# Bayesian Data Analysis
## 2020

**Week 4: Hierarchical models**

Jarno.Vanhatalo@helsinki.fi

# On previous weeks

- Markov chain Monte Carlo (MCMC)
  - $E[\theta|y] \approx \frac{1}{S}\sum_{i=1}^{S} \theta^{i}$
  - Convergence diagnostics
  - Stan
- Stan computing environment

# Aims of the week

- Theory
  - Hierarchical models
  - Exchangeability
  - Hyperprior
  - Population distribution
- Models
  - Hierarchical Binomial model

# Joint distribution

- Two random variables $\theta_1, \theta_2$
  - e.g. $\alpha$ and $\beta$ in the bioassay example
- The joint distribution

$$p(\theta_1, \theta_2) = \underbrace{p(\theta_1|\theta_2)}_{\text{conditional}}\underbrace{p(\theta_2)}_{\text{marginal}}$$

- Independent variables

$$p(\theta_1, \theta_2) = \underbrace{p(\theta_1)}_{\text{marginal}}\underbrace{p(\theta_2)}_{\text{marginal}}$$

# Joint distribution

- Prior independence does not imply posterior independence!
  - The likelihood typically ties the parameters
  - For example, Gaussian observation model excercise

$$p(\mu, \sigma^2 | y) \ \alpha \ p(y|\mu, \sigma^2) p(\mu) \mathrm{p}(\sigma^2) \neq p(\mu|y) \mathrm{p}(\sigma^2|y)$$

# conditionally independently and identically distributed (i.i.d)

- e.g. i.i.d. Gaussian observations $y = \{y_1, \dots y_n\}$
- Joint Distribution

$$p(y|\mu, \sigma^2) = N(y_1|\mu, \sigma^2) \times \dots \times N(y_n|\mu, \sigma^2)$$

$$= \prod_{i=1}^{n} N(y_i|\mu, \sigma^2)$$

Identical Gaussian distributions

# Conditionally independently distributed

- Observations $y = \{y_1, \dots y_n\}$ from

$$y_i | \mu_i, \sigma^2 \sim N(\mu_i, \sigma^2)$$
$$\mu_i = c + a x_{i,1} + b x_{i,2}$$

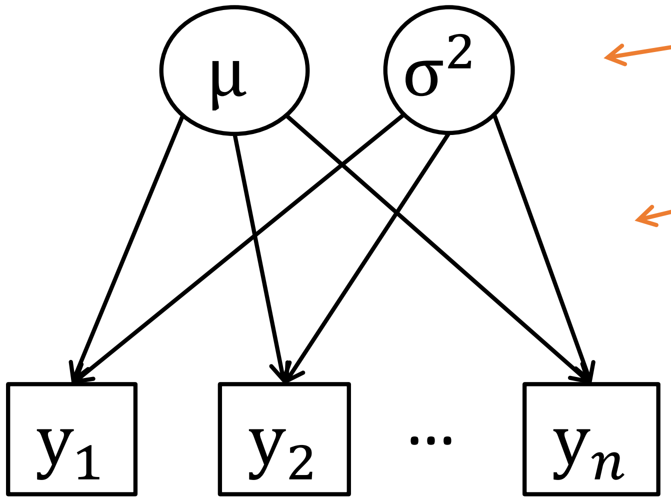
- Joint distribution

$$p(y|\mu, \sigma^2) = N(y_1|\mu_1, \sigma^2) \times \dots \times N(y_n|\mu_n, \sigma^2)$$
$$= \prod_{i=1}^{n} N(y_i|\mu_i, \sigma^2)$$

Non-Identical but independent distributions given $\boldsymbol{x}$

# Graphical description

i.i.d. observations

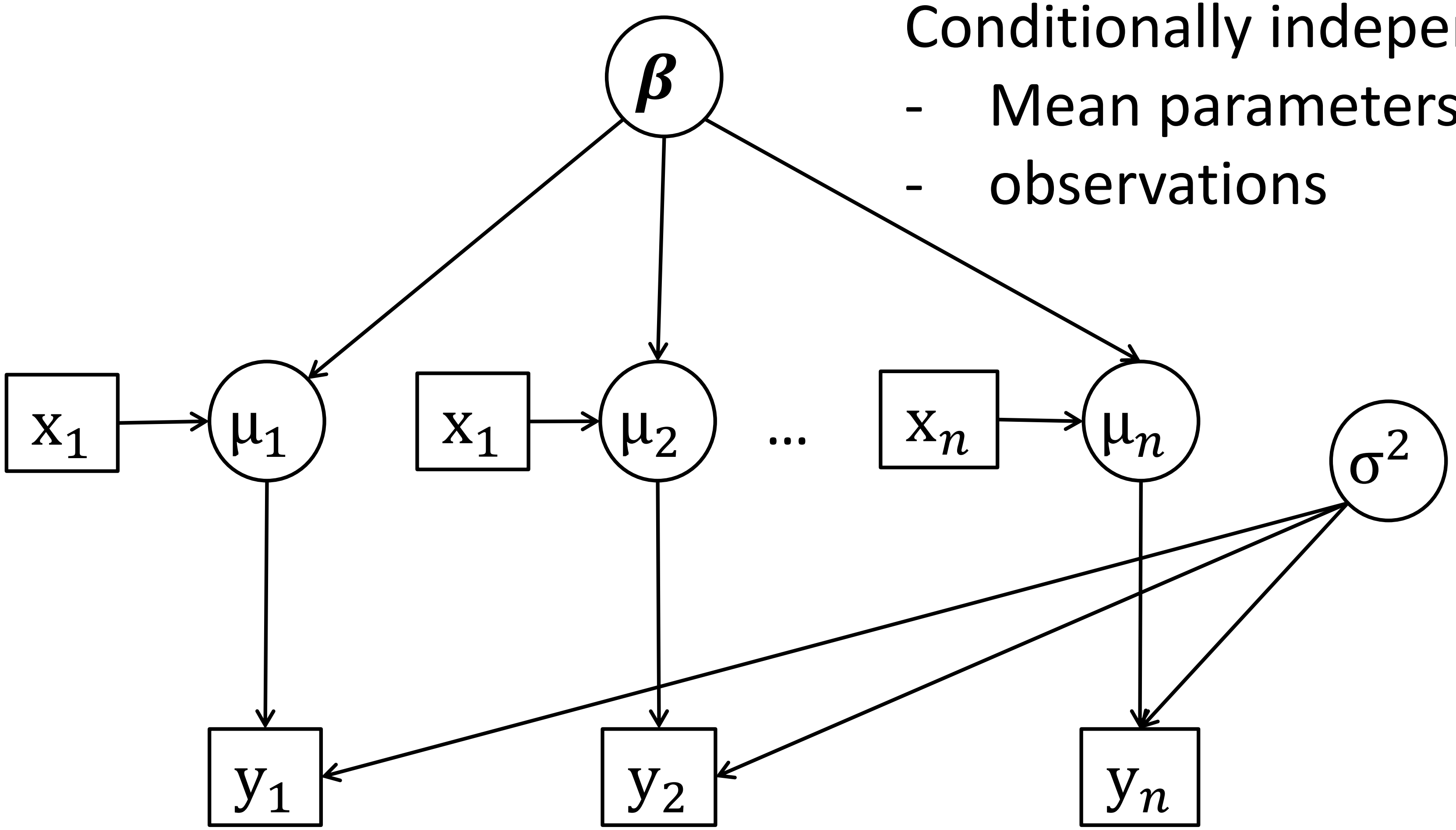



Directed acyclic graph (DAG)

- Unobserved variables (circle): parameters of the model
- Direction of conditional independence
  - Arrow points from parent to child
  - Children are independent <u>given</u> their parents
  - Tells which way the information flows
- Observed variables (boxes): data or fixed parameter values

# Graphical description

Conditionally independent
- Mean parameters
- observations

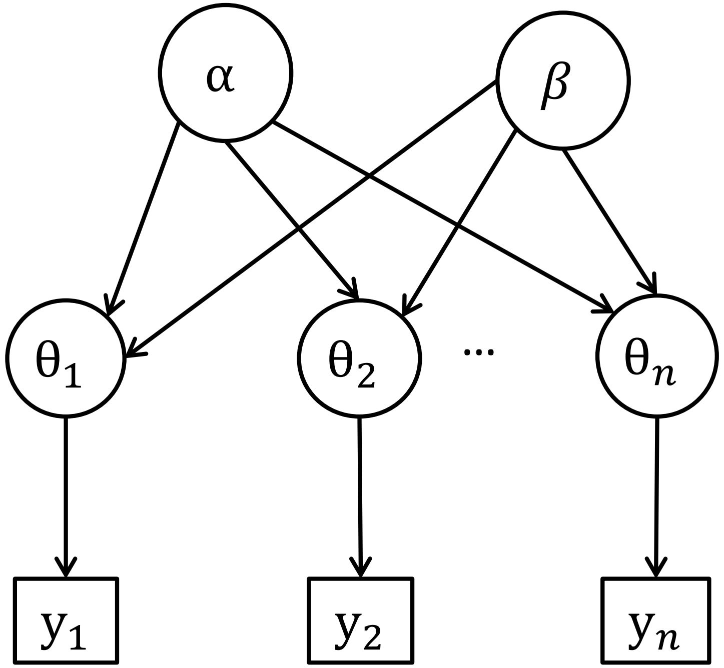


# Hierarchical models

- Simplest form
  - Likelihood: $p(y|\theta)$
  - Prior: $p(\theta|\gamma)$ 1st stage prior
  - Hyperprior: $p(\gamma)$ 2nd stage prior
- Can go further and further by extending hyperparameters

# Hierarchical model

- Example: white fish larval presence (see also Rat tumor experiments in BDA3, p. 102)
  - The same survey set-up in n=19 different areas which may differ in their environmental properties




Hyperparameters / Hyperprior:

$$p(\alpha, \beta) = p(\alpha)p(\beta)$$

Parameters / Prior:

$$p(\theta|\alpha, \beta) = \prod p(\theta_i|\alpha, \beta)$$

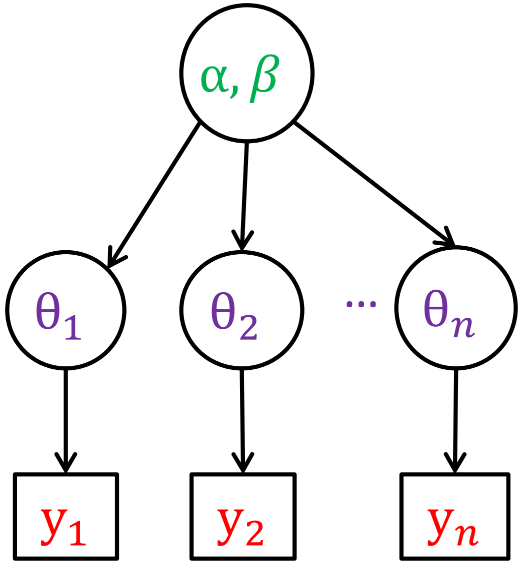
Observations / Observation model:

$$p(y|\theta) = \prod p(y_i|\theta_i)$$

# Hierarchical model

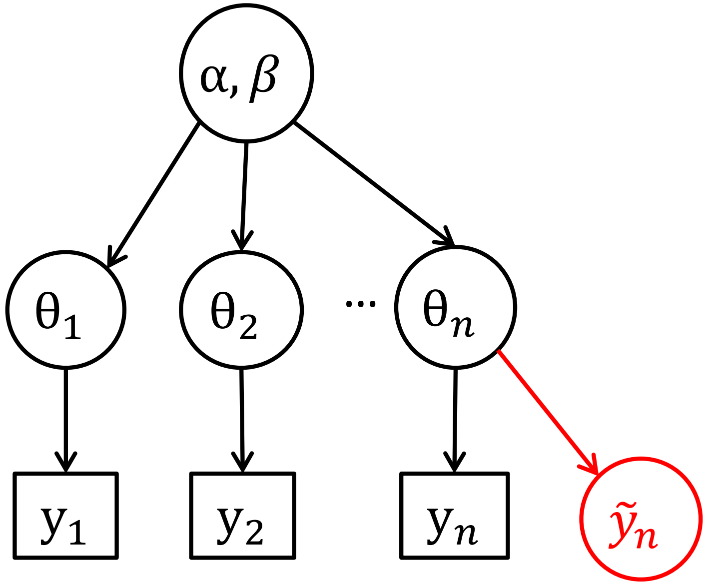
- Posterior through chain rule

$$p(\theta, \alpha, \beta | y) \propto p(y|\theta) p(\theta|\alpha, \beta) p(\alpha, \beta)$$



# Prediction with hierarchical model

- Prediction within a group from where we have data
  - e.g. new observation in the n'th area
- $p(\tilde{y}_n | y_1 \dots y_n) = \int p(\tilde{y}_n | \theta_n) p(\theta_n \,|\, y_1 \dots y_n) d\theta_n$

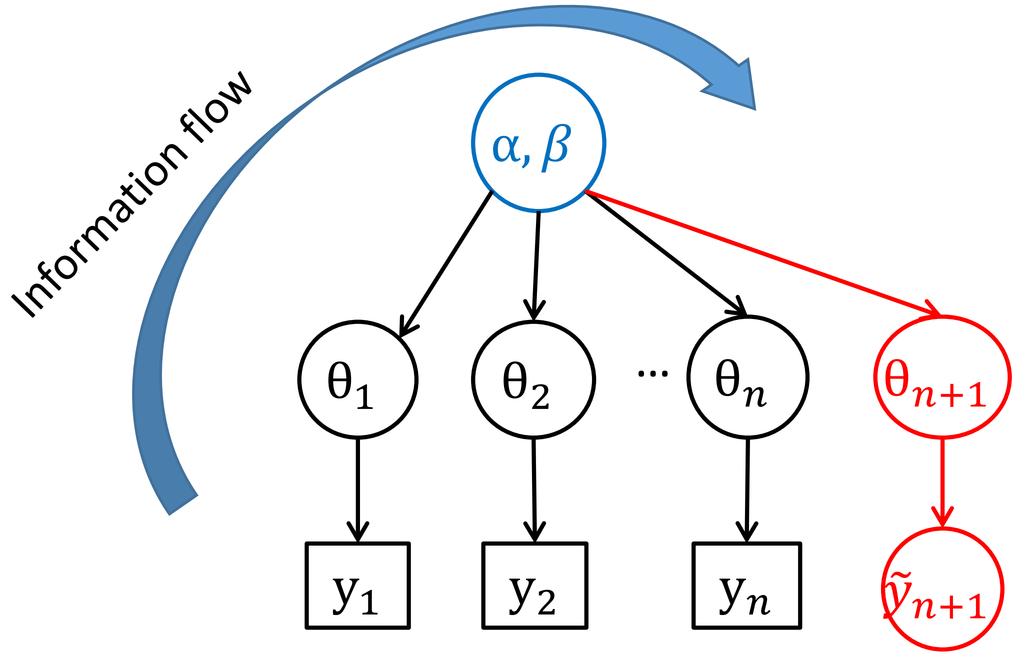



Monte Carlo approximation
- Sample $\theta_n^s \sim p(\theta_n \,|\, y_1 \dots y_n)$
- Sample $\tilde{y}_n^s \sim p(\tilde{y}_n \,|\, \theta_n^s)$
- Repeat

# Prediction with hierarchical model

- Prediction for new area
  - e.g. new survey in a new area

$$p(\tilde{y}_n|y_1 \dots y_n) = \int p(\tilde{y}_{n+1}|\theta_{n+1})p(\theta_{n+1}|\alpha,\beta)p(\alpha,\beta|y_1 \dots y_n)d\theta_{n+1}d\alpha d\beta$$




Monte Carlo approximation

- Sample $\alpha^s, \beta^s \sim p(\alpha, \beta | y_1 \dots y_n)$
- Sample $\theta^s_{n+1} \sim p(\theta_{n+1} | \alpha^s, \beta^s)$
- Sample $\tilde{y}^s_{n+1} \sim p(\tilde{y}_{n+1} | \theta^s_{n+1})$
- Repeat

# Exchangeability

- First about random variables:
  - $Y$ denotes random variable which has distribution

$$Y \sim p(Y)$$

    - Thus, $Y$ is ”something we have not observed”
  - $y$ is an outcome/realization of a random variable
    - Does not have distribution
  - $p(Y = y)$ is
    - the probability that random variable gets value $y$ (discrete)
    - the probability density at y (continuous)

# Exchangeability

- *If no prior information is available to distinguish any of the random variables $Y_i$ from any of the others, one must assume symmetry among the variables in their prior distribution.*
- This symmetry is represented by exchangeability:
  - Inability to order the variables a priori
  - ”Ignorance implies exchangeability”

# Exchangeability

- Mathematically
  - The $n$ variables $Y_i$ are exchangeable if their joint distribution
    $$\mathrm{p}(Y_1, \ldots, Y_n)$$
    is invariant to permutations of the indices (1,…,n).

# Exchangeable

- Consider the whole Finnish population
  - Pick up 100 individuals randomly and measure their lenghts $Y_i$
    - No prior information to distinguish between individuals
    - → the lenghts are exchangeable
  - Pick randomly 50 females and 50 males
    - No prior information to distinguish which measurements correspond to male and female
    - → the lenghts are exchangeable
  - Pick first 50 females and then 50 males
    - There is prior information to distinguish between the first and second half of sample
    - → $Y_1 \ldots Y_{100}$ are not exchangeable
    - → however, $Y_1 \ldots Y_{50}$ and $Y_{51} \ldots Y_{100}$ are exchangeable

# Example: exchangeability

1. Consider a bag of 1 white and 1 black ball. Let the random variables be

    - $Y_1$ the outcome of drawing one ball from the bag
    - $Y_2$ the outcome of drawing second ball from the bag after replacing the first ball

- Are $Y_1$ and $Y_2$ exchangeable?
- Are $Y_1$ and $Y_2$ independent?

# Example: exchangeability

2. Consider a bag of 1 white and 1 black ball. Let the random variables be

  - $Y_1$ the outcome of drawing one ball from the bag
  - $Y_2$ the outcome of drawing second ball from the bag <u>without replacing the first ball</u>

- Are $Y_1$ and $Y_2$ exchangeable?
- Are $Y_1$ and $Y_2$ independent?

# Example: exchangeability

3. Consider a bag of 10000 white and 10000 black balls. Let the random variables be

    - $Y_1$ the outcome of drawing one ball from the bag
    - $Y_2$ the outcome of drawing second ball from the bag <u>without replacing the first ball</u>

- Are $Y_1$ and $Y_2$ exchangeable?
- Are $Y_1$ and $Y_2$ independent?
- Can we act as if $Y_1$ and $Y_2$ were independent?

# Example: exchangeability

4. Consider a bag of white and black balls (n is known but the proportion is unknown). Let the random variables be

    - $Y_1$ the outcome of drawing one ball from the bag
    - $Y_2$ the outcome of drawing second ball from the bag after replacing the first ball

- Are $Y_1$ and $Y_2$ exchangeable?
- Are $Y_1$ and $Y_2$ independent?
- Can we act as if $Y_1$ and $Y_2$ were independent?

# Example: exchangeability

5. Consider a bag of white and black balls (n is known but the proportion is unknown). Let the random variables be

  - $Y_1$ the outcome of drawing one ball from the bag
  - $Y_2$ the outcome of drawing second ball from the bag without replacing the first ball

- Are $Y_1$ and $Y_2$ exchangeable?
- Are $Y_1$ and $Y_2$ independent?
- Can we act as if $Y_1$ and $Y_2$ were independent?

# Example: exchangeability

- 6. Consider a bag of white and black balls (n is unknown and the proportion is unknown). Let the random variables be
  - $Y_1$ the outcome of drawing one ball from the bag
  - $Y_2$ the outcome of drawing second ball from the bag without replacing the first ball
- Are $Y_1$ and $Y_2$ exchangeable?
- Are $Y_1$ and $Y_2$ independent?
- Can we act as if $Y_1$ and $Y_2$ were independent?

# Bag of balls – reflection

- Compare the bag of balls examples to the earlier exercises in the course
  - Mark recapture
  - Female birth rate

# Earlier exercises – reflection

- Discuss the exchangeability within the earlier experiments:
  - Populations and population parameters exercise
  - Effect of bottom coverage on white fish larvae
  - Newcombs speed of light

# Why exchangeability is important?

- Assume a superpopulation of variables $Y_i$ that are exchangeable
  - e.g. a characteristic in Finnish population, repeated measurements, all the balls in the bag
- Take a sample from the superpopulation
  - We can act as if we had a random sample from the superpopulation
  - We can extrapolate for the whole population
- Exchangeability can often be represented through conditional independence
  - We can act as if we had a random, conditionally independent sample from the superpopulation

# Exchangeability in practice

- Stating that variables are exchangeable does not say anything about the exact form of their distribution.

- How do we model the exchangeability in practice?

# Exchangeability in practice

- The simplest form of an exchangeable distribution is conditionally i.i.d. from a population distribution

$$p(Y_1, \dots, Y_n|\theta) = \prod_{i=1}^{n} p(Y_i\,|\theta)$$

- The population parameter θ is usually unknown, and thus the marginal (exchangeable) distribution is

$$p(Y_1, \dots, Y_n) = \int \prod_{i=1}^{n} p(Y_i\,|\theta)\, d\theta$$

  - Mixture of independent and identical distributions

Example
- Bag of balls
- Speed of light

# Exchangeability and hierarchical models

- Exchangeable observations within a group $Y_j = (Y_{j,1}, \dots, Y_{j,n})$

$$p\left(Y_j \middle| \theta_j\right) = \prod_{i=1}^{n} p(Y_{i,j} \,|\theta_j)$$

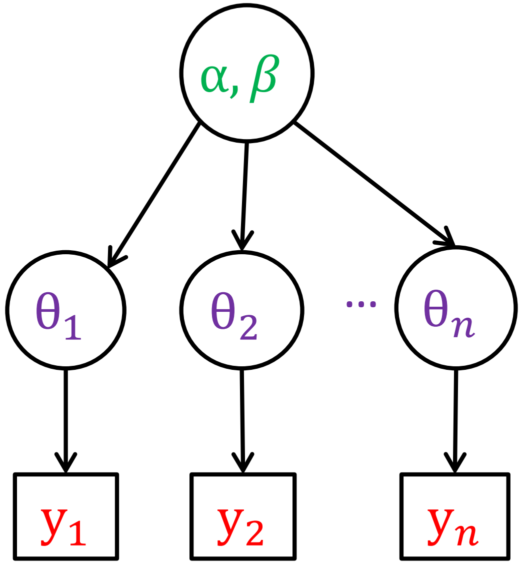
- Exchangeable group specific parameters $\theta = (\theta_1, \dots, \theta_J)$

$$p(\theta|\varphi) = \prod_{j=1}^{J} p(\theta_j|\varphi)$$

- Exchangeable hyperparameters
- … etc.

# Exchangeability and hierarchical models

- Rat tumor example:
  - We know there are 71 laboratories
    - The results, $y_i$, within each laboratory are exchangeable
    - The laboratories are exchangeable




Hyperprior: $p(\alpha, \beta)$

Prior: $p(\theta|\alpha,\beta) = \prod p(\theta_i|\alpha,\beta)$

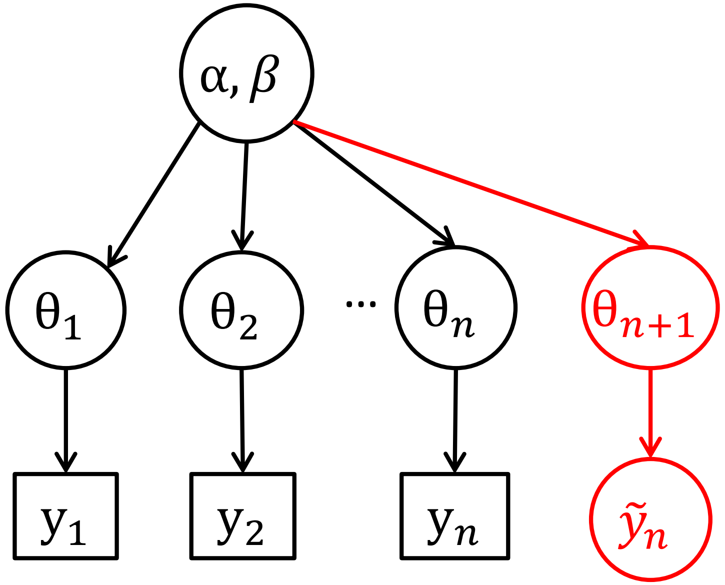
Observation model: $p(y|\theta) = \prod p(y_i|\theta_i)$

# Exchangeability and de Finetti

de Finetti's theorem:

- All exchangeable distributions of infinite number of variables can be written as a mixture of conditionally i.i.d. distributions

$$p(Y_1, Y_2, \dots) = \int \prod_{i=1}^{\infty} p(Y_i \mid \theta)\, d\theta$$



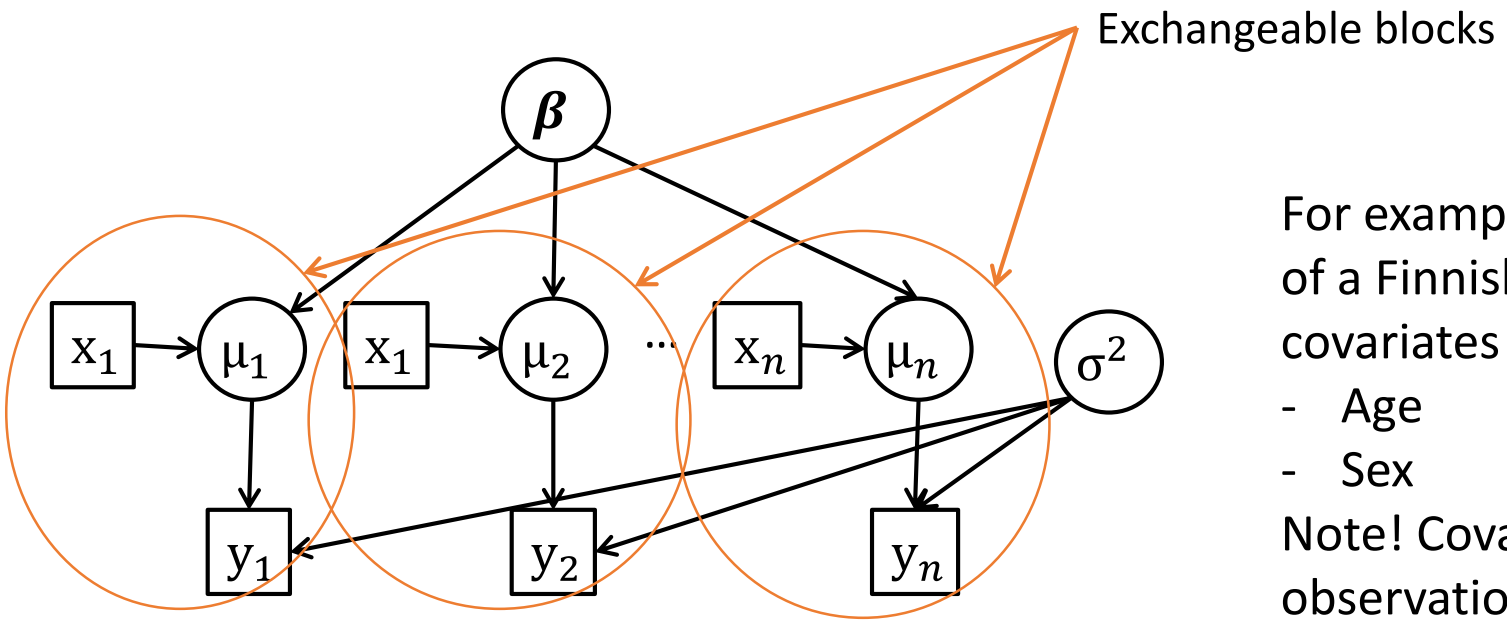

→ often good enough approximation for finite (and large) number of variables

→ theoretical justification for hierarchical models

→ theoretical justification for extrapolation with posterior predictive distribution

# Exchangeability and covariates

- If covariates $x_i$ can be attached to observations $y_i$ we can build exchangeable model for pairs $(x_i, y_i)$




For example, the length of a Finnish person with covariates of:
- Age
- Sex

Note! Covariates are observations as well.

# This week

- ”Ignorance implies exchangeability”
  - Exchangeability does not induce independence
- Exchangeable distribution can be constructed from conditionally independent distributions
  - Hierarchical models
- Exchangeable parameters
  - → theoretical justification for conditionally i.i.d. model blocks
  - → Theoretical justification for extrapolation

# Next week

- Generalized linear models