Summarizing explained variance and partial pooling

Patrick Breheny

April 16

Introduction

- It is often helpful to understand and summarize how much explanatory power a model has
- For normally distributed outcomes, a natural definition of explanatory power is the proportion of variance explained (R^2)
- This concept can be extended to multilevel modeling, although it is worth discussing, as each level of the multilevel model can have its own R^2

Radon model

- Let's go back to our hierarchical model with uranium as a group-level predictor for radon levels:

$$Y_{ij} \sim \mathrm{N}(\alpha_j + \beta x_{ij}, \sigma_y^2)$$
$$\alpha_j \sim \mathrm{N}(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2)$$

- Here we have two models predicting two different quantities, each with its own distribution
- Certainly, it may be possible that we can explain a great deal of the variability among the $\{\alpha_j\}$, but little variation among the $\{Y_{ij}\}$, or vice versa

Definition

- The classical definition of the proportion of variance explained is

$$R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{RSS}_0},$$

where RSS and RSS_0 are the residual sums of squares in the model and the null (intercept-only) model, respectively:

$$\mathrm{RSS} = \sum_i (y_i - \hat{\mu}_i)^2$$

$$\mathrm{RSS}_0 = \sum_i (y_i - \bar{y})^2$$

- The same basic idea extends directly to multilevel models, although with the slight caveat that the equivalent to "y" is not directly observed

Code

Although possible to calculate directly in R, it is easier to just calculate the RSS for each iteration in JAGS:

```
## House level
for (i in 1:n) {
  y[i] ~ dnorm(a[cid[i]] + b*x[i], sigma[1]^(-2))
  r1[i] <- y[i] - a[cid[i]] - b*x[i]
}
RSS[1] <- sum(r1^2)

## County level
for (j in 1:J) {
  a[j] ~ dnorm(g[1] + g[2]*u[j], sigma[2]^(-2))
  r2[j] <- a[j] - g[1] - g[2]*u[j]
}
RSS[2] <- sum(r2^2)
```

Code

Then, in R,

```
rss0 <- function(x) crossprod(x-mean(x))

## Data level
1-mean(RSS[,1])/rss0(y)

## Group level
1-mean(RSS[,2])/mean(apply(a, 1, rss0))
```

Alternatively, we could work with variances (as opposed to sums)

Results: House level

Model	R^2
Identical	0.07
Independent	0.29
Independent (adj)	0.21
Multilevel	0.21

Note that the adjusted R^2 for the independent parameters model and the R^2 for the multilevel model are not required to be identical, although conceptually, they are making a similar sort of adjustment

Results: County level

- At the county level, our model explains 74% of the variability among county intercepts
- Uranium is far more successful at explaining variability among county intercepts than county-specific intercepts and floor are at explaining variability among individual houses
- This seems correct – compare slide 15, 3-19 notes with slide 7, 3-21 notes

Results: Varying-intercept, varying-slope model

For the varying-intercept, varying-slope model, we have three models and three values for R^2:

Level	R^2
House	0.23
County (intercept)	0.82
County (slope)	0.18

Introduction

- We have remarked several times that multilevel models are a way of striking a balance between an "identical parameters" model and an "independent parameters" model
- Often, it is helpful to quantify where, exactly, that balance is being struck; *i.e.*, to quantify the degree of pooling

Shrinkage factor

- Suppose $Y_i \sim N(\alpha_{j[i]}, \sigma_y^2)$, $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$, and let n_j denote the number of observations in group j
- Recall from our original lecture on the normal distribution that the posterior mean for α_j is

$$\bar{\alpha}_j = \frac{\omega_\alpha \mu_\alpha + n\tau_y \bar{y}_j}{\omega_\alpha + n\tau_y} = \lambda_j \mu_\alpha + (1 - \lambda_j)\bar{y}_j$$

$$\lambda_j = \frac{1/\sigma_\alpha^2}{1/\sigma_\alpha^2 + n/\sigma_y^2}$$

- The quantity λ_j, then, may be thought of as a "shrinkage factor" that equals 1 if a parameter is shrunk all the way to the "identical parameters" value and 0 if it equals the "independent parameters" estimate

Connections

- Note that

$$\frac{1/\sigma_\alpha^2}{1/\sigma_\alpha^2 + n/\sigma_y^2} = 1 - \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_y^2/n_j},$$

  which bears a striking resemblance to our formula for R^2
- This makes intuitive sense, as the more accurately our model fits the $\{\alpha_j\}$'s, the more heavily we would want to shrink outliers towards what the model predicts

Connections (cont'd)

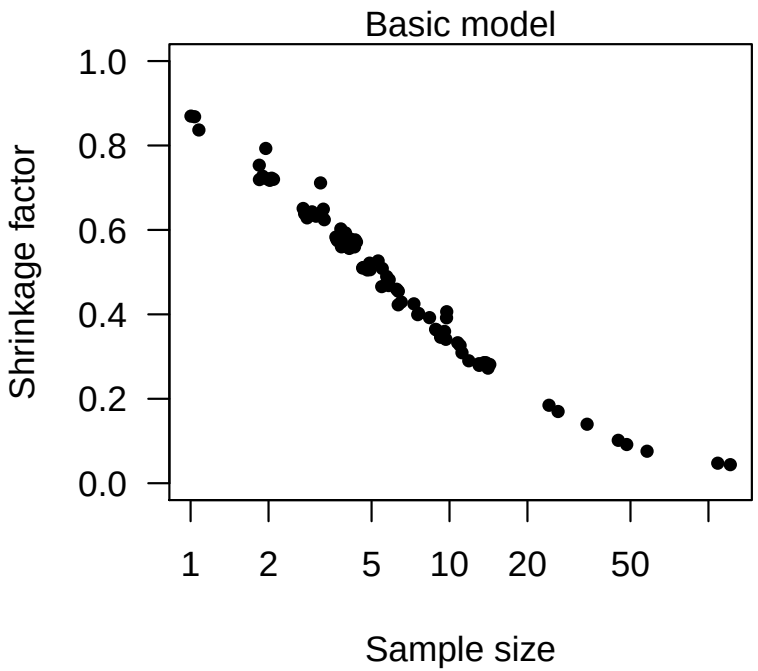
- Also note that

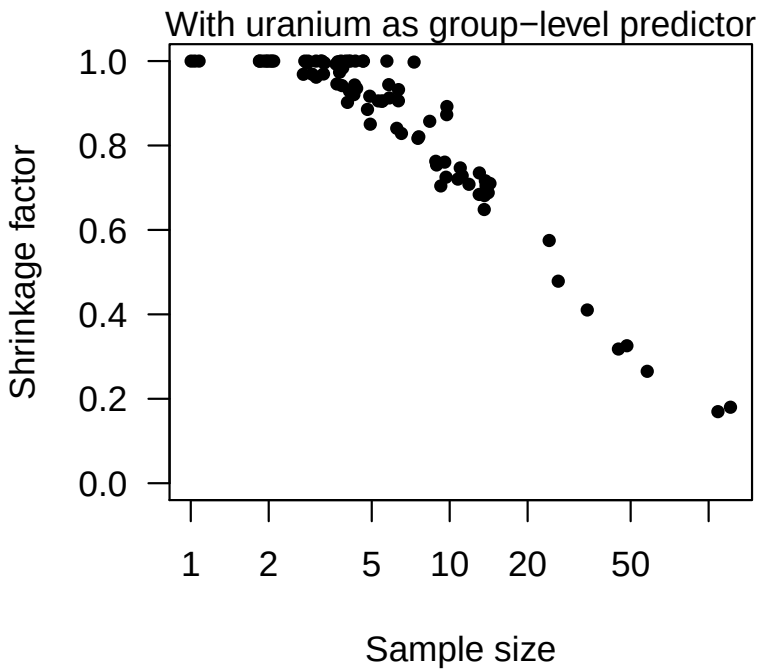
$$\lambda_j = \frac{\text{Var}(\alpha_j|\mathbf{y})}{\sigma^2_\alpha}$$

- In words, the shrinkage factor is equal to posterior variance divided by the group-level variability
- This is a convenient form to work with from a computational perspective:

```
lam <- apply(a, 2, var)/mean(sigma[,2]^2)
```

Shrinkage factors for two radon models





Overall shrinkage

- Finally, it is useful to summarize the individual-parameter shrinkage factors into a summary measure describing the overall extent of shrinkage at a given level of a model
- Let $\epsilon_j = \alpha_j - \hat{\alpha}_j$ denote the residual for the jth value of α; for a simple model, $\hat{\alpha}$ is simply μ_α, but group-level predictors could be involved
- Gelman and Pardoe (2006) define the following summary measure for the shrinkage factor:

$$\lambda = 1 - \frac{\text{Var}(\bar{\epsilon}_j)}{\text{Var}(\epsilon_j)}$$

Overall shrinkage: results

- For the basic model, $\lambda = 0.54$; for the model that includes uranium as a group-level predictor, $\lambda = 0.78$
- Certainly, both would seem to be reasonable summaries of the figures of slide 14
- It is worth noting that, like R^2 itself, these quantities are reasonably well-defined for linear models following normal distributions; extending these measures to generalized linear models, on the other hand, is not as straightforward