Hierarchical (multilevel) models

August 16, 2017

UNIVERSITÉ DE SHERBROOKE

What is a hierarchical model?

Global Model

$$P(y = 1) = \beta x + \varepsilon$$

Seperate model

$$P(y = 1) = \beta_1 x + \varepsilon_1$$

$$P(y = 1) = \beta_2 x + \varepsilon_2$$

$$P(y = 1) = \beta_3 x + \varepsilon_3$$

Hierarchical model

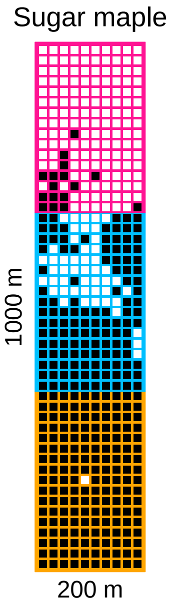
$$P(y = 1) = \beta_i x + \varepsilon$$

where

$$\beta_i \sim \mathcal{D}(\mu_\beta, \sigma_\beta)$$

y is the distribution of sugar maple

x is elevation



Why are hierarchical model worth studying?

To learn about the effect of a treatment that vary

Because it makes it possible to perform inferences using all the data for groups with small sample size

To make prediction of a new unsampled group

Allows to inherently analyse structured data

It is more efficient in making inferences for regression parameters than classical regressions

Makes it possible to include predictors at multiple levels

It accurately acounts for uncertainty in prediction and estimation

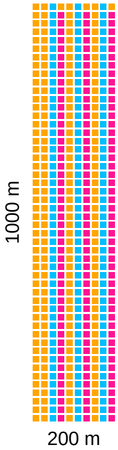
It is a bridge for multivariate modelling

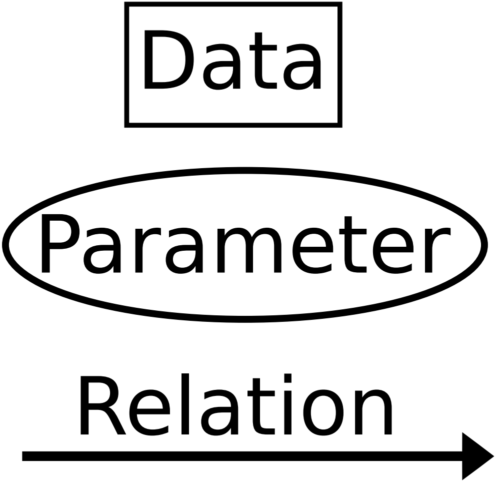
Problem

Jonathan *Brass* Brassard (our team research professional) had two helpers (Steve *Overflow* Vissault and Amaël *Lemalin* LeSquin) working with him and he wants to know if their effort is affected by the climb of Mont Sutton.

Data

```
sutton <- read.csv("sutton.csv",sep=";")
tree <- rowSums(sutton[,3:9])
field <- sutton[,10]
```

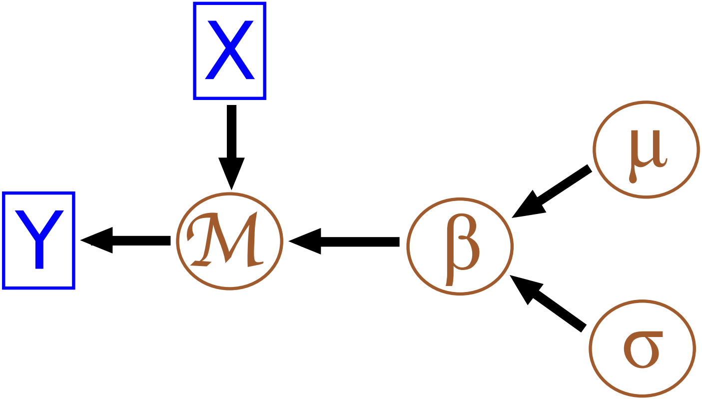




Data
Parameter
Relation

Building a hierarchical model

Direct Acyclic Graph



Model definition

$$\text{Poisson}(\mathbf{y}_i) = \beta \mathbf{X}_i$$

$$\log(E(\mathbf{y}_i)) = \beta \mathbf{X}_i$$

$$E(\mathbf{y}_i) = e^{\beta \mathbf{X}_i}$$

$$\beta \sim \mathcal{N}(\mu, \sigma^2)$$

Prior definition

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$$\frac{1}{\sigma^2} \sim \mathcal{G}(\tau_0, \varphi_0)$$

Data

$\mathbf{y}_i$ The number of trees at site i

$\mathbf{X}_i$ Elevation at site i

Parameters

β The importance of elevation for species

μ Average response of the species to elevation

σ^2 How a species varies in its response to elevation

Priors

μ_0 Mean prior about how μ is distributed

σ_0^2 Variance prior about how μ is distributed

τ_0 Scale prior about how σ is distributed

φ_0 rate prior about how σ is distributed

Write a Gibbs sampler to estimate all parameters of the hierarchical model described previously

Recall that

$$\underbrace{P(\text{Model} \mid \text{Data})}_{\text{Posterior}} \propto \underbrace{P(\text{Data} \mid \text{Model})}_{\text{Likelihood}} \underbrace{P(\text{Model})}_{\text{Prior}}$$

$$P(\boldsymbol{\theta} \mid \mathbf{Y}) \propto P(\mathbf{Y} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})$$

Define the different parts... Mathematically

Likelihood

$$P(\mathbf{y}_i|\beta, \mathbf{X}) = \prod_{i=1}^{n} \frac{e^{\mathbf{y}_i \beta \mathbf{X}_i} e^{-e^{\beta \mathbf{X}_i}}}{y_i!}$$

$$P(\beta|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta-\mu)^2}{2\sigma^2}}$$

Define the different parts... Mathematically

Prior

$$P(\mu, \sigma^2) = (\mu | \mu_0, \sigma_0)(\sigma^2 | \tau_0, \varphi_0)$$

$$= \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \times \frac{{\varphi_0}^{\tau_0}}{\Gamma(\tau_0)} \left(\frac{1}{\sigma^2}\right)^{\tau_0 - 1} e^{-\varphi_0 \frac{1}{\sigma^2}}$$

How to sample each parameter independently

β

$$P(\beta|\mathbf{y}, \mathbf{X}, \mu, \sigma^2) \propto \prod_{i=1}^{n} \frac{e^{\mathbf{y}_i \beta \mathbf{X}_i} e^{-e^{\beta \mathbf{X}_i}}}{y_i!} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta-\mu)^2}{2\sigma^2}}$$

μ

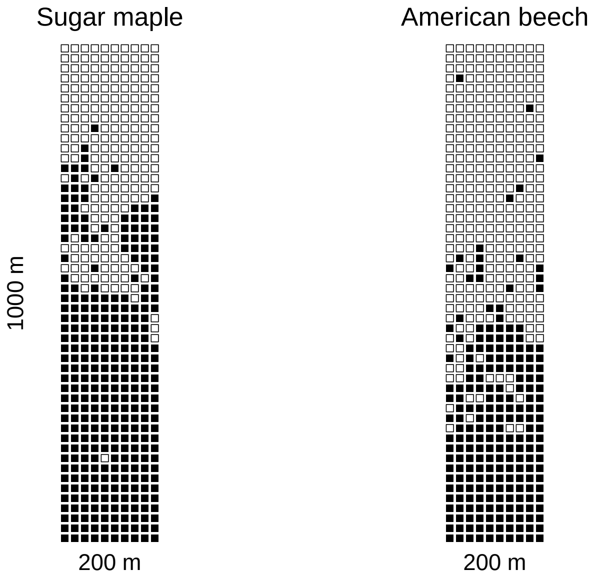
$$P(\mu|\mathbf{y}, \mathbf{X}, \sigma, \beta) \propto \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta-\mu)^2}{2\sigma^2}} \times e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$$

σ

$$P(\sigma|\mathbf{y}, \mathbf{X}, \mu, \beta) \propto \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta-\mu)^2}{2\sigma^2}} \times \frac{1}{\sigma^2}^{\tau_0-1} e^{-\varphi_0 \frac{1}{\sigma^2}}$$

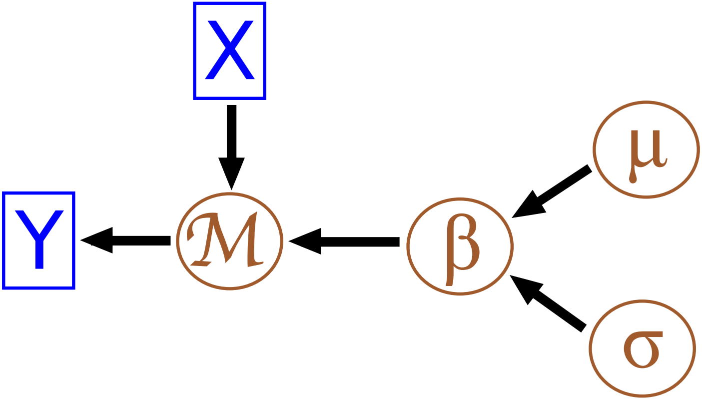
Building a hierarchical model

How are the sugar maple and the american beech influenced by elevation?



Building a hierarchical model

Direct Acyclic Graph



Model definition

$$\text{logit}(P(\mathbf{Y}_{ij} = 1)) = \beta_j \mathbf{X}$$

$$P(\mathbf{Y}_{ij} = 1) = \frac{e^{\beta_j \mathbf{X}}}{1 + e^{\beta_j \mathbf{X}}}$$

$$\beta_j \sim \mathcal{N}(\mu, \sigma^2)$$

Prior definition

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$$\frac{1}{\sigma^2} \sim \mathcal{G}(\tau_0, \varphi_0)$$

Data

$\mathbf{Y}_{ij}$ The presence (or absence) of species j at location i

$\mathbf{X}$ Elevation

Parameters

β_j The importance of elevation for species j

μ Average response of the species to elevation

σ^2 How a species varies in its response to elevation

Priors

μ_0 Mean prior about how μ is distributed

σ_0^2 Variance prior about how μ is distributed

τ_0 Scale prior about how σ is distributed

φ_0 rate prior about how σ is distributed

Write a Gibbs sampler to estimate all parameters of the hierarchical model described previously

Recall that

$$\underbrace{P(\text{Model}|\text{Data})}_{\text{Posterior}} \propto \underbrace{P(\text{Data}|\text{Model})}_{\text{Likelihood}} \underbrace{P(\text{Model})}_{\text{Prior}}$$

$$P(\boldsymbol{\theta}|\mathbf{Y}) \propto P(\mathbf{Y}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

Define the different parts... Mathematically

Likelihood

$$P(\mathbf{Y}_j|\beta, \mathbf{X}) = \prod_{i=1}^{n}\left(\frac{e^{\beta_j \mathbf{X}_i}}{1+e^{\beta_j \mathbf{X}_i}}\right)^{\mathbf{Y}_{ij}}\left(\frac{1}{1+e^{\beta_j \mathbf{X}_i}}\right)^{1-\mathbf{Y}_{ij}}$$

$$P(\beta_j|\mu, \sigma^2) = \prod_{j=1}^{m}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\beta_j-\mu)^2}{2\sigma^2}}$$

Define the different parts... Mathematically

Prior

$$P(\mu, \sigma^2) = (\mu | \mu_0, \sigma_0)(\sigma^2 | \tau_0, \varphi_0)$$

$$= \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \times \frac{{\varphi_0}^{\tau_0}}{\Gamma(\tau_0)} \left(\frac{1}{\sigma^2}\right)^{\tau_0-1} e^{-\varphi_0 \frac{1}{\sigma^2}}$$

How to sample each parameter independently

β_j

$$P(\beta_j|\mathbf{Y}_j, \mathbf{X}, \mu, \sigma^2) \propto \prod_{i=1}^{n} \left(\frac{e^{\beta_j \mathbf{X}_i}}{1 + e^{\beta_j \mathbf{X}_i}}\right)^{\mathbf{Y}_{ij}} \left(\frac{1}{1 + e^{\beta_j \mathbf{X}_i}}\right)^{1-\mathbf{Y}_{ij}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta_j - \mu)^2}{2\sigma^2}}$$

μ

$$P(\mu|\mathbf{Y}_j, \mathbf{X}, \sigma, \beta_j) \propto \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta_j - \mu)^2}{2\sigma^2}} \times e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}}$$

σ

$$P(\sigma|\mathbf{Y}_j, \mathbf{X}, \mu, \beta_j) \propto \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\beta_j - \mu)^2}{2\sigma^2}} \times \frac{1}{\sigma^2}^{\tau_0 - 1} e^{-\varphi_0 \frac{1}{\sigma^2}}$$