# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.

# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.
- Remember: compute acceptance probabilities on a log scale!!

# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.
- Remember: compute acceptance probabilities on a log scale!!
- How to check that an MCMC algorithm is working ok?/ If you suspect a bug, how to find the problem?

# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.
- Remember: compute acceptance probabilities on a log scale!!
- How to check that an MCMC algorithm is working ok?/ If you suspect a bug, how to find the problem?
- Some advises when coding a Metropolis–Hastings algorithm:

# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.
- Remember: compute acceptance probabilities on a log scale!!
- How to check that an MCMC algorithm is working ok?/ If you suspect a bug, how to find the problem?
- Some advises when coding a Metropolis–Hastings algorithm:
  - test the code on your own simulated data first
    - choose/simulate parameter values, simulate data
    - then you know the true parameter values

# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.
- Remember: compute acceptance probabilities on a log scale!!
- How to check that an MCMC algorithm is working ok?/ If you suspect a bug, how to find the problem?
- Some advises when coding a Metropolis–Hastings algorithm:
  - test the code on your own simulated data first
    - choose/simulate parameter values, simulate data
    - then you know the true parameter values
  - implement and test one update type at a time
    - make one function for each update type
    - keep other parameters equal to their true values
    - if not working properly: print out and look at proposed value, acceptance probabilities, factors in acceptance probabilities

# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.
- Remember: compute acceptance probabilities on a log scale!!
- How to check that an MCMC algorithm is working ok?/ If you suspect a bug, how to find the problem?
- Some advises when coding a Metropolis–Hastings algorithm:
  - test the code on your own simulated data first
    - choose/simulate parameter values, simulate data
    - then you know the true parameter values
  - implement and test one update type at a time
    - make one function for each update type
    - keep other parameters equal to their true values
    - if not working properly: print out and look at proposed value, acceptance probabilities, factors in acceptance probabilities
  - then combine two update types, three update types and so on

# Some comments to Exercise 2

- No natural ordering of Problem A and Problem B.
- Remember: compute acceptance probabilities on a log scale!!
- How to check that an MCMC algorithm is working ok?/ If you suspect a bug, how to find the problem?
- Some advises when coding a Metropolis–Hastings algorithm:
  - test the code on your own simulated data first
    - choose/simulate parameter values, simulate data
    - then you know the true parameter values
  - implement and test one update type at a time
    - make one function for each update type
    - keep other parameters equal to their true values
    - if not working properly: print out and look at proposed value, acceptance probabilities, factors in acceptance probabilities
  - then combine two update types, three update types and so on
  - if you have a bug in your code:
    - don't ask the TA: I have a bug in my code, where is it?
    - first try to locate approximately where the problem is, then ask the TA for help (if you don't sort it out yourself)

# Bayesian modelling

- A Bayesian model is specified by:
  - observed data: $x$
  - likelihood: $f(x|\theta)$
  - prior: $f(\theta)$
- Distribution of interest:
  - posterior distribution: $f(\theta|x) \propto f(\theta)f(x|\theta)$

# Bayesian modelling

- A Bayesian model is specified by:
  - observed data: $x$
  - likelihood: $f(x|\theta)$
  - prior: $f(\theta)$
- Distribution of interest:
  - posterior distribution: $f(\theta|x) \propto f(\theta)f(x|\theta)$
- The prior $f(\theta)$:
  - ideally: reflect our prior knowledge about $\theta$ before observing $x$
  - improper prior
  - conjugate prior

# Bayesian modelling

- A Bayesian model is specified by:
  - observed data: $x$
  - likelihood: $f(x|\theta)$
  - prior: $f(\theta)$
- Distribution of interest:
  - posterior distribution: $f(\theta|x) \propto f(\theta)f(x|\theta)$
- The prior $f(\theta)$:
  - ideally: reflect our prior knowledge about $\theta$ before observing $x$
  - improper prior
  - conjugate prior
- How to characterise the posterior $f(\theta|x)$?
  - by mean values, $\mathrm{E}[g(\theta)]$
  - by intervals for $g(\theta)$ — this will not be confidence intervals!

# $k$-parameter exponential family

- A distribution belongs to the $k$-parameter exponential family if

$$f(x|\theta) = a(x)e^{\sum_{i=1}^{k} \phi_i(\theta)t_i(x)+b(\theta)}$$

for some functions $a(x)$, $\phi_1(\theta), \ldots, \phi_k(\theta)$, $t_1(x), \ldots, t_k(x)$ and $b(\theta)$

# $k$-parameter exponential family

- A distribution belongs to the $k$-parameter exponential family if

$$f(x|\theta) = a(x)e^{\sum_{i=1}^{k} \phi_i(\theta)t_i(x)+b(\theta)}$$

  for some functions $a(x)$, $\phi_1(\theta), \ldots, \phi_k(\theta)$, $t_1(x), \ldots, t_k(x)$ and $b(\theta)$

- Example: $x \sim \mathrm{N}(\mu, \sigma^2)$ has this form for $k = 2$ and (for example)
  - $a(x) = \frac{1}{\sqrt{2\pi}}$
  - $\phi_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2}$ and $t_1(x) = x^2$
  - $\phi_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$ and $t_2(x) = x$
  - $b(\mu, \sigma^2) = -\frac{\mu}{2\sigma^2} - \frac{1}{2}\ln(\sigma^2)$

# $k$-parameter exponential family

- A distribution belongs to the $k$-parameter exponential family if

$$f(x|\theta) = a(x)e^{\sum_{i=1}^{k} \phi_i(\theta)t_i(x)+b(\theta)}$$

for some functions $a(x)$, $\phi_1(\theta), \ldots, \phi_k(\theta)$, $t_1(x), \ldots, t_k(x)$ and $b(\theta)$

- Example: $x \sim \mathrm{N}(\mu, \sigma^2)$ has this form for $k = 2$ and (for example)
  - $a(x) = \frac{1}{\sqrt{2\pi}}$
  - $\phi_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2}$ and $t_1(x) = x^2$
  - $\phi_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$ and $t_2(x) = x$
  - $b(\mu, \sigma^2) = -\frac{\mu}{2\sigma^2} - \frac{1}{2}\ln(\sigma^2)$

- Conjugate prior to a $k$-parameter exponential family is

$$f(\theta) = k(\alpha, \beta)e^{\sum_{i=1}^{k} \alpha_i\phi_i(\theta)+\beta b(\theta)}$$

# Conditional conjugacy — example

- Let $x_1, \ldots, x_n | \mu, \sigma^2 \sim \mathrm{N}(\mu, \sigma^2)$

# Conditional conjugacy — example

- Let $x_1, \ldots, x_n | \mu, \sigma^2 \sim \mathrm{N}(\mu, \sigma^2)$
- Prior: $f(\mu, \sigma^2) = f(\mu) f(\sigma^2)$, where

$$\begin{aligned} \mu &\sim \mathrm{N}(\mu_0, \tau^2) && \text{conjugate prior when } \sigma^2 \text{ is known} \\ \sigma^2 &\sim \mathrm{IG}(\alpha, \beta) && \text{conjugate prior when } \mu \text{ is known} \end{aligned}$$

# Conditional conjugacy — example

- Let $x_1, \ldots, x_n | \mu, \sigma^2 \sim \mathrm{N}(\mu, \sigma^2)$
- Prior: $f(\mu, \sigma^2) = f(\mu) f(\sigma^2)$, where

$$\begin{array}{lcll} \mu & \sim & \mathrm{N}(\mu_0, \tau^2) & \text{conjugate prior when } \sigma^2 \text{ is known} \\ \sigma^2 & \sim & \mathrm{IG}(\alpha, \beta) & \text{conjugate prior when } \mu \text{ is known} \end{array}$$

- Posterior distribution:

$$f(\mu, \sigma^2 | x_1, \ldots, x_n) \propto e^{-\frac{1}{2\tau^2}(\mu - \mu_0)^2} \cdot \frac{e^{-\frac{1}{\beta \sigma^2}}}{(\sigma^2)^{\alpha+1}} \cdot \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

# Conditional conjugacy — example

- Let $x_1, \ldots, x_n | \mu, \sigma^2 \sim \mathrm{N}(\mu, \sigma^2)$
- Prior: $f(\mu, \sigma^2) = f(\mu) f(\sigma^2)$, where

$$\begin{aligned} \mu &\sim \mathrm{N}(\mu_0, \tau^2) && \text{conjugate prior when } \sigma^2 \text{ is known} \\ \sigma^2 &\sim \mathrm{IG}(\alpha, \beta) && \text{conjugate prior when } \mu \text{ is known} \end{aligned}$$

- Posterior distribution:

$$f(\mu, \sigma^2 | x_1, \ldots, x_n) \propto e^{-\frac{1}{2\tau^2}(\mu - \mu_0)^2} \cdot \frac{e^{-\frac{1}{\beta\sigma^2}}}{(\sigma^2)^{\alpha+1}} \cdot \frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

- Full conditionals:

$$\begin{aligned} \mu | x_1, \ldots, x_n, \sigma^2 &\sim \mathrm{N}\left( \frac{\frac{\mu_0}{\tau^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \right) \\ \sigma^2 | x_1, \ldots, x_n, \mu &\sim \mathrm{IG}\left( \alpha + \frac{n}{2}, \frac{1}{\frac{1}{\beta} + \frac{1}{2}\sum_{i=1}^n (x_i - \mu)^2} \right) \end{aligned}$$