Introduction to hierarchical models: linear models, generalized linear models, and random effects

Marc Kéry

marc.kery@vogelwarte.ch

Occupancy Workshop,
Patuxent,
12–14 Nov 2013

Outline of talk

- What are hierarchical models ? --- Why are they useful/beautiful ?
- What is a GLM ? --- What is a linear model and how can we describe it ?
- What are random effects ? --- Why/when do we specify parameters as random variables ?
- See also Word document with same content

Hierarchical model: sequence of random variables

- Hierarchical model (HM): sequence of random variables (RVs), observed (y) or unobserved (x)

$$x \sim f(\omega)$$
$$y \sim g(x, \theta)$$

- Factorization of joint distribution [x,y] into marginal ([x]) and conditional distribution, [y|x]:

$$[x,y] = [x]\,[y|x]$$

- Estimands: parameters ω and θ; latent variables x
- x: latent variables = unobserved RVs = **random effects**
- Large class of models can be described as HMs

Hierarchical model: sequence of random variables

- HMs powerful manner of building statistical models
- Typically reflect our thinking about mechanistic structure of process studied
- Describe stochastic processes with multiple layers/levels of variability and uncertainty
- (Note: variability *vs.* uncertainty)
- HMs can be fitted using classical (likelihood) methods (e.g., R packages **lme4** and **unmarked**)
- Bayesian analysis more straightforward
- BUGS language ideal for fitting HMs: indeed, developed especially for HMs !
- Models naturally described hierarchically in BUGS language

Why hierarchical models ?

- HMs make the fitting of complex models easier (especially bayesianly)
- ... foster mechanistic, cleaner thinking about a problem
- ... lead to mechanistic modeling
- ... foster synthetic understanding of large array of models
- ... enable honest accounting for all components of variability and uncertainty in a system
- ... ideal for rigorous modeling of measurement-error processes in ecological field data

HMs foster mechanistic thinking about system and synthetic understanding of large class of models

- e.g., state-space model for time-series of observations (y)

Ecological process

[y] = initial system state z_0 *

system state z_{t+1} given system state z_t *

observation y given system state z

Measurement-error process, observation process

- y counts at 1 site:
  state-space model (BPA chap. 5)
- y individual detection histories, multiple individuals:
  JS model (BPA chap. 10)
- **y site-detection histories, multiple sites:**
  dynamic occupancy model (BPA chap. 13)

HMs foster synthetic understanding of models

- **moreover ...**

  [y] = initial system state z_0 *

  system state z_{t+1} given system state z_t *

  observation y given system state z

- y counts, no dynamics:
  random-effects model (BPA chap. 4)
- y site counts, multiple sites:
  Dail-Madsen Nmix model (BPA chap. 12)
- y individual detection histories, no dynamics:
  closed-population model (BPA chap. 6)
- y individual detection histories, no initial state:
  CJS model (BPA chap. 7), multi-state models (BPA chap. 8)
- **y site-detection histories or counts, multiple sites, but no dynamics: static occupancy model (BPA chap. 13), static N-mixture model (BPA chap. 12)**

vogelwartech

Components of HMs: GLMs

- HMs consist of two or more random variables
- How do we describe structure in random variables ? --- > typically as GLMs
- **HMs typically consist of a sequence of two or more GLMs !**

Components of HMs: GLMs

- statistical model:

  response ~ deterministic part + stochastic part

  e.g. $y_i = \alpha + \varepsilon_i$

  with $\varepsilon_i \sim Normal(0, \sigma^2)$
- stochastic/random part: statistical distribution
- deterministic/systematic part: non-linear or linear function
- non-linear may be more mechanistic and better for extrapolation, but (much) more difficult to fit
- linear model easiest by far for deterministic part of response
- For hierarchical modeling, must understand GLM
- For GLM, must understand linear modeling
- Chapter 3 in BPA book; also Kéry (2010)

Generalized linear models (GLMs)

- Have some random variable y (response);
  some explanatory variable(s) x
- Linear model in terms of x applied to transformation g
  of expected response E(y),
  i.e., g(E(y)) = some linear model in x
- Three components of GLM:
  (1) random part of response: statistical distribution
  (2) link function g
  (3) systematic part of response (linear predictor):
  some linear function of x

Three most frequent GLMs

- Normal response:

  Random part: $y \sim \text{Normal}(\mu, \sigma^2)$

  (typical) Link function: none ("identity link")

  Systematic part: SLM (some linear model)

- Poisson response:

  Random part: $y \sim \text{Poisson}(\lambda)$

  (typical) Link function: log

  Systematic part: SLM

- Binomial response:

  Random part: $y \sim \text{Binomial}(p, N)$

  (typical) Link function: $\text{logit} = \log(p / (1-p))$

  Systematic part: SLM

Types of descriptions of linear model in GLM

- words: e.g., “A and x act additively ....”
- graphs: e.g., lines and bars
- specific labels: e.g., ANOVA, ANCOVA, t-test, regression
- algebra: e.g., $y_i \sim Normal(\alpha_j + \beta * xi, \sigma^2)$
- matrix algebra: e.g, **y** = **Xb** + **e**
- system of equations: see example later
- R language: e.g.,

```
lm(y ~ A + x)
```

- BUGS language: e.g.,

```
y[i] ~ dnorm(mu[i], tau)
mu[i] <- alpha[A[i]] + beta * x[i]
```

- alternative, equivalent descriptions: parameterisations

Specification of linear models in BUGS language

- one of the main stumbling blocks of becoming a BUGS modeler at first
- **BUT:** wonderful side effect that finally *enforces* an understanding of linear models fit by issuing R commands such as `lm(y ~ A*x)`

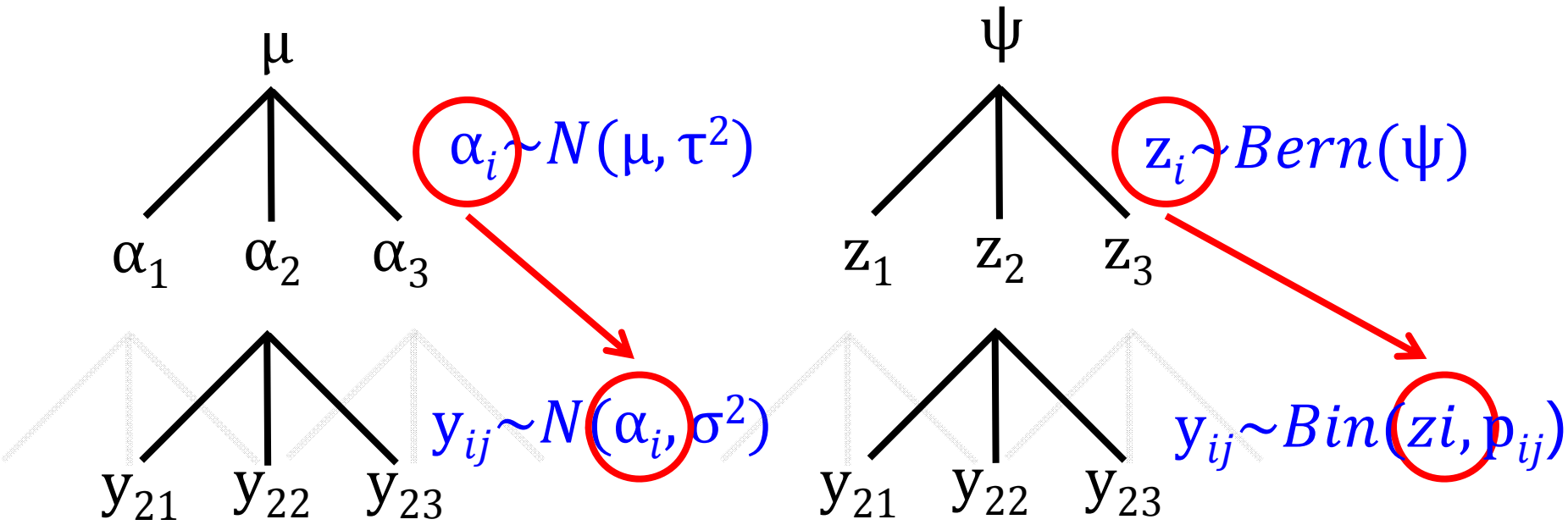
Some important linear models in glm/lm and BUGS

- e.g., linear models with factor `pop` and covariate `length`

No.	Model in R	Model in algebra	Model in BUGS	Traditional name of technique based on that linear model	Number of parameters	Meaning
1	`1`	α	`alpha`	"Model of the mean"	1	Constant term (intercept) only
2	`pop`	α_j	`alpha[pop[i]]`	One-way ANOVA	3	Three constants, one for each population (called the level of the factor). Called a t-test if factor has only two levels.
3	`pop-1`	$\mu+\alpha_j$	`mu + alpha[pop[i]]`	One-way ANOVA	3	"Subtract the intercept"; this is a mere reparameterisation of model 2. with an intercept (= the value for the first population) and two constants that are the differences between the values of population 2 and 1 and 3 and 1. In BUGS, the first level of the vector alpha must be manually set to zero to avoid overparameterisation. In R this is done automatically.
4	`length`	$\alpha+\beta*x_i$	`alpha + beta * length[i]`	Simple linear regression	2	An intercept plus a slope, common to all three populations (i.e., no effect of pop)
5	`length-1`	$\beta*x_i$	`beta * length[i]`	Simple linear regression through the origin	1	"Subtract the intercept": this is NOT a mere reparameterisation of model 4. Regression through the origin; not usually a meaningful model.
6	`pop+length`	$\alpha_j+\beta*x_i$	`alpha[pop[i]] + beta * length[i]`	Main-effects ANCOVA	4	One separate intercept for each population and a common slope
7	`pop*length`	$\alpha_j+\beta_j*x_i$	`alpha[pop[i]] + beta[pop[i]] * length[i]`	Interaction-effects ANCOVA	6	Three separate intercepts and three separate slopes. That is, fully separate regression of wing on length for each population.

The 2nd thing in HMs: random effects

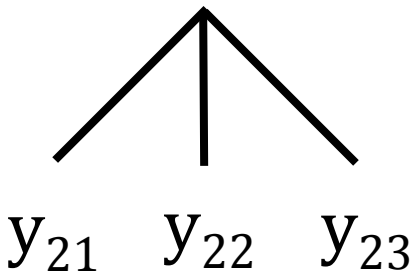
- **Hierarchical models: Dependent sequence of random variables (observed & unobserved)**
- e.g., randomised block ANOVA, „hierarchical logistic regression“



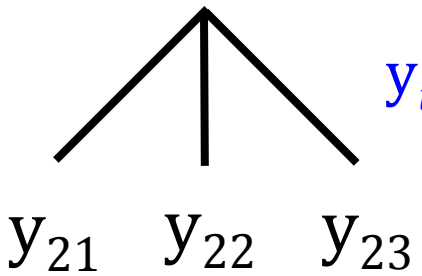
- Many other names: state-space, random-, mixed-effects, latent variables, unobserved components, **mixed models** ...

Modeling of measurement errors in HMs

- Last level of HM: measurement error = observation process



$y_{ij} \sim N(\alpha_i, \sigma^2)$



$y_{ij} \sim Bin(zi, pij)$

- <u>Qualitatively very different:</u>

(1) cancel out
$E(y_{i.}) = \alpha_i$

(2) do **NOT** cancel out
$E(y_{i.}) \neq z_i$

=> Should care for measurement errors

vogelwarte.ch

So what are random effects ?

- Realized values of unobserved random variables in a HM !
- two or more latent variables that are given a (prior) distribution; **that's all !**
- prior distribution assumed Normal in most classical GLMMs
- **BUT:** can be non-normal, e.g., Poisson (discrete) or even Bernoulli (discrete and binary): cf. Nmix and site-occupancy models
- can estimate parameters of prior distribution (also called mixing distribution), e.g., τ^2 or ψ, **and** realized values α and z !
- parameter, prior/mixing distribution, hyperparameters, hyperpriors, ...

$$\alpha_i \sim N(\mu, \tau^2) \qquad z_i \sim Bern(\psi)$$

$$y_{ij} \sim N(\alpha_i, \sigma^2) \qquad y_{ij} \sim Bin(zi, pij)$$

vogelwarte.ch

So what are random effects ?

- Realized values of unobserved random variables in a HM !
- replication of study would (usually) yield different effects
- can estimate values of “new” draws from same prior distribution
- in analysis, treatment as fixed or random is a judgment; cannot be “tested”
- in Bayes, effects exchangable: “similar but not identical”

  -> don’t treat “dissimilar” effects as random (e.g., captive vs. wild populations)
- (also: bad estimates when few levels)
- (HMs computationally more expensive)
- (HMs more difficult to explain/understand)

Motivations for treating set of params random

- extend scope of inference
- assess/partition system variability
- proper assessment of uncertainty
- model random effects, e.g., as a GLM, correlations
- spatial and temporal correlations
- avoid pseudo-replication
- borrow strength/improve individual random effects estimates/shrinkage
- combine information (meta-analysis)
- others

See numerical example in Word doc with R code

- see there ....