

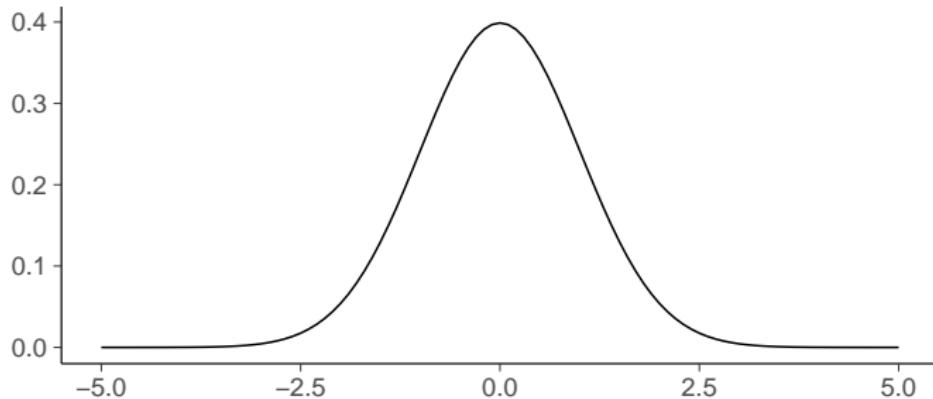
# Chapter 3

- 3.1 Marginalization
- 3.2 Normal distribution with a noninformative prior (important)
- 3.3 Normal distribution with a conjugate prior (important)
- 3.4 Multinomial model (can be skipped)
- 3.5 Multivariate normal with known variance (useful for chapter 4)
- 3.6 Multivariate normal with unknown variance (glance through)
- 3.7 Bioassay example (very important, related to one of the exercises)
- 3.8 Summary (summary)

## Normal / Gaussian

observation  $y$ , and parameters  $\mu$  and  $\sigma$

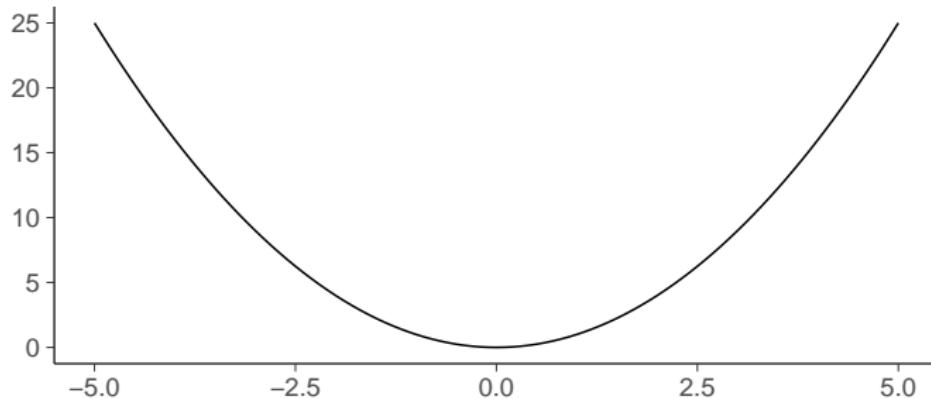
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



## Normal / Gaussian

observation  $y$ , and parameters  $\mu$  and  $\sigma$

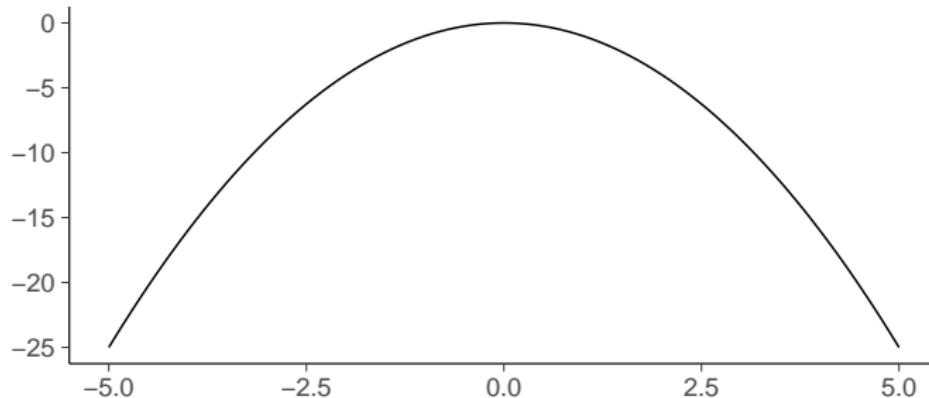
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



## Normal / Gaussian

observation  $y$ , and parameters  $\mu$  and  $\sigma$

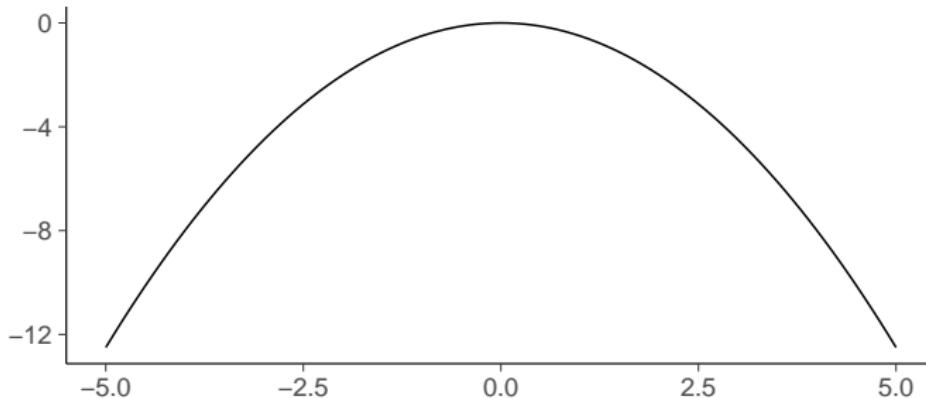
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



## Normal / Gaussian

observation  $y$ , and parameters  $\mu$  and  $\sigma$

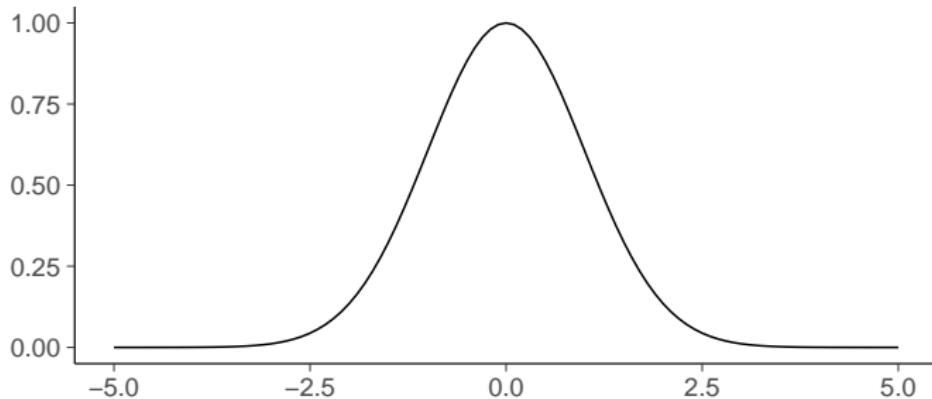
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



## Normal / Gaussian

observation  $y$ , and parameters  $\mu$  and  $\sigma$

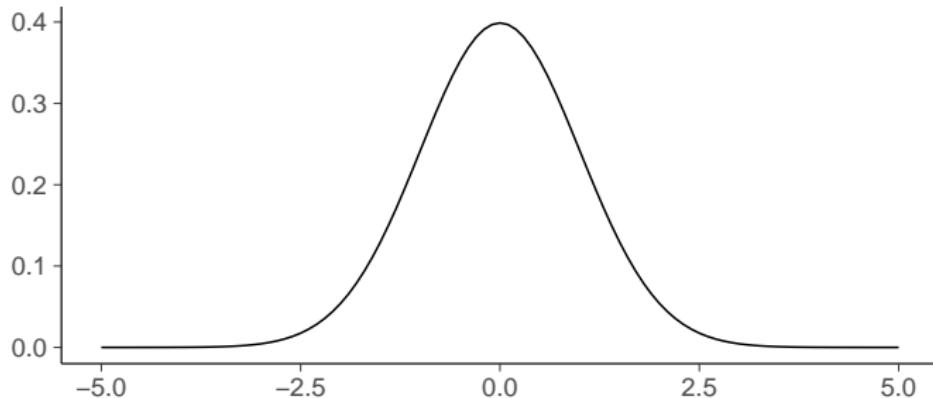
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



## Normal / Gaussian

observation  $y$ , and parameters  $\mu$  and  $\sigma$

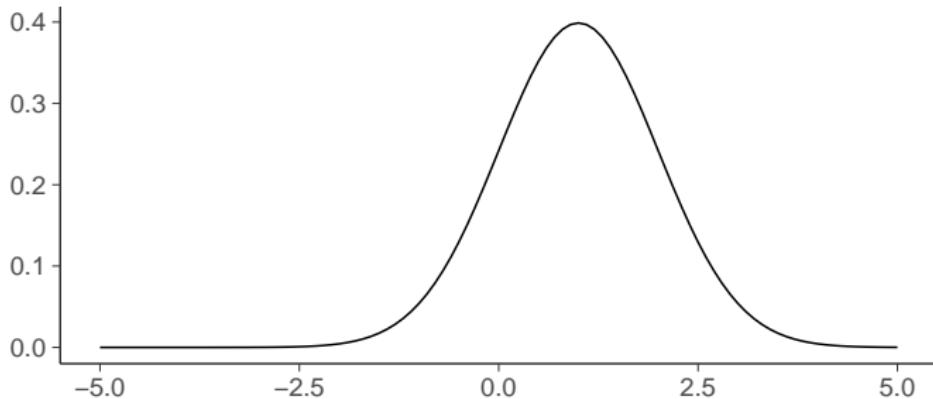
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



## Normal / Gaussian

observation  $y$ , and parameters  $\mu = 1$  and  $\sigma = 1$

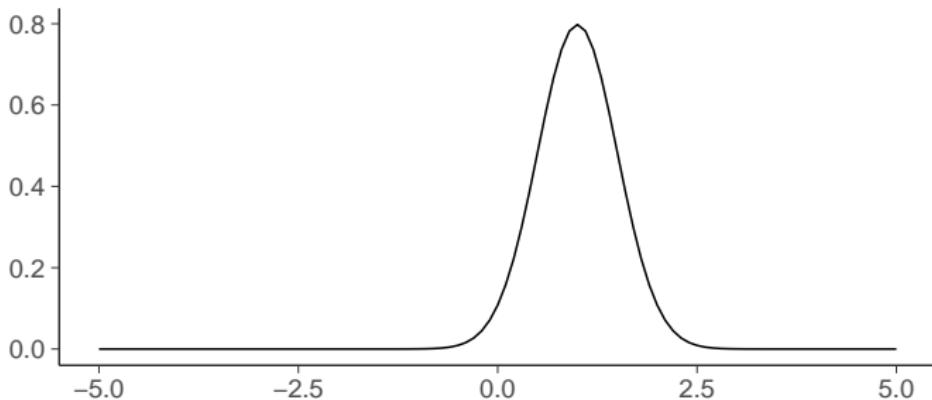
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



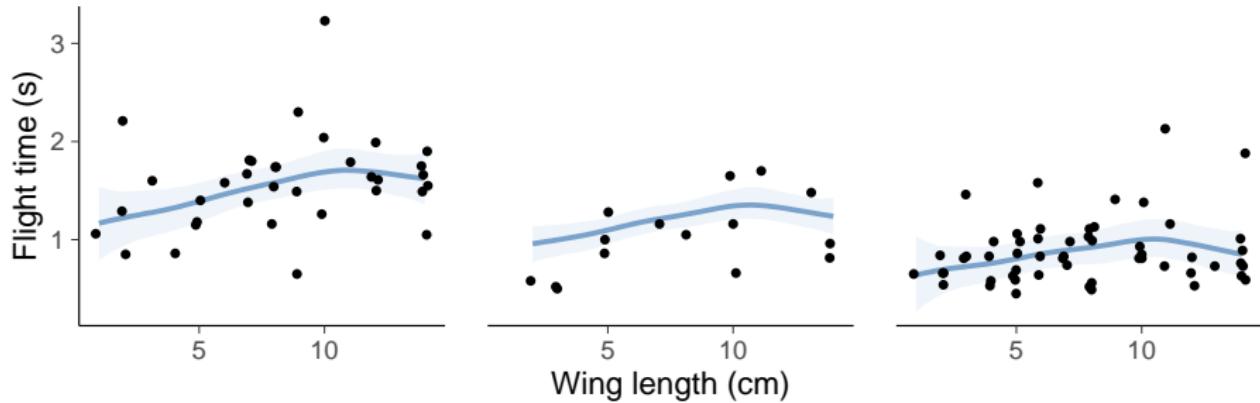
## Normal / Gaussian

observation  $y$ , and parameters  $\mu = 1$  and  $\sigma = 1/2$

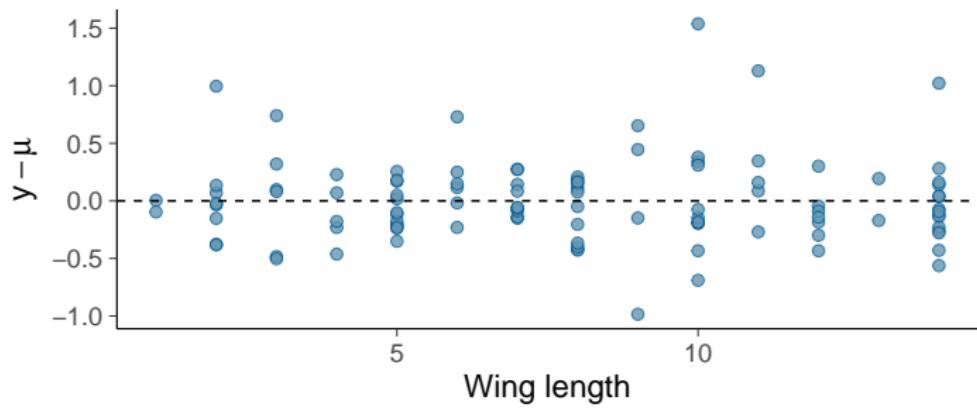
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



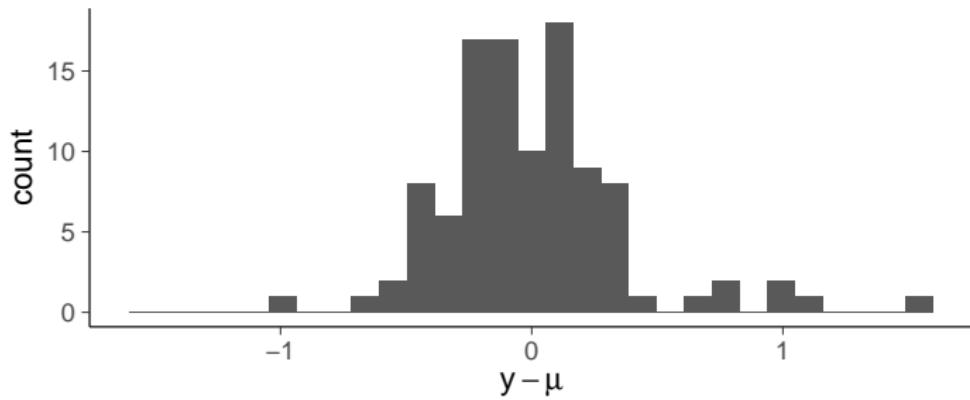
# Variation in helicopter flight times



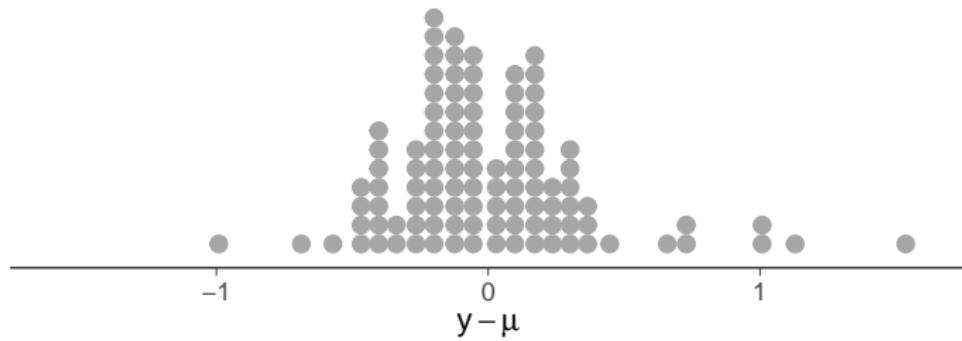
# Variation in helicopter flight times



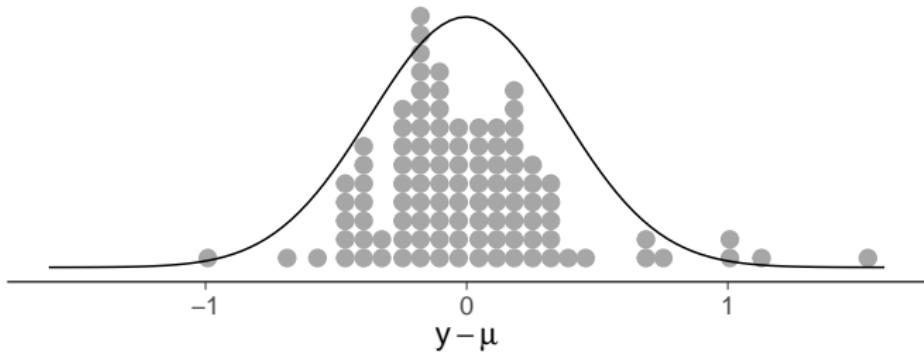
# Variation in helicopter flight times



## Variation in helicopter flight times

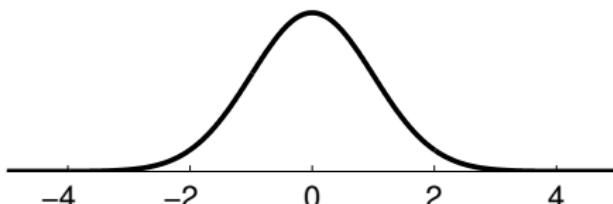


# Variation in helicopter flight times



## Normal / Gaussian

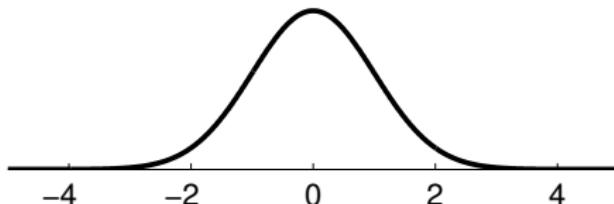
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



- **Normal:** some prominent statistician started to use this term systematically in the end of 19th century (but it's a bit of exaggeration)

## Normal / Gaussian

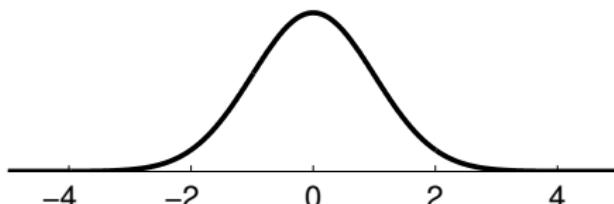
$$p(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right)$$



- **Normal:** some prominent statistician started to use this term systematically in the end of 19th century (but it's a bit of exaggeration)
- **Gaussian:** favored by Germans (but it was studied earlier by De Moivre and Laplace, and Gauss avoided using it)

## Normal / Gaussian

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



- **Normal:** some prominent statistician started to use this term systematically in the end of 19th century (but it's a bit of exaggeration)
- **Gaussian:** favored by Germans (but it was studied earlier by De Moivre and Laplace, and Gauss avoided using it)
- Shorthand notation
  - $y \sim N(\mu, \sigma^2)$  with variance  $\sigma^2$   
(useful in derivations)
  - $y \sim \text{normal}(\mu, \sigma)$  with deviation  $\sigma$   
(useful for interpreting prior and posterior scales, used in Stan)

## Normal / Gaussian

- Normal linear regression

## Normal / Gaussian

- Normal linear regression
- Analysis of real valued observations
  - part of Assignment 3

## Normal / Gaussian

- Normal linear regression
- Analysis of real valued observations
  - part of Assignment 3
- Sometimes convenient approximation for discrete observations

## Normal / Gaussian

- Normal linear regression
- Analysis of real valued observations
  - part of Assignment 3
- Sometimes convenient approximation for discrete observations
- Convenient as a prior distribution

## Normal / Gaussian

- Normal linear regression
- Analysis of real valued observations
  - part of Assignment 3
- Sometimes convenient approximation for discrete observations
- Convenient as a prior distribution
- Gaussian processes are in practice multivariate normals

## Normal / Gaussian

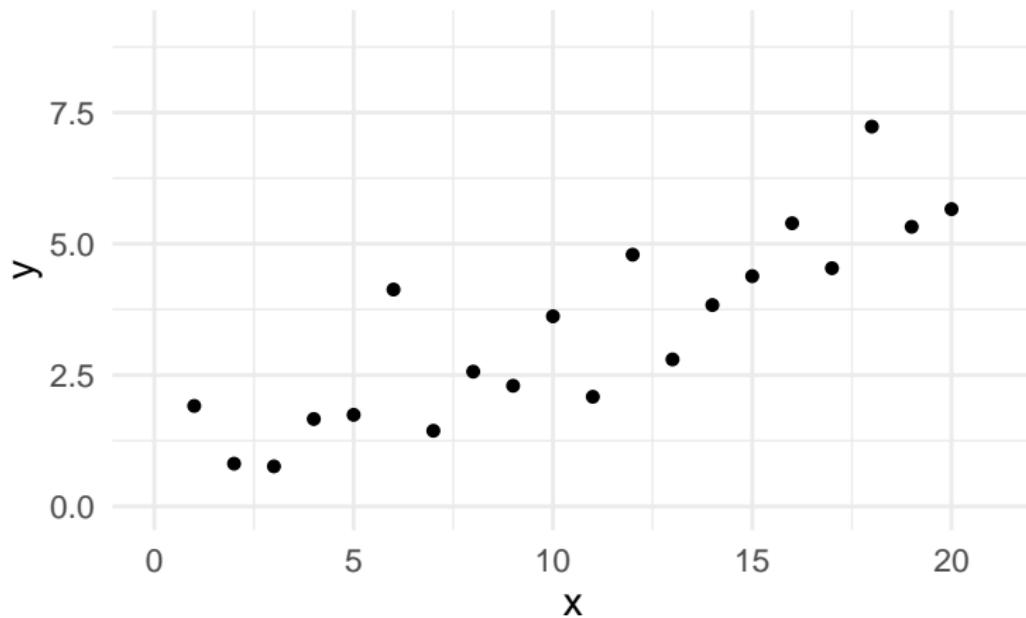
- Normal linear regression
- Analysis of real valued observations
  - part of Assignment 3
- Sometimes convenient approximation for discrete observations
- Convenient as a prior distribution
- Gaussian processes are in practice multivariate normals
- Kalman filters are normals plus chain rule

## Normal / Gaussian

- Normal linear regression
- Analysis of real valued observations
  - part of Assignment 3
- Sometimes convenient approximation for discrete observations
- Convenient as a prior distribution
- Gaussian processes are in practice multivariate normals
- Kalman filters are normals plus chain rule
- Posterior distribution approximation with Laplace, variational inference, expectation propagation

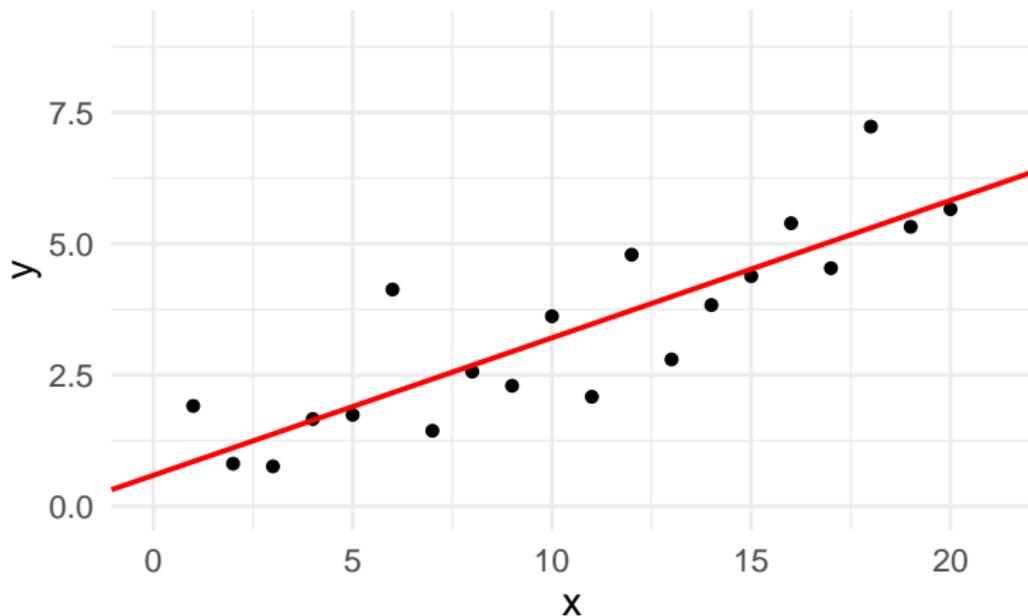
# Example of uncertainty in modeling

Data



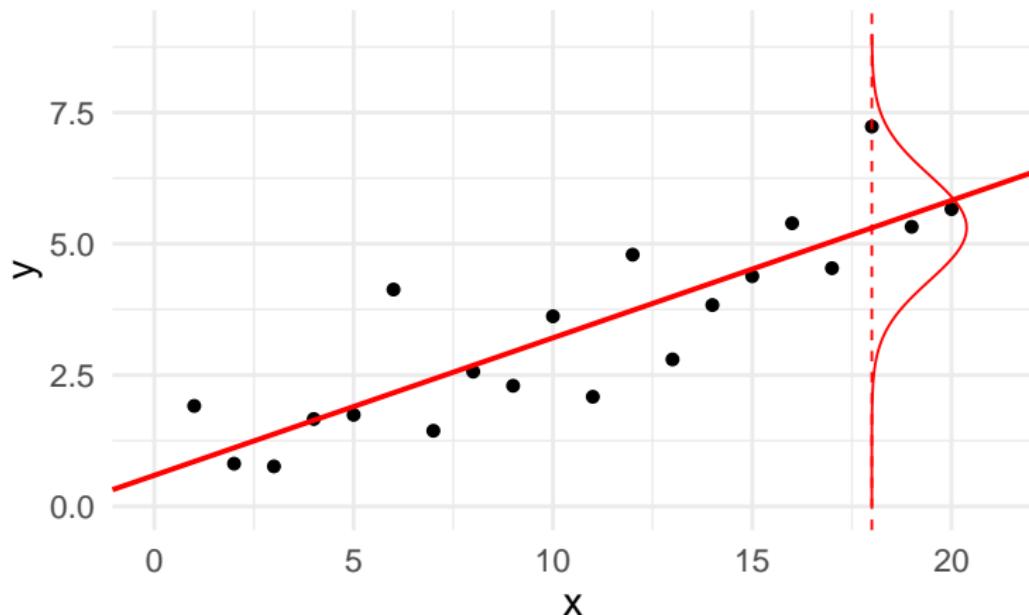
# Example of uncertainty in modeling

Posterior mean



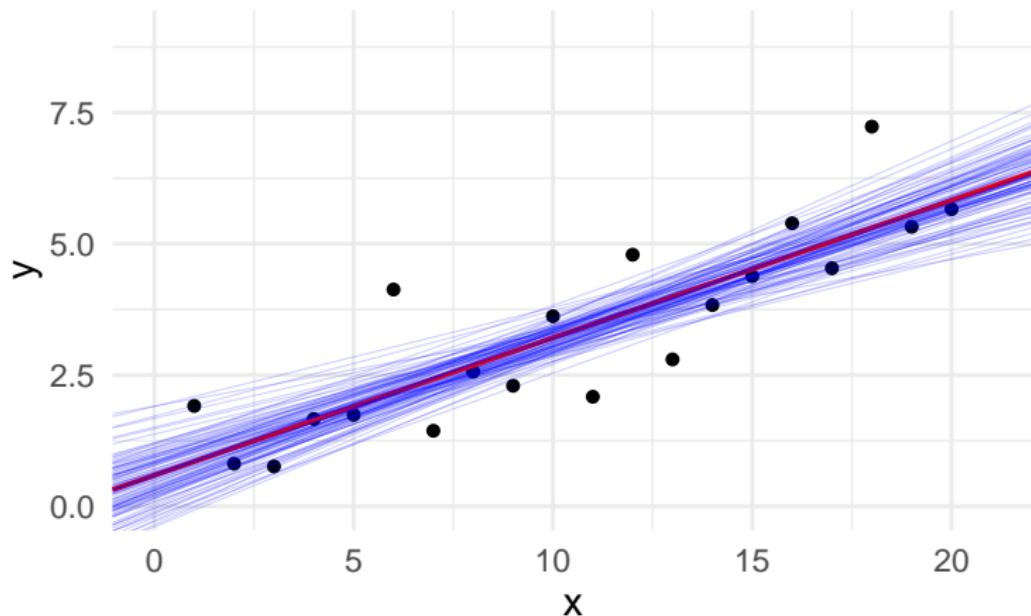
# Example of uncertainty in modeling

Predictive distribution given posterior mean



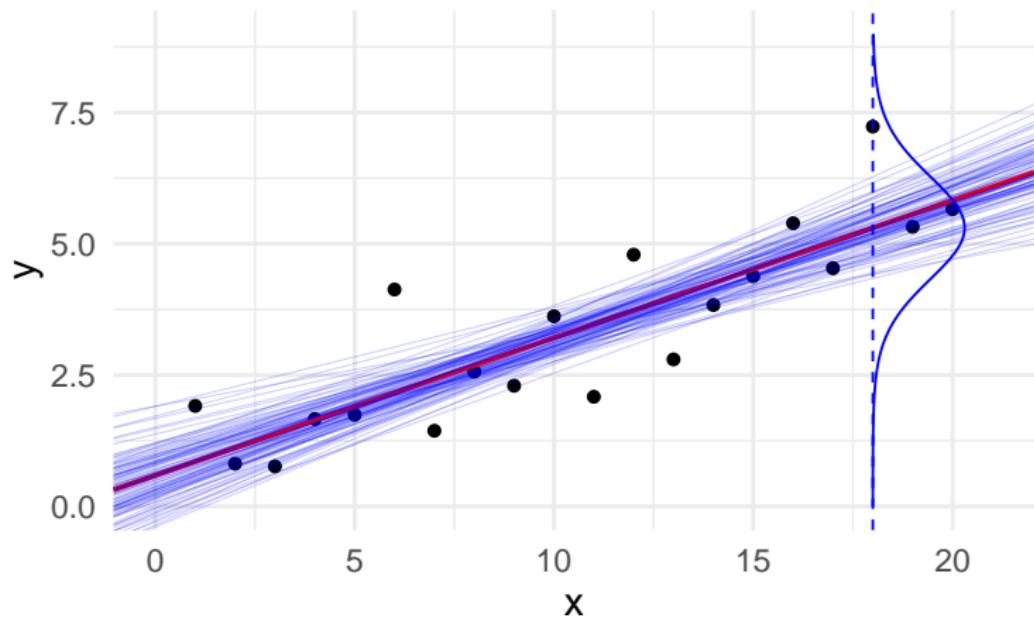
# Example of uncertainty in modeling

Posterior draws



# Example of uncertainty in modeling

Posterior draws and predictive distribution



## Posterior for $\mu$ and $\sigma$ given $y$

- Data model for  $y$   $p(y | \mu, \sigma)$

## Posterior for $\mu$ and $\sigma$ given $y$

- Data model for  $y$   $p(y | \mu, \sigma)$
- Likelihood for parameters given observed  $y$   $p(\mu, \sigma | y)$

## Posterior for $\mu$ and $\sigma$ given $y$

- Data model for  $y$   $p(y | \mu, \sigma)$
- Likelihood for parameters given observed  $y$   $p(\mu, \sigma | y)$
- Prior  $p(\mu, \sigma)$

## Posterior for $\mu$ and $\sigma$ given $y$

- Data model for  $y$   $p(y | \mu, \sigma)$
- Likelihood for parameters given observed  $y$   $p(\mu, \sigma | y)$
- Prior  $p(\mu, \sigma)$
- Posterior  $p(\mu, \sigma | y)$

## Posterior for $\mu$ and $\sigma$ given $y$

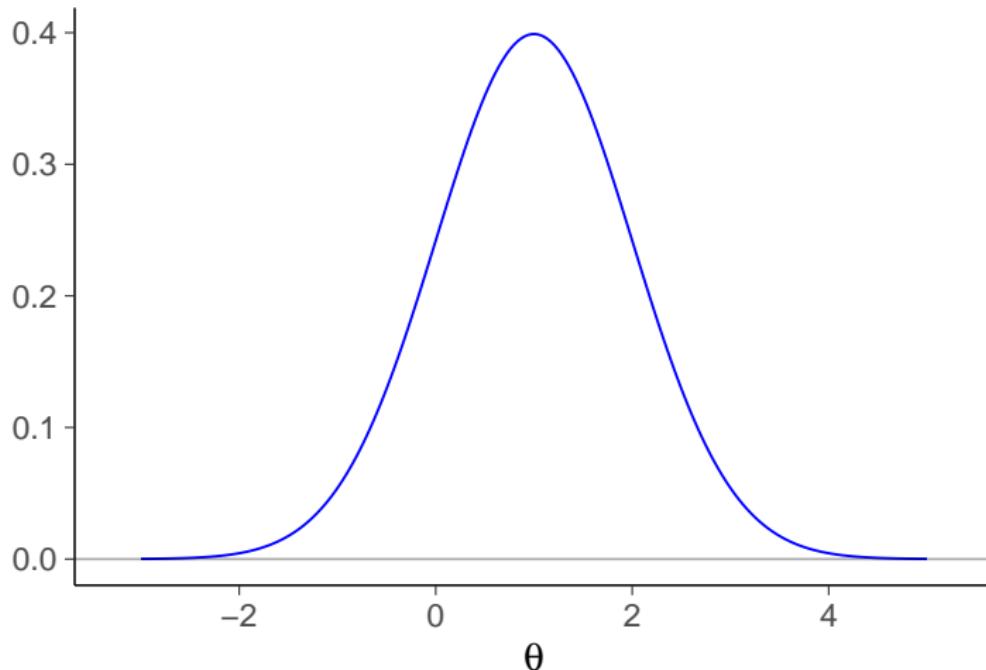
- Data model for  $y$   $p(y | \mu, \sigma)$
- Likelihood for parameters given observed  $y$   $p(y | \mu, \sigma)$
- Prior  $p(\mu, \sigma)$
- Posterior  $p(\mu, \sigma | y)$
- For some priors there is analytic solution for posterior, and Monte Carlo is a general approach

## Posterior for $\mu$ and $\sigma$ given $y$

- Data model for  $y$   $p(y | \mu, \sigma)$
- Likelihood for parameters given observed  $y$   $p(\mu, \sigma | y)$
- Prior  $p(\mu, \sigma)$
- Posterior  $p(\mu, \sigma | y)$
- For some priors there is analytic solution for posterior, and Monte Carlo is a general approach
- Before looking at the posterior  $p(\mu, \sigma | y)$ , connection between mean, cumulative density and Monte Carlo

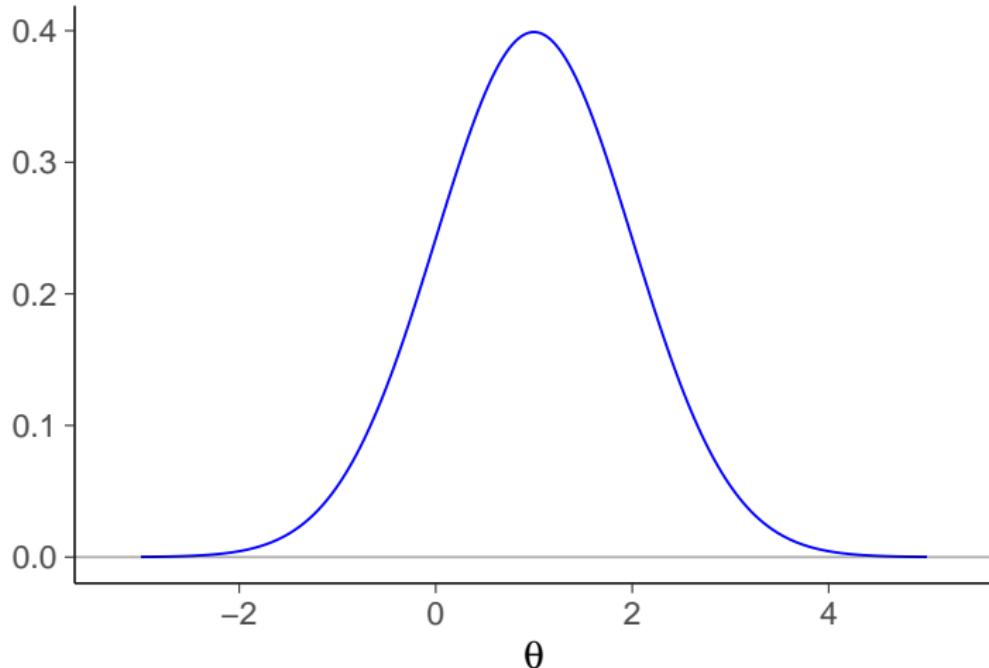
## Monte Carlo and posterior draws

$$\text{Density } p(\theta|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\theta - \mu)^2\right) \quad (\text{dnorm}(\cdot))$$



## Monte Carlo and posterior draws

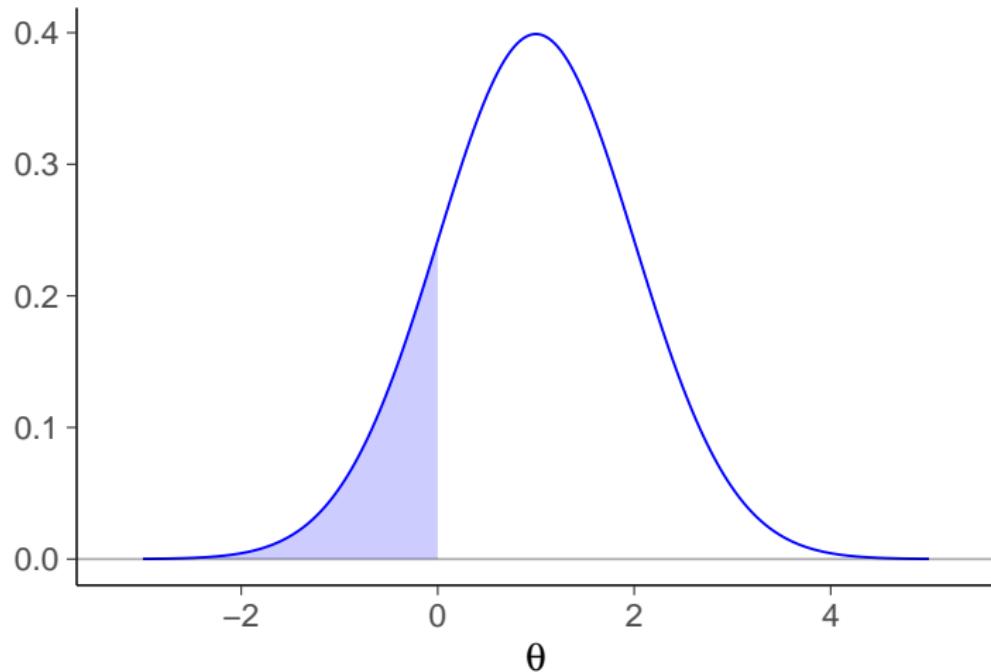
$$\text{Density } p(\theta|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\theta - \mu)^2\right) \quad (\text{dnorm}(\cdot))$$



$$E(\theta) = \int \theta p(\theta|\mu, \sigma) d\theta = \mu$$

## Monte Carlo and posterior draws

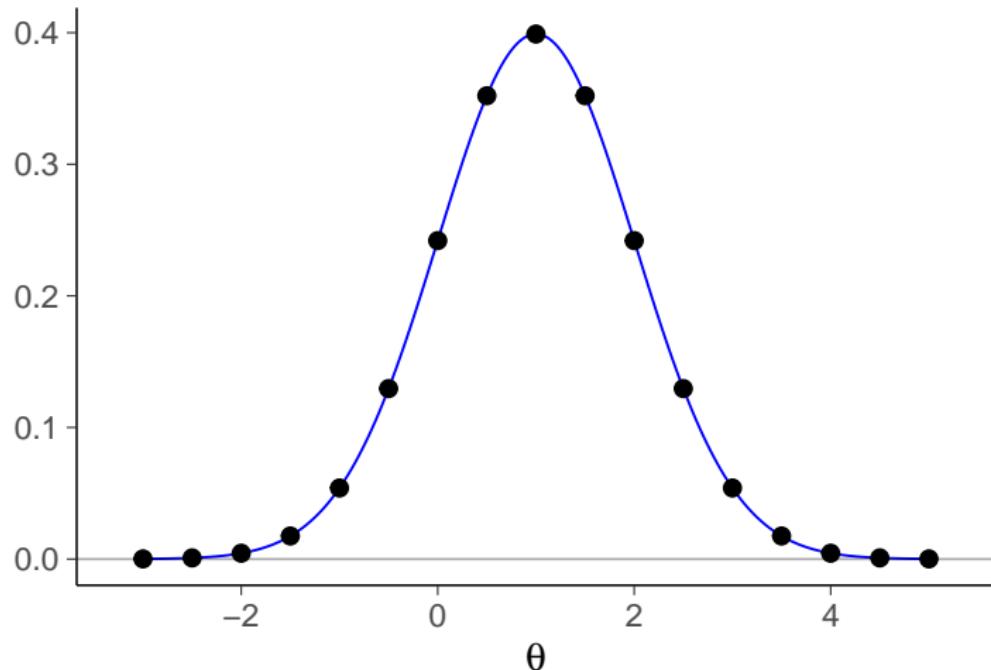
$$\text{Density } p(\theta|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\theta - \mu)^2\right)$$



$p(\theta \leq 0) = \int_{-\infty}^0 p(\theta|\mu, \sigma) d\theta$ ,  
many numerical approximations (pnorm())

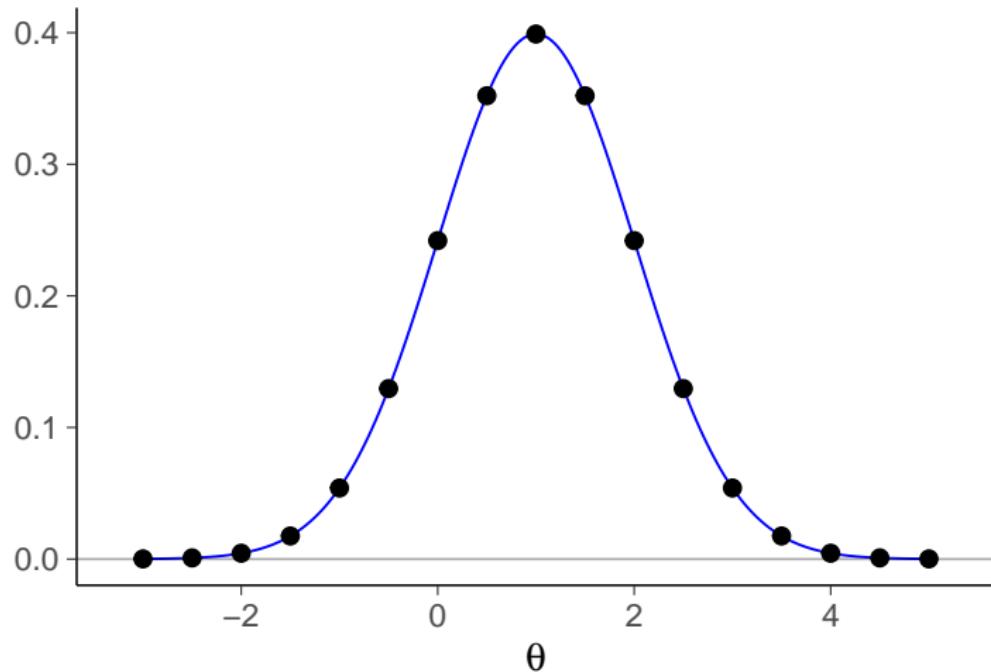
# Monte Carlo and posterior draws

In practice evaluate in finite number of locations (`dnorm()`)



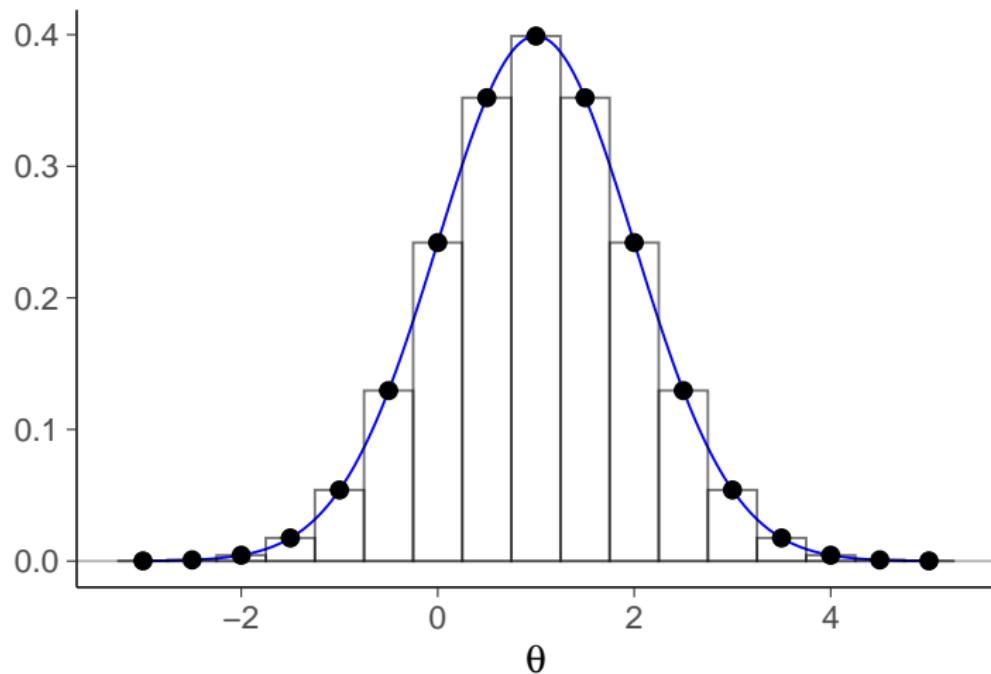
# Monte Carlo and posterior draws

Here evaluated in grid with bin width 0.5



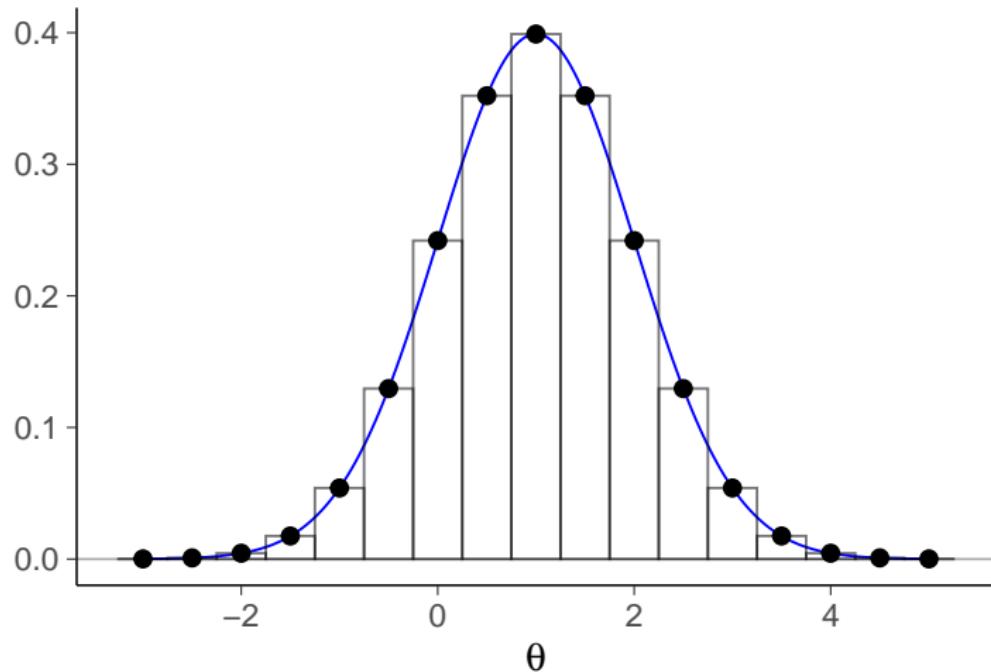
# Monte Carlo and posterior draws

Here evaluated in grid with bin width 0.5



# Monte Carlo and posterior draws

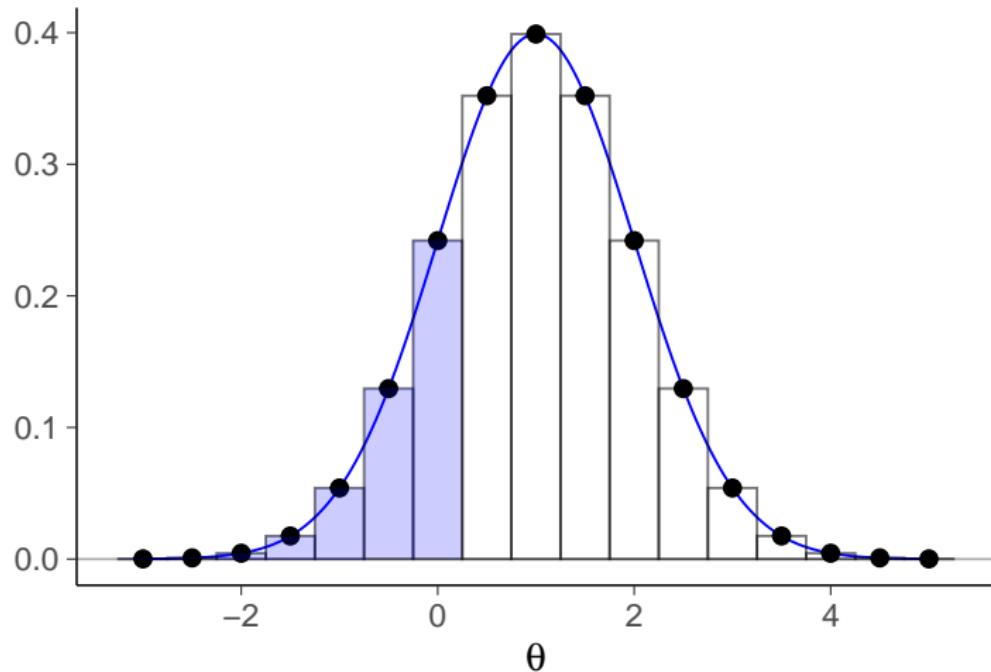
Here evaluated in grid with bin width 0.5



$$E(\theta) = \int \theta p(\theta) d\theta \approx \sum_s^S \theta^{(s)} w_s \approx 1, \text{ where } w_s = 0.5p(\theta)$$

# Monte Carlo and posterior draws

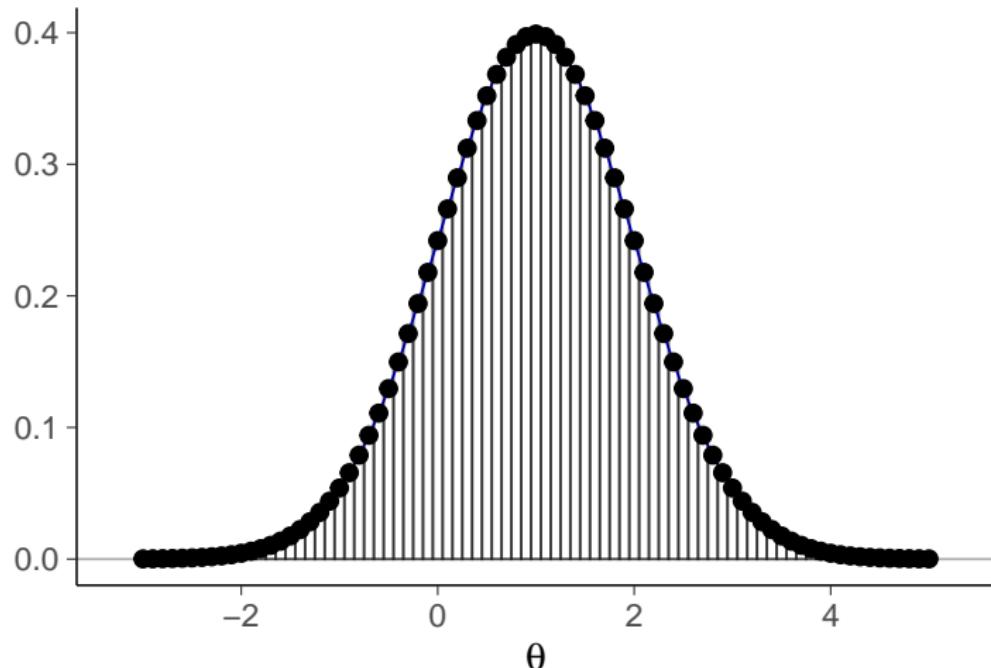
Here evaluated in grid with bin width 0.5



$$p(\theta \leq 0) = \int_{-\infty}^0 p(\theta) d\theta \approx \sum_s^S I(\theta^{(s)} \leq 0) w_s \approx 0.22$$

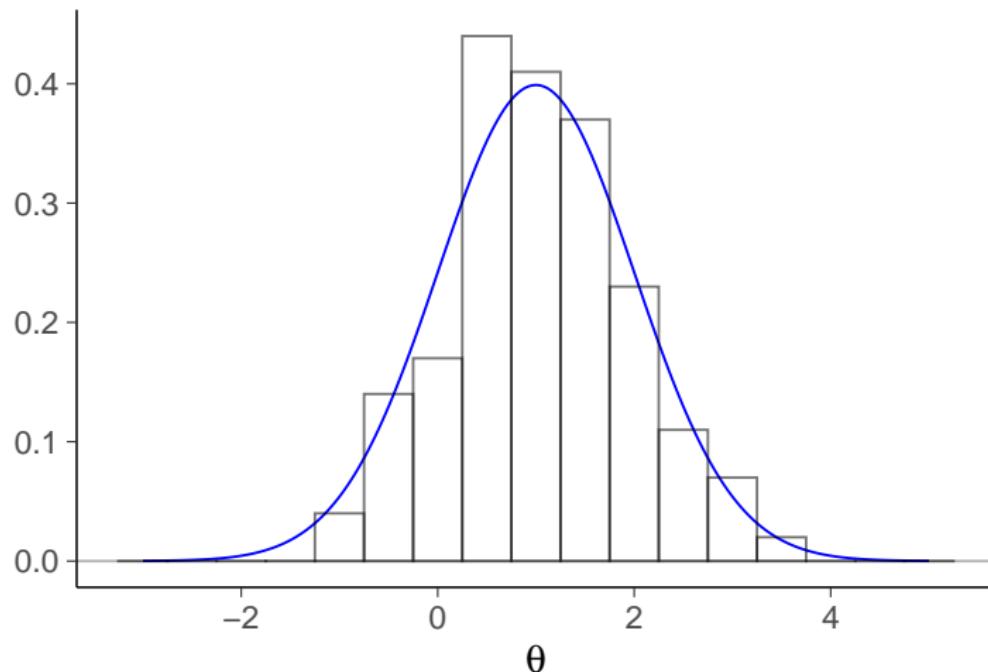
# Monte Carlo and posterior draws

Here evaluated in grid with bin width 0.1



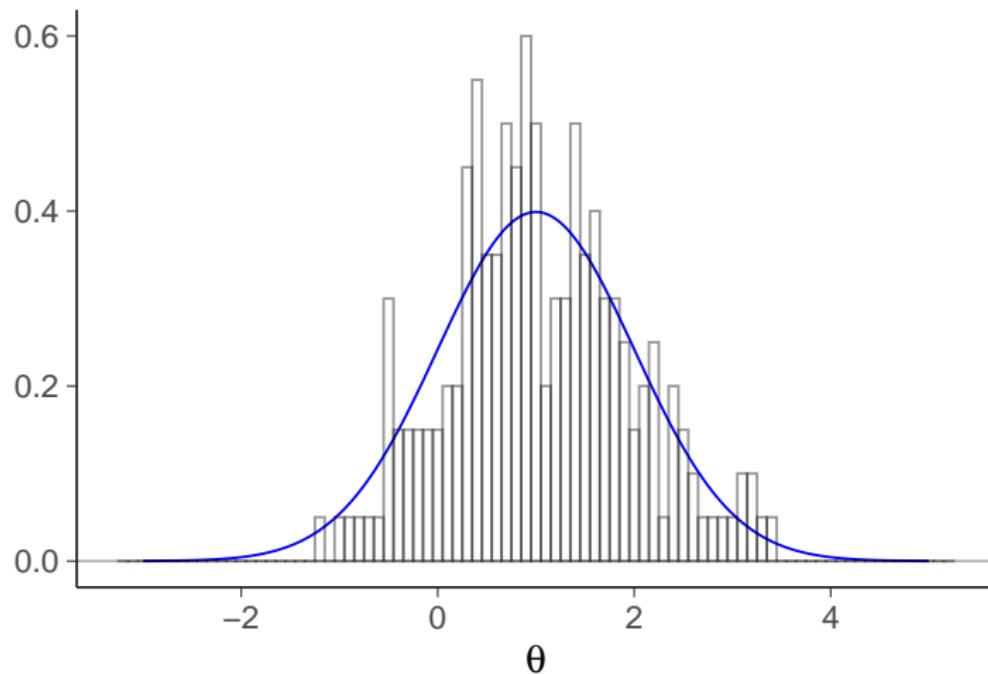
# Monte Carlo and posterior draws

Histogram of 200 random draws (`rnorm()`), bin width 0.5



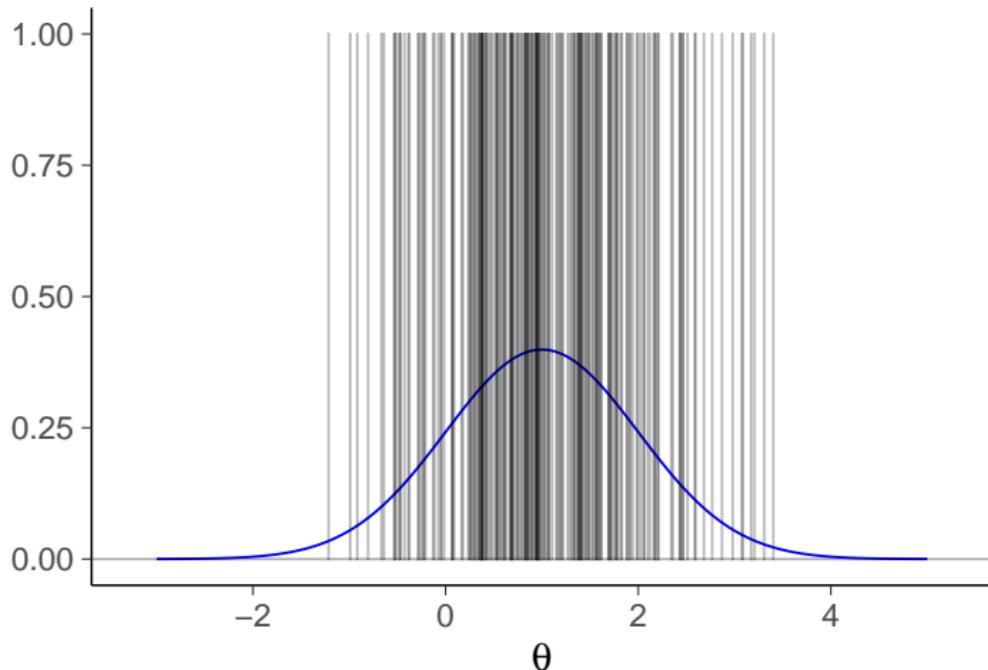
# Monte Carlo and posterior draws

Histogram of 200 random draws (`rnorm()`), bin width 0.1



## Monte Carlo and posterior draws

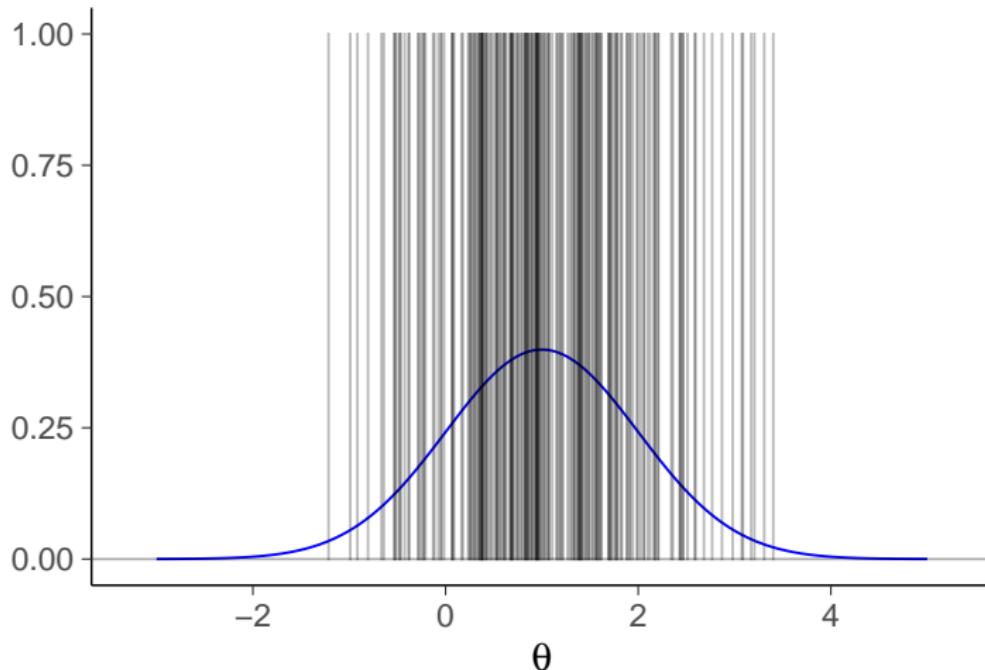
Histogram of 200 random draws (`rnorm()`), bin width 0



each bin has either 0 or 1 draw (and 0's can be ignored)

# Monte Carlo and posterior draws

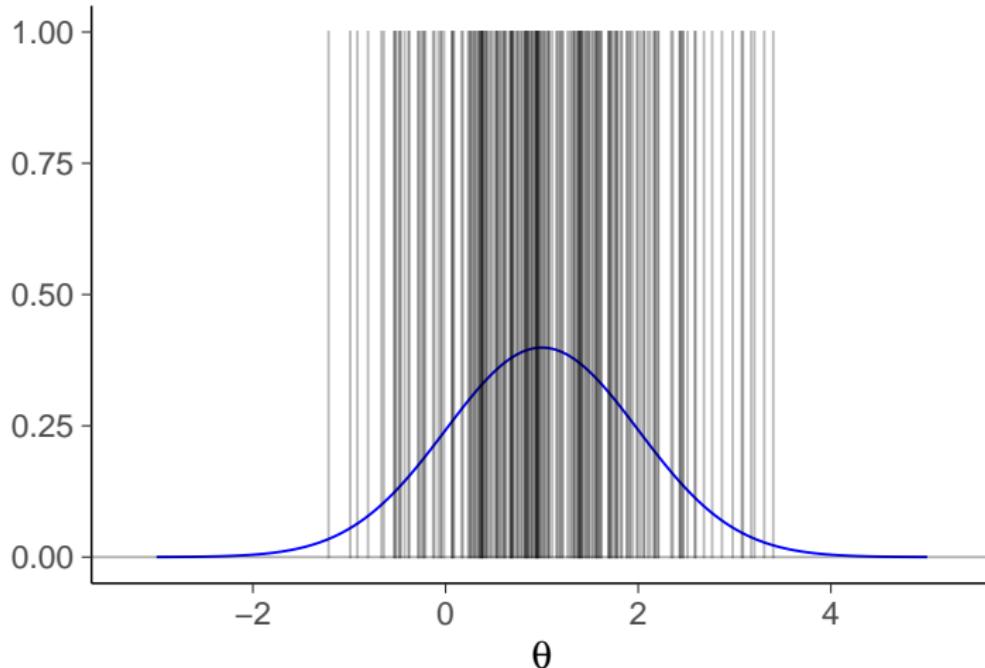
Histogram of 200 random draws (`rnorm()`), bin width 0



each bin with 1 draw has weight  $1/S$

# Monte Carlo and posterior draws

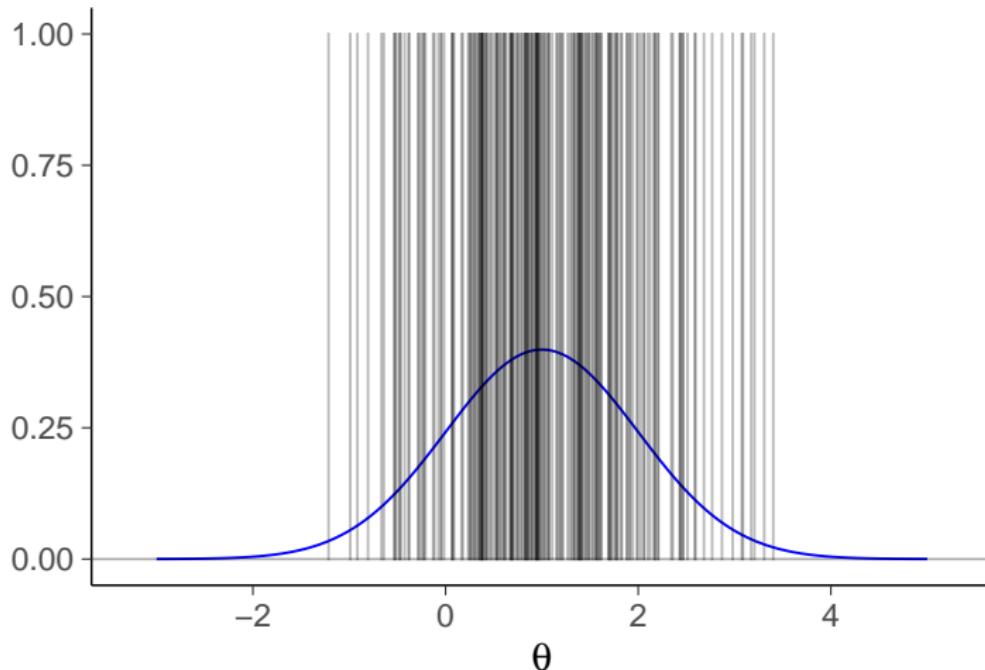
Histogram of 200 random draws (`rnorm()`), bin width 0



$$E(\theta) \approx \frac{1}{S} \sum_s^S \theta^{(s)} \approx 1$$

# Monte Carlo and posterior draws

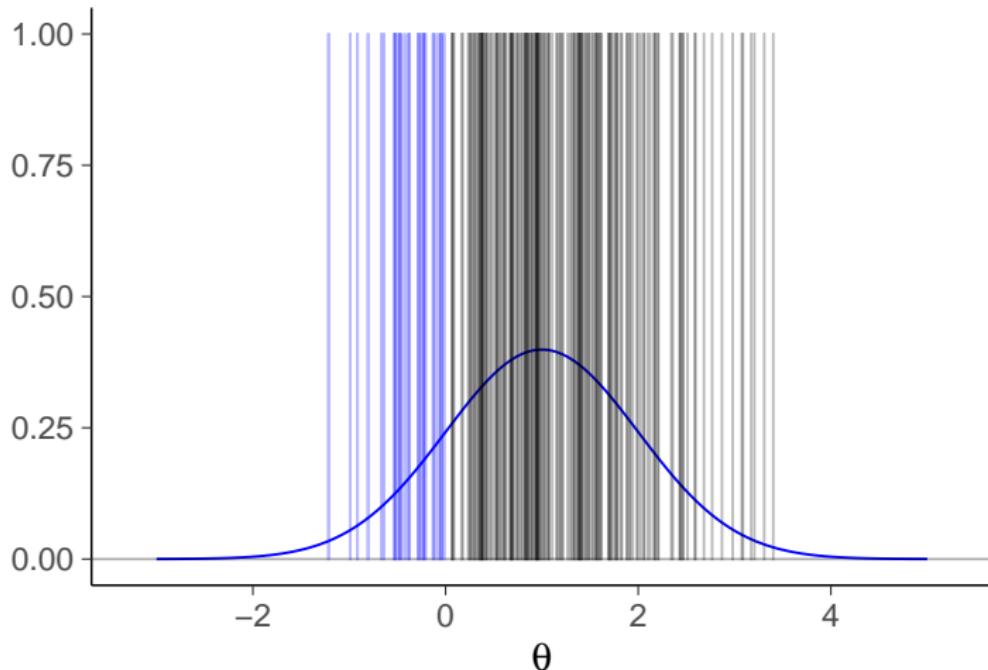
Histogram of 200 random draws (`rnorm()`), bin width 0



$$E(\theta) \approx \frac{1}{S} \sum_s^S \theta^{(s)} \approx 1, \text{ Monte Carlo estimate}$$

# Monte Carlo and posterior draws

Histogram of 200 random draws, bin width 0



$$p(\theta \leq 0) \approx \frac{1}{S} \sum_s^S I(\theta^{(s)} \leq 0) \approx 0.14$$

## Monte Carlo and posterior draws

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

# Monte Carlo and posterior draws

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

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

# Monte Carlo and posterior draws

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

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

- easy to approximate expectations of functions (push forward)

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

# Monte Carlo and posterior draws

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

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

- easy to approximate expectations of functions (push forward)

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

- If  $p(g(\theta))$  has finite variance, then the Monte Carlo estimate is unbiased and the error approaches 0 with increasing  $S$  based on the central limit theorem (CLT)
  - more about this later

# Marginalization

- Joint distribution of parameters

$$p(\mu, \sigma \mid y) \propto p(y \mid \mu, \sigma)p(\mu, \sigma)$$

# Marginalization

- Joint distribution of parameters

$$p(\mu, \sigma \mid y) \propto p(y \mid \mu, \sigma)p(\mu, \sigma)$$

- Marginalization

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

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

# Marginalization

- Joint distribution of parameters

$$p(\mu, \sigma | y) \propto p(y | \mu, \sigma)p(\mu, \sigma)$$

- Marginalization

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

$p(\mu | y)$  is a marginal distribution

- Analytic solution
  - sometimes the integral has an analytic solution

# Marginalization

- Joint distribution of parameters

$$p(\mu, \sigma | y) \propto p(y | \mu, \sigma)p(\mu, \sigma)$$

- Marginalization

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

$p(\mu | y)$  is a marginal distribution

- Analytic solution
  - sometimes the integral has an analytic solution
- Monte Carlo approximation

$$\begin{aligned} & \text{if } (\mu^{(s)}, \sigma^{(s)}) \sim p(\mu, \sigma | y) \\ & \text{then } \mu^{(s)} \sim p(\mu | y) \end{aligned}$$

## Marginalization - predictive distribution

- Posterior predictive distribution for a future  $\tilde{y}$  is obtained by marginalizing the joint distribution of unknowns

$$p(\tilde{y}, \mu, \sigma | y)$$

## Marginalization - predictive distribution

- Posterior predictive distribution for a future  $\tilde{y}$  is obtained by marginalizing the joint distribution of unknowns

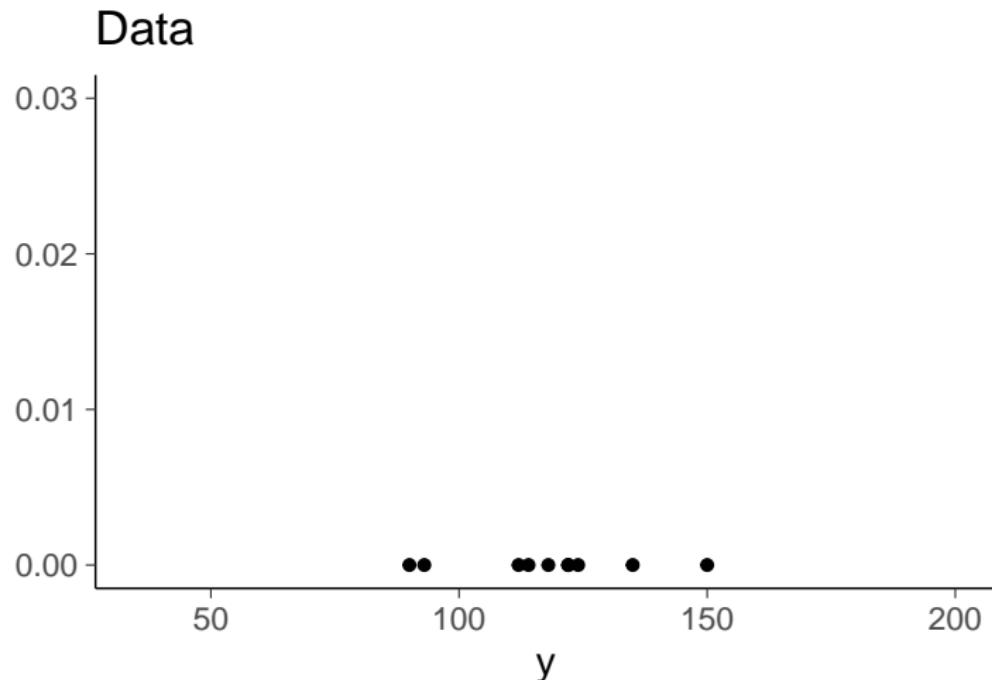
$$p(\tilde{y} | y) = \int p(\tilde{y}, \mu, \sigma, | y) d\mu \sigma$$

## Marginalization - predictive distribution

- Posterior predictive distribution for a future  $\tilde{y}$  is obtained by marginalizing the joint distribution of unknowns

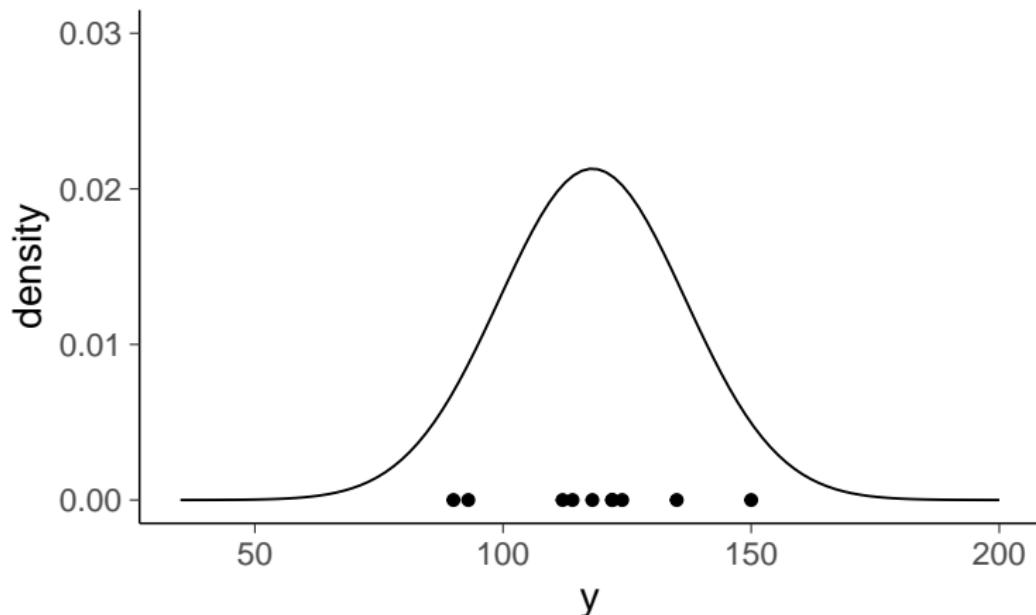
$$\begin{aligned} p(\tilde{y} \mid y) &= \int p(\tilde{y}, \mu, \sigma, \mid y) d\mu \sigma \\ &= \int p(\tilde{y} \mid \mu, \sigma) p(\mu, \sigma \mid y) d\mu \sigma \end{aligned}$$

# Normal distribution example



# Normal distribution example

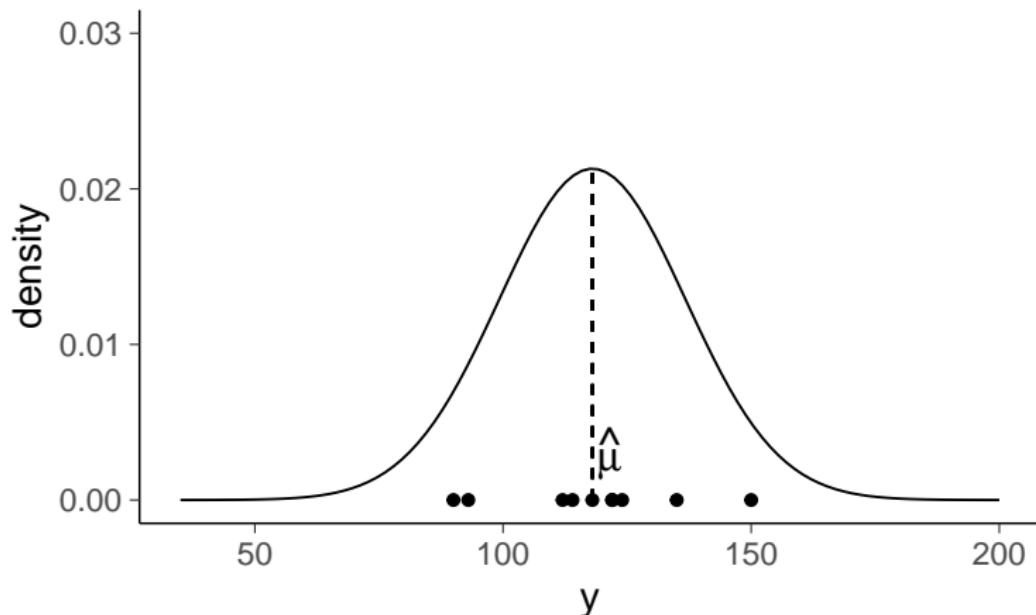
Normal fit with posterior mean



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

# Normal distribution example

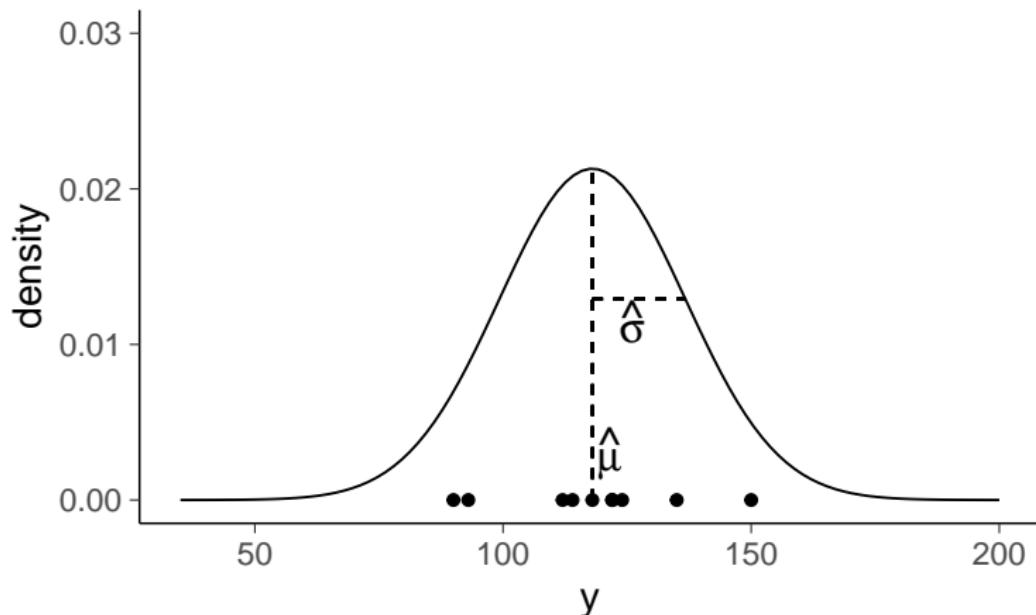
Normal fit with posterior mean



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

# Normal distribution example

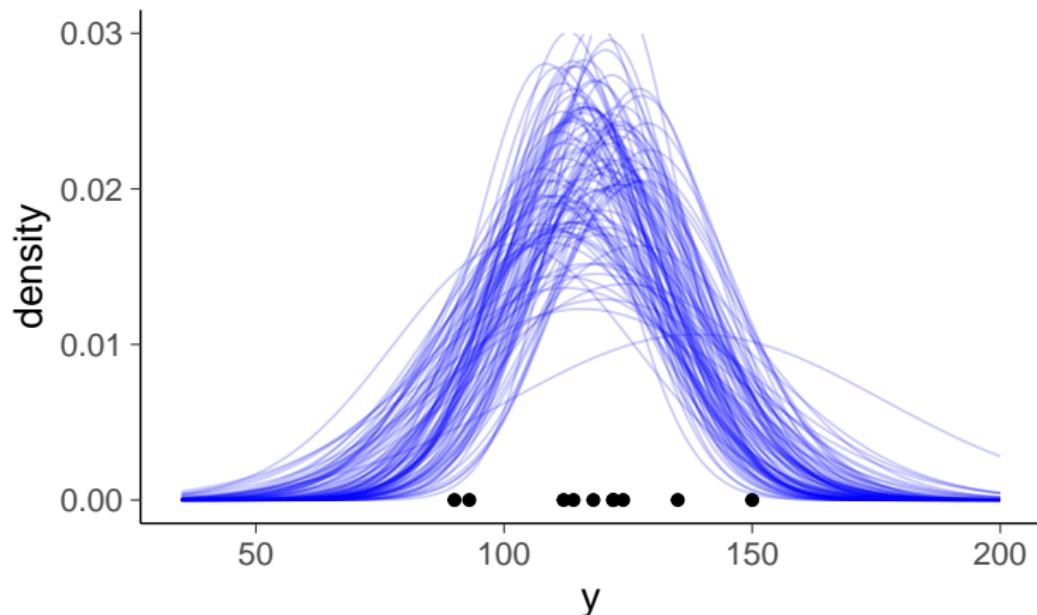
Normal fit with posterior mean



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

# Normal distribution example

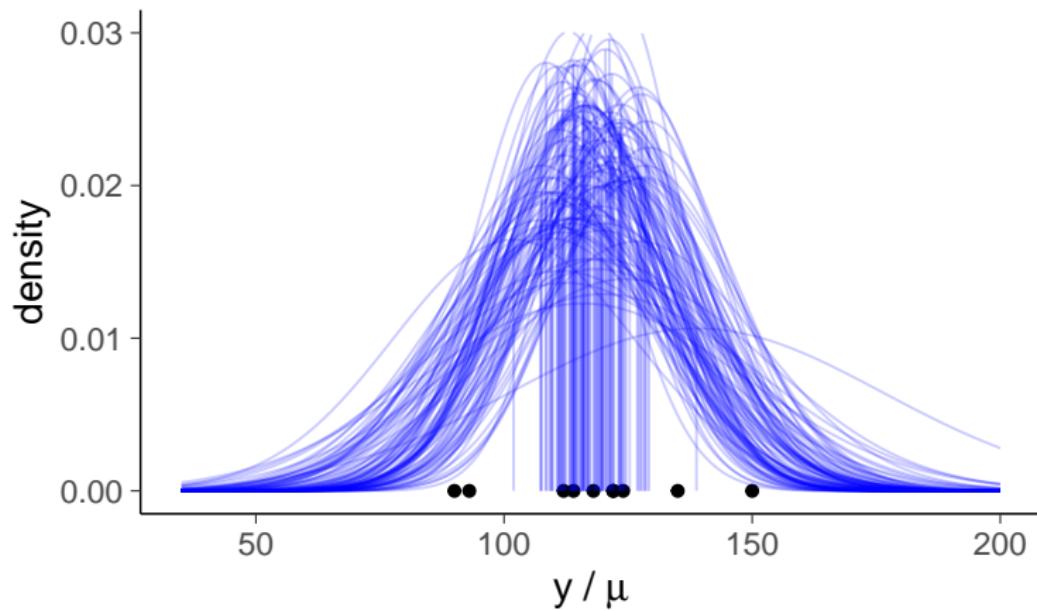
Normals with posterior draw parameters



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

# Normal distribution example

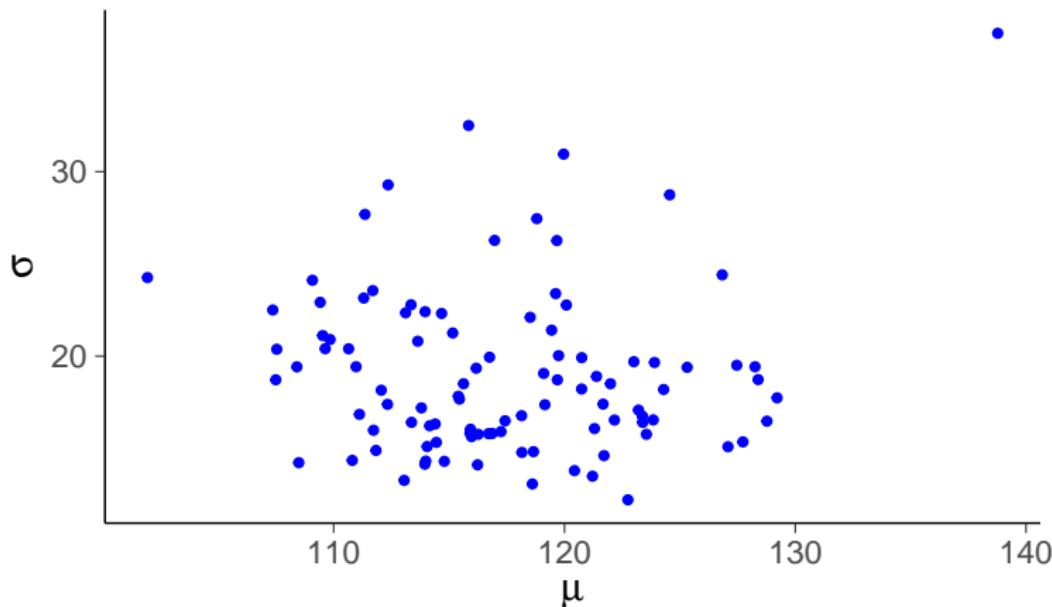
Normals with posterior draw parameters



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

# Normal distribution example

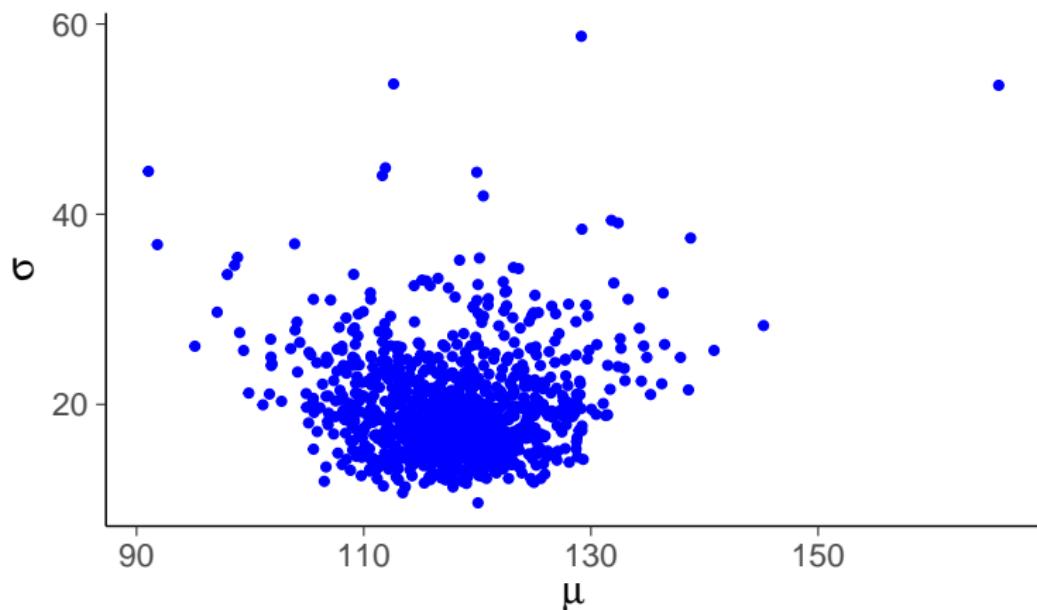
Draws from the joint posterior distribution



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

# Normal distribution example

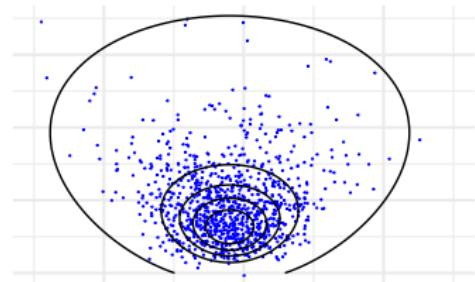
Draws from the joint posterior distribution



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

Joint posterior

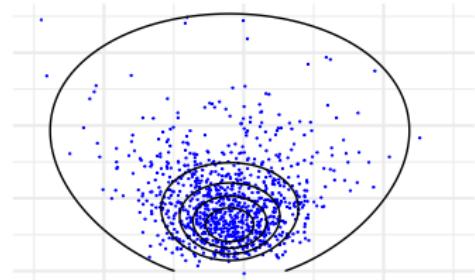
$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma | y)$$



Joint posterior

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

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

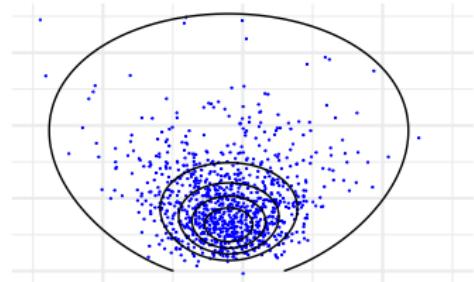


Joint posterior

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

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

with  $p(\mu, \sigma) \propto \sigma^{-1}$  (see BDA3 p. 21 transformation of variables)

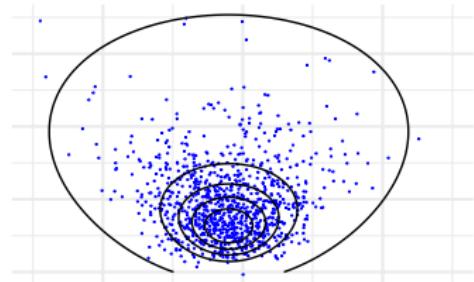


Joint posterior

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

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

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

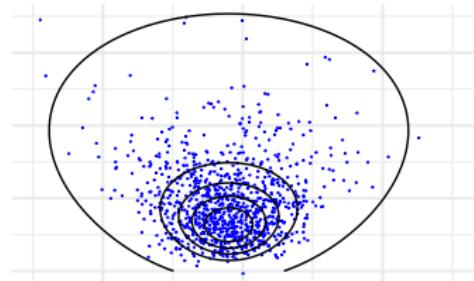


Joint posterior

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

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

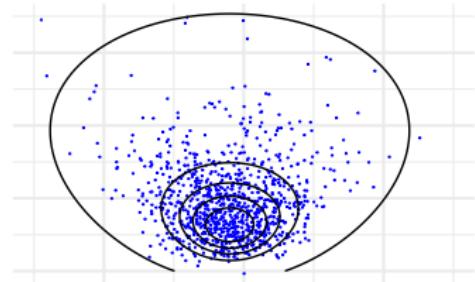
$$p(\mu, \sigma^2 | y) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$



## Joint posterior

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

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



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

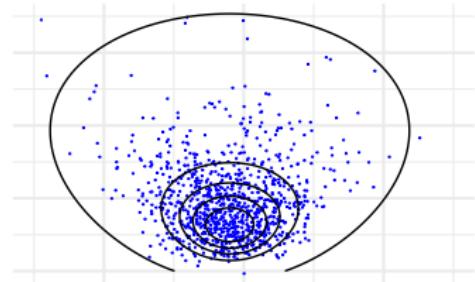
$$= \sigma^{-n-2} \exp \left( -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right] \right)$$

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

## Joint posterior

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

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



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

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

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

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

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

## Normal - non-informative prior

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

## Normal - non-informative prior

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

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

## Normal - non-informative prior

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

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

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

## Normal - non-informative prior

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

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

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

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

## Normal - non-informative prior

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

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

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

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

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

## Normal - non-informative prior

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

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

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

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

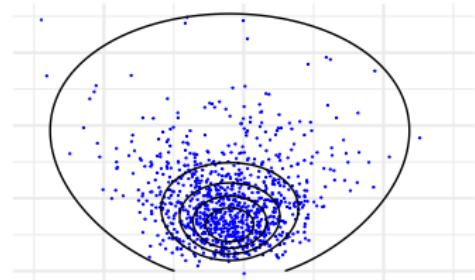
$$\sum_{i=1}^n (y_i - \bar{y})^2 + n(\mu^2 - 2\bar{y}\mu - \bar{y}^2 + 2\bar{y}\bar{y})$$

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

## Joint posterior

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

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



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

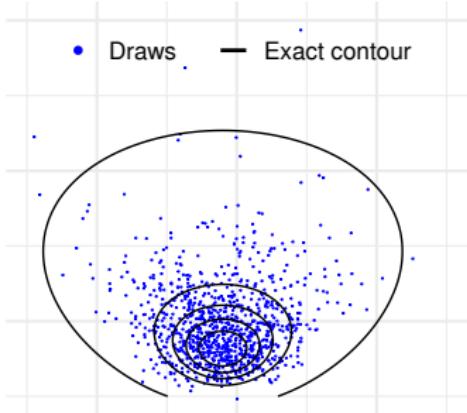
$$= \sigma^{-n-2} \exp \left( -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right] \right)$$

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

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

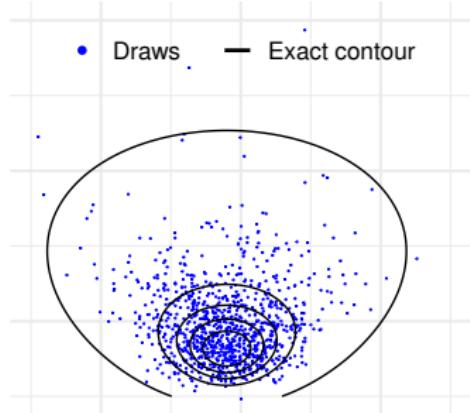
$$\text{where } s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

## Joint posterior

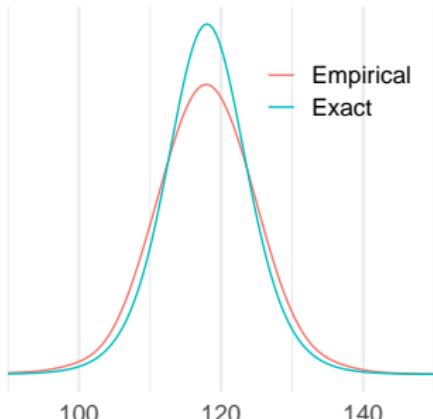


$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma | y)$$

## Joint posterior



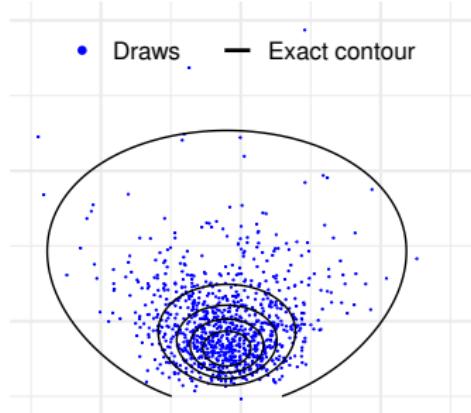
## Marginal of mu



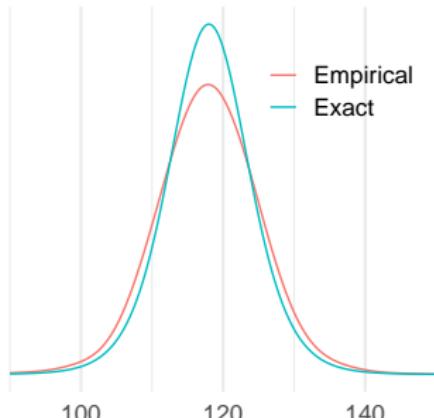
$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma | y)$   
marginals

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

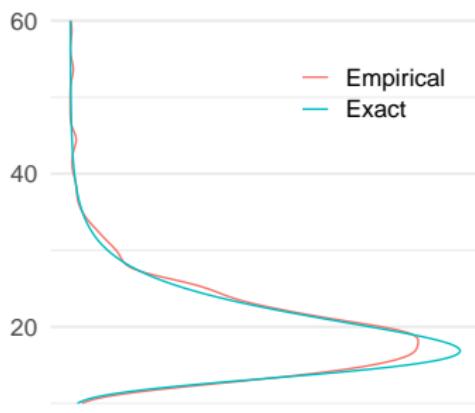
Joint posterior



Marginal of mu



Marginal of sigma



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

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

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

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

$$p(\sigma^2 | y) \propto \int p(\mu, \sigma^2 | y) d\mu$$

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

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

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

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

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

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

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

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

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

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

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

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

## Normal - non-informative prior

Known mean

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

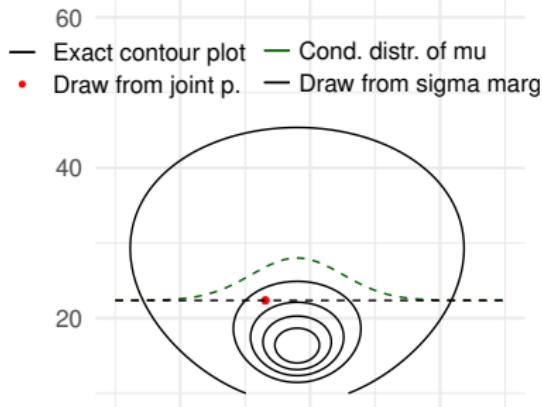
where  $v = \frac{1}{n} \sum_{i=1}^n (y_i - \theta)^2$

Unknown mean

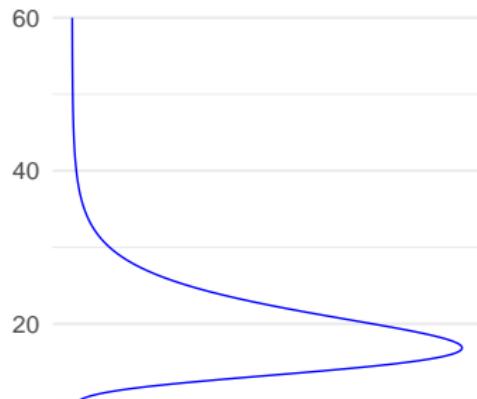
$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(n - 1, s^2)$$

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

## Joint posterior



## Marginal of sigma

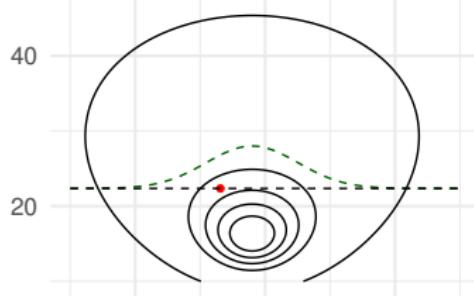


## Factorization

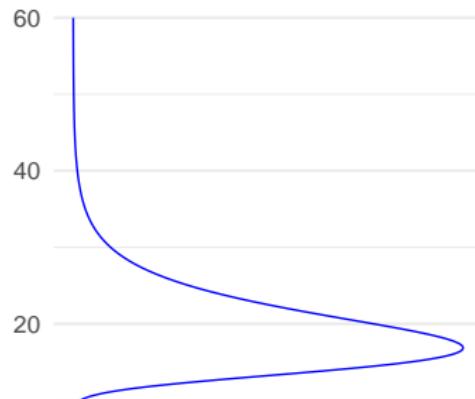
$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y) p(\sigma^2 | y)$$

## Joint posterior

- Exact contour plot
- Draw from joint p.
- Cond. distr. of mu
- Draw from sigma marg



## Marginal of sigma



## Factorization

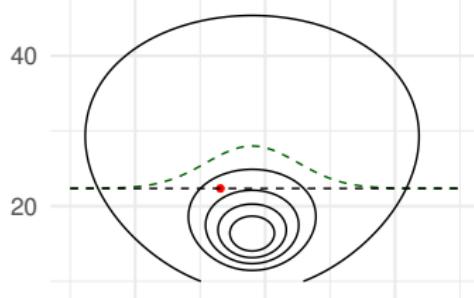
$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y) p(\sigma^2 | y)$$

$$p(\sigma^2 | y) = \text{Inv-}\chi^2(\sigma^2 | n - 1, s^2)$$

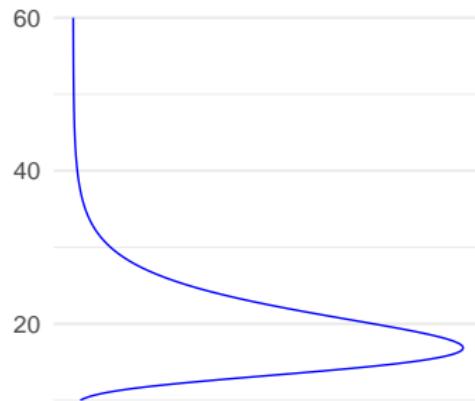
$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

## Joint posterior

- Exact contour plot
- Draw from joint p.
- Cond. distr. of mu
- Draw from sigma marg



## Marginal of sigma



## Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y) p(\sigma^2 | y)$$

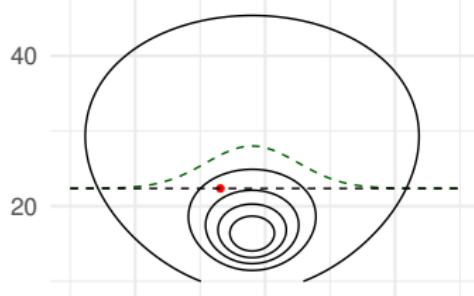
$$p(\sigma^2 | y) = \text{Inv-}\chi^2(\sigma^2 | n - 1, s^2)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

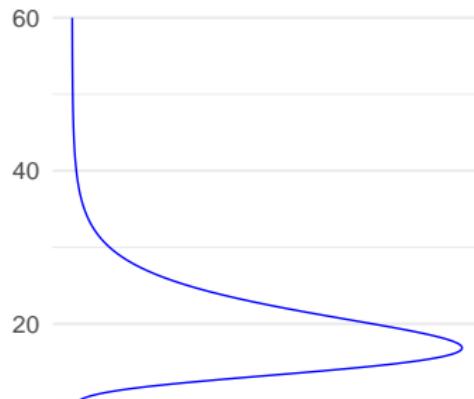
$$p(\mu | \sigma^2, y) = N(\mu | \bar{y}, \sigma^2/n)$$

## Joint posterior

- Exact contour plot
- Draw from joint p.
- Cond. distr. of mu
- Draw from sigma marg



## Marginal of sigma



## Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y) p(\sigma^2 | y)$$

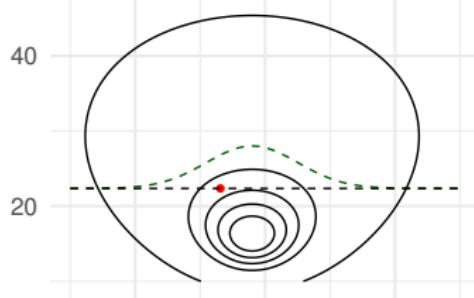
$$p(\sigma^2 | y) = \text{Inv-}\chi^2(\sigma^2 | n - 1, s^2)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

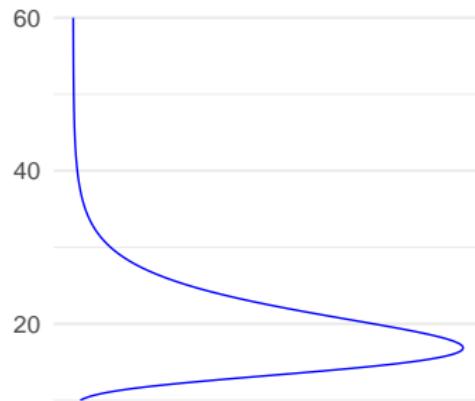
$$p(\mu | \sigma^2, y) = N(\mu | \bar{y}, \sigma^2/n) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{y} - \mu)^2\right)$$

## Joint posterior

- Exact contour plot
- Draw from joint p.
- Cond. distr. of mu
- Draw from sigma marg



## Marginal of sigma



## Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y) p(\sigma^2 | y)$$

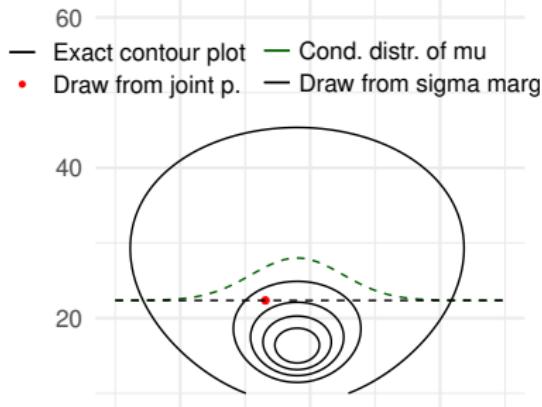
$$p(\sigma^2 | y) = \text{Inv-}\chi^2(\sigma^2 | n - 1, s^2)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

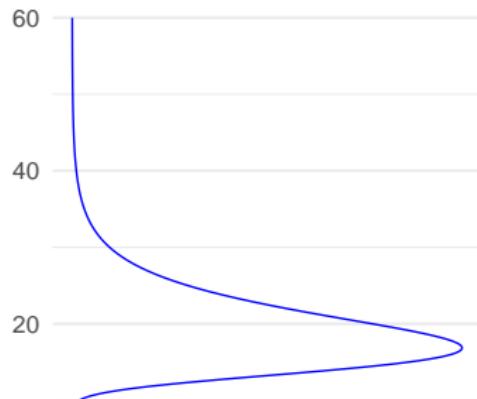
$$p(\mu | \sigma^2, y) = N(\mu | \bar{y}, \sigma^2/n)$$

$$\mu^{(s)} \sim p(\mu | (\sigma^2)^{(s)}, y)$$

## Joint posterior



## Marginal of sigma



## Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y) p(\sigma^2 | y)$$

$$p(\sigma^2 | y) = \text{Inv-}\chi^2(\sigma^2 | n - 1, s^2)$$

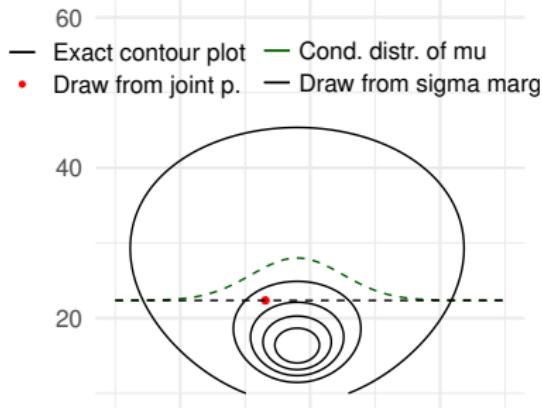
$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

$$p(\mu | \sigma^2, y) = N(\mu | \bar{y}, \sigma^2/n)$$

$$\mu^{(s)} \sim p(\mu | (\sigma^2)^{(s)}, y)$$

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

## Joint posterior



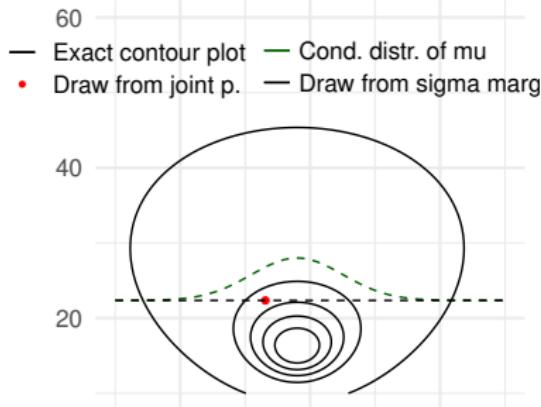
## Marginal of $\sigma$



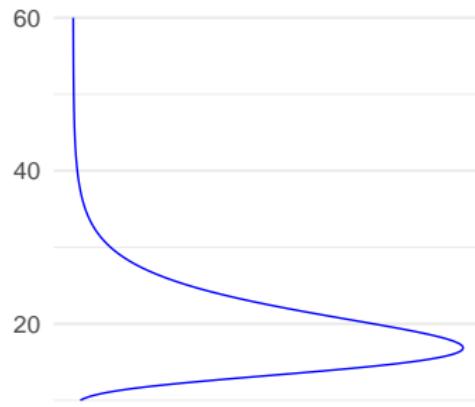
## Factorization

$$p(\mu, \sigma^2 | y) = \textcolor{green}{p}(\mu | \sigma^2, y) \textcolor{blue}{p}(\sigma^2 | y)$$

## Joint posterior



## Marginal of sigma

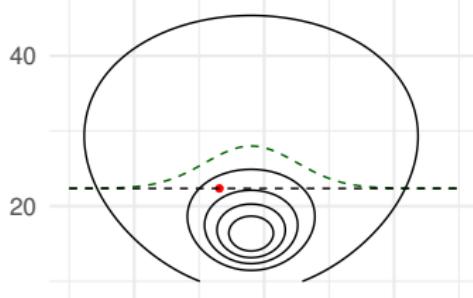


## Factorization

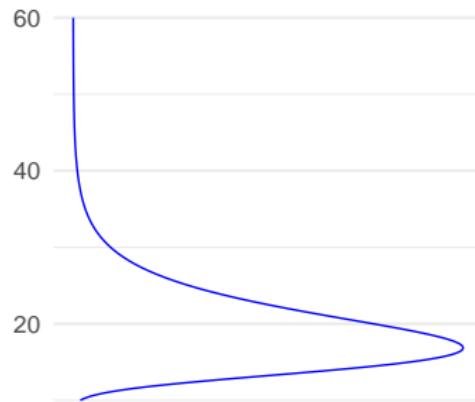
$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y)p(\sigma^2 | y)$$
$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

## Joint posterior

- Exact contour plot
- Draw from joint p.
- Cond. distr. of  $\mu$
- Draw from sigma marg



## Marginal of sigma



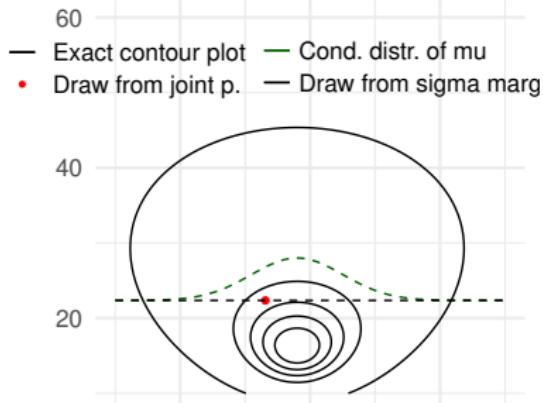
## Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y)p(\sigma^2 | y)$$

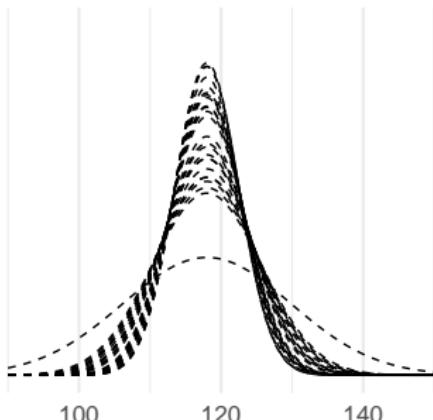
$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

$$p(\mu | (\sigma^2)^{(s)}, y) = N(\mu | \bar{y}, (\sigma^2)^{(s)}/n)$$

## Joint posterior



Cond distr of  $\mu$  for 25 draws



## Marginal of sigma



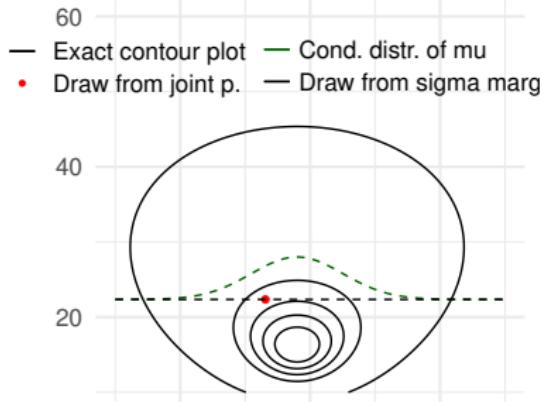
Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y)p(\sigma^2 | y)$$

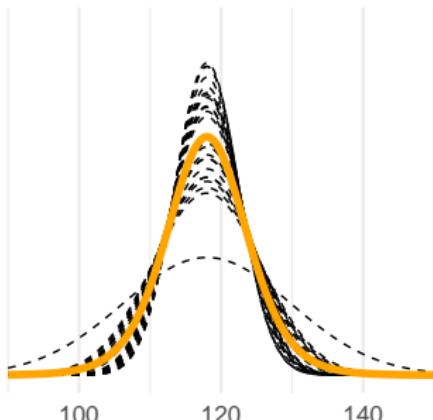
$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

$$p(\mu | (\sigma^2)^{(s)}, y) = N(\mu | \bar{y}, (\sigma^2)^{(s)}/n)$$

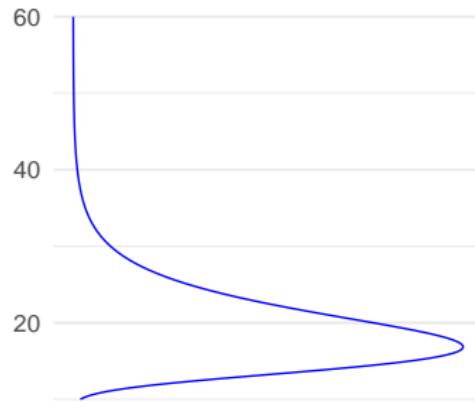
## Joint posterior



Cond distr of mu for 25 draws



## Marginal of sigma



Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y)p(\sigma^2 | y)$$

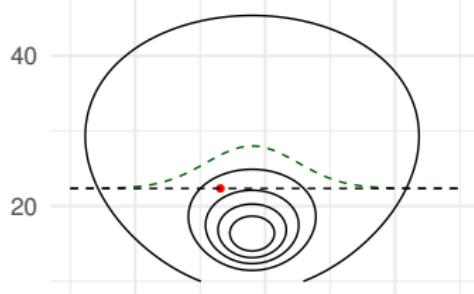
$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

$$p(\mu | (\sigma^2)^{(s)}, y) = N(\mu | \bar{y}, (\sigma^2)^{(s)}/n)$$

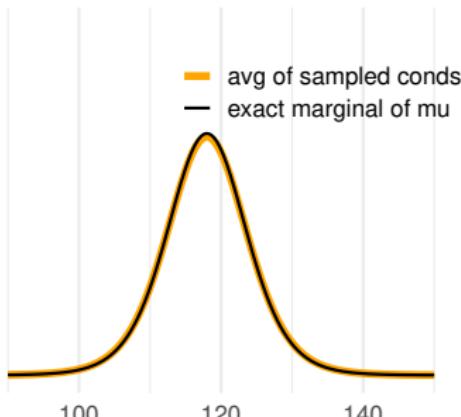
$$p(\mu | y) \approx \frac{1}{S} \sum_{s=1}^S N(\mu | \bar{y}, (\sigma^2)^{(s)}/n)$$

## Joint posterior

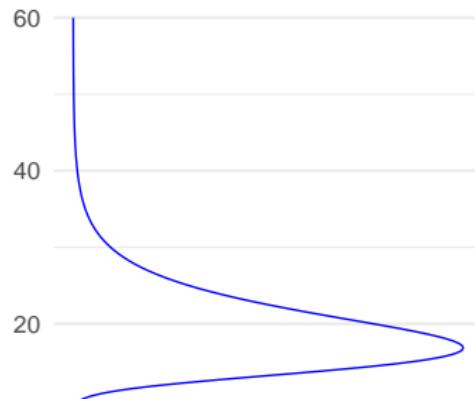
- Exact contour plot
- Draw from joint p.
- Cond. distr. of  $\mu$
- Draw from sigma marg



Cond. distr of  $\mu$



## Marginal of sigma



Factorization

$$p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y)p(\sigma^2 | y)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 | y)$$

$$p(\mu | (\sigma^2)^{(s)}, y) = N(\mu | \bar{y}, (\sigma^2)^{(s)}/n)$$

$$p(\mu | y) \approx \frac{1}{S} \sum_{s=1}^S N(\mu | \bar{y}, (\sigma^2)^{(s)}/n)$$

Marginal posterior  $p(\mu | y)$

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

## Marginal posterior $p(\mu | y)$

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

Marginal posterior  $p(\mu | y)$

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

$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right) d\sigma^2$$

Transformation

$$A = (n-1)s^2 + n(\mu - \bar{y})^2$$

Marginal posterior  $p(\mu | y)$

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

Transformation

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

Marginal posterior  $p(\mu | y)$

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

Transformation

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

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

## Marginal posterior $p(\mu | y)$

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

Transformation

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

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

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

Marginal posterior  $p(\mu | y)$

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

Transformation

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

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

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

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

## Marginal posterior $p(\mu | y)$

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

Transformation

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

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

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

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

## Marginal posterior $p(\mu | y)$

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

Transformation

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

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

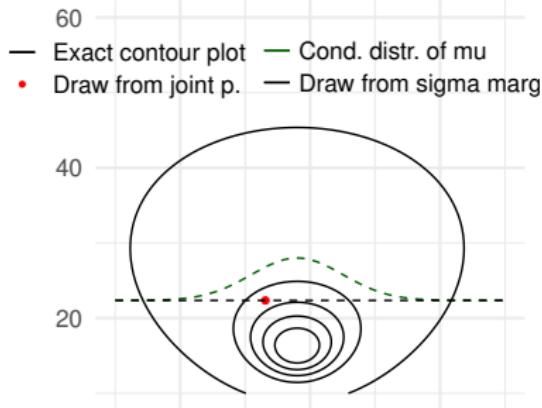
Recognize gamma integral  $\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$

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

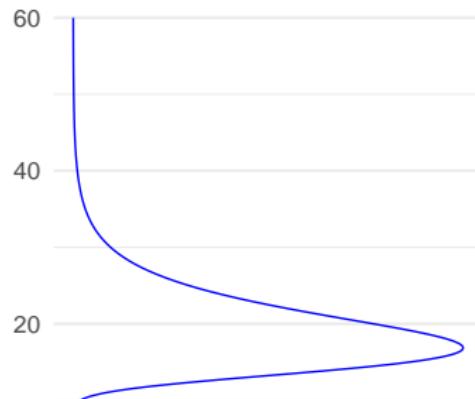
$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

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

### Joint posterior



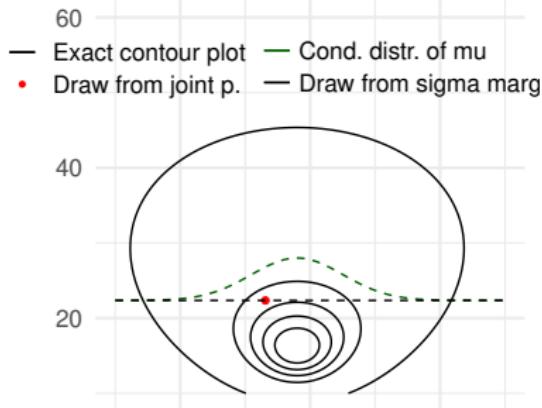
### Marginal of sigma



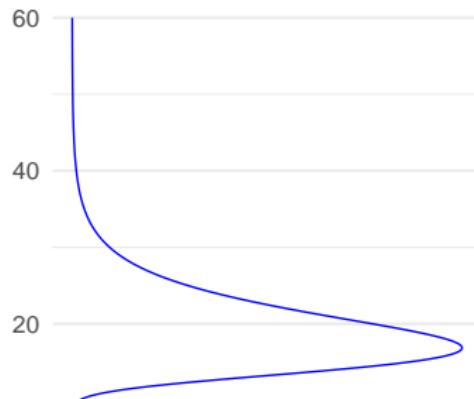
Predictive distribution for new  $\tilde{y}$

$$p(\tilde{y} | y) = \int p(\tilde{y} | \mu, \sigma) p(\mu, \sigma | y) d\mu \sigma$$

## Joint posterior



## Marginal of sigma

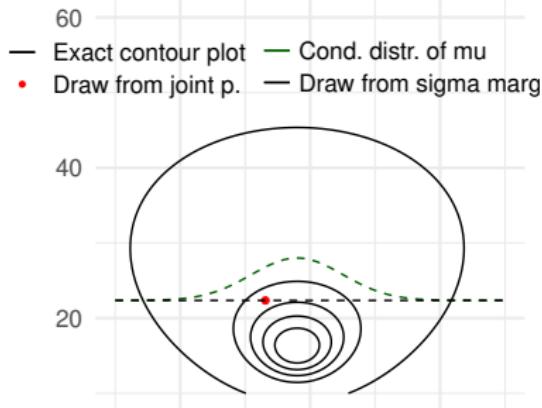


Predictive distribution for new  $\tilde{y}$

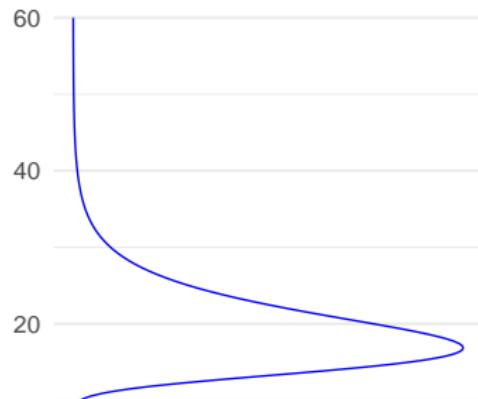
$$p(\tilde{y} | y) = \int p(\tilde{y} | \mu, \sigma) p(\mu, \sigma | y) d\mu d\sigma$$

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

## Joint posterior



## Marginal of sigma



## Predictive distribution for new $\tilde{y}$

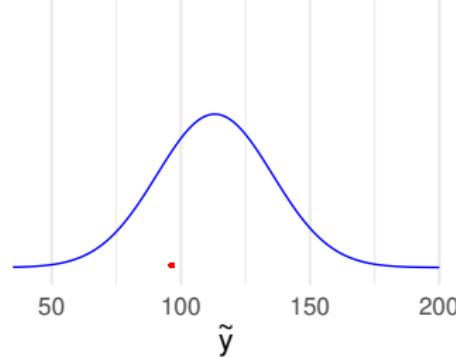
$$p(\tilde{y} | y) = \int p(\tilde{y} | \mu, \sigma) p(\mu, \sigma | y) d\mu d\sigma$$

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

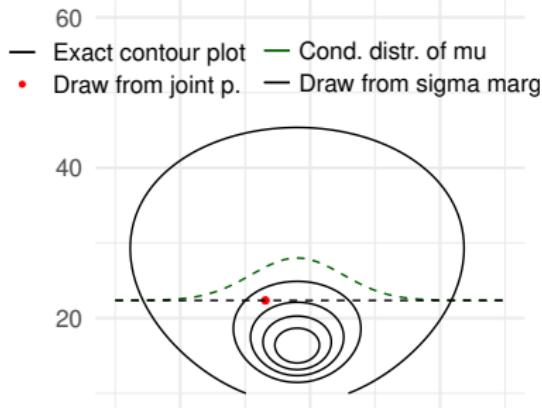
$$\tilde{y}^{(s)} \sim p(\tilde{y} | \mu^{(s)}, \sigma^{(s)})$$

## Posterior predictive distribution

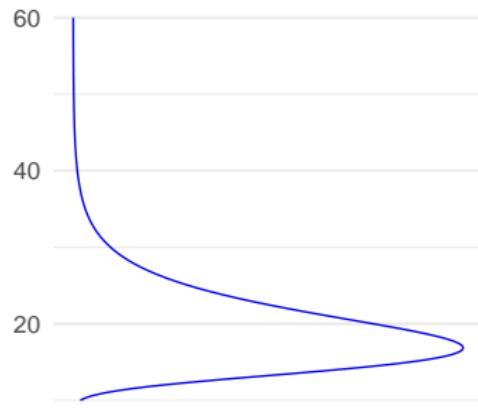
- A draw from the predictive distribution
- Predictive distribution given posterior draws



## Joint posterior



## Marginal of sigma



## Predictive distribution for new $\tilde{y}$

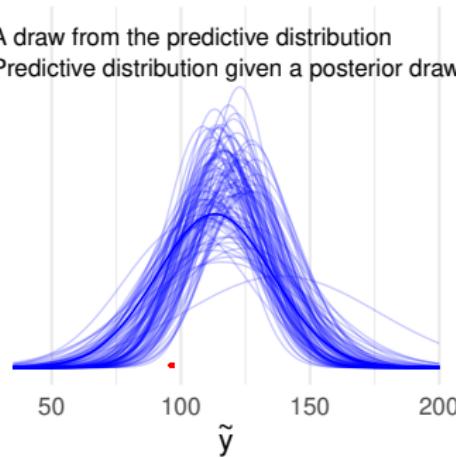
$$p(\tilde{y} | y) = \int p(\tilde{y} | \mu, \sigma) p(\mu, \sigma | y) d\mu d\sigma$$

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

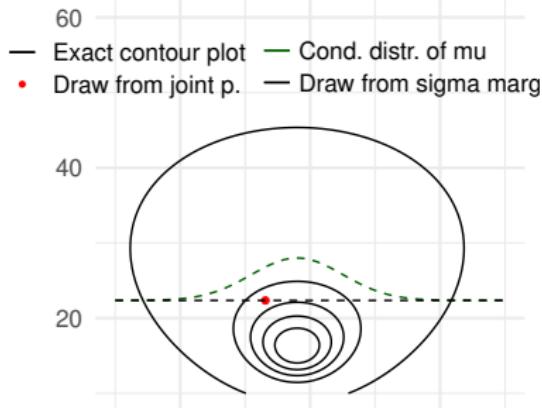
$$\tilde{y}^{(s)} \sim p(\tilde{y} | \mu^{(s)}, \sigma^{(s)})$$

## Posterior predictive distribution

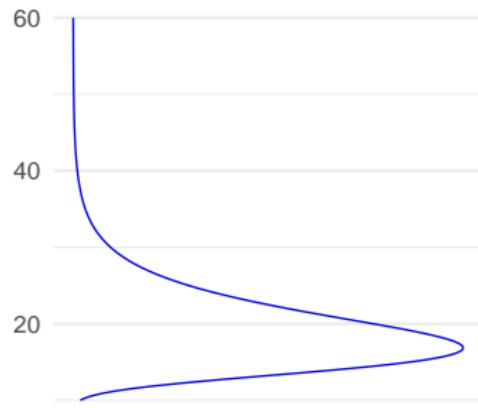
- A draw from the predictive distribution
- Predictive distribution given a posterior draw



## Joint posterior



## Marginal of sigma



## Predictive distribution for new $\tilde{y}$

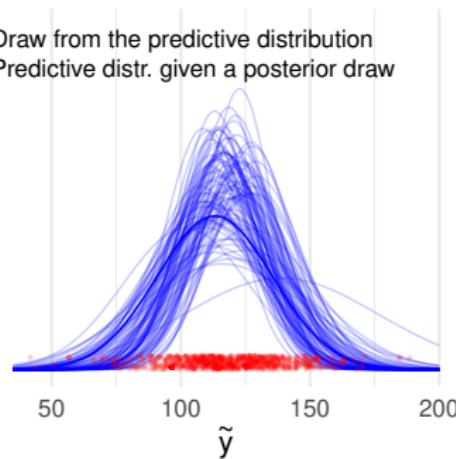
$$p(\tilde{y} | y) = \int p(\tilde{y} | \mu, \sigma) p(\mu, \sigma | y) d\mu d\sigma$$

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

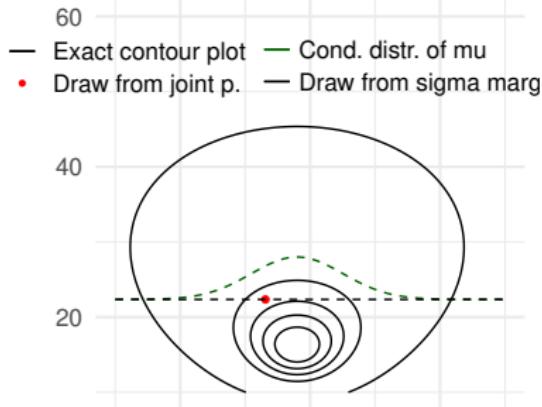
$$\tilde{y}^{(s)} \sim p(\tilde{y} | \mu^{(s)}, \sigma^{(s)})$$

## Posterior predictive distribution

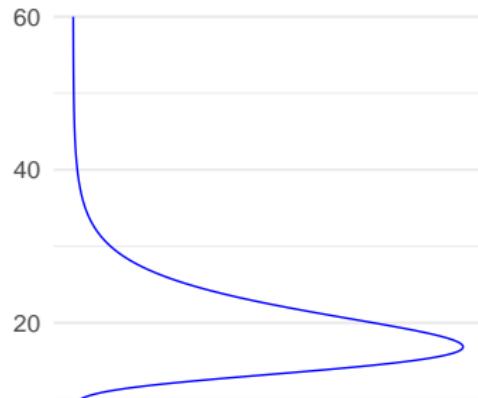
- Draw from the predictive distribution
- Predictive distr. given a posterior draw



## Joint posterior



## Marginal of sigma



## Predictive distribution for new $\tilde{y}$

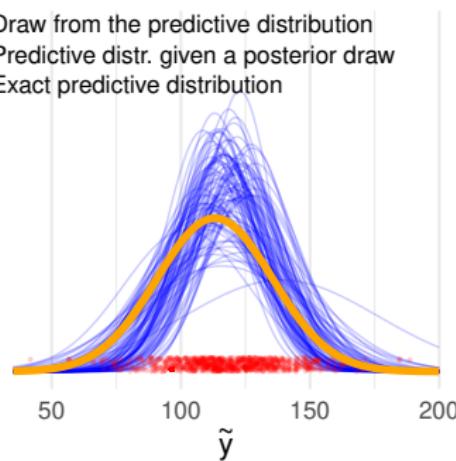
$$p(\tilde{y} | y) = \int p(\tilde{y} | \mu, \sigma) p(\mu, \sigma | y) d\mu d\sigma$$

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

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

## Posterior predictive distribution

- Draw from the predictive distribution
- Predictive distr. given a posterior draw
- Exact predictive distribution



## Normal - posterior predictive distribution

Posterior predictive distribution given known variance

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

## Normal - posterior predictive distribution

Posterior predictive distribution given known variance

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

## Normal - posterior predictive distribution

Posterior predictive distribution given known variance

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

## Normal - posterior predictive distribution

Posterior predictive distribution given known variance

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

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

## Normal - posterior predictive distribution

Posterior predictive distribution given known variance

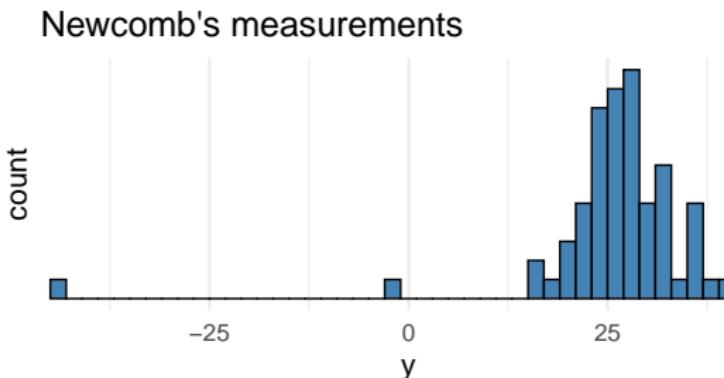
$$\begin{aligned} p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\ &= \int N(\tilde{y} \mid \mu, \sigma^2) N(\mu \mid \bar{y}, \sigma^2/n) d\mu \\ &= N(\tilde{y} \mid \bar{y}, (1 + \frac{1}{n})\sigma^2) \end{aligned}$$

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

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

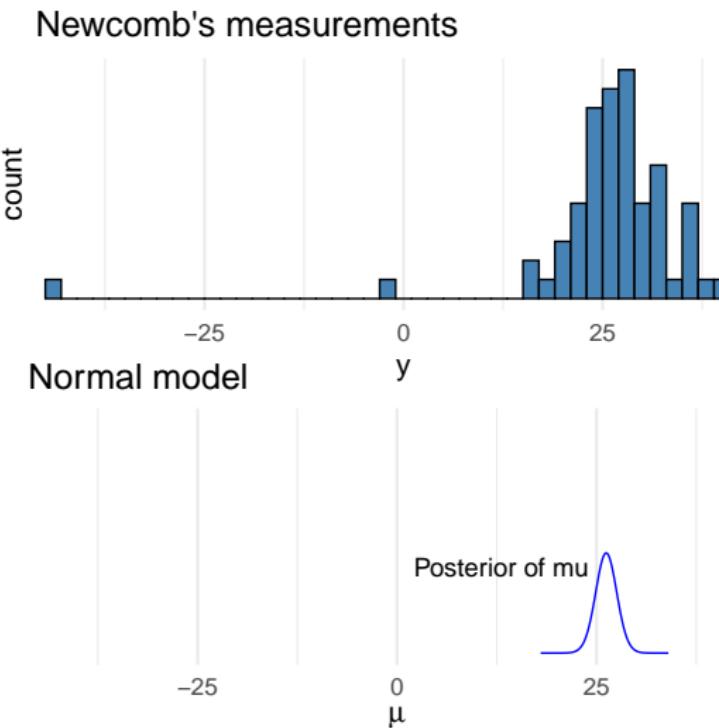
## Simon Newcomb's light of speed experiment in 1882

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.



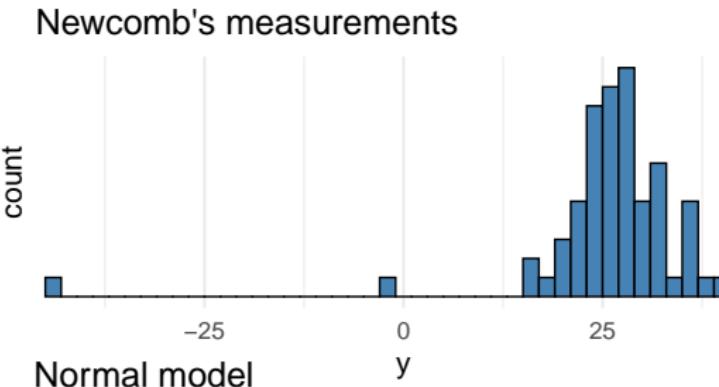
## Simon Newcomb's light of speed experiment in 1882

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.



## 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.

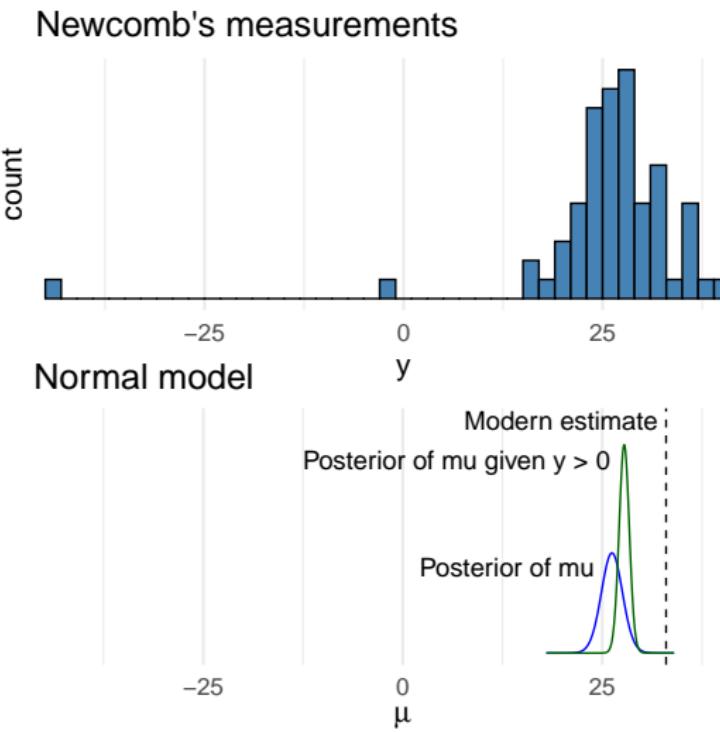


Posterior of  $\mu$  given  $y > 0$

Posterior of  $\mu$

## 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.



## Normal - conjugate prior

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

## Normal - 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

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

## Normal - 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

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

- $\mu$  and  $\sigma^2$  are a priori dependent
  - if  $\sigma^2$  is large, then  $\mu$  has wide prior

# Normal - conjugate prior

Joint posterior (exercise 3.9)

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

where

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

$$\kappa_n = \kappa_0 + n$$

$$\nu_n = \nu_0 + n$$

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

## Comparison of means of two normals

- The difference of two normally distributed variables is normally distributed
- The difference of two  $t$  distributed variables with different variances and degrees of freedom doesn't have a closed form
  - but easy to sample from the two distributions, and obtain draws of the differences

$$\text{if } \mu_1^{(s)} \sim p(\mu_1 | y_1)$$

$$\mu_2^{(s)} \sim p(\mu_2 | y_2)$$

$$\delta^{(s)} = \mu_1^{(s)} - \mu_2^{(s)}$$

$$\text{then } \delta^{(s)} \sim p(\delta | y_1, y_2)$$

# Multivariate normal

- Observation model

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

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

# Multivariate normal

- Observation model

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

- BDA3 p. 72–
- Recommended LKJ-prior mentioned in Appendix A, see more in Stan manual
- Gaussian process and Gaussian Markov random field models are in practice computed with multivariate normals
  - GPs in BDA3 Chapter 21, and a course in spring
  - GPs and GMRFs often used also as priors for latent functions and combined with non-normal observation models

## Normal linear regression

- $y_i \sim N(\alpha + \beta x_i, \sigma^2), \quad i = 1, \dots, N$

## Normal linear regression

- $y_i \sim N(\alpha + \beta x_i, \sigma^2)$ ,  $i = 1, \dots, N$
- with normal fixed scale prior on  $\alpha$  and  $\beta$ , and known  $\sigma^2$  the posterior is multivariate normal

## Normal linear regression

- $y_i \sim N(\alpha + \beta x_i, \sigma^2)$ ,  $i = 1, \dots, N$
- with normal fixed scale prior on  $\alpha$  and  $\beta$ , and known  $\sigma^2$  the posterior is multivariate normal
- with unknown  $\sigma^2$ , the posterior is multivariate N-Inv- $\chi^2$

## Normal linear regression

- $y_i \sim N(\alpha + \beta x_i, \sigma^2)$ ,  $i = 1, \dots, N$
- with normal fixed scale prior on  $\alpha$  and  $\beta$ , and known  $\sigma^2$  the posterior is multivariate normal
- with unknown  $\sigma^2$ , the posterior is multivariate N-Inv- $\chi^2$
- with unknown prior scales and  $\sigma^2$ , numerical integration needed

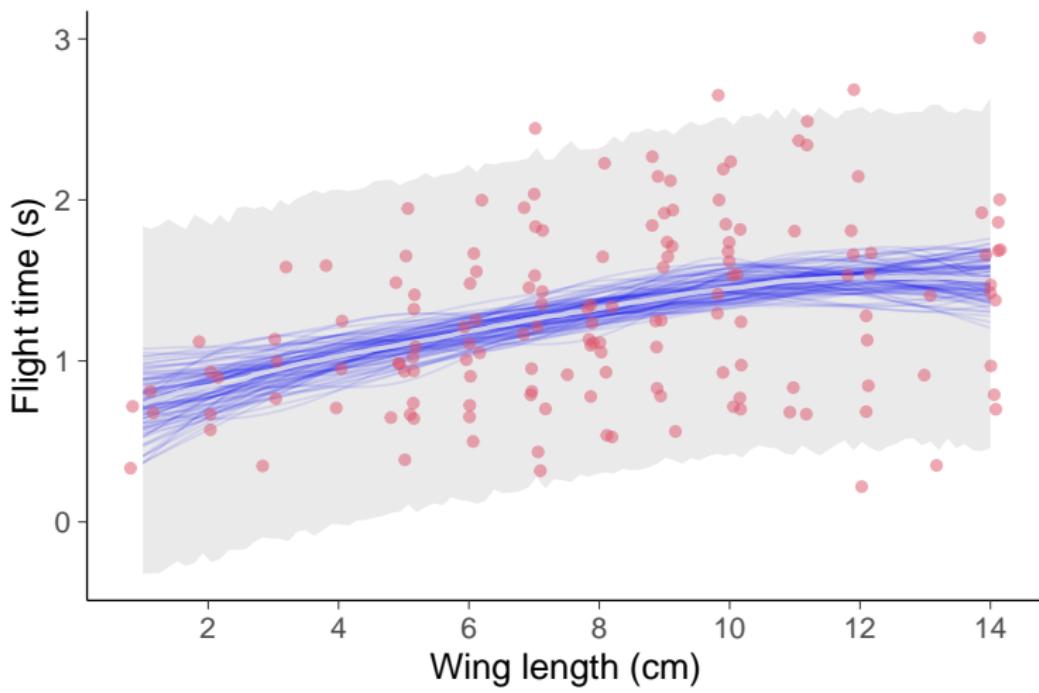
## Normal linear regression

- $y_i \sim N(\alpha + \beta x_i, \sigma^2)$ ,  $i = 1, \dots, N$
- with normal fixed scale prior on  $\alpha$  and  $\beta$ , and known  $\sigma^2$  the posterior is multivariate normal
- with unknown  $\sigma^2$ , the posterior is multivariate N-Inv- $\chi^2$
- with unknown prior scales and  $\sigma^2$ , numerical integration needed
- more in BDA3 Chapter 14 (not part of the course) and Regression and Other Stories book

# Paper helicopter flight time

$$y \sim \text{normal}(f, \sigma)$$

$$f \sim GP(0, K(x, \theta))$$

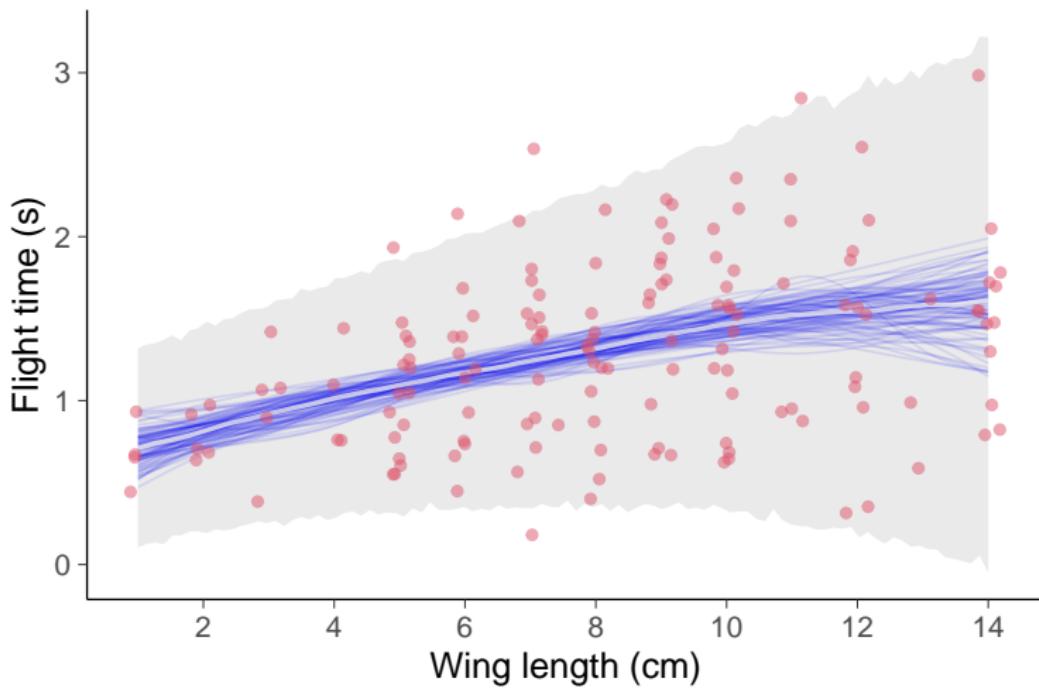


# Paper helicopter flight time

$$y \sim \text{normal}(f, \sigma)$$

$$f \sim GP(0, K_f(x, \theta_f))$$

$$\log(\sigma) \sim GP(0, K_g(x, \theta_g))$$



## Scale mixture of normals

- Many useful distributions can be presented as scale mixture of normals, e.g.
  - Student's  $t$
  - Cauchy
  - Double exponential aka Laplace
  - Horseshoe
  - R2-D2

## Multinomial model for categorical data

- Extension of binomial
- Observation model

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

- BDA3 p. 69–

## Generalized linear model (GLM)

- $y_i \sim p(g^{-1}(\alpha + \beta x_i), \phi), \quad i = 1, \dots, N$ 
  - where  $p$  is non-normal (in original definition in exponential family)
  - and  $g$  is a link function

## Generalized linear model (GLM)

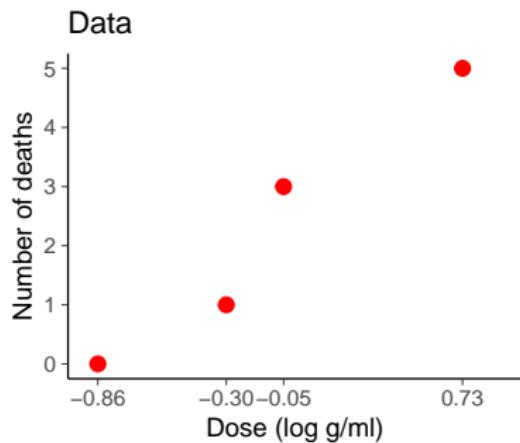
- $y_i \sim p(g^{-1}(\alpha + \beta x_i), \phi)$ ,  $i = 1, \dots, N$ 
  - where  $p$  is non-normal (in original definition in exponential family)
  - and  $g$  is a link function
- Bioassay analysis is used as an example

## Generalized linear model (GLM)

- $y_i \sim p(g^{-1}(\alpha + \beta x_i), \phi)$ ,  $i = 1, \dots, N$ 
  - where  $p$  is non-normal (in original definition in exponential family)
  - and  $g$  is a link function
- Bioassay analysis is used as an example
- More in BDA3 Chapter 16 and Regression and other stories book

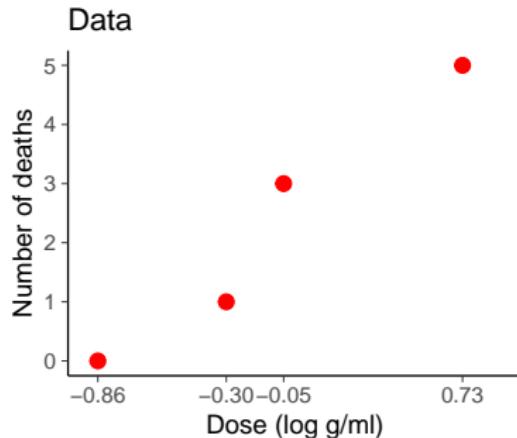
# Bioassay

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



# 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

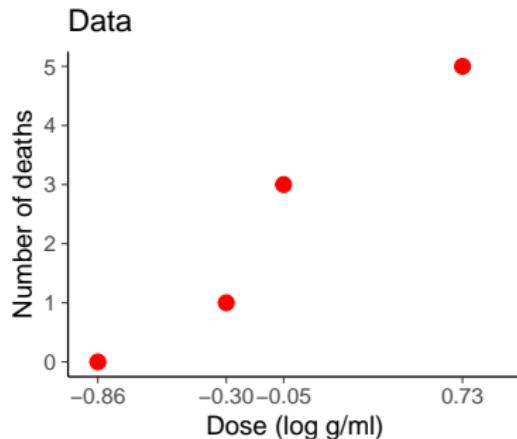


Find out lethal dose 50% (LD50)

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

# 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)

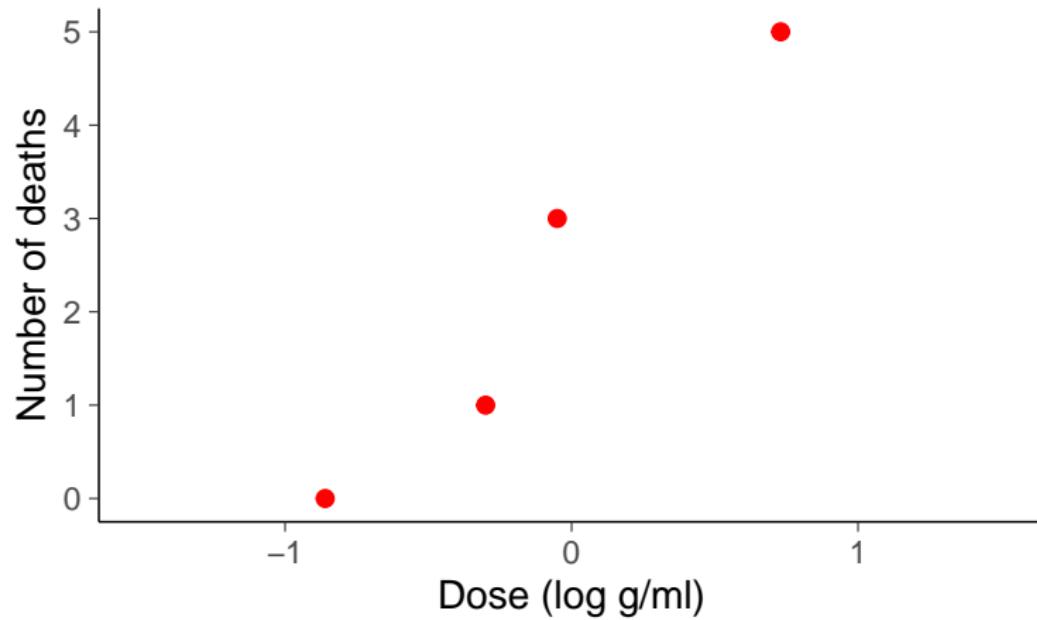
- used to classify how hazardous chemical is
- 1984 EEC directive has 4 levels (see the chapter notes)

Bayesian methods help to

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

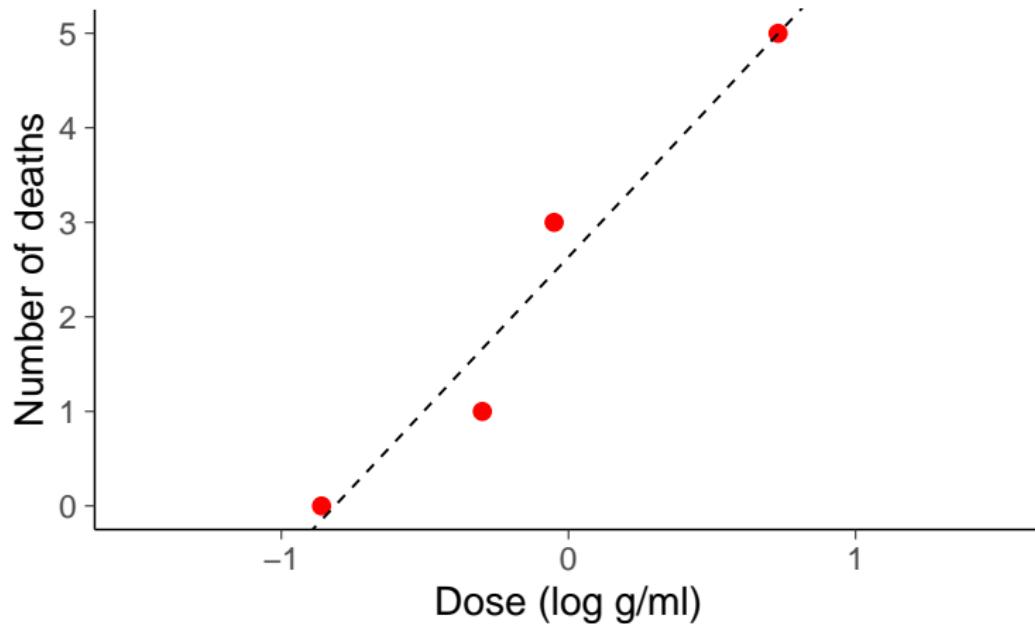
# Bioassay

## Data



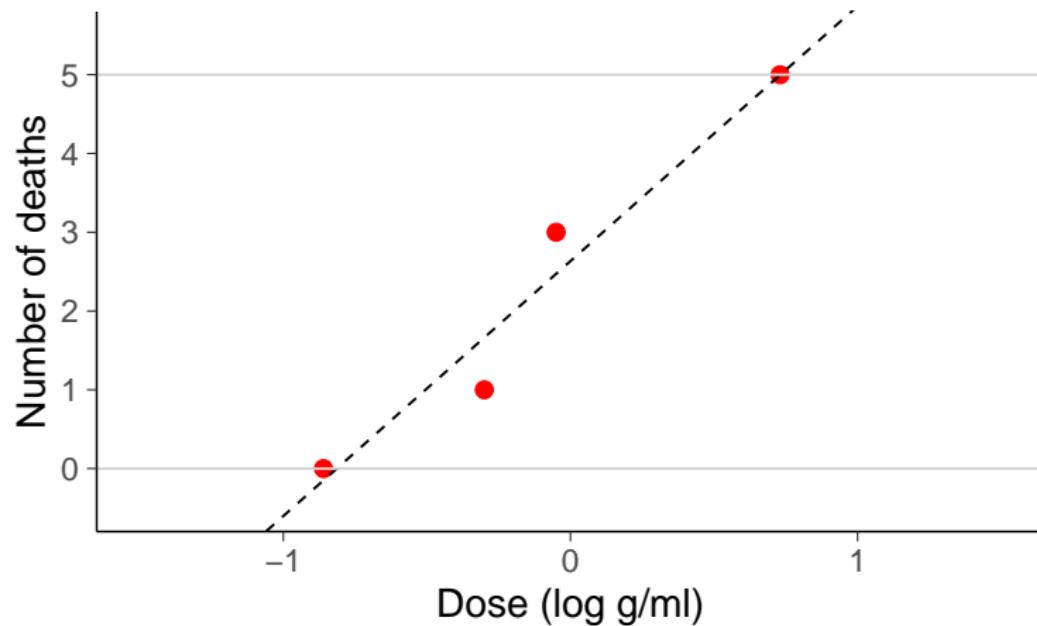
# Bioassay

## Linear fit

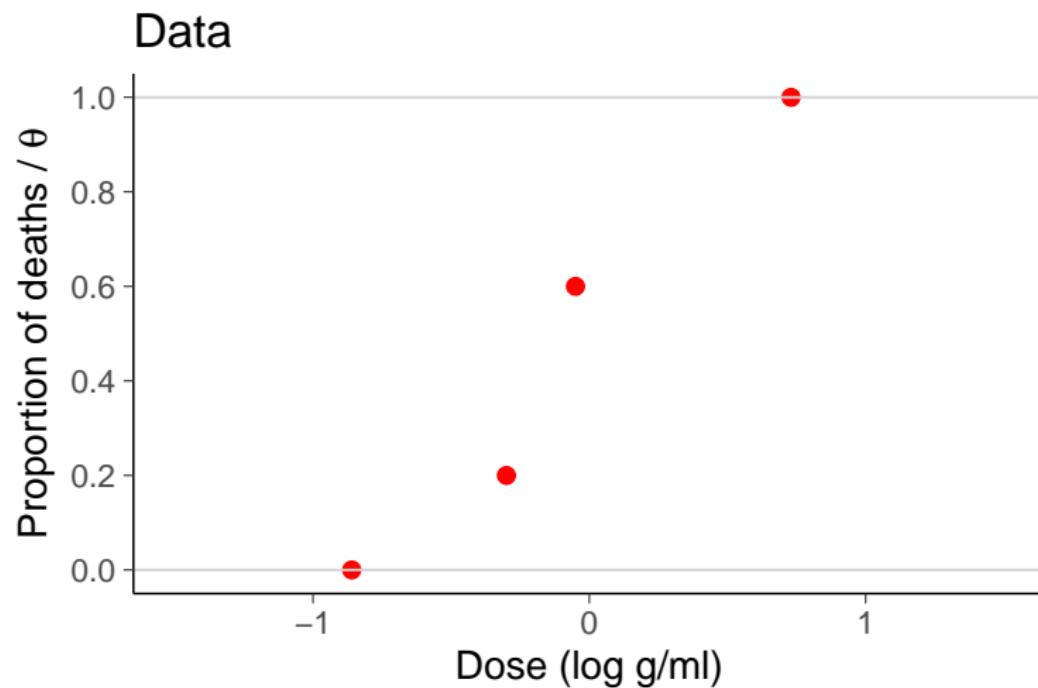


# Bioassay

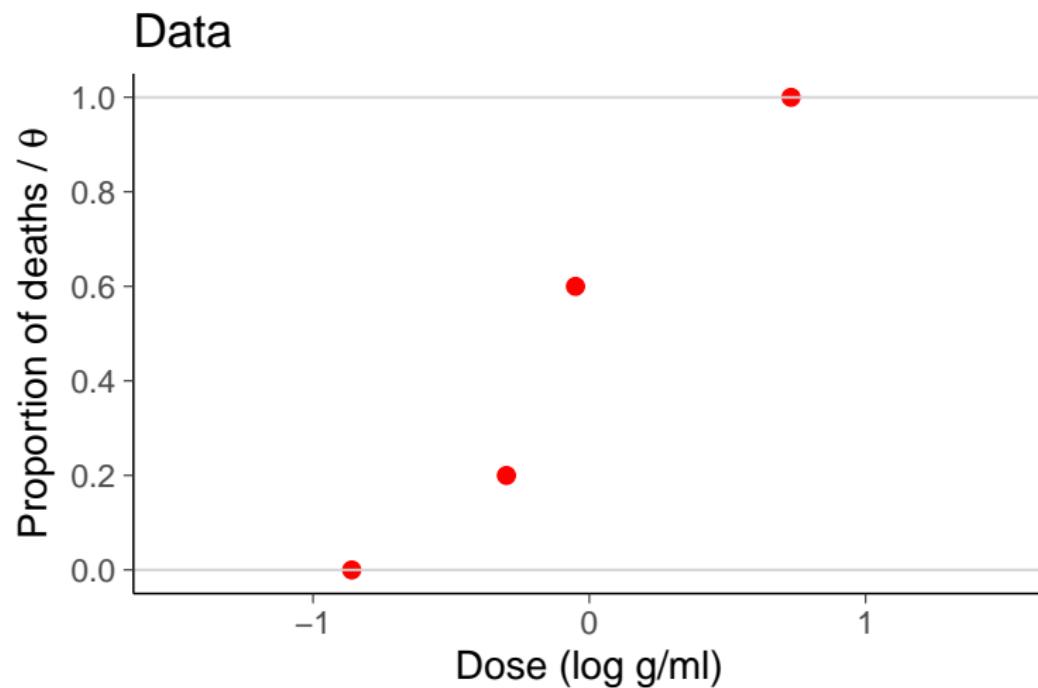
## Linear fit



# Bioassay



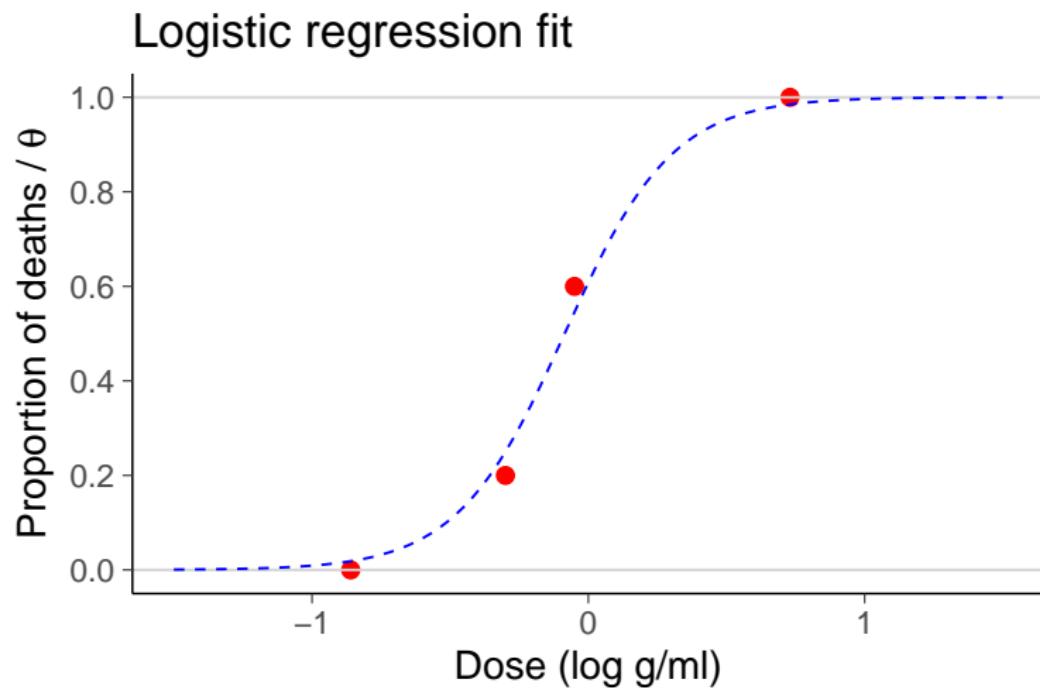
# Bioassay



Binomial model

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

# Bioassay



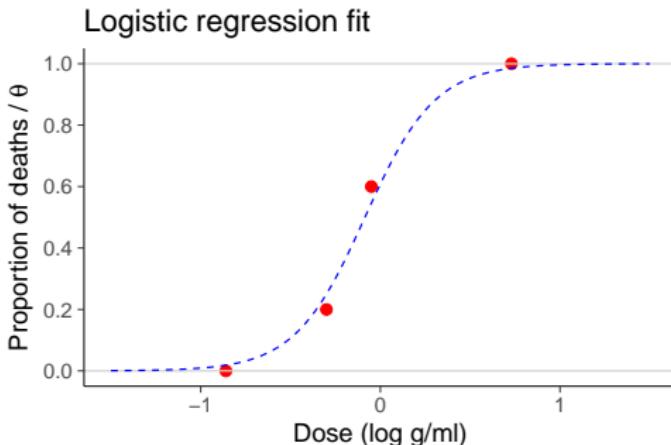
Binomial model

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

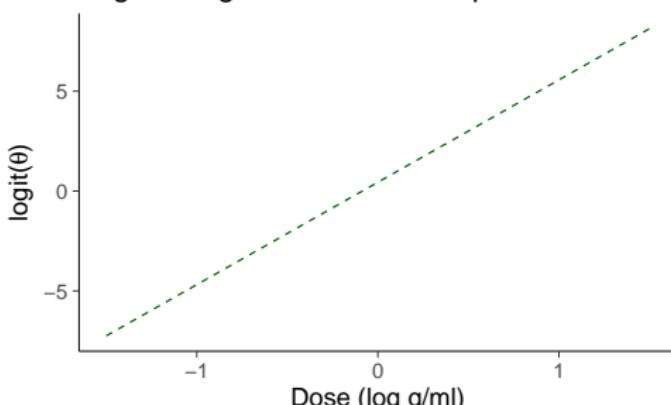
# Bioassay

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

$$\begin{aligned}\text{logit}(\theta_i) &= \log\left(\frac{\theta_i}{1 - \theta_i}\right) \\ &= \alpha + \beta x_i\end{aligned}$$



Logistic regression in latent space

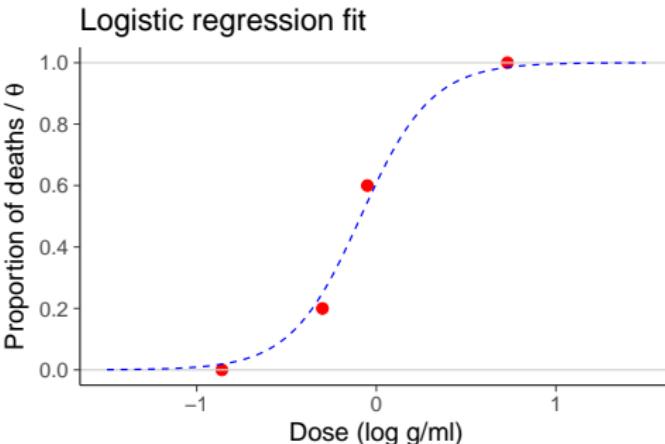


# Bioassay

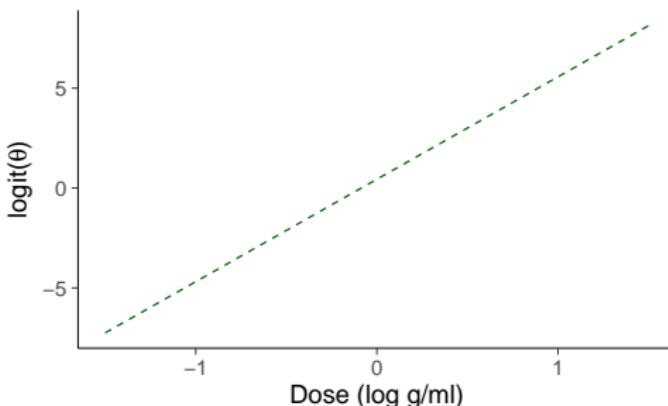
$$y_i \mid \theta_i \sim \text{Bin}(\theta_i, n_i)$$

$$\begin{aligned}\text{logit}(\theta_i) &= \log\left(\frac{\theta_i}{1 - \theta_i}\right) \\ &= \alpha + \beta x_i\end{aligned}$$

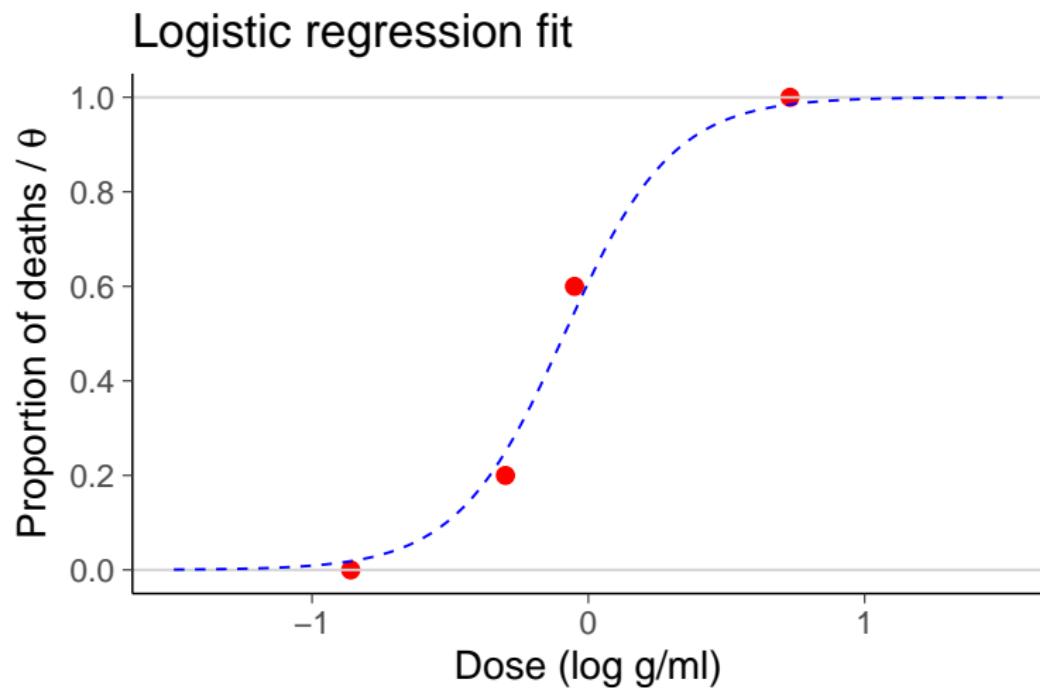
$$\theta_i = \frac{1}{1 + \exp(-(\alpha + \beta x_i))}$$



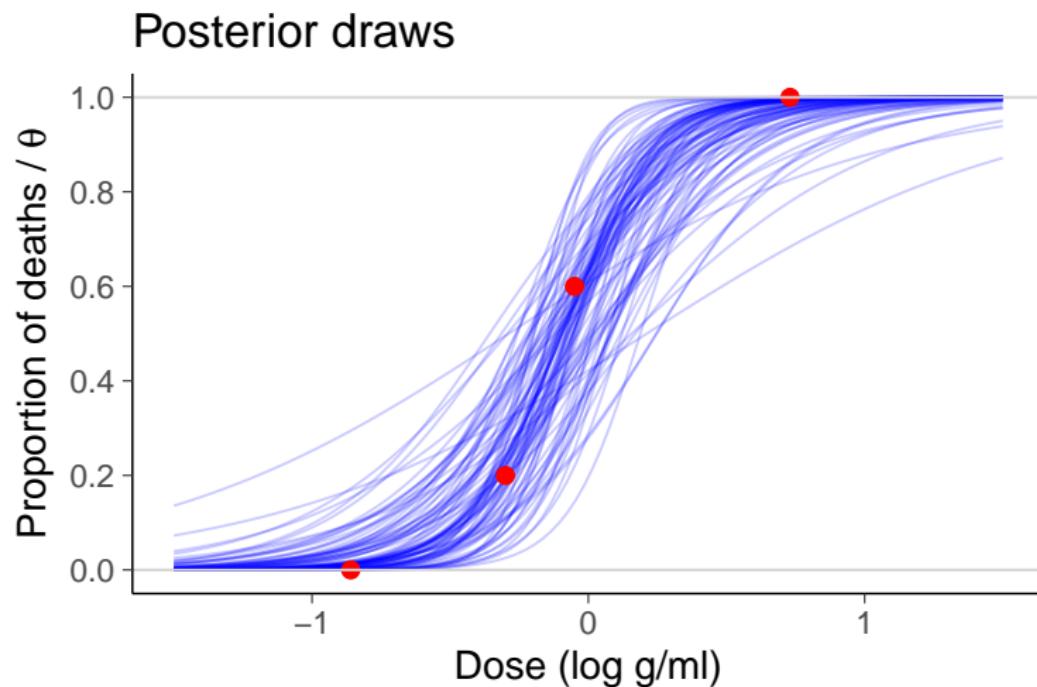
Logistic regression in latent space



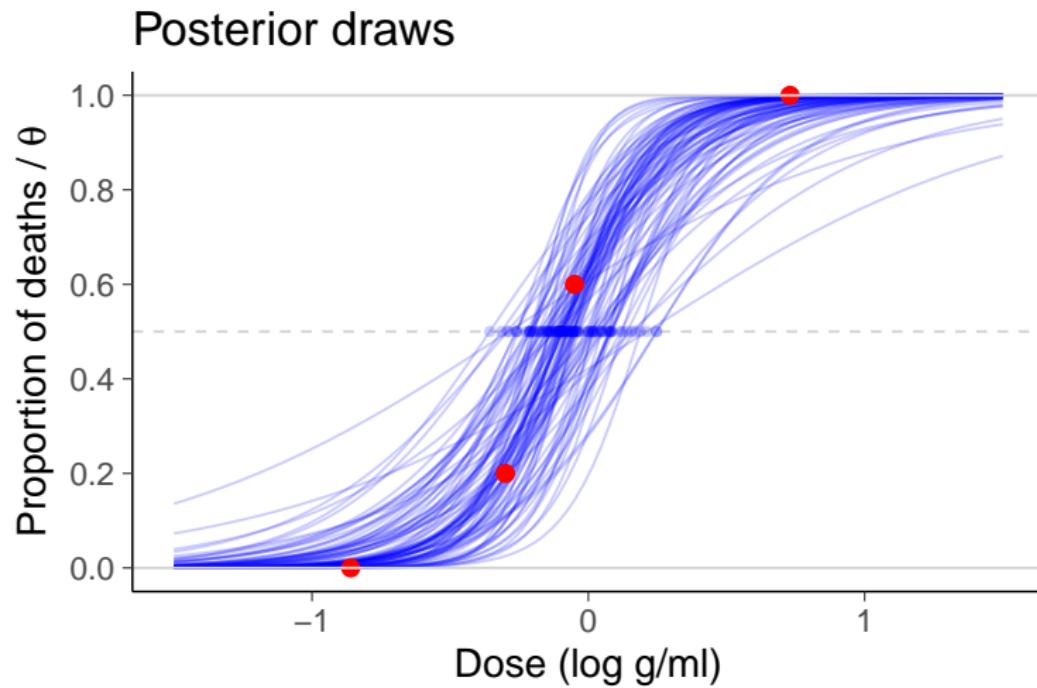
# Bioassay



# Bioassay

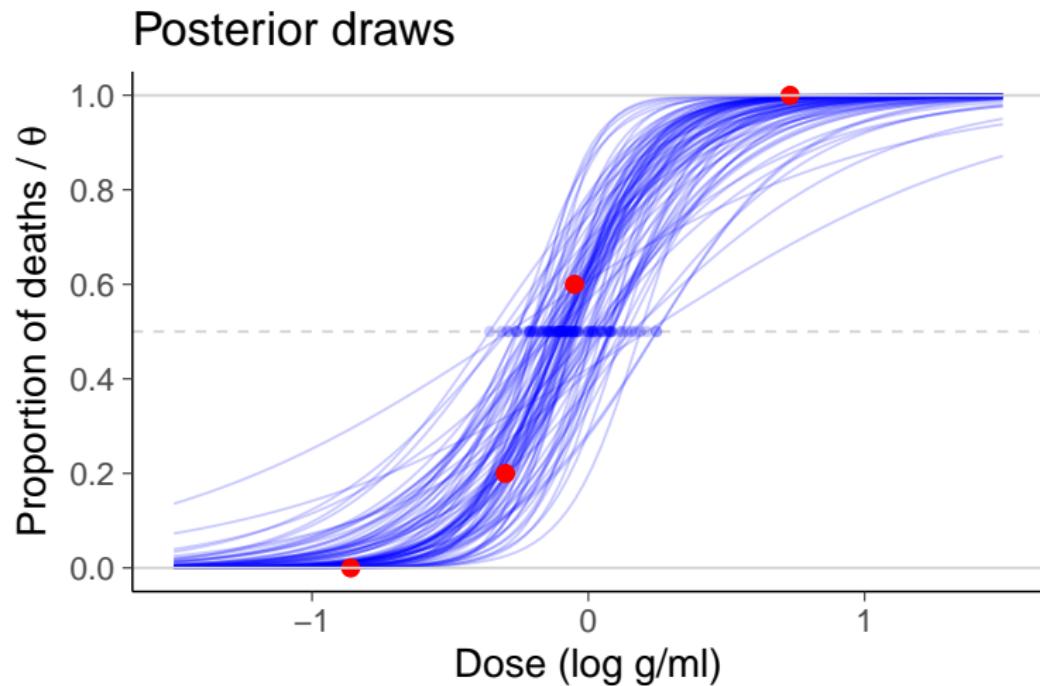


# Bioassay



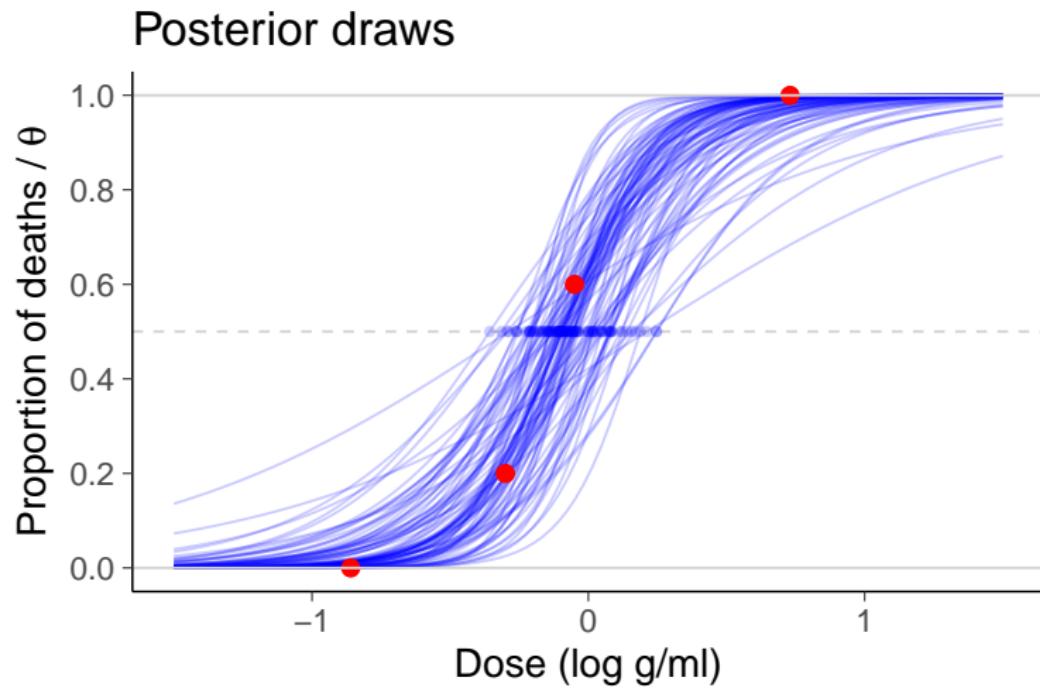
$$\text{LD50: } E\left(\frac{y}{n}\right) = \text{logit}^{-1}(\alpha + \beta x) = 0.5$$

# Bioassay



$$\text{LD50: } E\left(\frac{y}{n}\right) = \text{logit}^{-1}(\alpha + \beta x) = 0.5 \quad \Rightarrow \quad x_{\text{LD50}} = -\alpha/\beta$$

# Bioassay

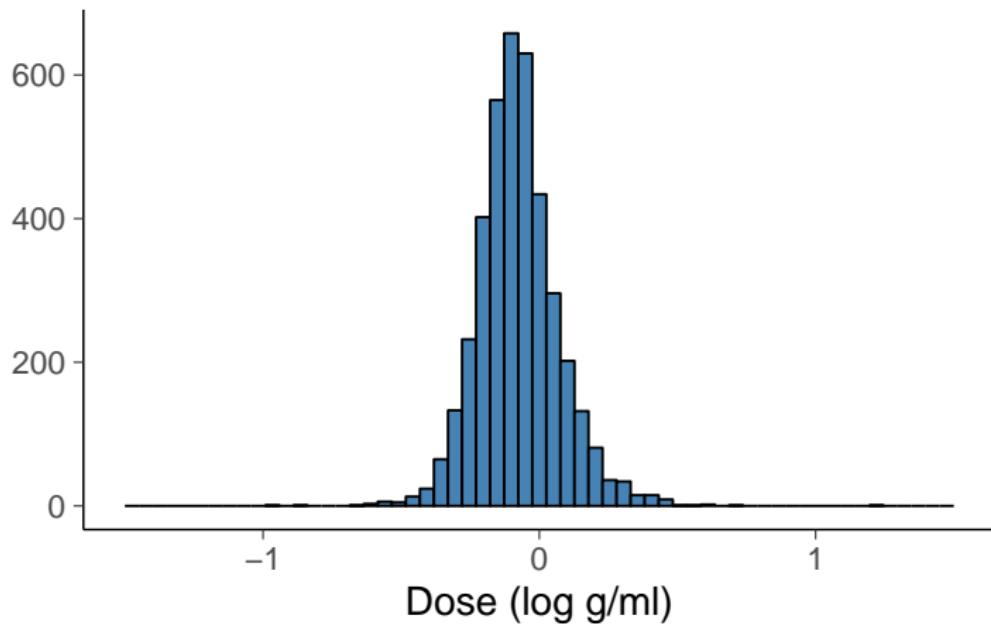


$$\text{LD50: } E\left(\frac{y}{n}\right) = \text{logit}^{-1}(\alpha + \beta x) = 0.5 \quad \Rightarrow \quad x_{\text{LD50}} = -\alpha/\beta$$

$$x_{\text{LD50}}^{(s)} = -\alpha^{(s)}/\beta^{(s)}$$

# Bioassay

## Bioassay LD50



$$\text{LD50: } E\left(\frac{y}{n}\right) = \text{logit}^{-1}(\alpha + \beta x) = 0.5 \quad \Rightarrow \quad x_{\text{LD50}} = -\alpha/\beta$$

$$x_{\text{LD50}}^{(s)} = -\alpha^{(s)}/\beta^{(s)}$$

# Bioassay posterior

Binomial model

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

Link function

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

# 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

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

# 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

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

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

# 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

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

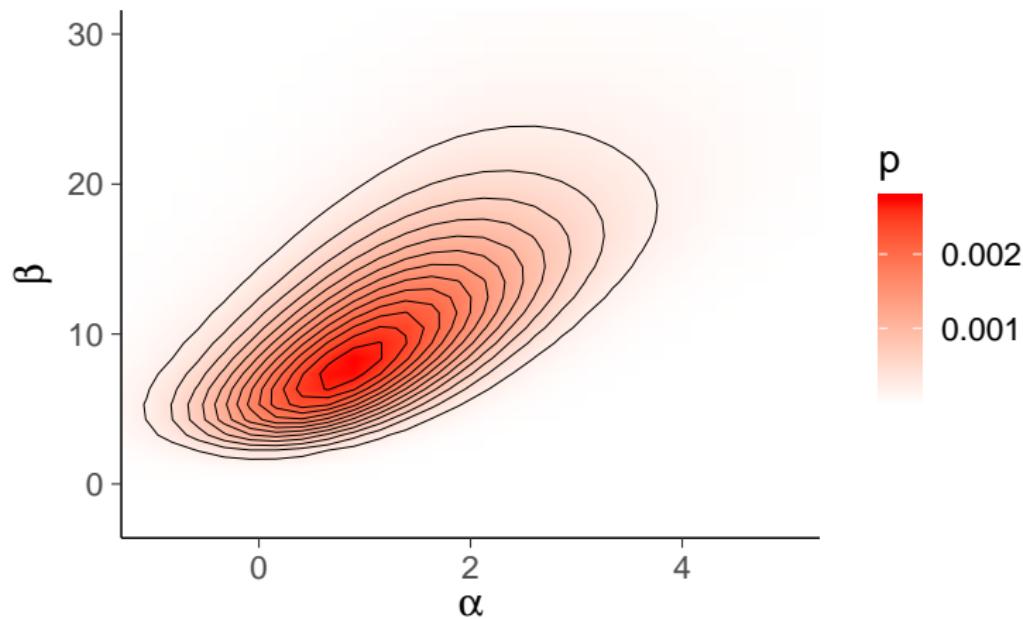
$$p(y_i \mid \alpha, \beta, n_i, x_i) \propto [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i} [1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}$$

Posterior (with uniform prior on  $\alpha, \beta$ )

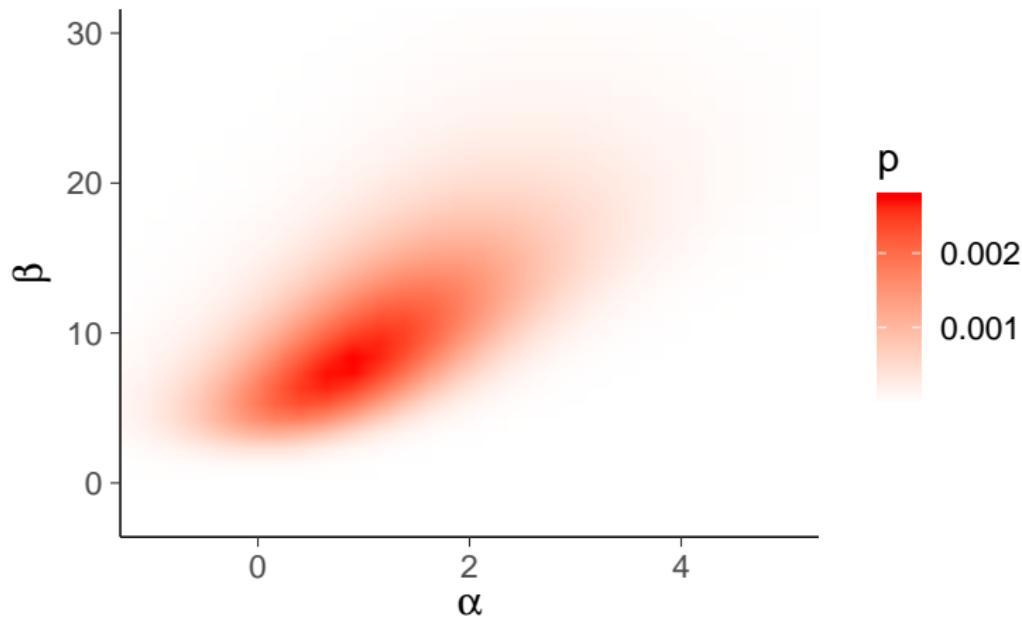
$$p(\alpha, \beta \mid y, n, x) \propto p(\alpha, \beta) \prod_{i=1}^n p(y_i \mid \alpha, \beta, n_i, x_i)$$

# Bioassay

Posterior density evaluated in a grid



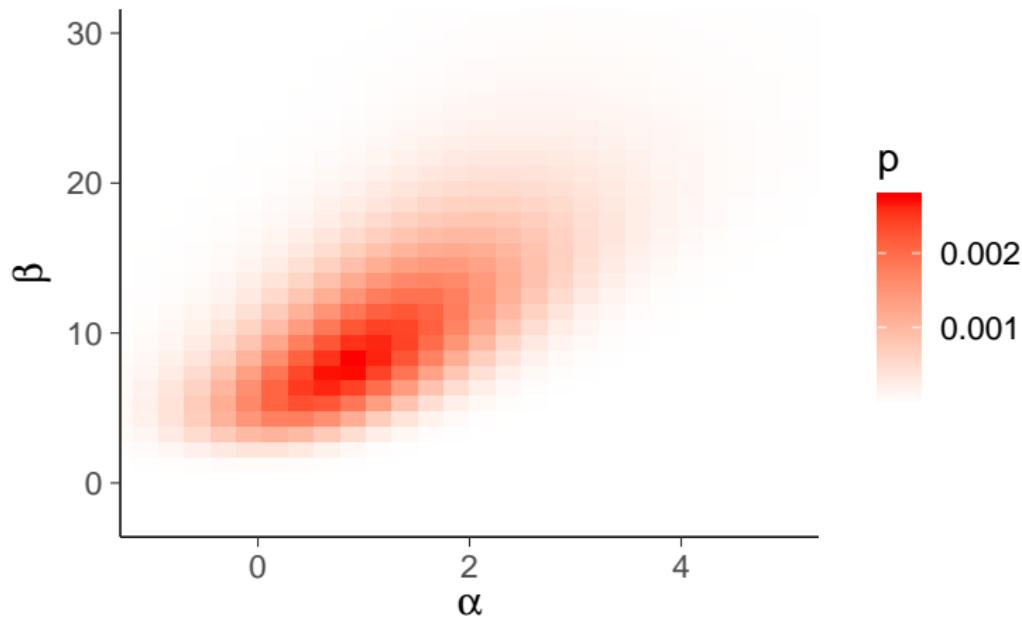
## Posterior density evaluated in a grid



Density evaluated in grid, but plotted using interpolation

# Bioassay

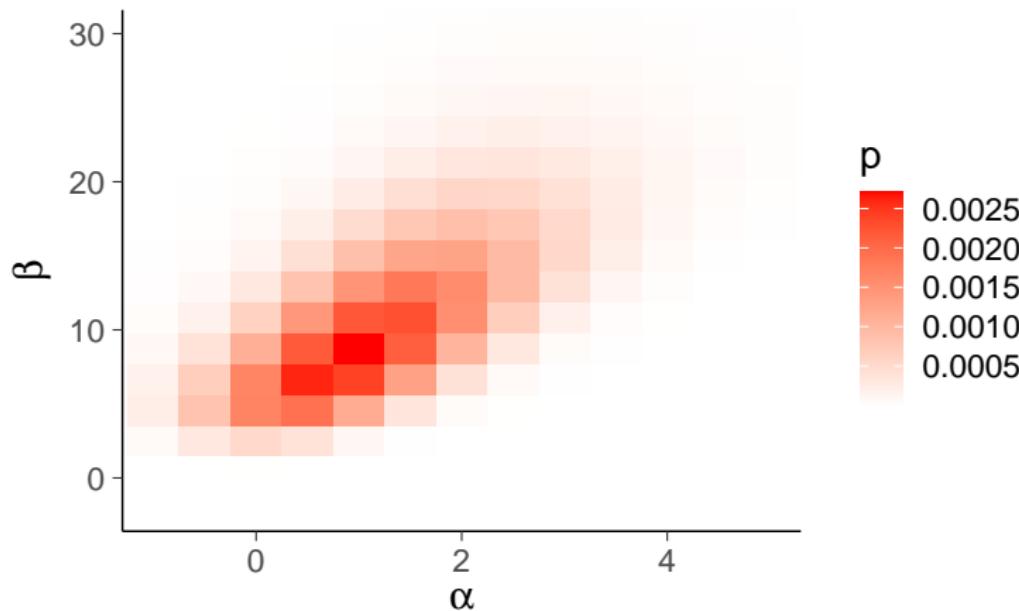
Posterior density evaluated in a grid



Density evaluated in grid, and plotted without interpolation

# Bioassay

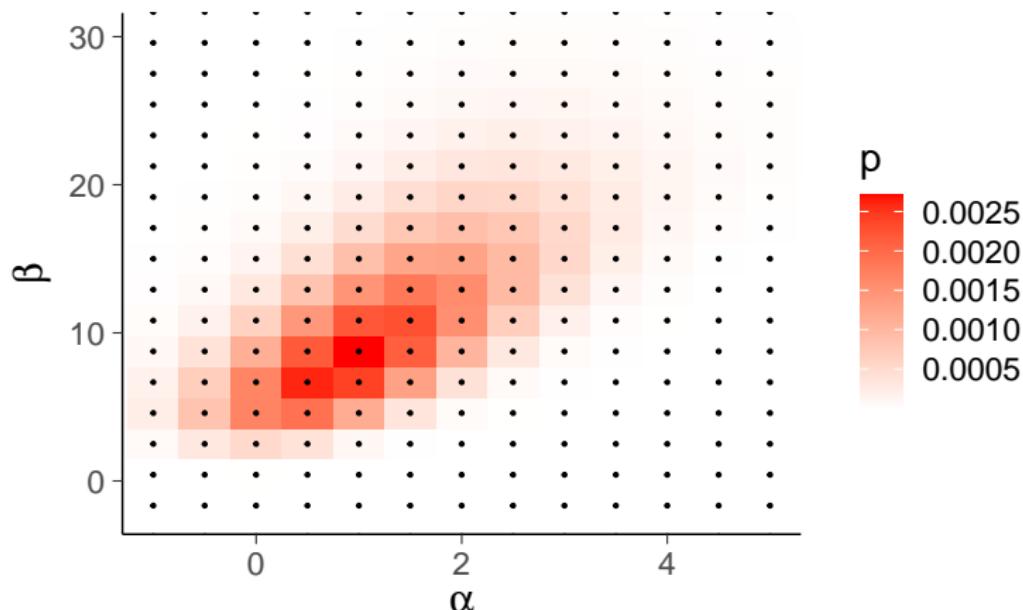
Posterior density evaluated in a grid



Density evaluated in a coarser grid

# Bioassay

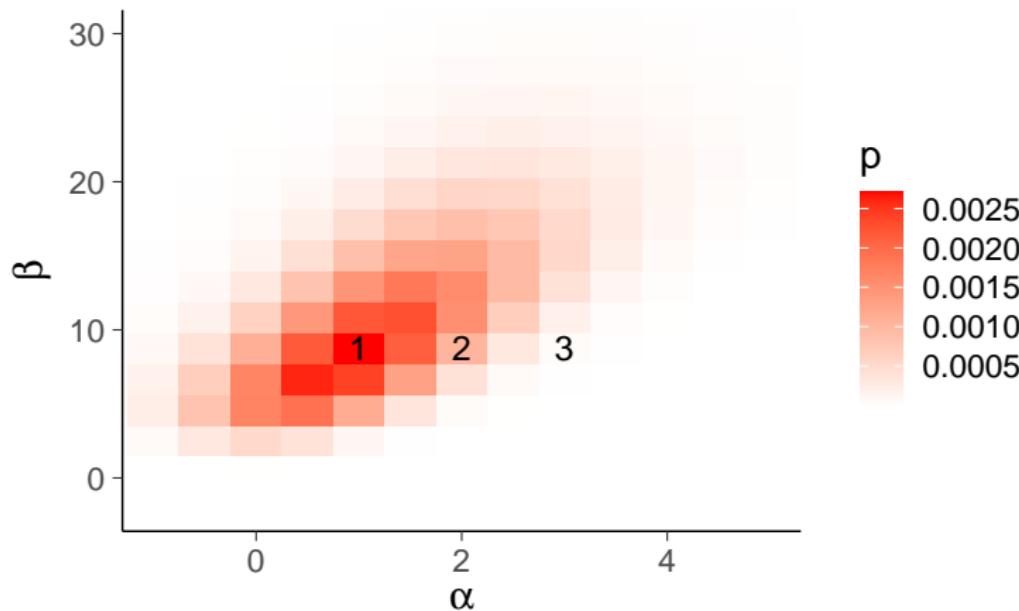
Posterior density evaluated in a grid



- Approximate the density as piecewise constant function
- Evaluate density in a grid over some finite region
- Density times cell area gives probability mass in each cell

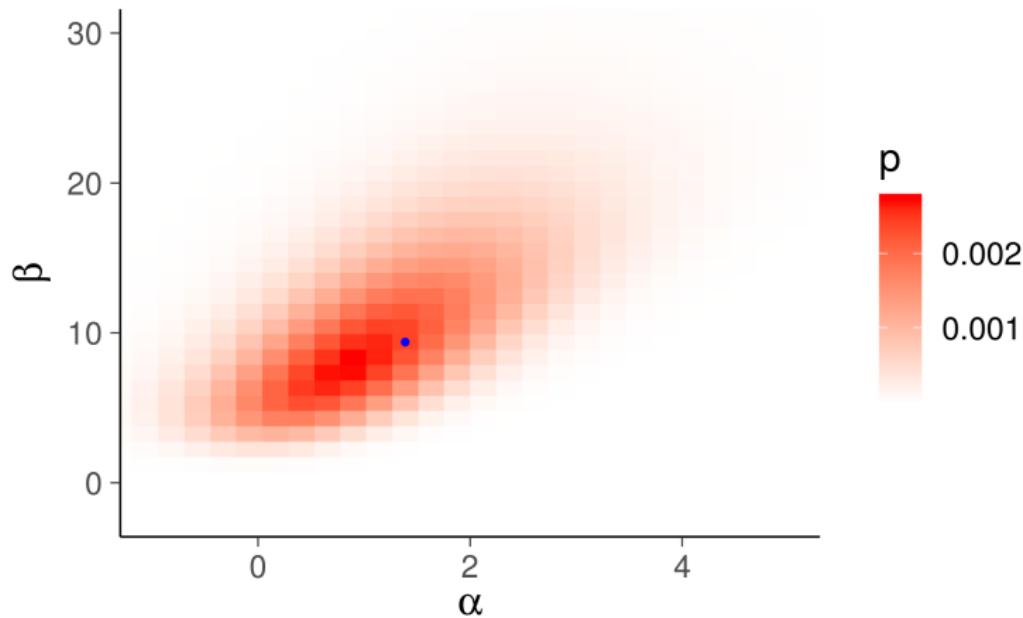
# Bioassay

Posterior density evaluated in a grid



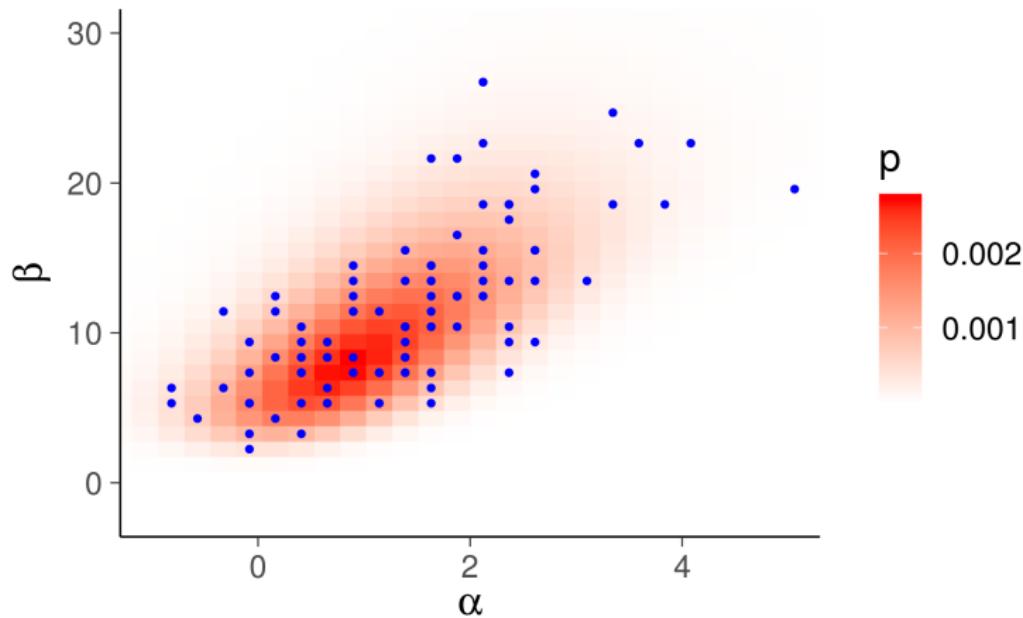
- Densities at 1, 2, and 3: 0.0027 0.0010 0.0001
- Probabilities of cells 1, 2, and 3: 0.0431 0.0166 0.0010
- Probabilities of cells sum to 1

## Posterior density and draws in a grid



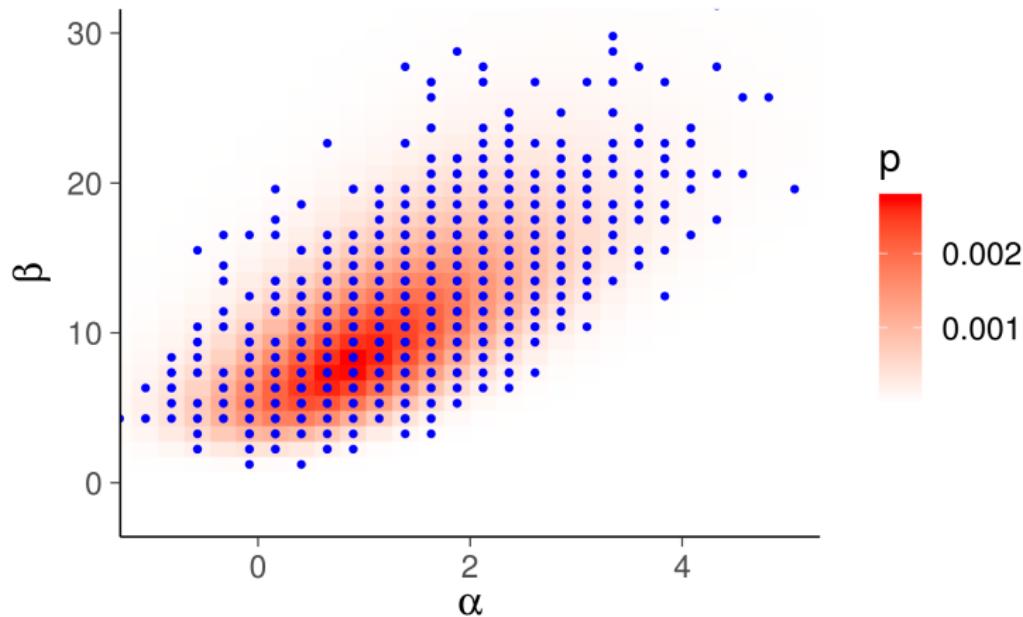
- Sample according to grid cell probabilities

## Posterior density and draws in a grid



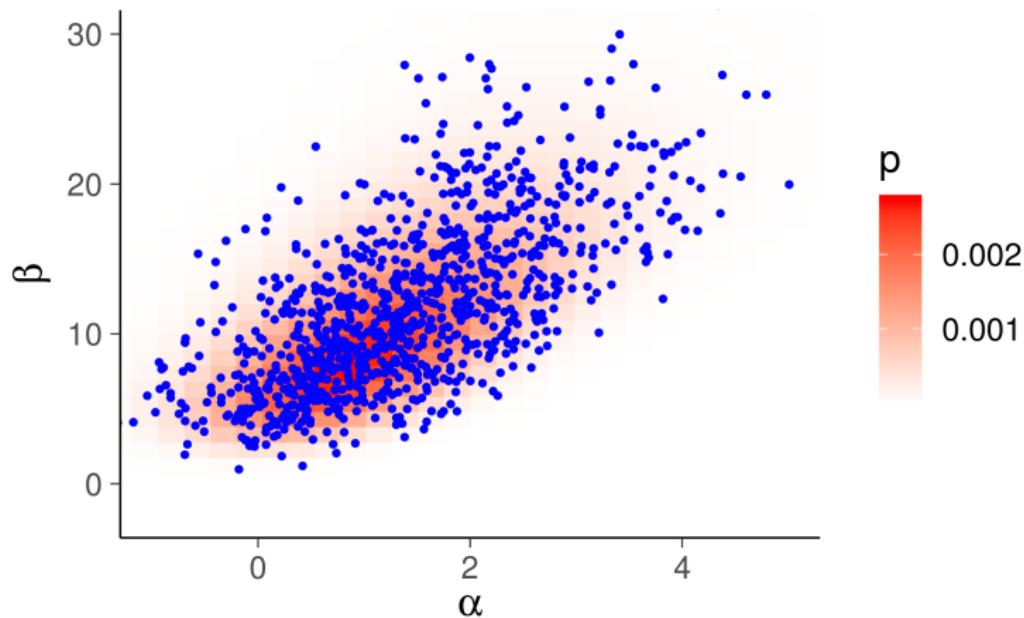
- Sample according to grid cell probabilities

## Posterior density and draws in a grid



- Sample according to grid cell probabilities
- Several draws can be from the same grid cell

## Posterior density in a grid and jittered draws



- Jitter can be added to improve visualization

## Grid sampling

- Draws can be used to estimate expectations, for example

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

## 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

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

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

- 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

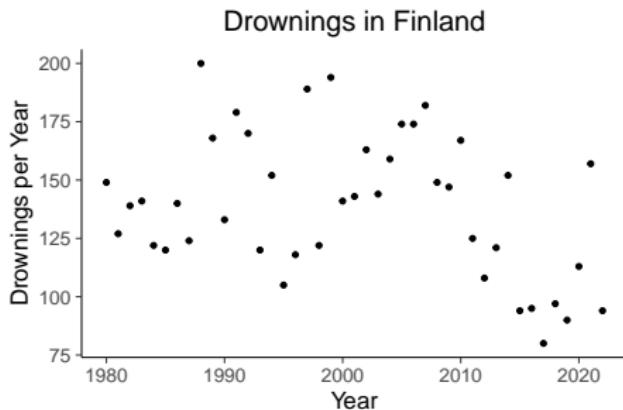
$$E[-\alpha/\beta] \approx \sum_{t=1}^T -\frac{\alpha^{(t)}}{\beta^{(t)}} w_{\text{cell}}^{(t)},$$

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

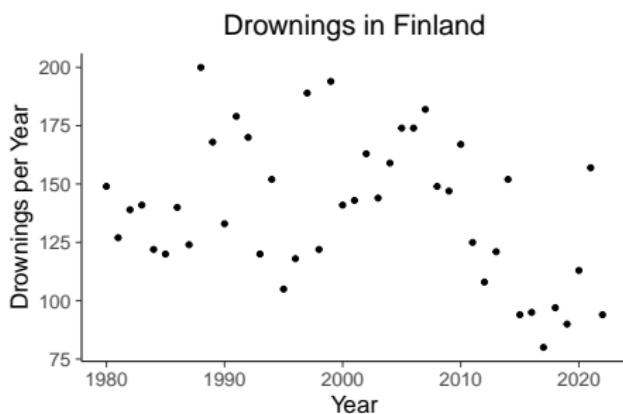
# Example GLM

Count of deaths, $y_i$	Year
149	1980
127	1981
139	1982
:	:
157	2021
94	2022



# Example GLM

Count of deaths, $y_i$	Year
149	1980
127	1981
139	1982
:	:
157	2021
94	2022

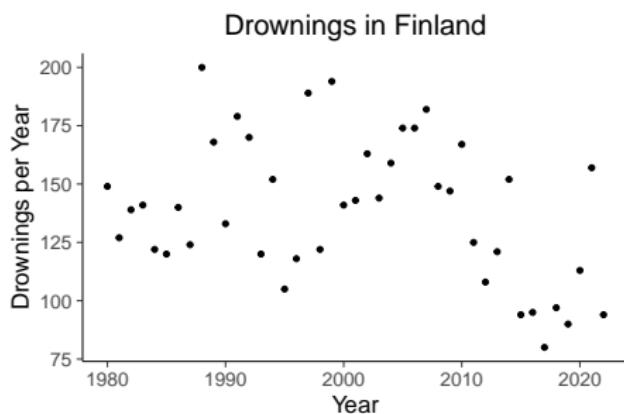


Swimming is popular in Finland, but also hazardous

- On average  $\sim 140$  drownings per year
- Finnish government has invested in measures for reducing deaths
- Recent narrative based on effectiveness of education

# Example GLM

Count of deaths, $y_i$	Year
149	1980
127	1981
139	1982
:	:
157	2021
94	2022



Swimming is popular in Finland, but also hazardous

- On average  $\sim 140$  drownings per year
- Finnish government has invested in measures for reducing deaths
- Recent narrative based on effectiveness of education

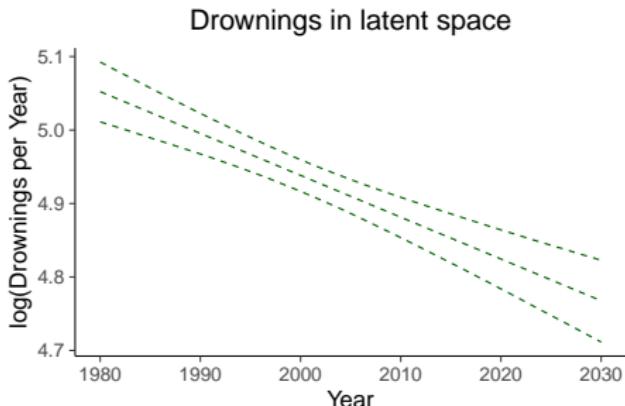
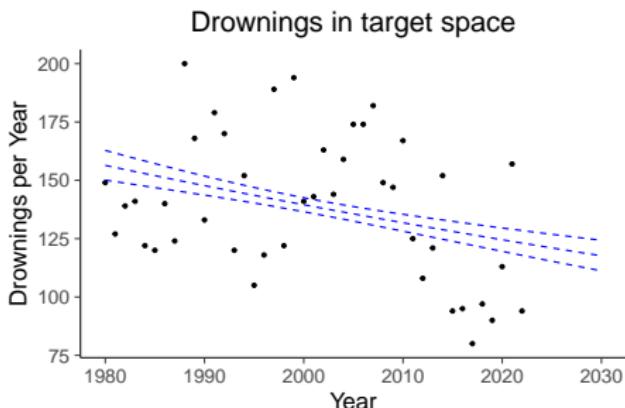
Bayesian methods help

- Describe trends over time
- Evaluate uncertainty

# Example GLM: Poisson Linear Model

$$y_i \mid \mu_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i = e^{\alpha + \beta x_i}$$

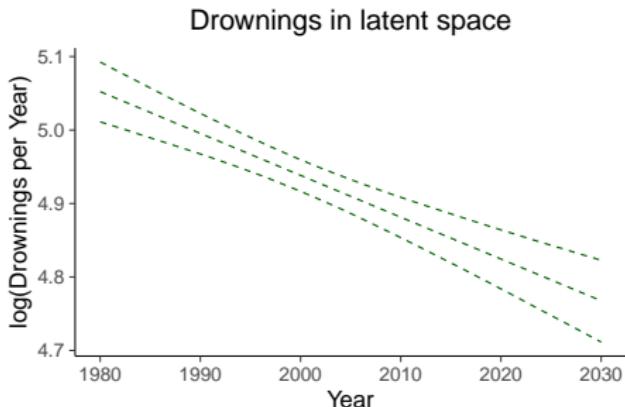
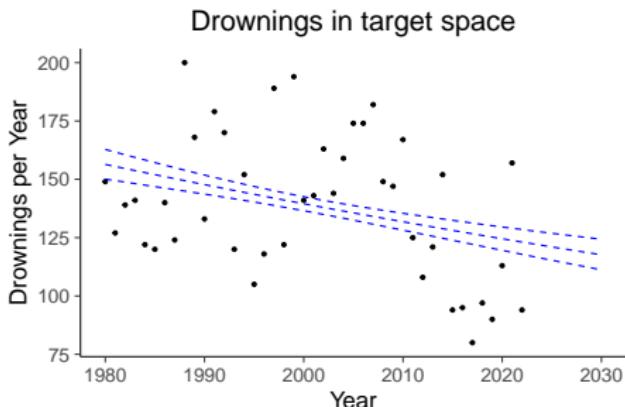


# Example GLM: Poisson Linear Model

$$y_i \mid \mu_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i = e^{\alpha + \beta x_i}$$

$$\text{Poisson}(y_i | \mu_i) = \frac{1}{y_i!} \mu_i^{y_i} e^{-\mu_i}$$



# Example GLM: Poisson Linear Model

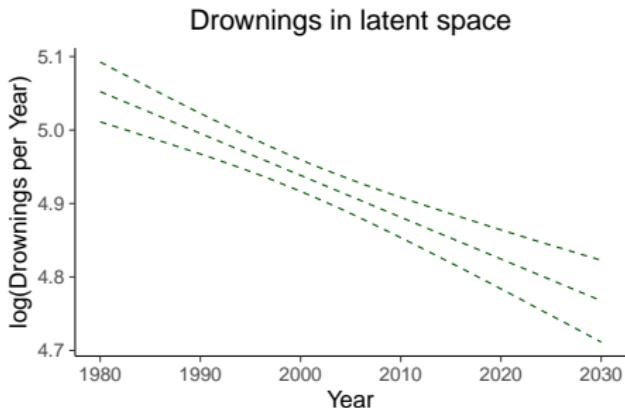
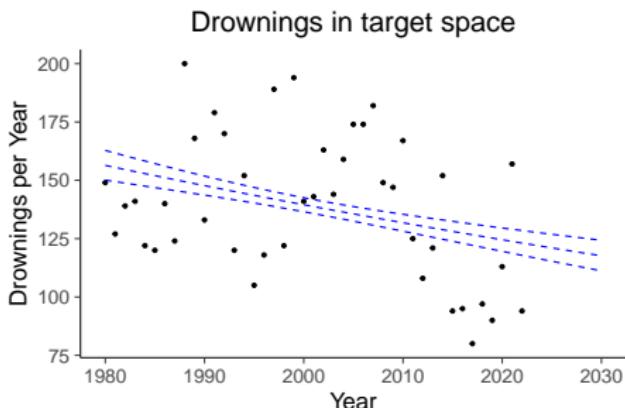
$$y_i \mid \mu_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i = e^{\alpha + \beta x_i}$$

$$\text{Poisson}(y_i | \mu_i) = \frac{1}{y_i!} \mu_i^{y_i} e^{-\mu_i}$$

Alternatively :

$$y_i \mid \mu_i, \phi \sim \text{Neg-bin}(y_i | \mu_i, \phi)$$



# Example GLM: Poisson Linear Model

$$y_i \mid \mu_i \sim \text{Poisson}(\mu_i)$$

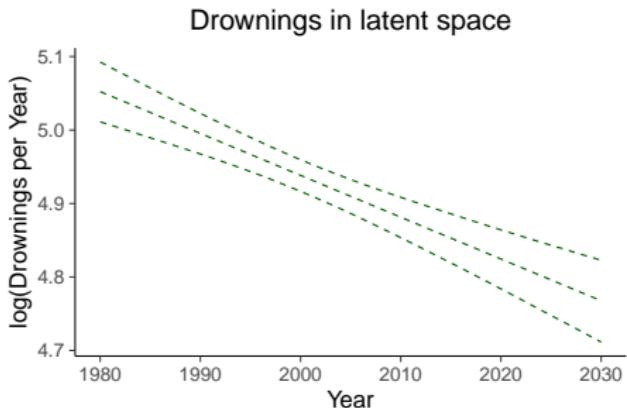
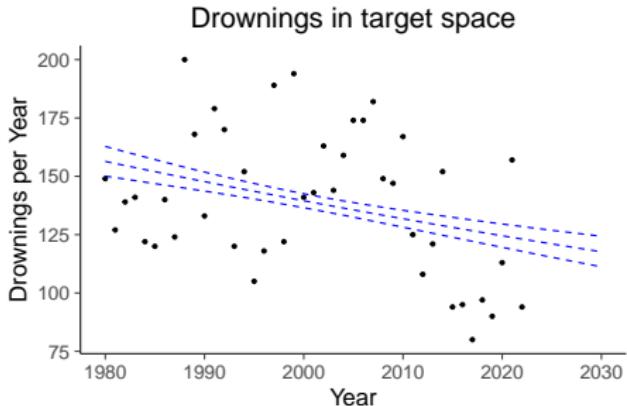
$$\mu_i = e^{\alpha + \beta x_i}$$

$$\text{Poisson}(y_i | \mu_i) = \frac{1}{y_i!} \mu_i^{y_i} e^{-\mu_i}$$

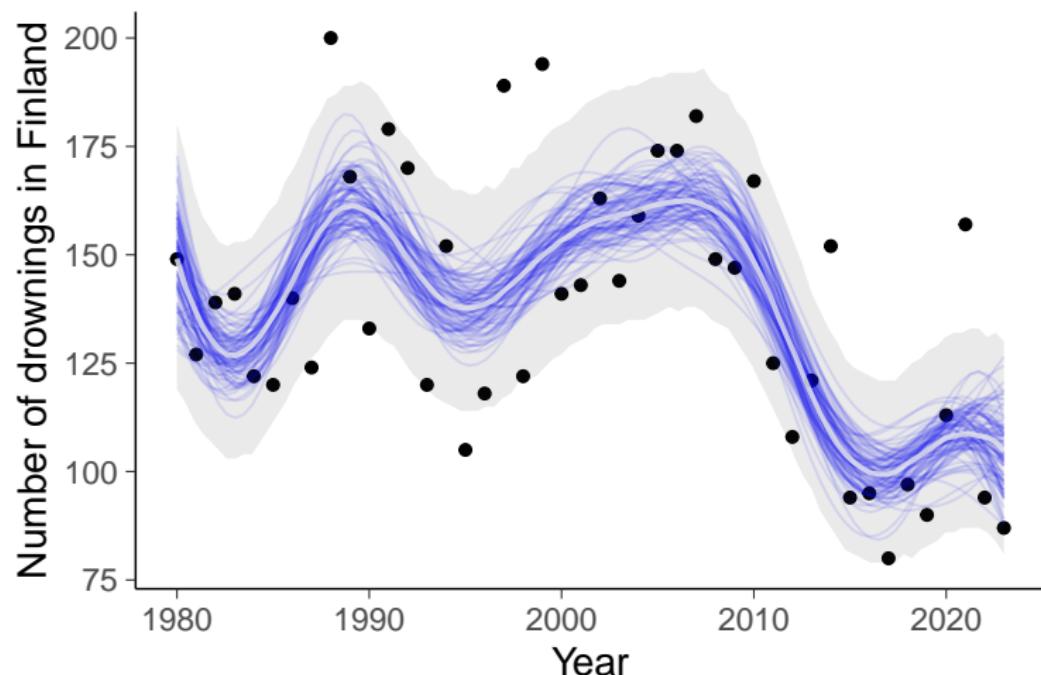
Alternatively :

$$y_i \mid \mu_i, \phi \sim \text{Neg-bin}(y_i | \mu_i, \phi)$$

$$\text{Neg-bin}(y_i \mid \mu_i, \phi) = \frac{\Gamma(y_i + \phi)}{y_i! \Gamma(\phi)} \left( \frac{\mu_i}{\mu_i + \phi} \right)^{y_i} \left( \frac{\phi}{y_i + \phi} \right)^\phi$$



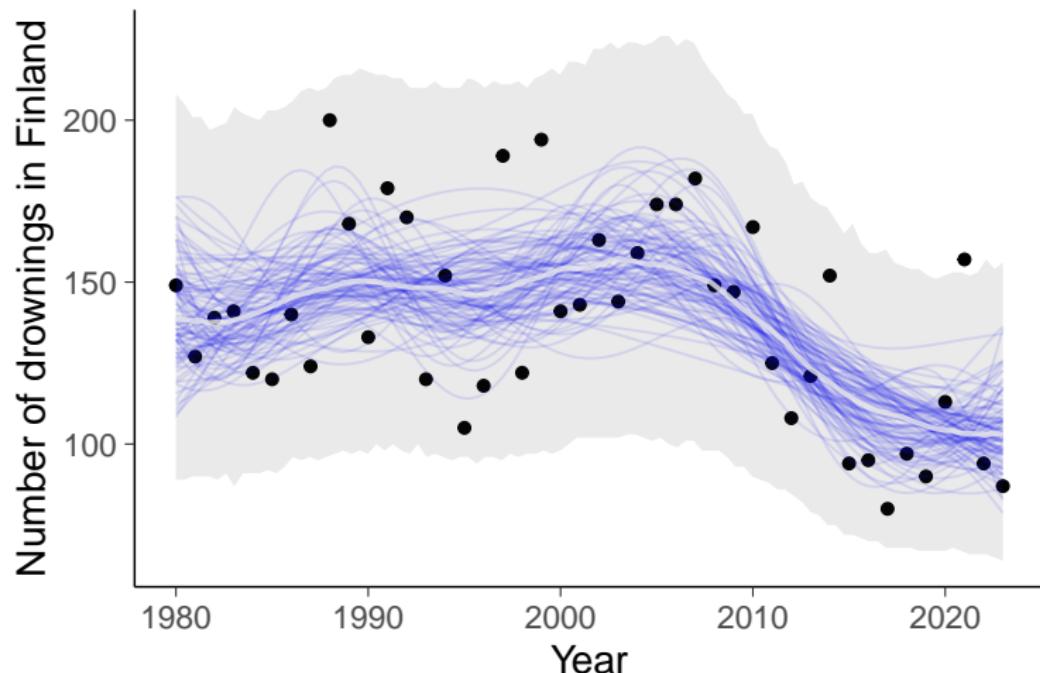
# Example GLM: Gaussian Process Models



$$y_i \mid \mu_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i \sim e^{f_i}, f \sim \text{GP}(0, k(\text{Year}, \theta))$$

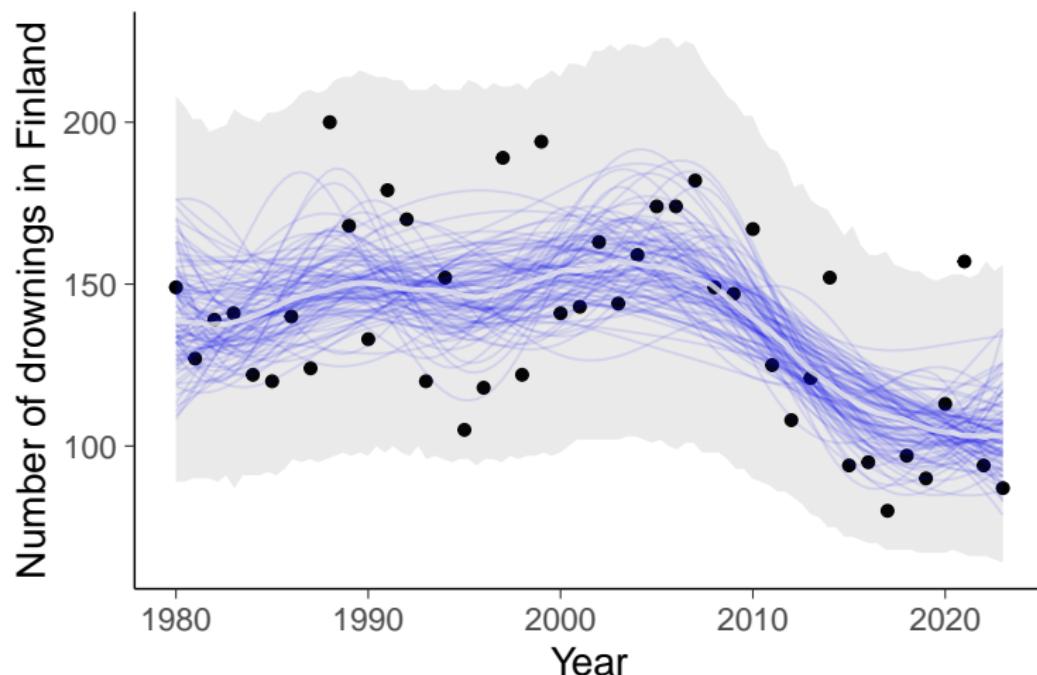
# Example GLM: Gaussian Process Models



$$y_i \mid \mu_i \sim \text{Neg-bin}(\mu_i, \phi)$$

$$\mu_i \sim e^{f_i}, f \sim \text{GP}(0, k(\text{Year}, \theta))$$

# Example GLM: Gaussian Process Models



- Clear overdispersion
  - later we use posterior predictive checking and cross-validation to confirm this
- Trend interpretations shouldn't be based on one observation

## Thinking counts

- For simplicity of exposition, we often start learning with normal observation models
- But we observe count data on a daily basis
- Very relevant in industry (number of sold products, ad views, customer count, etc.)
- Can you think of such examples from the class room?
  - Think of how many students attend BDA lectures over the course
  - Number of students who report getting sick over time until Christmas
  - Number of dropouts
  - Would you expect overdispersion?