Hierarchical Dependence in Meta-Analysis: Methods

John R. Stevens
6 June 2008
SAMSI Program on Meta-Analysis
Outline

- Meta-analysis example: language learning
- Sampling dependence
- Hierarchical dependence
- Gene expression example: empirical results
- Future directions
Meta-analysis example

- L1: native language
- L2: non-native language
- gloss: reading aid
  (definitions in margins, etc.)
- 18 “study reports” from 13 research groups
  - each evaluated effect of L1 glossing on L2 reading comprehension (test scores)
  - compared “treatment” (glossing) to “control” (non-glossing) students
- fundamental differences:
  - test type: recall or multiple choice
  - test time: time limit or not
  - participant level: year of L2 study
  - percent: of text glossed

(Stevens and Taylor 2008 JEBS)
Effect size estimates

\[ \hat{\theta}_j = c_j \frac{\bar{Y}_{2,j} - \bar{Y}_{1,j}}{S_p} \]

\[ c_j = 1 - \frac{3}{4df_E - 1} \]
Meta-analysis notation

- Multiple studies report standardized estimates of the same “treatment effect”

\[ \hat{\theta}_j = c_j \frac{\bar{Y}_{2,j} - \bar{Y}_{1,j}}{S_p} \]

- Linear model

\[ \tilde{\theta} = X\beta + \delta + \varepsilon, \quad Var[\varepsilon] = V, \quad Var[\delta] = \tau^2 I \]

- Methods for estimating parameters \((\tau, \beta)\)
  - Fixed effects \((\tau^2 = 0)\)
  - Random effects (method of moments, or ML)
  - Hierarchical Bayes (prior on \(\tau\))

(Glass 1976 Educational Research; DerSimonian & Laird 1986 Controlled Clinical Trials; DuMouchel & Harris 1983 JASA; Cooper & Hedges 1994)
Sampling Dependence

- One study’s results in this format:

\[ Y_{ijk} = \mu + T_i + S_j + TS_{ij} + \varepsilon_{ijk} \]

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>S</td>
<td>2</td>
<td>.</td>
</tr>
<tr>
<td>TS</td>
<td>2</td>
<td>.</td>
</tr>
<tr>
<td>Error</td>
<td>83</td>
<td>57.68</td>
</tr>
<tr>
<td></td>
<td>88</td>
<td></td>
</tr>
</tbody>
</table>

- Only variance summary is in MSE
Sampling Dependence, cont’d.

- Here, $S_p = \sqrt{MSE}$
- For semesters $j$ and $h$:
  \[
  \hat{\theta}_j = c_E \frac{\bar{Y}_{2,j} - \bar{Y}_{1,j}}{\sqrt{MSE}}, \quad \hat{\theta}_h = c_E \frac{\bar{Y}_{2,h} - \bar{Y}_{1,h}}{\sqrt{MSE}}
  \]
- Can work out covariance (similar to variance calculation in Hedges 1981):
  \[
  V_{j,h} = \text{Cov}[\hat{\theta}_j, \hat{\theta}_h] \approx (p_E - 1) \cdot \hat{\theta}_j \cdot \hat{\theta}_h, \quad p_E = \frac{c_E^2 \cdot df_E}{df_E - 2}
  \]

(Hedges 1981 J. of Educ. Statistics; Stevens and Taylor 2008 JEBS)
Hierarchical dependence: delta-splitting

- When studies are related, split $\delta$:
  - study (researcher) component
  - substudy-within-study component

- Equivalently:
  \[ \tau^2 = \text{level of inter - study variability} = \text{hierarchical variance}\]
  \[ \zeta = \text{co - variability between related sub - studies} = \text{hierarchical covariance}\]
Hierarchical dependence in meta-analysis

\[ \Delta = \text{Var}[\delta] = \tau^2 I + \varsigma M = \begin{bmatrix} \tau^2 \\ \tau^2 & \varsigma & \varsigma & \varsigma \\ \varsigma & \tau^2 & \varsigma & \varsigma \\ \varsigma & \varsigma & \tau^2 & \varsigma \\ \varsigma & \varsigma & \varsigma & \tau^2 \\ \vdots \\ \tau^2 \end{bmatrix} \]

- Off-diagonal entries nonzero iff corresponding study reports are hierarchically dependent
- Estimate \( \tau^2 \) and \( \varsigma \)
  - Random Effects (iterative method of moments)
  - Hierarchical Bayes (priors on \( \tau \) and \( \varsigma \)
Random Effects: Startup

\[ \tilde{\theta} \sim N(X\beta, \Psi) \]

\[ \Psi = V + \Delta \]

\[ = V + (\tau^2 I + \zeta M) \]

\[ \hat{\beta}_{(k)} = (X^T \Psi^{-1}_{(k)} X)^{-1} X^T \tilde{\theta} \]

\[ RSS_{(k)} = (\tilde{\theta} - X\hat{\beta}_{(k)})^T \Psi^{-1}_{(k)} (\tilde{\theta} - X\hat{\beta}_{(k)}) \]

\[ = \tilde{\theta}^T A_{(k)} \tilde{\theta} \]

\[ A_{(k)} = \Psi^{-1}_{(k)} - \Psi^{-1}_{(k)} X \left( X^T \Psi^{-1}_{(k)} X \right)^{-1} X^T \Psi^{-1}_{(k)} \]

\[ E[RSS_{(k)}] = tr(A_{(k)} V) + \tau^2 \cdot tr(A_{(k)}) + \zeta \cdot tr(A_{(k)} M) \]
Random Effects: Iterative Procedure

1. \( \hat{\tau}_{(0)} = \hat{\zeta}_{(0)} = 0 \)

2. \( \Psi_{(k)} = V + \hat{\tau}_{(k)} I + \hat{\zeta}_{(k)} M \Rightarrow A_{(k)} \Rightarrow RSS_{(k)} \)

3. \( \hat{\tau}_{(k+1)} = \left\{RSS_{(k)} - tr(A_{(k)} V) - \hat{\zeta}_{(k)} \cdot tr(A_{(k)} M)\right\} / tr(A_{(k)}) \)

4. \( \Psi_{(k+1)} = V + \hat{\tau}_{(k+1)} I + \hat{\zeta}_{(k)} M \Rightarrow A_{(k+1)} \Rightarrow RSS_{(k+1)} \)

5. \( \hat{\zeta}_{(k+1)} = \left[RSS_{(k+1)} - tr(A_{(k+1)} V) - \hat{\tau}_{(k+1)}^2 \cdot tr(A_{(k+1)} M)\right] / tr(A_{(k+1)} M) \)

6. Iterate steps 2 - 5 to convergence of \( \hat{\tau}_{(k)} \) and \( \hat{\zeta}_{(k)} \) \( \Rightarrow \hat{\Psi} \)

7. \( \hat{\beta} = \left(X^T \hat{\Psi}^{-1} X \right)^{-1} X^T \hat{\Psi}^{-1} \hat{\theta}, \hat{\Sigma} = \left(X^T \hat{\Psi}^{-1} X \right)^{-1} \)

F = fixed effects; R = random effects (D-L); * = iterate ML
Random Effects: Numerical Constraints

- Dersimonian-Laird (1986): $\tau^2 \geq 0$

- Let $K =$ size of largest block on diagonal of $\Delta$:

$$
\begin{bmatrix}
\tau^2 & \varsigma \\
\vdots & \ddots \\
\varsigma & \tau^2
\end{bmatrix}
$$

  - $K$-by-$K$ compound-symmetric matrix will have positive determinant if

$$
- \tau^2 / (K - 1) \leq \varsigma \leq \tau^2
$$

  - A sufficient condition to make $\Delta$ (and $\Psi$) positive definite
Hierarchical Bayes: Startup

\[
\tilde{\theta} = X\beta + \delta + \varepsilon
\]
\[
\delta \sim N(0, \Delta)
\]
\[
\Delta(\tau, \zeta) = \tau^2 I + \zeta M
\]
\[
\varepsilon \sim N(0, V)
\]

\[
\beta \sim N(b, \text{diag}(d_1^2, \ldots, d_N^2))
\]
\[
\zeta | \tau \sim \pi^{(1)}(\zeta | \tau)
\]
\[
\tau \sim \pi(\tau)
\]

(DuMouchel & Harris 1983 JASA; DuMouchel & Normand 2000, in Stangl & Berry, Eds.)
Hierarchical Bayes: Priors

- Let $d_i \to \infty$  (diffuse prior on $\beta$)

- Right-skewed distribution for $\tau$
  - Empirical evidence for unif. distn. of variance ratio:
    \[ I^2 = \frac{\tau^2}{\left(\tau^2 + \hat{\sigma}_c^2\right)} \]
    ($\sigma_c^2$ is sampling variance common to all studies)
  - Contextual argument
  - log-logistic prior
    ($c_0$ harmonic mean of sampling variances)

(DuMouchel & Normand 2000, in Stangl & Berry, Eds.; Higgins et al. 2003 British Medical Journal; Stevens & Taylor 2008 JEBS)
Hierarchical Bayes: Priors and Estimation

- Non-informative prior based on numerical constraint:
  \[ \zeta | \tau \sim \text{Uniform}\left(\frac{-\tau^2}{K-1}, \tau^2\right) \]

- Estimation of interest (Simpson approx.):
  - Posterior means:
    \[ E[\beta | \tilde{\theta}], \ E[\tau^2 | \tilde{\theta}], \ E[\zeta | \tilde{\theta}] \]
  - Posterior probability
    \[ P(\beta_j > 0 | \tilde{\theta}), \quad P_j = 1 - 2\left|0.5 - P(\beta_j > 0 | \tilde{\theta})\right| \]

(DuMouchel & Normand 2000, in Stangl & Berry, Eds.; Stevens & Taylor 2008 JEBS; Louis & Zelteman 1994, in Cooper & Hedges, Eds.)
Sensitivity to Hierarchical Parameters

Random effects estimate of $\beta_0$
Random Effects vs. Hierarchical Bayes (conclusions from simulation study)

- 1. RE model can grossly overestimate $\tau^2, \varsigma$ (more likely when $\varsigma$ is close to zero)
- 2. Var. of $\beta_0$ estimates greater for RE (more so for larger $\beta_0$ and $\varsigma$)
- 3. RE est. of $\varsigma$ near interval endpoints (more at lower endpoint for larger $\tau^2$ and $\varsigma$)
- 4. (HB) Ignoring hierarchical dependence results in loss of power for testing $\beta_0$

(Stevens & Taylor 2008 JEBS)
Effect of delta-splitting

Effect Size Estimate +/- 2 SE

(HB) (Traditional) (Delta-split)
Effect of delta-splitting – exaggerated

(HB)
Comparison with Literature

- **Multivariate meta-analysis**
  - studies report estimates for multiple effect sizes (possibly correlated), such as effect of coaching on math SAT and verbal SAT scores

- **Hierarchical dependence**
  - studies report estimates of the same effect size (possibly at different covariate levels)
  - term initially used by Gurevitch and Hedges 1999 Ecology; example: experiments nested within laboratories → variance components model
Gene expression example

- **Experimental Autoimmune Encephalomyelitis (EAE)**
  - Mouse model for multiple sclerosis
  - Deterioration of covering (myelin) of nerve fibers
  - Impaired motor skills

- **Six gene expression study reports**
  - Differential expression from healthy (control) to EAE (treatment), based on signal log-ratio (SLR) as effect size
  - Differences in tissue site and mouse strain
  - Three study reports from one lab → hierarchical dependence

(microarrays) 21
Meta-analysis of gene expression studies

- **Effect size models**  
  (Choi et al. 2003 Bioinformatics; Stevens and Doerge 2005 BMC Bioinformatics; Conlon et al. 2006 BMC Bioinformatics; Conlon et al. 2007 BMC Bioinformatics)
  - Integrative correlations  
    (Parmigiani et al. 2004 Clinical Cancer Research)
  - Non-optimal and study-quality weights  
    (Feri et al. 2003 PNAS; Hu et al. 2005 BMC Bioinformatics)

- **Combining probabilities**  
  (Rhodes et al. 2002 Cancer Research; Ghosh et al. 2003 Func. and Int. Genomics; Shen et al. 2004 BMC Genomics; Choi et al. 2007 BMC Bioinformatics)

- **List-based approaches**  
  (Rhodes et al. 2004 PNAS; Pan et al. 2006 Bioinformatics; Hong et al. 2006 Bioinformatics; DeConde et al. 2006 Stat. Appl. in Gen. and Mol. Biol.)

- **Pooling raw data**  
  (Morris et al. 2003 CAMDA; Park et al. 2006 Bioinformatics)

- **Repositories**  
  (Stokes et al. 2008 BMC Bioinformatics)
Gene expression example: key results

Hierarchical Bayes results

Effect size est. using δ-splittings

Effect size est. from traditional meta-analysis

P-value using δ-splittings

P-value from traditional meta-analysis
Key summaries

- Log-logistic prior on $\tau$ looks [mostly] reasonable
- Non-informative uniform prior on $\varsigma | \tau$ could be modified based on empirical results
Future Directions

- Alternative priors
  - empirical evidence for both $\varsigma$ and $\tau$
- Alternative covariance structures
  - $\varsigma$ for each hierarchical group
- R package: metahep

Acknowledgements

- Alan M. Taylor, BYU-Idaho
- Cooperating source study authors
- USU New Faculty Research Grant
Standardized Estimates

\[ \zeta | \tau \sim \text{Uniform} \left[ -\frac{\tau^2}{K-1}, \tau^2 \right] \]

\[ -\frac{\tau^2}{K-1} \leq \zeta \leq \tau^2 \]

\[ -1 \leq \frac{\zeta}{\tau^2} \leq 1 \]

\[ -1 \leq (K-1) \zeta \leq K-1 \]

\[ 0 \leq 1 + \frac{(K-1)\zeta}{\tau^2} \leq K \]

\[ 0 \leq \frac{1}{K} + \frac{(K-1)\zeta}{K\tau^2} \leq 1 \]

\[ 0 \leq \frac{1}{K} + \left(1 - \frac{1}{K}\right) \frac{\zeta}{\tau^2} \leq 1 \]

\[ \pi(\tau) = \frac{c_0}{(c_0 + \tau)^2} \]

\[ c_0 = \sqrt{\frac{N}{\text{tr}\{\text{diag}(V)^{-1}\}}} = \text{harmonic mean of } V_{1,1}, \ldots, V_{N,N} \]

\[ Y = \log \tau \]

\[ \Rightarrow Y \sim \text{logistic(loc = log } c_0, sc = 1) \]

\[ \log \frac{\tau}{c_0} \sim \text{logistic(loc = 0, sc = 1)} \]
Gene Expression Technology: Crashcourse

Each gene has multiple probes (spots or features) on array; call this collection a probeset

mRNA from expressed genes in a sample hybridize to array for sample

(Color images courtesy affymetrix.com)
Array is scanned; spots with greater hybridization (mRNA) fluoresce more

For each probeset (or gene), look for systematic differences in fluorescence between control and treatment samples

General goal: find these differentially expressed genes

(Images courtesy affymetrix.com)
QQplots of standardized estimates

Random Effects

Hierarchical Bayes