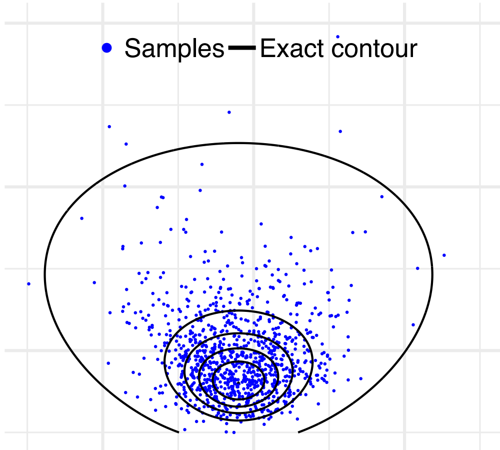
Joint posterior draws

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

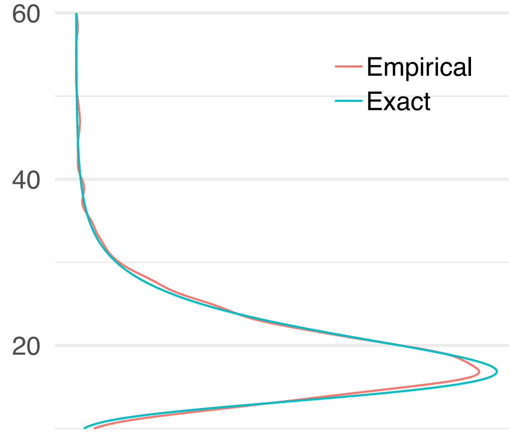


# Location-scale Normal model, noninformative prior

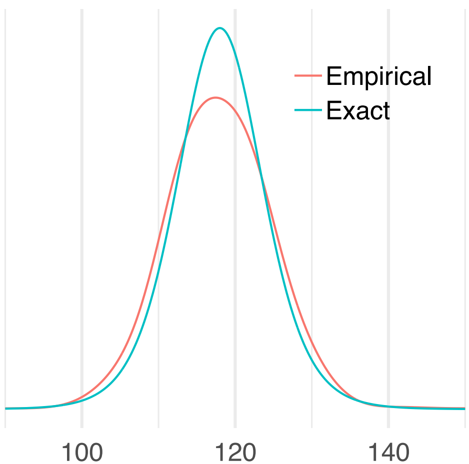
Joint posterior




Marginal of sigma




Marginal of mu




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

marginals

$$p(\mu \mid y) = \int p(\mu, \sigma \mid y) d\sigma$$

$$p(\sigma \mid y) = \int p(\mu, \sigma \mid y) d\mu$$

# Location-scale Normal model, noninformative prior

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

Univariate **factorization** of the joint posterior:

- $\mu \mid \sigma^2, y \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$
- $\sigma^2 \mid y \sim \text{Inv-}\chi^2(n-1, s^2)$

# Location-scale Normal model, noninformative prior

$$p(\sigma^2 \mid y) \propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu$$

$$\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n}$$

$$\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)$$

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

# Location-scale Normal model, noninformative prior

**Marginal posterior** for $\mu$:

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$

$$\propto \int_0^\infty \sigma^{-n-2} \exp\Bigg( \underbrace{-\frac{1}{2\sigma^2}\Big[\underbrace{(n-1)s^2 + n(\bar{y}-\mu)^2}_{A}\Big]}_{z} \Bigg) d\sigma^2$$

$$\propto A^{-n/2} \underbrace{\int_0^\infty z^{(n-2)/2} \exp(-z)\, dz}_{\text{unnormalized Gamma density}}$$

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

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

# Location-scale Normal model, noninformative prior

**Posterior predictive** distribution for a future observation $\tilde{y}$:

$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \mu, \sigma^2, y) p(\mu, \sigma^2 \mid y) d\mu d\sigma^2.$$

A more clever arrangement:

$$p(\tilde{y} \mid y) = \int \underbrace{p(\tilde{y} \mid \sigma^2, y)}_{\parallel} p(\sigma^2 \mid y) d\sigma^2$$

$$\int p(\tilde{y} \mid \mu, \sigma^2, y) p(\mu \mid \sigma^2, y) d\mu$$

$$\sim N\left(\bar{y}, \left(1 + \tfrac{1}{n}\right)\sigma^2\right)$$

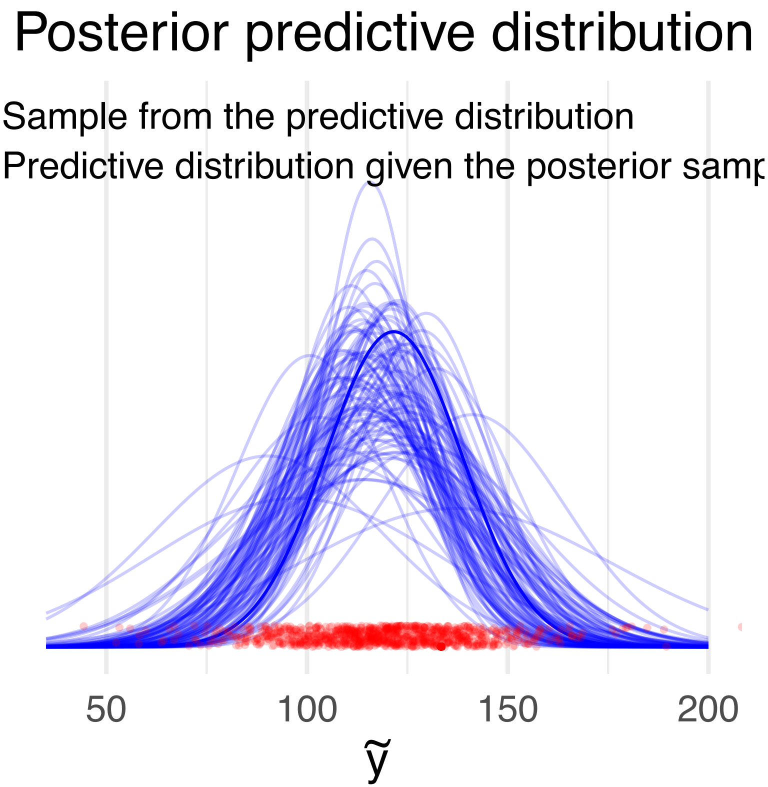
We have

$$\tilde{y} \mid y \sim t_{n-1}\left(\bar{y}, \left(1 + \frac{1}{n}\right) s^2\right).$$

# Location-scale Normal model, noninformative prior

Monte Carlo strategy for posterior predictive draws:

$$
\begin{aligned}
\mu^{(s)}, \sigma^{(s)} &\sim p(\mu, \sigma \mid y) \\
\tilde{y}^{(s)} \mid \mu^{(s)}, \sigma^{(s)} &\sim p(\tilde{y} \mid \mu^{(s)}, \sigma^{(s)})
\end{aligned}
$$



# Location-scale Normal model, conjugate prior

Parameterize the conjugate prior for the Normal location-scale model as

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

jointly denoted as

$$\left(\mu, \sigma^2\right) \sim \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2).$$

Notice that the two parameters are *apriori* dependent.

# Location-scale Normal model, conjugate prior

Joint posterior

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

where

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y}, \\
\kappa_n &= \kappa_0 + n, \\
\nu_n &= \nu_0 + n, \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2.
\end{aligned}$$

# Location-scale Normal model, conjugate prior

Joint posterior

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

which upon factorization,

$$\begin{aligned} \mu \mid \sigma^2, y &\sim N(\mu_n, \sigma^2/\kappa_n), \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2). \end{aligned}$$

The marginal posterior for $\mu$

$$\mu \mid y \sim t_{\nu_n}\left(\mu_n, \frac{\sigma_n^2}{\kappa_n}\right).$$

# Example: Bioassay

An experiment administers various dose levels of some substance to batches of animals, and record the outcomes. Data is of the form

$$(x_i, n_i, y_i), i = 1, \cdots, k$$

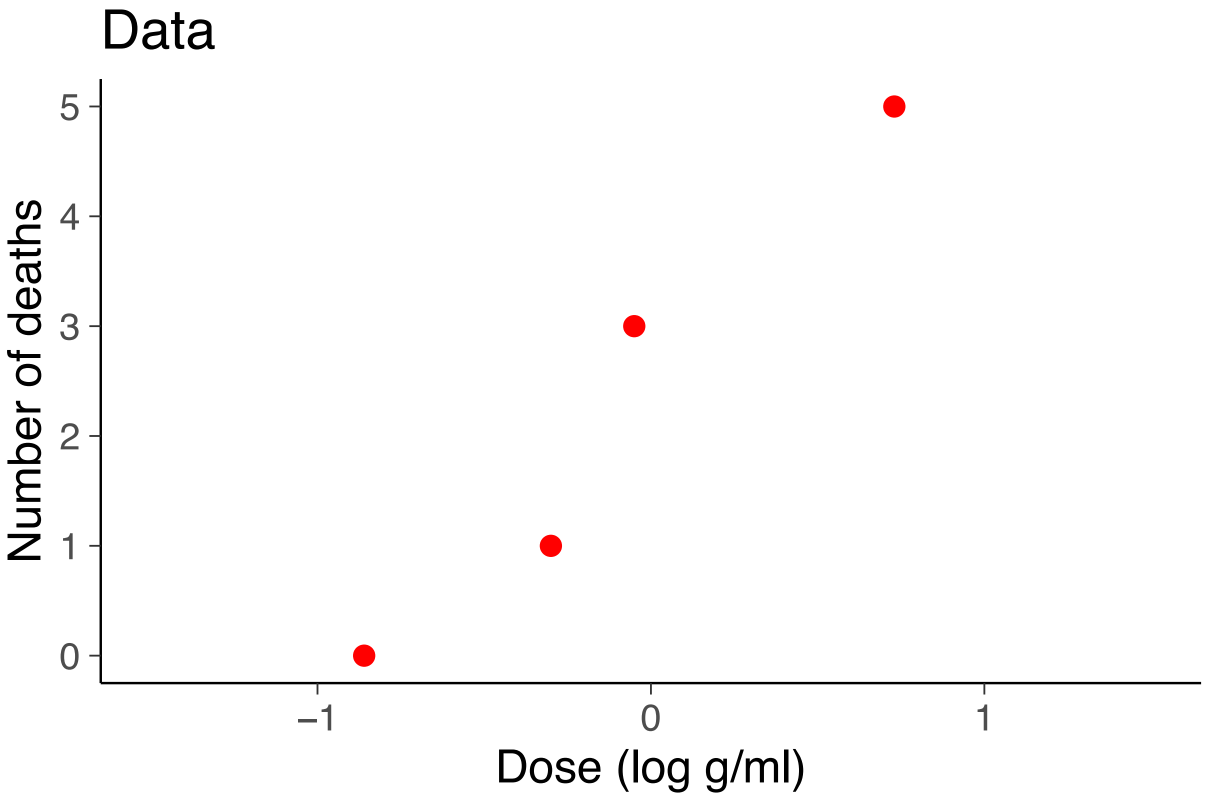
where

- $x_i$: the $i$th of $k$ dose levels,
- $n_i$: # animals in batch $i$;
- $y_i$: # animals in batch $i$ with positive outcome, e.g. death:'( .

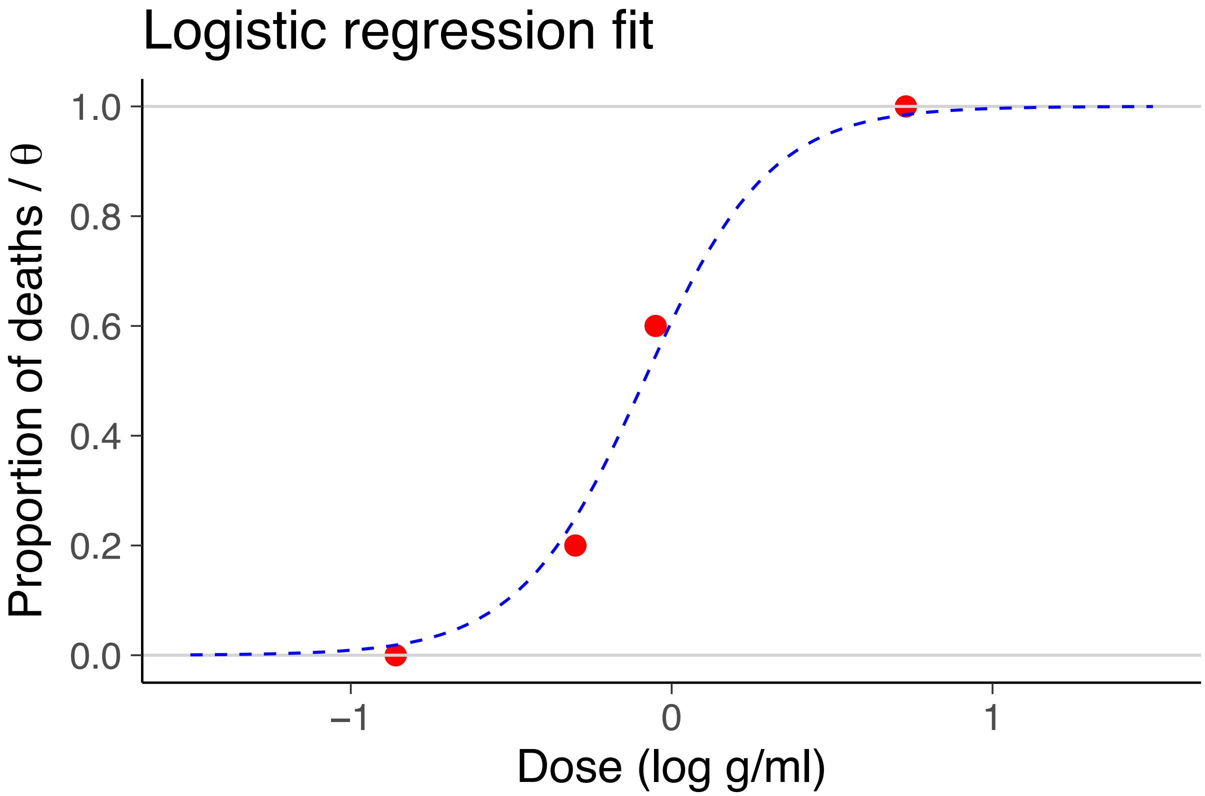
| Dose, $x_i$ (log g/ml) | Number of animals, $n_i$ | Number of deaths, $y_i$ |
|---|---|---|
| -0.86 | 5 | 0 |
| -0.30 | 5 | 1 |
| -0.05 | 5 | 3 |
| 0.73 | 5 | 5 |

Table: Table 3.1 of BDA3

# Example: Bioassay



# Example: Bioassay



# Example: Bioassay

Consider the **logistic regression** model (with one predictor per observation). Sampling model:

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

# Example: Bioassay

**Likelihood**:

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

**Posterior**:

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

With uniform prior $p(\alpha, \beta) \propto 1$, this is not simply analytical. We will compute the joint posterior at a grid of points $(\alpha, \beta)$, and use it to produce posterior draws $(\alpha^{(s)}, \beta^{(s)})$.

# Example: Bioassay - LD50

**LD50** is the dose level, denote as $x_{\text{LD50}}$, at which the probability of death is 50%. For the logistic model,

$$E\left(\frac{y_i}{n_i}\right) = \text{logit}^{-1}(\alpha + \beta x_{\text{LD50}}) = 0.5,$$

that is,

$$x_{\text{LD50}} = -\alpha/\beta.$$

Posterior draws $(\alpha^{(s)}, \beta^{(s)})$ can be used, e.g. to estimate

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

See `R`.