Hieararchical models: 1

So you're having a hard time choosing priors...

So you're having a hard time choosing priors...

- Not surprising!
- Takes practice

Useful tips for prior selection

1. Any constraints on parameter?

- variance parameters: $\sigma > 0$
- probabilities: $0 \leq p \leq 1$
- correlations: $-1 \leq \rho \leq 1$

Useful tips for prior selection

1. Any constraints on parameter?
2. Prior predictive distribution:

$$[y]$$

Review: posterior predictive distribution

Distribution of predicted data, given the observations

$$[\tilde{y} \mid y]$$

Concept:
For a *good* model, predicted data resembles the real data

Prior predictive distribution

Distribution of predicted data, given your priors

$$[y]$$

Concept:
For *good* priors, predicted data resembles your expectations for the data

Prior predictive distribution simulations

1. Simulate parameter draws from prior
2. Simulate data using these parameters

- how different from posterior predictive simulation?

Useful tips for prior selection

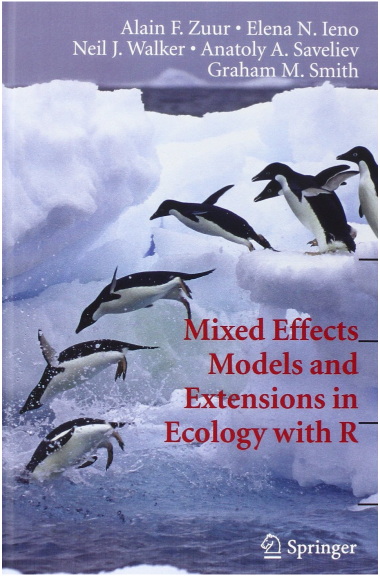
1. Constraints
2. Prior predictive distribution
3. Expert recommendations https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

Useful tips for prior selection

1. Constraints
2. Prior predictive distribution
3. Expert recommendations
4. Treat the prior parameters as unknown!

- aka use a hierarchical model

Hierarchical models: why bother?

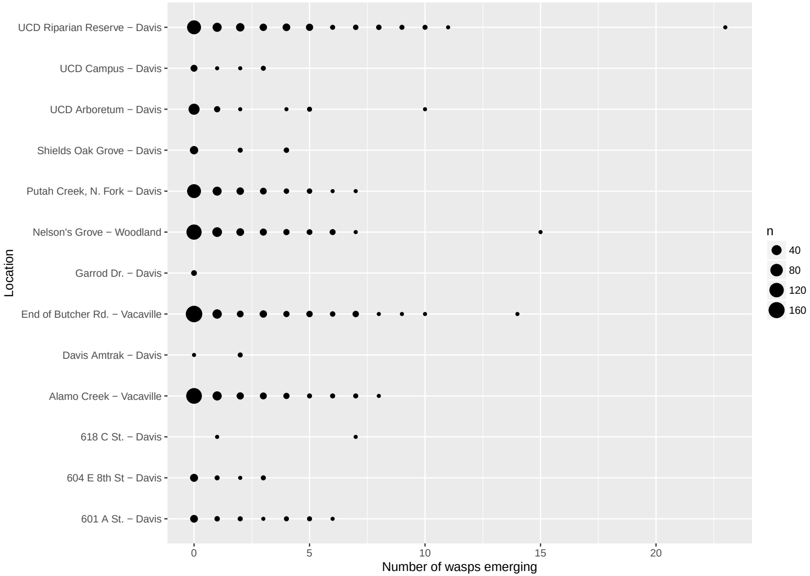


Gall wasp example

Goal: Estimate mean number of wasps for each location

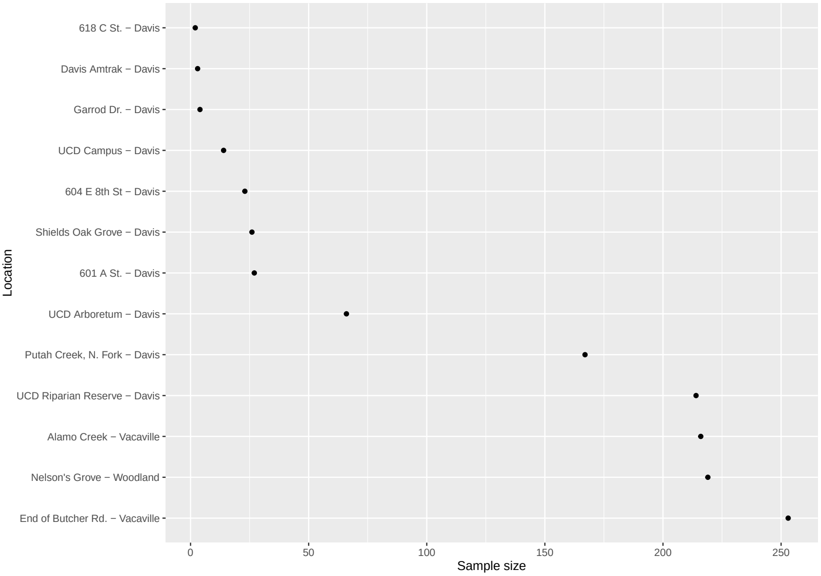
1. Sample locations $j = 1, ..., J$
2. Sample galls at each location
3. Gall i is from site j

The data



UCD Riparian Reserve − Davis
UCD Campus − Davis
UCD Arboretum − Davis
Shields Oak Grove − Davis
Putah Creek, N. Fork − Davis
Nelson's Grove − Woodland
Garrod Dr. − Davis
End of Butcher Rd. − Vacaville
Davis Amtrak − Davis
Alamo Creek − Vacaville
618 C St. − Davis
604 E 8th St − Davis
601 A St. − Davis
Location
0
5
10
15
20
Number of wasps emerging
n
40
80
120
160

Sample sizes by location



Two extreme choices to estimate means

1. Complete pooling: all locations are the same
2. No pooling: locations have different means

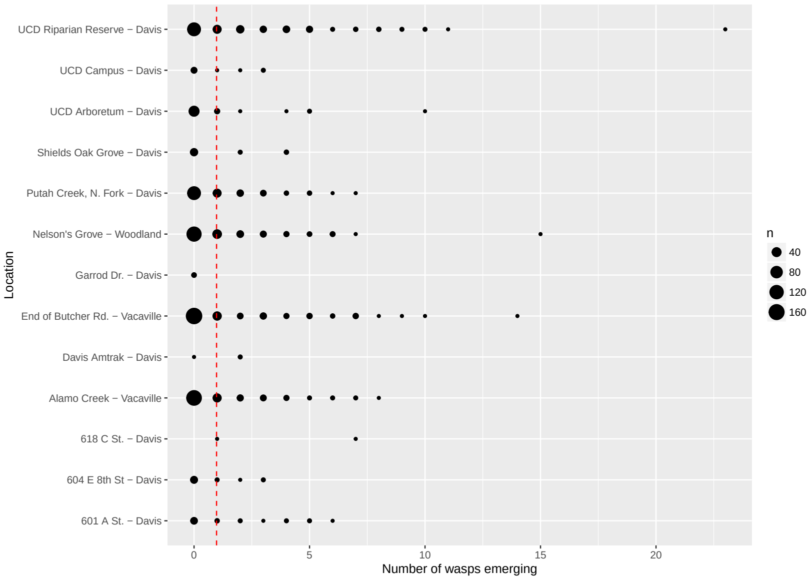
Complete pooling

```
complete_pool <- glm(n_cynip ~ 1,
                     data = d, family = poisson)
```

$$y_i \sim Poisson(\lambda)$$

$$log(\lambda) = \beta_0$$

Complete pooling



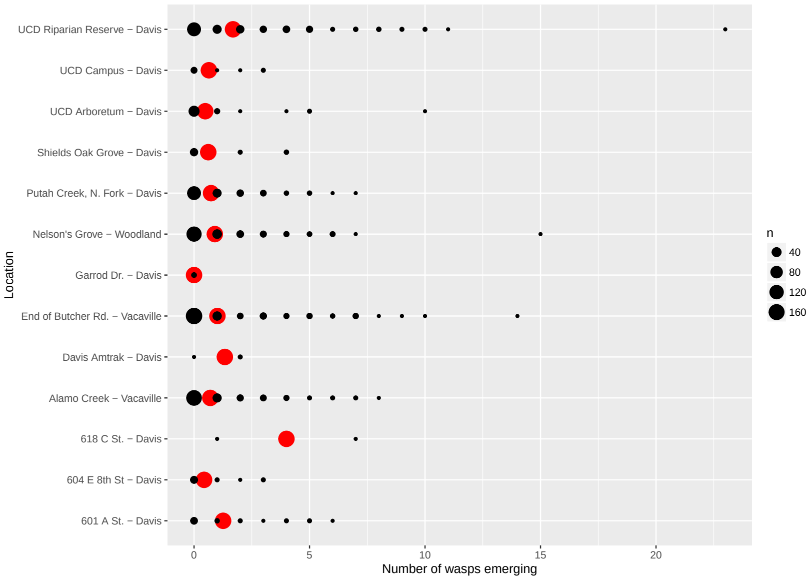
No pooling: locations different and independent

```
no_pool <- glm(n_cynip ~ 0 + gall_locality,
                     data = d, family = poisson)
```

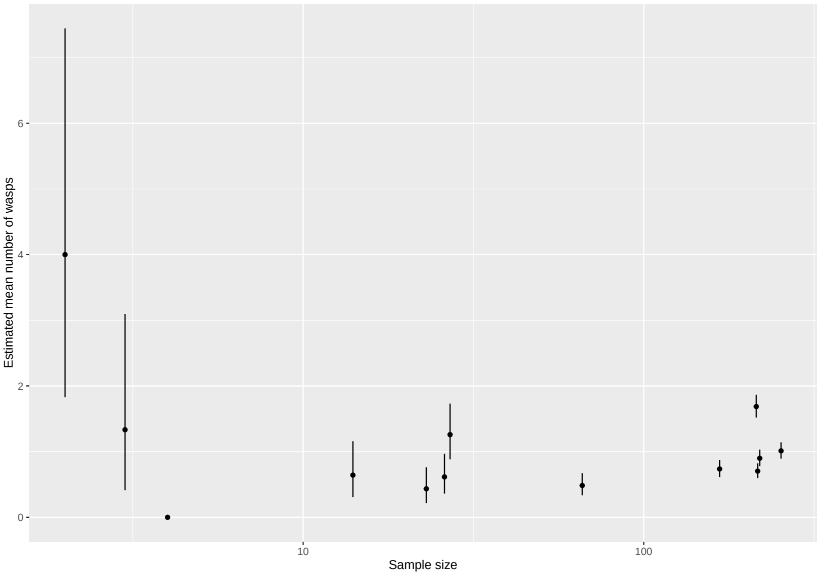
$$y_i \sim Poisson(\lambda_i)$$

$$log(\lambda_i) = \beta_{j[i]}$$

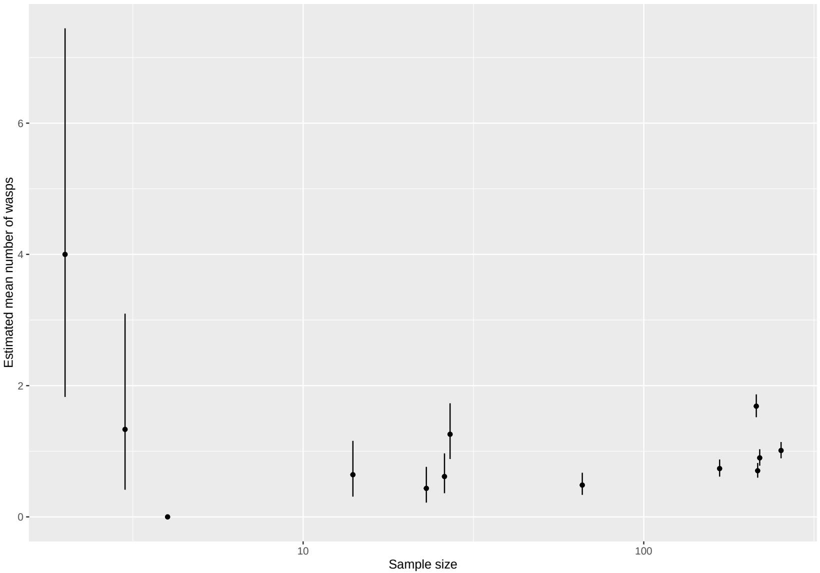
No pooling


UCD Riparian Reserve − Davis
UCD Campus − Davis
UCD Arboretum − Davis
Shields Oak Grove − Davis
Putah Creek, N. Fork − Davis
Nelson's Grove − Woodland
Garrod Dr. − Davis
End of Butcher Rd. − Vacaville
Davis Amtrak − Davis
Alamo Creek − Vacaville
618 C St. − Davis
604 E 8th St − Davis
601 A St. − Davis
Location
0
5
10
15
20
Number of wasps emerging
n
40
80
120
160

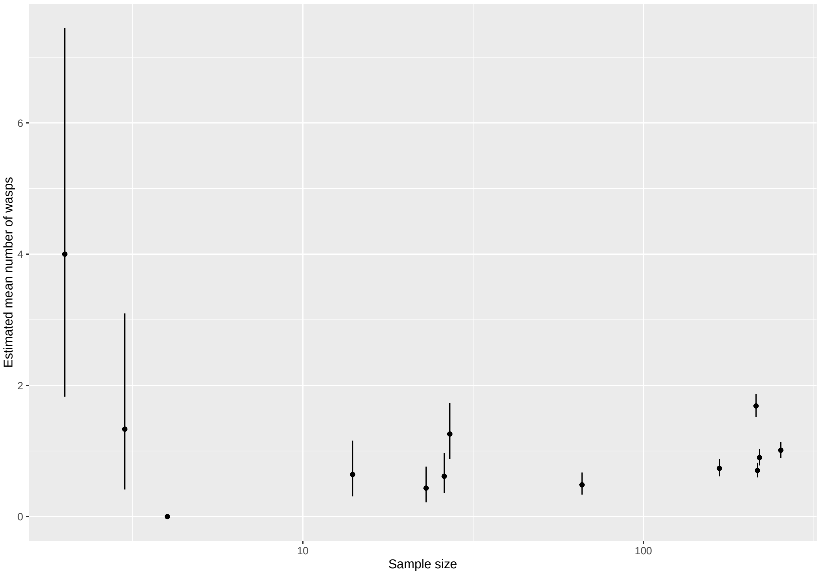
Uncertainty and sample size



Which estimates do we trust?



How can we improve estimates with small n?



Gall wasp hierarchical model

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$log(\lambda_i) = \alpha_0 + \alpha_{j[i]}$$

$$\alpha_j \sim Normal(0, \sigma_\alpha)$$

Parameter interpretation

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$log(\lambda_i) = \alpha_0 + \alpha_{j[i]}$$

$$\alpha_j \sim Normal(0, \sigma_\alpha)$$

Fitting a hierarchical model

```r
library(lme4)
partial_pool <- glmer(n_cynip ~(1 | gall_locality),
                data = d, family = poisson)
```

Understanding the model object

```
partial_pool
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplac
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: n_cynip ~ (1 | gall_locality)
##    Data: d
##       AIC       BIC    logLik  deviance  df.resid
##  4235.610  4245.846 -2115.805  4231.610      1232
## Random effects:
##  Groups        Name        Std.Dev.
##  gall_locality (Intercept) 0.4773
## Number of obs: 1234, groups:  gall_locality, 13
## Fixed Effects:
## (Intercept)
##     -0.1692
```

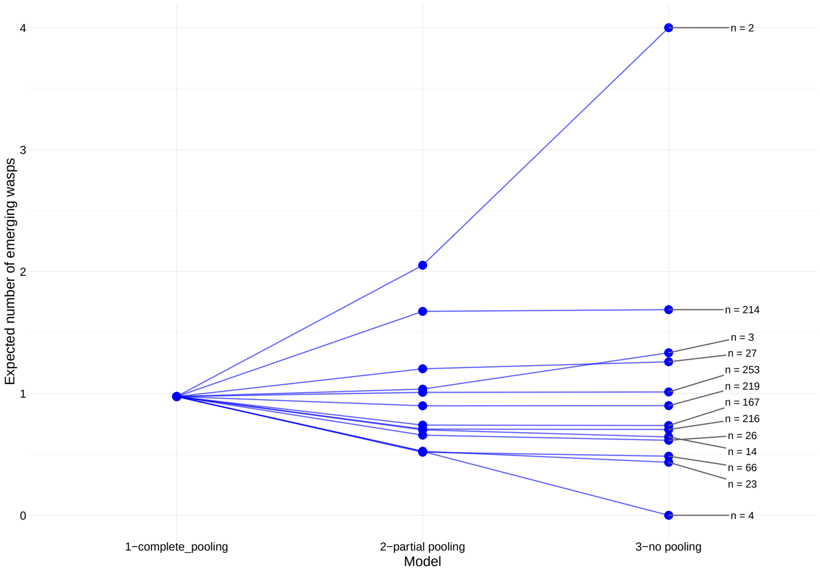
Is this a Bayesian model?

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$log(\lambda_i) = \alpha_0 + \alpha_{j[i]}$$

$$\alpha_j \sim Normal(0, \sigma_\alpha)$$

Comparing estimates: which estimates were shrunk?



Bayesian connections

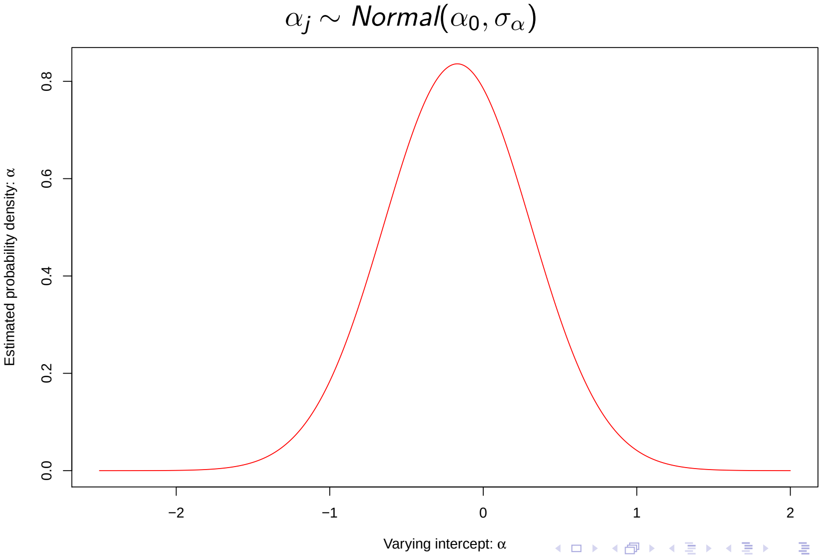
$$y_i \sim \text{Poisson}(\lambda_i)$$

$$log(\lambda_i) = \alpha_{j[i]}$$

$$\alpha_j \sim Normal(\alpha_0, \sigma_\alpha)$$

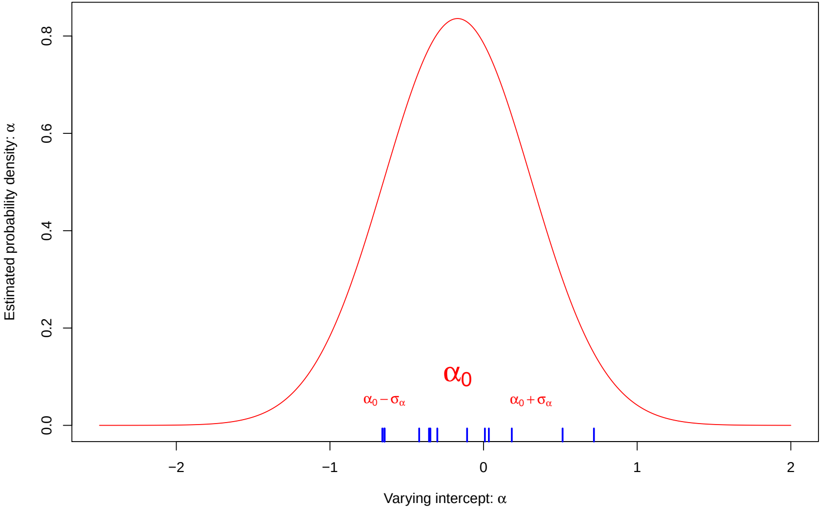
Bayesian connections

Estimated distribution of intercepts



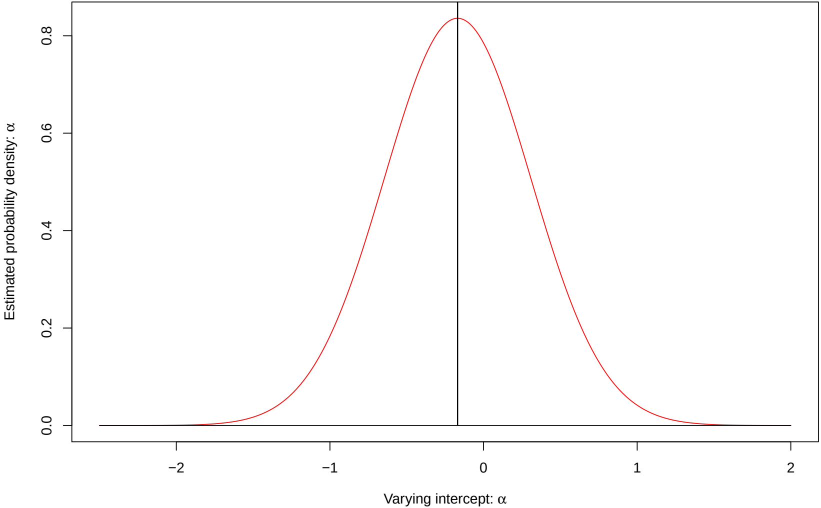
Bayesian connections

Estimated distribution of intercepts



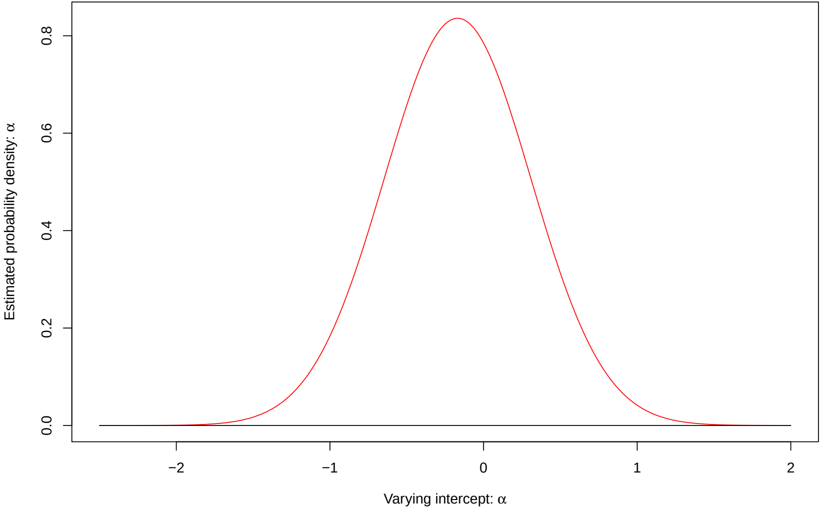
Bayesian connections

Complete pooling: $\sigma_\alpha \to 0$



Bayesian connections

No pooling: $\sigma_\alpha \to \infty$



Partial pooling: a reasonable compromise

Complete pooling: $\sigma_\alpha \to 0$
No pooling: $\sigma_\alpha \to \infty$
Partial pooling: $0 < \sigma_\alpha < \infty$

Hierarchical models

Why bother?

1. Shrinkage & partial pooling
   - sharing information among groups

Hierarchical models

Why bother?

1. Shrinkage & partial pooling
   - sharing information among groups

How many groups do we need to justify hierarchical modeling?

Hierarchical models

Why bother?

1. Shrinkage & partial pooling
2. Predictions for new groups

This week

Amniotes & free throws redux