Lab 2: Mixed-effects models

April 5, 2018

Laws of probability

1. Axiom of conditional probability

$$\Pr(X, Y) = \Pr(Y|X)\Pr(X)$$

   – Often easier to specify conditional probabilities than joint probabilities

2. Law of total probability

$$\Pr(X) = \int \Pr(X, Y)\mathrm{d}Y$$

   – Used when justifying hierarchical models

Likelihood statistics

Laplace approximation

– Define joint log-likelihood:

$$f(\theta, \varepsilon; y) = \log(\Pr(y|\theta_1, \varepsilon)\Pr(\varepsilon|\theta_2))$$

– Taylor series expansion of joint log-likelihood

$$f(\varepsilon|\theta, y) \approx f(\hat{\varepsilon}|\theta) + f'(\hat{\varepsilon}|\theta)(\hat{\varepsilon} - \varepsilon) + \frac{1}{2} f''(\hat{\varepsilon}|\theta)(\hat{\varepsilon} - \varepsilon)^2$$

– Evaluate Taylor series around "inner maximum"

$$\hat{\varepsilon} = \mathrm{argmax}_{\varepsilon}\big(f(\theta, \varepsilon)\big)$$

- Implies that $f'(\hat{\varepsilon}|\theta) = 0$

– Approximate joint likelihood via Taylor series expansion

$$\Pr(y|\theta_1, \varepsilon)\Pr(\varepsilon|\theta_2) = e^{f(\varepsilon|\theta)} \approx e^{f(\hat{\varepsilon}|\theta) - \frac{1}{2}|f''(\hat{\varepsilon})|(\hat{\varepsilon} - \varepsilon)^2}$$

Likelihood statistics

Laplace approximation

- Approximate joint likelihood via Taylor series expansion

$$\Pr(y|\theta_1,\varepsilon)\Pr(\varepsilon|\theta_2) = e^{f(\varepsilon|\theta)} \approx e^{f(\hat{\varepsilon}|\theta)-\frac{1}{2}|f''(\hat{\varepsilon})|(\hat{\varepsilon}-\varepsilon)^2}$$

- Integrate both sides

$$\int \Pr(y|\theta_1,\varepsilon)\Pr(\varepsilon|\theta_2)\,\mathrm{d}\varepsilon = \int e^{f(\varepsilon|\theta)}\mathrm{d}\varepsilon$$

$$\int \Pr(y|\theta_1,\varepsilon)\Pr(\varepsilon|\theta_2)\,\mathrm{d}\varepsilon \approx \int e^{f(\hat{\varepsilon}|\theta)}e^{-\frac{1}{2}|f''(\hat{\varepsilon})|(\hat{\varepsilon}-\varepsilon)^2}\mathrm{d}\varepsilon$$

  - And $e^{f(\hat{\varepsilon}|\theta)}$ is a constant so:

$$\int e^{f(\hat{\varepsilon}|\theta)}e^{-\frac{1}{2}|f''(\hat{\varepsilon})|(\hat{\varepsilon}-\varepsilon)^2}\mathrm{d}\varepsilon = e^{f(\hat{\varepsilon}|\theta)}\int e^{-\frac{1}{2}|f''(\hat{\varepsilon})|(\hat{\varepsilon}-\varepsilon)^2}\mathrm{d}\varepsilon$$

- Looks like a normal distribution
  - $\hat{\varepsilon}$ is the mean of the normal distribution
  - $f''(\hat{\varepsilon})$ is the hessian of the normal distribution ($f''(\hat{\varepsilon}) = \Sigma^{-1}$)

$$\mathit{Normal\ PDF}\colon\ \Pr(\varepsilon|\mu,\Sigma) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}}\exp\left(\frac{-(\varepsilon-\mu)^{\mathrm{T}}\Sigma^{-1}(\varepsilon-\mu)}{2}\right)$$

Chi-squared example

$$X \sim Chi.Squared(k)$$

$$\Pr(X = x) \propto x^{\frac{k}{2}-1} e^{\frac{-x}{2}}$$

Defining the log-likelihood

$$\log(\Pr(x)) \propto f(x)$$

Taking derivatives:

$$f(x) \propto \left(\frac{k}{2} - 1\right) \log(x) - \frac{x}{2}$$

$$f'(x) \propto \left(\frac{k}{2} - 1\right) x^{-1} - \frac{1}{2}$$

$$f''(x) \propto -\left(\frac{k}{2} - 1\right) x^{-2}$$

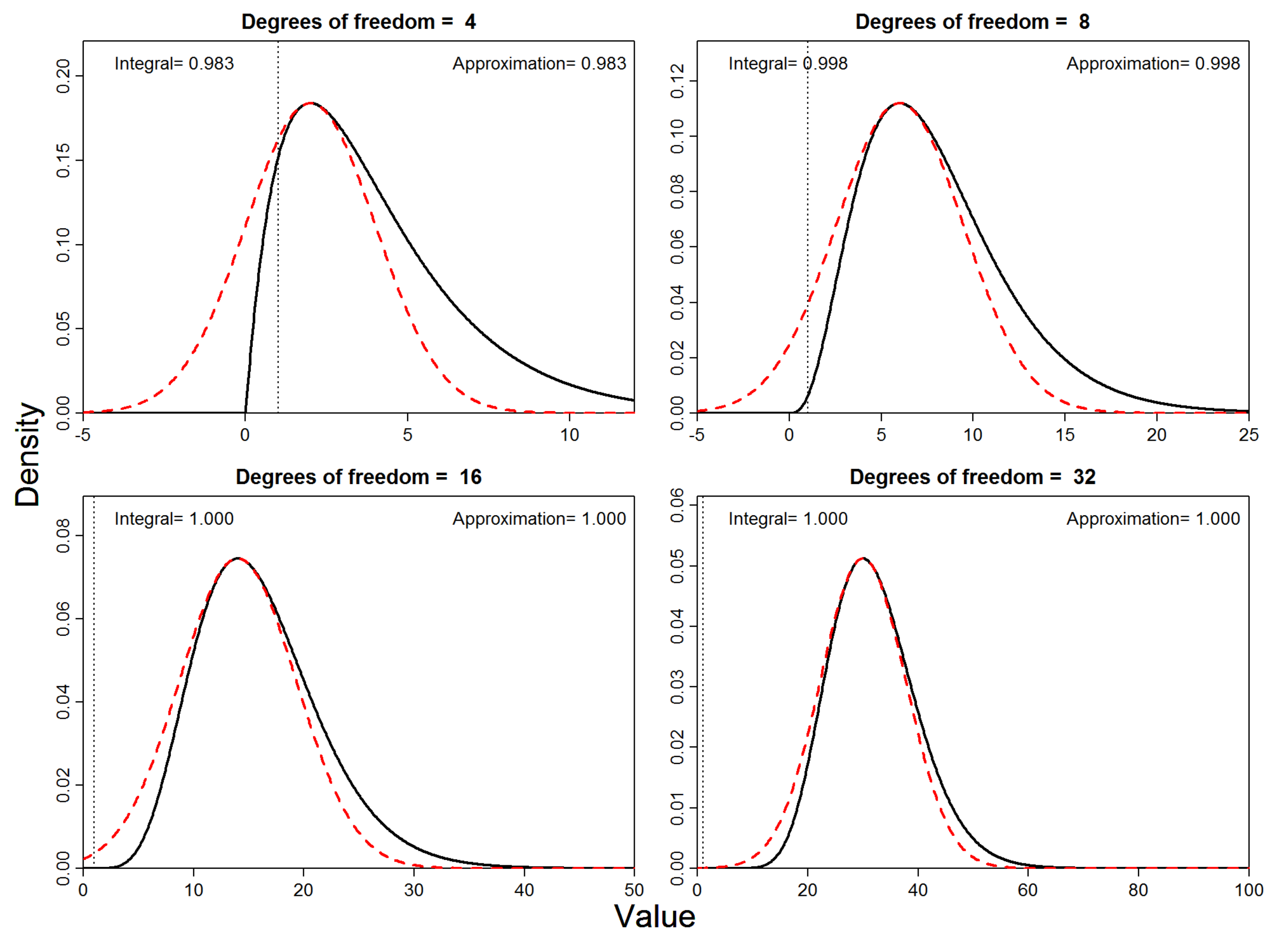
Solving for mode and Hessian:

$$f'(x) = 0 \quad \rightarrow \quad \hat{x} = k - 2$$

$$f''(\hat{x}) = -\left(\frac{1}{2(k-2)}\right)$$

Hence:

$$\Pr(x) \propto Normal(k - 2, 2(k - 2))$$



Degrees of freedom = 4
Integral= 0.983
Approximation= 0.983
Degrees of freedom = 8
Integral= 0.998
Approximation= 0.998
Degrees of freedom = 16
Integral= 1.000
Approximation= 1.000
Degrees of freedom = 32
Integral= 1.000
Approximation= 1.000
Density
Value

Likelihood statistics

Bottom line

$$\ln L(\theta; y) \equiv \int \Pr(y, \varepsilon|\theta)\, d\varepsilon \cong \log(\Pr(y, \varepsilon|\theta)) - \frac{1}{2}\log(|\mathbf{H}|)$$

- Where

$$\Pr(y, \varepsilon|\theta) = \Pr(y|\theta_1, \varepsilon)\Pr(\varepsilon|\theta_2)$$

- And

$$\mathbf{H} = \frac{\partial^2}{\partial \varepsilon^2}(\log(\Pr(y, \varepsilon|\theta)))$$

- Definitions
  - $\log(L(\theta; y))$ is the marginal log-likelihood
  - $\Pr(y, \varepsilon|\theta)$ is the joint likelihood
  - $|\mathbf{H}|$ is the determinant of the Hessian matrix

Mixed-effects models

Generalized linear mixed model

1. Specify distribution for response variable

$$c_i \sim Poisson(\lambda_i)$$

2. Specify function for expected value

$$g^{-1}(\lambda_i) = x_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\varepsilon}$$

3. Specify a link function

$$g^{-1}(a) = \log(a) \quad \rightarrow \quad g(a) = \exp(a)$$

4. Specify distribution for random effects

$$\varepsilon \sim Normal(0, \sigma_\varepsilon^2)$$

= General linear model + mixed effect(s)

Mixed-effects models

How to estimate standard errors?

– Estimate the "Hessian" at the log-marginal likelihood

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

– Calculate its inverse

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

– Extract element and take square root

$$\widehat{\mathrm{SE}}(\theta_i; \mathbf{y}) = \sqrt{\widehat{\mathbb{V}}(\boldsymbol{\theta}; \mathbf{y})_{i,i}}$$

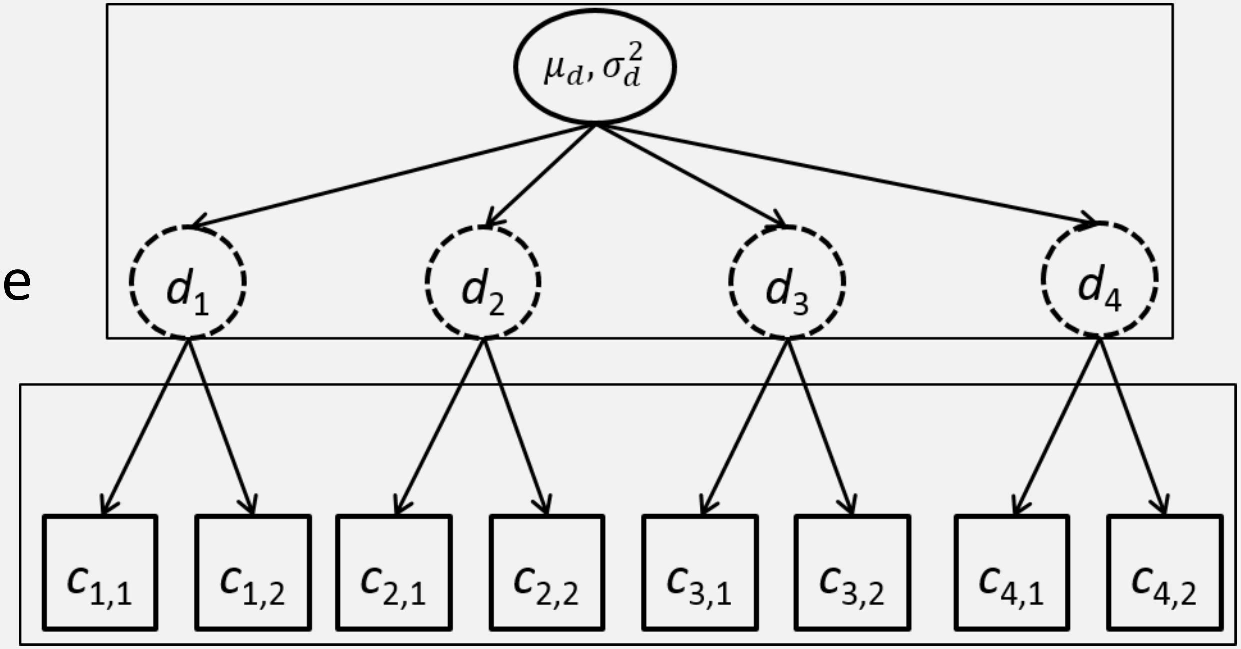
Mixed-effects models

Example – Hierarchical count samples

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

$$c_{i,j} \sim Poisson(d_j)$$

- Counts
  - 4 sites
  - 2 observations/site
  - 3 fixed effects
  - 4 random effects



Mixed-effects models

Example – Hierarchical count samples

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

$$c_{i,j} \sim Poisson(d_j)$$

Questions:

1. What is the mean of d_j across all sites j?
2. What is the variance of d_j across all sites?
3. What is the mean of $c_{i,j}$ across all sites j and samples i?
4. What is the variance of $c_{i,j}$ across all sites and samples?

Mixed-effects models

- Simulating data
  - [See R code]

Mixed-effects models

Fit using R

- Using *lme4* package
- *formula*: way to specify model

1. Linear model – *lm(formula= …)*

- Count ~ 0 + factor(Site)
- "Count" – response variable
- "0" – Don't include intercept
- "factor(Site)" – Include a fixed effect for each site

2. Linear mixed model – *lm(formula = … | …)*

- Count ~ (1 | factor(Site))
- "(1 | factor(Site))" – Include a random effect for each site

Mixed-effects models

Fit using R

- **[See R code]**

Mixed-effects models

Fit using TMB

Steps during optimization

1. Write joint log-likelihood $\Pr(y, \varepsilon|\theta)$ in CPP file

$$f(\theta, \varepsilon) = \log(\Pr(y|\theta_1, \varepsilon)\Pr(\varepsilon|\theta_2))$$

2. Choose initial values for fixed θ_0 and random ε_0
3. "Inner optimization" – Optimize random effects with $\theta_0$held constant

$$\hat{\varepsilon} = \mathrm{argmax}_\varepsilon\big(f(\theta_0, \varepsilon)\big)$$

4. Calculate Laplace approx. for marginal likelihood of fixed effects

$$\ln L(\theta_0; y) \cong f(\theta_0, \hat{\varepsilon}) - \frac{1}{2}\log(|\mathbf{H}|)$$

   - TMB also provides the gradient of the penalized likelihood with respect to fixed effects

5. "Outer optimization" – Repeat steps 2-3
   - Outer optimization is done in R using the function value and gradient provided by TMB

Fit using TMB

[See R code]

Mixed-effects models

- Benefits of using linear mixed models
  - Separate estimate of measurement and between-site variability
  - Include covariates for either one
  - Improved precision
  - *“Shrinkage”*

- Draw-backs
  - Biased if random effects aren’t “exchangeable”

Mixed-effects models

Restricted maximum likelihood models (REML)

- Maximum likelihood (ML) estimates of variance parameters are biased
  - ML estimate ${\hat{\sigma}_{ML}}^2$

$${\hat{\sigma}_{ML}}^2 = \frac{1}{n_i} \sum_{i=1}^{n_i} (y - \widehat{\mu_i})^2$$

  - Expectation

$${\hat{\sigma}_{unbiased}}^2 = \frac{1}{n_i - 1} \sum_{i=1}^{n_i} (y - \widehat{\mu_i})^2$$

  - Same problem arises for variance estimates of random effects
- REML gives unbiased estimates of random-effect variances
  - Also sometimes helps convergence
  - Important when log-likelihood function is correlated with respect to random and fixed effects

Confidence interval:

- Parameter estimates are normally distributed

- Computation

$$CI_{x\%}(\hat{\theta}) = \hat{\theta} \pm \widehat{\mathrm{SE}}(\hat{\theta}) \times \Phi^{-1}\left(\frac{x}{2}\right)$$

  - Where $CI_{x\%}$ contains the true value x% of the time if the model is correct
  - Φ^{-1} is the inverse cumulative distribution for a normal distribution
  - $\hat{\theta}$ is the estimate for parameter θ
  - $\widehat{\mathrm{SE}}(\hat{\theta})$ is the estimated standard error for parameter θ

Confidence interval coverage

– *Coverage* – the expected proportion of times that an estimated x% confidence interval contains the true value given an estimation model and true "data-generating process"

Estimation:

1. Simulate data with a known value for parameter θ
2. Record true parameter values
3. Apply estimator
4. Record confidence interval $CI_{x\%}(\hat{\theta})$ for parameter θ
5. Repeat steps 1-4 hundreds of times
6. Compute the proportion of times where $CI_{x\%}(\hat{\theta})$ contains the true value for parameter θ

Likelihood statistics

Separability

- What if different components of the model are statistically independent?

$$\Pr(y|\theta_1, \varepsilon)\Pr(\varepsilon|\theta_2) = \prod_{i=1}^{N} \Pr(y|\theta_1, \varepsilon_i)\Pr(\varepsilon_i|\theta_2)$$

- Examples:
  - Overdispersed samples

$$C_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) \sim Normal(\mu, \sigma^2)$$

  - Each λ_i is independent conditional on μ, σ^2

$$\Pr(C|\lambda)\Pr(\lambda|\mu, \sigma^2) = \prod_{i=1}^{N} \Pr(C_i|\lambda_i)\Pr(\lambda_i|\mu, \sigma^2)$$

Likelihood statistics

Separability

- Then we can factor the integral

$$\int \Pr(y|\theta_1, \varepsilon) \Pr(\varepsilon|\theta_2) \, \mathrm{d}\varepsilon = \prod_{i=1}^{N} \int \Pr(y|\theta_1, \varepsilon_i) \Pr(\varepsilon_i|\theta_2) \, \mathrm{d}\varepsilon_i$$

  – Where we replace a *N*-dimensional integral with *N* 1-dimenstional integrals

Uses

1. Meta-analysis: species are often independent
2. Time series: years are often “conditionally” independent

Mixed-effects models

[Explore “map” argument to TMB]