3.1: Review of GLMs and GLMMs

How do we estimate things?

1. Specify a model
   – Function generating predictions
2. Identify plausible values for any unknown parameters
   – Maximize probability of observations given function
3. Assess uncertainty
   – Explore function around plausible values

Maximum likelihood estimation (MLE)

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}(L(\boldsymbol{\theta}; \mathbf{y}))$$

– Where $\hat{\boldsymbol{\theta}}$ is the MLE estimate of parameters

– $\text{argmax}_{\boldsymbol{\theta}}(L(\boldsymbol{\theta}; \mathbf{y}))$ is the maximum value for $L(\boldsymbol{\theta}; \mathbf{y})$ that can be achieved for any value of $\boldsymbol{\theta}$

- Usually we specify that each datum is independent

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\left(\sum_{i=1}^{n_i} L(\boldsymbol{\theta}; y_i)\right)$$

Why use maximum likelihood estimation?

$$\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\big(L(\boldsymbol{\theta}; \mathbf{y})\big)$$

Where $p(\mathbf{y}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \mathbf{y})$ is your specified probability distribution

1. Consistency (correct model)
2. Consistency (incorrect model)
3. Asymptotic normality

Specify a generalized linear model

- Step 1 – Specify a linear predictor for response variable

$$\mathbf{x}_{\mathrm{i}}\mathbf{b} = \sum_{j=1}^{n_j} x_{i,j} b_j$$

  - where $\mathbf{x}_{\mathrm{i}}$ is a row of a predictor matrix $\mathbf{X}$
  - $\mathbf{b}$ is a vector of parameters

- Step 2 – Specify a link function

$$y_i^* = f^{-1}(\mathbf{x}_{\mathrm{i}}\mathbf{b})$$

  - Linear predictor $\mathbf{x}_{\mathrm{i}}\mathbf{b}$ ranges from (-Inf, Inf)
  - Reponse variable often has different range

- Step 3 – Specify a probability distribution for your response variable

$$y_i \sim Normal(y_i^*, \sigma^2)$$

Common link functions

Name	Equation for *f()*	Equation for *f* $^{-1}$*()*	Range
Identity	$f(x) = 1$	$f(x) = 1$	$-\infty < f^{-1}(x) < \infty$
Logit	$f(x) = \log\left(\frac{x}{1-x}\right)$	$f(x) = \frac{\exp(x)}{1+\exp(x)}$	$0 < f^{-1}(x) < 1$
Log	$f(x) = \log(x)$	$f(x) = \log(x)$	$0 < f^{-1}(x) < \infty$

- Why use link functions
    1. Restrict range for variable
    2. Help to interpret parameters
        - E.g., if using a log-link, then a 0.1 increase in predictors causes a 10% increase in the response

Common distributions for data

– Discrete

Name	Notation	Domain	Range
Bernoulli	$B \sim Bernoulli(p)$	$0 \leq p \leq 1$	B = {0, 1}
Binomial	$N \sim Binomial(p, n)$	$0 \leq p \leq 1$	N = {0, 1, …, n}
Poisson	$N \sim Poisson(\lambda)$	$\lambda > 0$	N = {0, 1, …,∞}
Negative binomial	$N \sim NegBin(\lambda, \theta)$	$\lambda > 0$ $\theta > 0$	N = {0, 1, …,∞}

– Continuous

Name	Notation	Domain	Range
Normal	$Y \sim Normal(\mu, \sigma^2)$	$\sigma^2 > 0$	Unrestricted
Lognormal	$\log(Y) \sim Normal(\mu, \sigma^2)$	$\sigma^2 > 0$	Y > 0
Gamma	$Y \sim Gamma(\mu, CV)$	$\mu > 0$ $CV > 0$	Y > 0

How to chose a distribution for data?

- Choice 1 – is it *continuous* or *discrete?*
  - Continuous: normal, lognormal, beta, gamma
  - Discrete: Bernoulli, binomial, poisson, negative binomial
- Choice 2 – what is the range of possible values?
  - E.g., if discrete:
    - If is is 0 or 1, then its Bernoulli
    - If its between 0 and N, where N is the number of trials, then its Binomial
- Choice 3 – How flexible do you want it?

- How to estimate standard errors?
  - Estimate the “Hessian” at the MLE

$$H(\boldsymbol{\theta};\mathbf{y}) = \begin{bmatrix} \dfrac{\partial^2 \ln L(\boldsymbol{\theta};\mathbf{y})}{\partial \theta_1^2} & \dfrac{\partial^2 \ln L(\boldsymbol{\theta};\mathbf{y})}{\partial \theta_1 \partial \theta_2} \\ \dfrac{\partial^2 \ln L(\boldsymbol{\theta};\mathbf{y})}{\partial \theta_1 \delta \theta_2} & \dfrac{\partial^2 \ln L(\boldsymbol{\theta};\mathbf{y})}{\partial \theta_2^2} \end{bmatrix}$$

  - Calculate its inverse

$$\widehat{Var}(\boldsymbol{\theta};\mathbf{y}) = \mathbf{H}^{-1}$$

  - Extract element and take square root

$$\widehat{SE}(\theta_1;\mathbf{y}) = \sqrt{\widehat{Var}(\boldsymbol{\theta};\mathbf{y})_{1,1}}$$

Definitions

Term	Definition
Random effect	Coefficient that is “exchangeable” with one or more other coefficients
Hyperdistribution	Distribution for “exchangeable” random effects
Exchangeable	No information is available to distinguish between residual variability in random effects
Fixed effect	Coefficient that is not exchangeable with others, and which hence is estimated without a hyperdistribution
Mixed-effect model	Model with both fixed and random effects

Generalized linear mixed model

1. Specify distribution for response variable
   - E.g: $c_i \sim \text{Poisson}(\lambda_i)$
2. Specify function for expected value
   - E.g., $\lambda_i = \exp(\beta_0 + \boldsymbol{\beta}\mathbf{x}_i + \mathbf{u}\mathbf{z}_i)$
3. Specify distribution for random effects
   - E.g., $u_i \sim Normal(0, \sigma_u^2)$

= General linear model + mixed effect(s)

Maximum likelihood estimation (MLE)

$$\widehat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}}(\mathcal{L}(\boldsymbol{\theta}|\mathbf{b}))$$

– Where $\widehat{\boldsymbol{\theta}}$ is the MLE estimate of parameters

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{b}) = \int \Pr\left(\mathbf{b}\middle|\mathbf{n}, \sigma_b^2\right) \Pr(\mathbf{n}|\mu_t, \sigma_n^2) d\mathbf{n}$$

– Where $\boldsymbol{\theta} = \left(\mu_t, \sigma_b^2, \sigma_n^2\right)^{\mathrm{T}}$

– $\mathbf{n}$ is a "random effect"

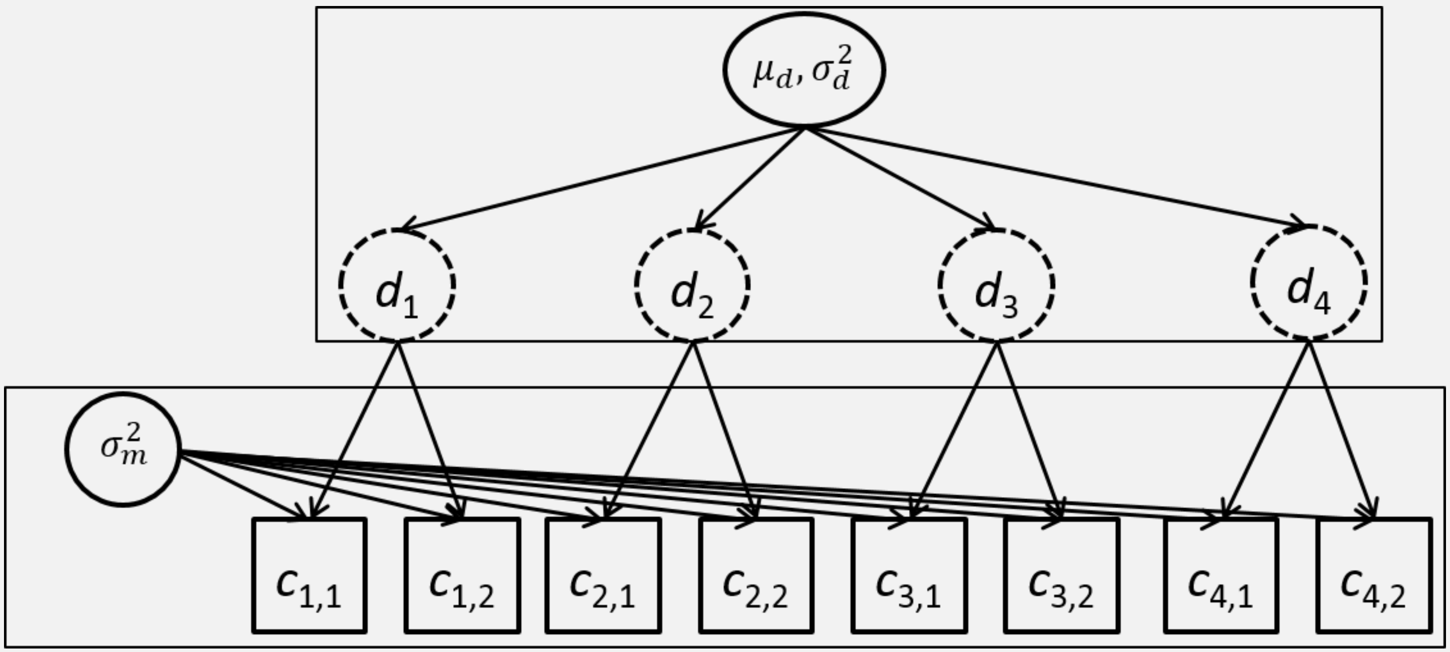
- Unobserved variable
- Not included in the marginal likelihood function

Maximum likelihood estimation (MLE)

- i.e., must “integrate” across random effects
  - They are never directly observed
  - They don’t count as “parameters” exactly
- Random effects help to calculate the marginal likelihood of parameters
  - Fixed effects can be estimated without ever interpreting random effects

- Say we’re sampling fishes near Seattle
  - Know from GIS maps that we have 120 streams
  - Can only conduct 60 samples
- Problems
  1. Can’t sample every stream!
  2. Have to deal with measurement errors
  3. Might have covariates to include

- Think about a hierarchy of parameters



- Mathematical notation
  - Local densities are exchangeable random effects

$$\log(d_s) \sim Normal(\mu_d, \sigma_d^2)$$

where μ_d is the log-mean density, and σ_d^2 is between-site variation

  - Local survey counts are also exchangeable within each site

$$\log(c_{s,i}) \sim Normal(\log(d_s), \sigma_c^2)$$

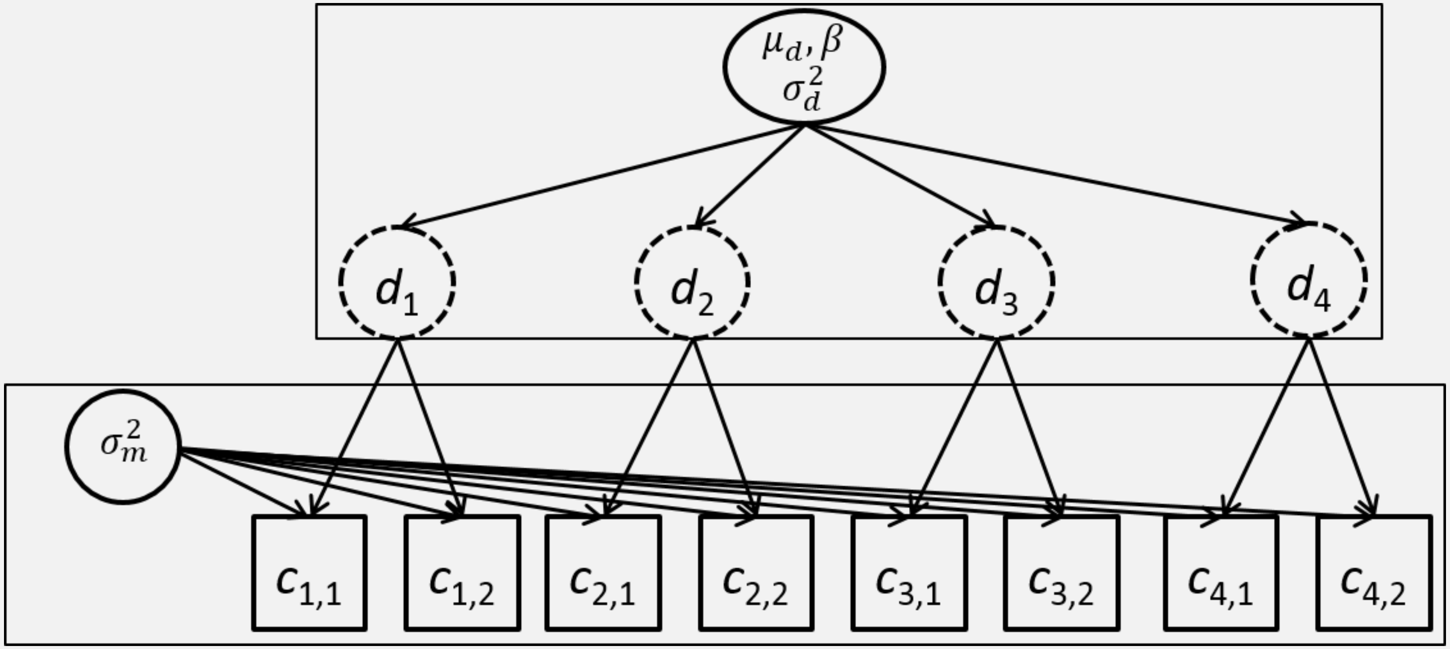
where σ_c^2 is between-sample, within-site variation

- We then just need priors

$$\mu_d \sim Uniform(-10, 10)$$
$$\sigma_d^2 \sim Uniform(0, 10)$$
$$\sigma_c^2 \sim Uniform(0, 10)$$

- What if there's a covariate?
  - Assume its measured at all chosen sites
  - Estimate its effect!

$$\log(d_s) \sim Normal(\mu_d + \beta x_s, \sigma_d^2)$$



How to extrapolate

- Un-sampled sites have an unobserved abundance
- Random sampling -> unobserved and observed are exchangeable

$$d_m \sim Normal(\mu_d + \beta x_m, \sigma_d^2)$$

- where D_m is abundance at unobserved site *m*
- Total abundance is easy to calculate!

$$d_{total} = \sum_{n=1}^{n_s} d_s + \sum_{n=1}^{n_m} d_m$$