Hierarchical Models

What is 'Bayesian' analysis?

- Bayesian analysis is a different way to learn from your data
- It is based around the application of Bayes' theorem
- It is a type of statistical inference
- Bayesian analysis has two inputs:
  – What we knew before the we analyzed the data

– The data itself

- Bayesian analysis tells us what the new state of knowledge is (including any uncertainty) after analyzing the data

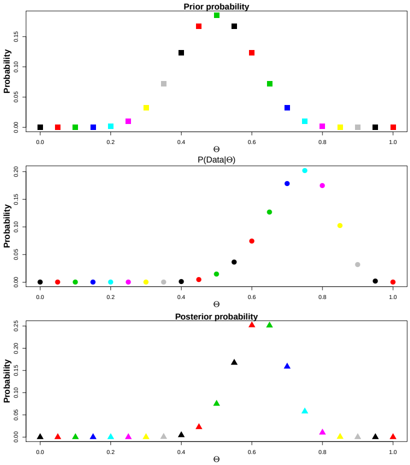
Statistical inference

- Nearly all forms of statistical analysis follow these basic steps:
  - Construct a mathematical simplification of the study system that encapsulates the phenomena that you are interested in. This is referred to as a 'model'. Linear models (used in linear regression) are an example of a type of model.
- Compare the model with the data
- Make conclusions based on this comparison
- Up until now you have encountered a few different types of statistical inference:
  - null-hypothesis testing
- (ordinary) least squares
- maximum likelihood
- Bayesian inference is another type of statistical inference; it is not a model.

Posterior: Normalized product of Likelihood and Prior

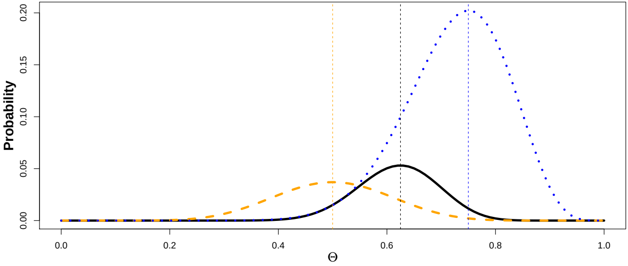
$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$

$$Posterior : Probability\ of\ parameter\ value\ given\ data$$



Posterior: Normalized product of Likelihood and Prior

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$



Hierarchical data structure

Question

Probability that (i) a child bought lunch from the school cafeteria, in (ii) different schools in (iii) different districts

Data:

- the number of days each child bought lunch (h)
- the number of days the child was at school (n)
- the school and district of the child

Hierarchical data structure

Question

Probability that (i) a child bought lunch from the school cafeteria, in (ii) different schools in (iii) different districts

Each child has an individual probability

Within a school probabilities of children are related

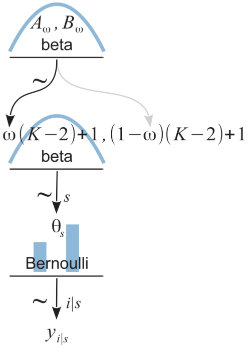
- food quality
- price
- tradition

School specific distribution of the probability of children buying food

Within each district school specific probabilities are related

- regulations
- subsidies

Hierarchical model



Probability of individual child: Θ_s

School specific probability: ω_{school}

District specific probability: $\omega_{district}$

Θ_s depends on ω_{school} which again depends on $\omega_{district}$

Hierarchical model:

Statistical model written in multiple levels in which values of parameters at one level depend on parameters at a higher level

$$P(\Theta, \omega_{school}, \omega_{district} | D) \propto$$
$$P(D|\Theta) P(\Theta|\omega_{school}) P(\omega_{school}|\omega_{district}) P(\omega_{district})$$

Hierachical model:

Coin flip example:

- Two coins produced in the same factory
  - the two coins have a slightly different probability of heads: Θ_1 *and* Θ_2
  - could estimate bias independently
  - these two **probabilities are related** (the factory aims to produce unbiased coins)

Hierachical model:

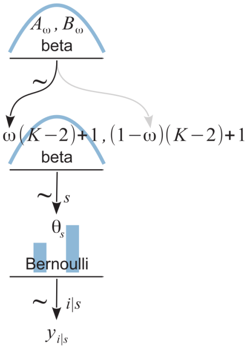
Coin flip example:

New parameterization:

ω : *mode* (*maximum*)

K : *concentration* (*variation around mean*)

Hierarchical model:

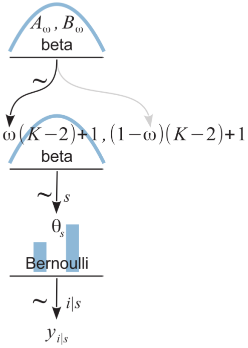


Θ_1 *and* Θ_2 depend on an other distribution on a higher hierarchical level (Hierarchy)

$$p(\omega, \Theta_1, \Theta_2 | y) = p(y|\Theta_1, \Theta_2) p(\Theta_1, \Theta_2 | \omega) p(\omega)$$

- data from both coins is used to estimate the factory bias (ω)

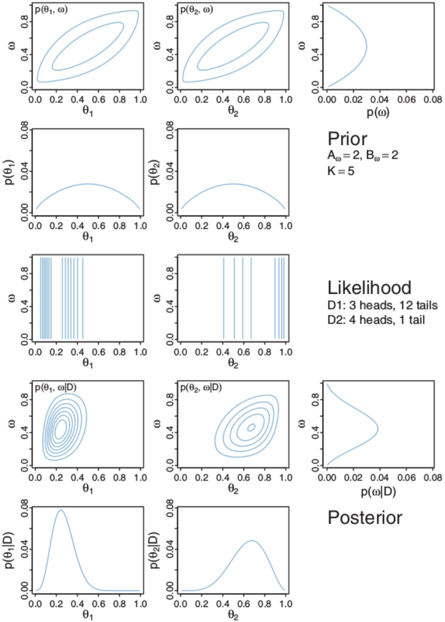
Hierarchical model:



$$p(\omega, \Theta_1, \Theta_2 | y) = p(y|\Theta_1, \Theta_2) p(\Theta_1, \Theta_2|\omega) p(\omega)$$

- data from both coins is used to estimate the factory bias (ω)
- coin bias is related to factory bias, the two coins influence each others bias estimate

Coin flip: Shrinkage



Hierarchical models: linear regression

Observations of the same variables from different locations (plots)

Model all locations **independently**

$y_s = a_s + b_s x + \epsilon$

Model all locations **together**

$y_i = a + b x_i + \epsilon_i$

Hierarchical models: linear regression

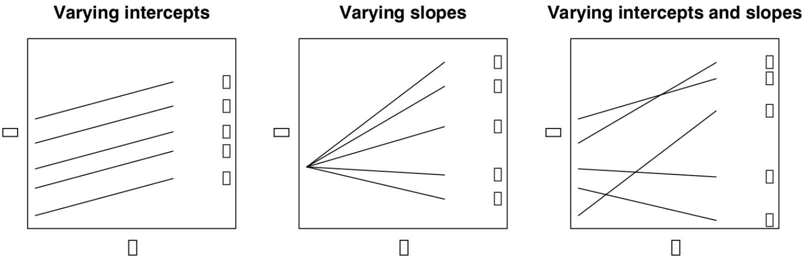
Each location has own intercept and slope but they are related through a hyperparameter for intercept and slope

$y_i = a_{j[i]} + bj[i]x_i + \epsilon_i$

$a_j \sim N(\bar{a}, s_a)$ $b_j \sim N(\bar{b}, s_b)$

Close relation to mixed effects models in classical statistics

Linear regression



Pollen vegetation model STEPPS

Pollen vegetation models are used to reconstruct vegetation based on fossil pollen data

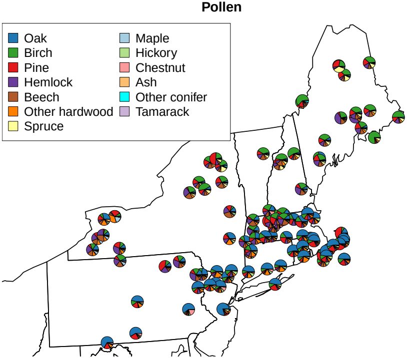
STEPPS consists of a pollen dispersal model and of a prediction model

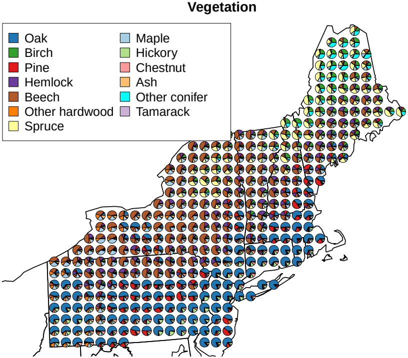
Parameter estimation in the pollen dispersal model

Pollen dispersal model: parameter estimation

$P(\text{Pollen}|\text{Vegetation},\Theta_{pollen})$

Calibration of pollen dispersal model: estimation of Θ_{pollen} based on simultaneously observed pollen and vegetation





Pollen dispersal model: structure

$$P(\Theta_{pollen}|Pollen, Vegetation) \propto P(Pollen|\Theta_{pollen}, Vegetation)P(\Theta_{pollen})$$

Vegetation observation at 8 km by 8 km grid

Pollen observation in same domain as vegetation

Pollen deposition modeled using two terms:

1. deposition within grid cell
2. pollen transport from other grid cells

Pollen dispersal model: Local deposition

$$deposition_{local} = \gamma_p \phi_p r_p(s_i)$$

γ_p: proportion of pollen deposited locally

ϕ_p: scaling factor accounting for taxon specific pollen productivity

$r_p(s_i)$: vegetation proportion of taxon p in grid cell s_i

Pollen dispersal model: Pollen transport

Pollen transport decreasing as a function of distance from depositional site

Inverse power law dispersal kernel:

$$w_{pl}(s_i s_k) = C(1 + \frac{d(s_i, s_k)}{a})^{-b}$$

a and b: parameters determining the rate of decline of the dispersal kernel

C: normalizing constant

$d(s_i, s_k)$: distance between two grid points

$$Transport_{pollen} = (1 - \gamma_p)\phi_p r_p(s_k) w_{pl}(s_i, s_k)$$

Pollen dispersal model: Total deposition

Total deposition: local and non-local contributions

$$r_{i,p}^{pol} = \gamma_p \phi_p r_p(s_i) + \frac{1}{\Omega}(1 - \gamma_p)\phi_p \sum_{s_k \neq s_i} r_p(s_k) w_{pl}(s_i, s_k)$$

$$\Omega = \sum_{s_k \neq s_i} w_{pl}(s_i, s_k)$$

Ω: Normalizing constant

Pollen dispersal model: Modeled and observed deposition

$$y_i \sim DM(n_i, r_{i,p}^{pol})$$

n_i: pollen counts at site i

y_i: observed pollen

DM: Dirichlet multinomial distribution: overdispersed multinomial distribution

Overdispersed: Excess variance

Model fitted using Stan 2.17.1.

Pollen dispersal model

$$P(\Theta|Pollen) \propto P(Pollen|\Theta)P(\Theta)$$

Estimates of pollen productivity

Taxon specific pollen productivity:

Pollen productivity is modeled as independent among taxa.

$\phi_k \sim U(0, 300)$

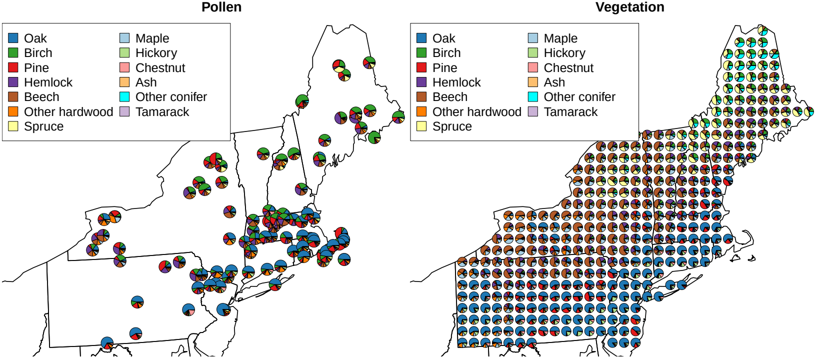
Estimates of local deposition

Taxon specific local deposition depends on a hyper parameter

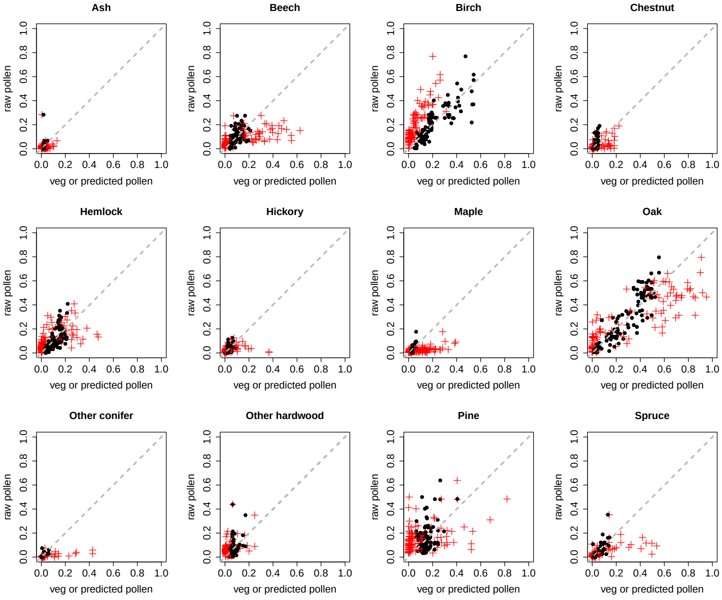
$P(\gamma_k, \mu_\gamma, \sigma_\gamma | Pollen) \propto P(Pollen|\gamma_k) P(\gamma_k | \mu_\gamma, \sigma_\gamma) P(\mu_\gamma) P(\sigma_\gamma)$

$\gamma_k \in [0, 1]$ logit prior

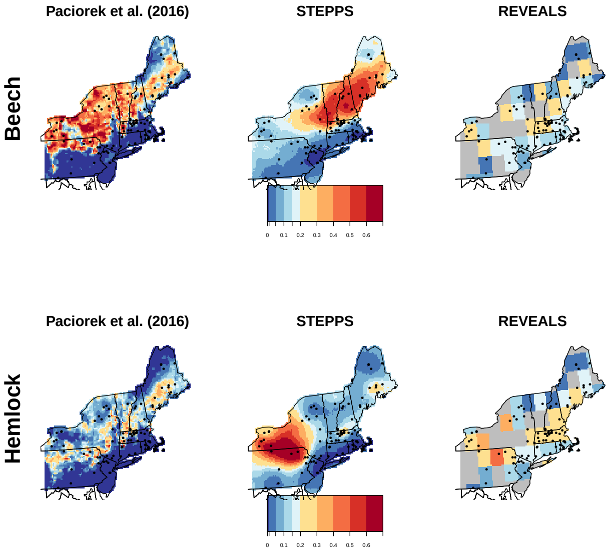
Pollen and vegetation data



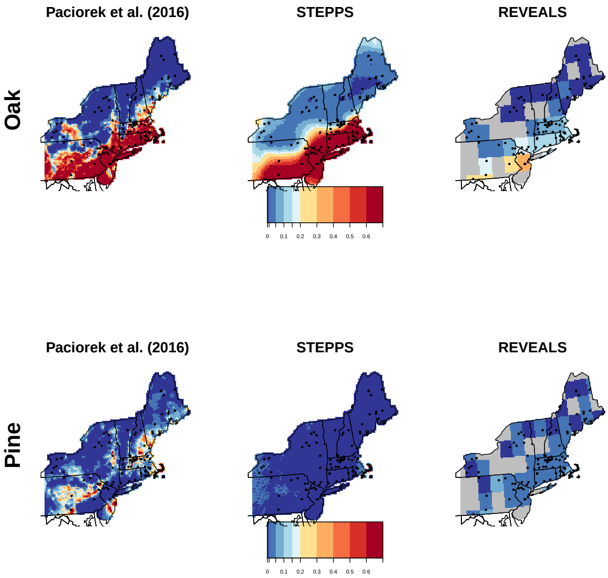
Comparison observed and modeled pollen



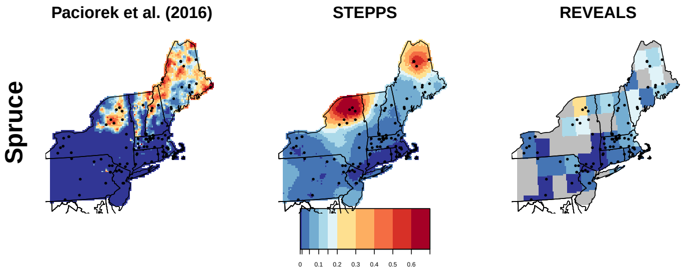
Settlement era predictions and observations: Beech and Hemlock



Settlement era predictions and observations: Oak and Pine



Settlement era predictions and observations: Spruce



Community level: Dissimilarity Paciorek et al. and STEPPS