Higher-Order Likelihood Inference in Meta-Analysis Using R

Annamaria Guolo and Cristiano Varin

University of Verona and Ca' Foscari University of Venice

annamaria.guolo@univr.it; sammy@unive.it

useR! 2011, University of Warwick

Outline

- Traditional meta-analysis and random-effects models
- First-order likelihood inference
- Higher-order asymptotics
  - second-order adjustment of the likelihood ratio statistic
  - meta-analysis and meta-regression problems
- R package `metaLik`
- Example
- Concluding remarks and open problems

Meta-analysis

- The process of combining and analyzing the results from K separate studies about the same issue of interest
- Aim: providing an overall estimation of a true effect β
- Typical in medical and epidemiological investigation
- Recent applications involve different areas of research, such as sociology, behavioral sciences, economics

Roberts (2005); Sutton & Higgins (2008)

Linear mixed-effects model for meta-analysis

- Y_i: measure of β from the i-th study (e.g., log-odds ratio)
- σ_i^2: within-study variance of the estimator of β
- Assumption: sample size of each study large enough to judge σ_i^2 as known, $\sigma_i^2 = \hat{\sigma}_i^2$
- Common meta-analysis model: linear random-effects model

$$Y_i = \beta_i + e_i,\ e_i \sim N(0, \hat{\sigma}_i^2)$$

with

$$\beta_i = \beta + \varepsilon_i,\ \varepsilon_i \sim N(0, \tau^2)$$

- e_i and ε_i independent; marginally, $Y_i \sim N(\beta, \hat{\sigma}_i^2 + \tau^2)$
- τ^2: variance component accounting for between-study heterogeneity

Meta-regression

- Extension of meta-analysis to include study-specific covariates
- Way of explaining sources of between-study heterogeneity
- X_i: vector of p covariates available at the aggregated meta-analysis level for study i, including the first value equal to one
- Meta-regression model

$$Y_i \sim N(X_i^\top \underline{\beta}, \hat{\sigma}_i^2 + \tau^2),$$

with $\underline{\beta}$ the fixed-effects p-dimensional vector

Thompson & Higgins (2002); Knapp & Hartung (2003)

- If X_i equal to one, $i = 1, \ldots, K$, the meta-regression model coincides with the meta-analysis model.

Standard approach to meta-analysis

- DerSimonian and Laird's (1986) approach
- Estimate of β as a weighted mean of Y_i

$$\hat{\beta}_{DL} = \frac{\sum_{i=1}^{K} Y_i/(\hat{\sigma}_i^2 + \hat{\tau}^2)}{\sum_{i=1}^{K} 1/(\hat{\sigma}_i^2 + \hat{\tau}^2)},$$

with $\hat{\tau}^2 = t$ (for $t > 0$), where

$$t = \frac{\hat{q} - (K-1)}{\sum_{i=1}^{K} \hat{\sigma}_i^{-2} - \sum_{i=1}^{K} \hat{\sigma}_i^{-4} / \sum_{i=1}^{K} \hat{\sigma}_i^{-2}}$$

and $\hat{\tau}^2 = 0$ otherwise. It is a biased estimate of τ^2.

- $var(\hat{\beta}_{DL}) = 1/\sum_{i=1}^{K}(\hat{\sigma}_i^2 + \hat{\tau}^2)^{-1}$
- Normal approximation of the distribution of the β estimator
- The method does not account for the uncertainty in estimating τ^2 $\rightarrow$ unreliable inferential conclusions
- Straightforward extension to meta-regression

First-order likelihood inference

- Whole parameter vector $\psi = (\theta, \lambda)^\top$
  - scalar interest component θ: one component of the fixed-effects vector $\underline{\beta}$
  - nuisance component λ: remaining elements of $\underline{\beta}$ plus τ^2
- For scalar θ, inference can rely on the signed profile log-likelihood ratio

$$r_P(\theta) = \text{sign}(\hat{\theta} - \theta)\sqrt{2\{\ell_\text{P}(\hat{\psi}) - \ell_\text{P}(\tilde{\psi})\}},$$

  - $\ell_P(\cdot)$: profile log-likelihood
  - $\hat{\psi} = (\hat{\theta}, \hat{\lambda})^\top$: MLE, not in closed-form in meta-analysis and meta-regression
  - $\tilde{\psi} = (\theta, \hat{\lambda}_{\hat{\theta}})^\top$: constrained MLE for a fixed θ
- Under mild regularity conditions, $r_P(\theta) \xrightarrow{d} N(0, 1)$, up to an error of order $O(n^{-1/2})$ Severini (2000)
- Questionable accuracy, especially in case of small sample sizes, as in meta-analysis or meta-regression problems

Higher-order inference

- Aim: improving the accuracy of the first-order asymptotic results

  Severini (2000); Reid (2003); Brazzale *et al.* (2007)

- Skovgaard's adjustment Skovgaard (1996)

$$\overline{r}_P(\theta) = r_P(\theta) + \frac{1}{r_P(\theta)} \log \frac{\overline{u}_P(\theta)}{r_P(\theta)},$$

where

$$\overline{u}_P(\theta) = [S^{-1}q]_\theta |\hat{j}|^{1/2} |\hat{i}|^{-1} |S| |\tilde{j}_{\lambda\lambda}|^{-1/2}$$

is a correction term, involving

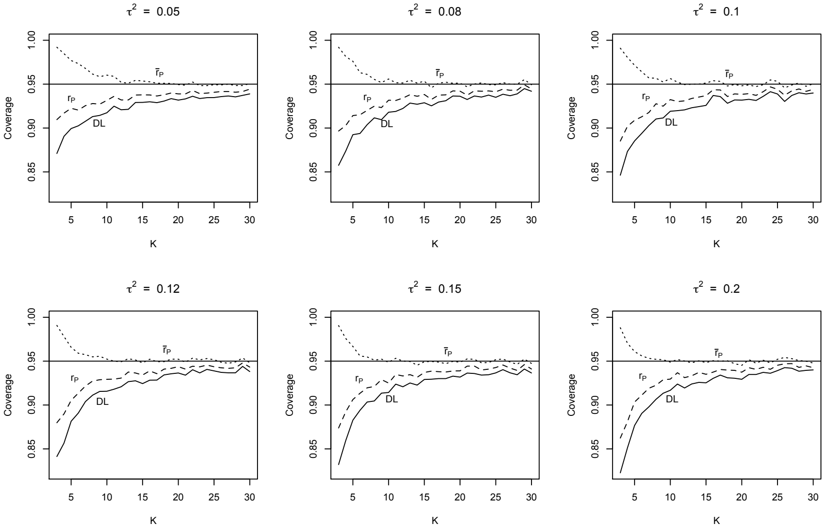
- $\hat{j}$: observed information matrix evaluated at $\hat{\psi}$;
- $\hat{i}$: expected information matrix evaluated at $\hat{\psi}$;
- $\tilde{j}_{\lambda\lambda}$: subblock of j corresponding to λ, evaluated at $\tilde{\psi}$;
- $[S^{-1}q]_\theta$: component of the vector $S^{-1}q$ corresponding to θ, where S and q are covariances of likelihood quantities.

Skovgaard's statistic $\bar{r}_P(\theta)$: properties

- $N(0,1)$ approximation up to an error of order $O(n^{-1})$
- Third-order accuracy in a full exponential family (unlikely, when $\hat{\sigma}_i^2 = \hat{\sigma}^2$)
- Well defined for a wide class of parametric models
- Invariant w.r.t. interest-respecting reparametrizations
- Components S and q with a compact form
- Complexity of components S and q similar to that of the expected information matrix
- Simulations studies indicate a performance superior to $r_P(\cdot)$ in terms of accuracy when approximating $N(0,1)$, under different scenarios.

Simulation study

Empirical coverages of confidence intervals; meta-analysis model with interest on $\beta = 0.5$; $10,000$ replications



metaLik package

- R package for likelihood inference in meta-analysis and meta-regression
- First-order likelihood inference based of the signed profile log-likelihood ratio statistic $r_P(\cdot)$ and its second-order Skovgaard's (1996) adjustment $\bar{r}_P(\cdot)$
- Comparison with DerSimonian and Laird's (1986) approach
- Hypothesis testing and confidence intervals for the fixed-effects components
- Extension to heterogeneity component τ^2 under development

Vaccine data

- $K = 13$ clinical studies on the efficacy of the Bacillus Calmette-Gurin (BCG) vaccine for preventing tuberculosis
  Berkey *et al.* (1995); Knapp & Hartung (2003)
- y_i: logarithm of the risk ratio in the i-th trial
- x_i: distance of the i-th study from the equator (latitude), surrogate for the presence of environmental mycobacteria providing a level of natural immunity against tuberculosis
- Meta-regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \varepsilon_i \sim N(0, \hat{\sigma}_i^2 + \tau^2)$$

- Inferential interest is mainly on β_1
- DerSimonian and Laird's test for $\beta_1 = 0$: -2.475 (p-value=0.013)

metaLik function

Implements first- and second-order likelihood inference

```
library(metaLik)
data(vaccine)
m <- metaLik(y~latitude, data=vaccine, sigma2=vaccine$sigma2)
m
Call:
metaLik(formula=y~latitude, data=vaccine, sigma2=vaccine$sigma2)

Coefficients:
(Intercept)     latitude        tau^2
   -0.30500     -0.01542      0.16756
Variance/covariance matrix:
             (Intercept)  latitude    tau^2
(Intercept)   5.020e-02  -1.101e-03  -2.629e-03
latitude     -1.101e-03   4.056e-05   3.600e-05
tau^2        -2.629e-03   3.600e-05   1.050e-02
Maximized log-likelihood:
[1]  1.121
```

Vaccine data analysis: summary

Adds information about the significance of the parameters

```
summary(m)
```

```
Likelihood inference in random effects meta analysis models

Call:
metaLik(formula=y~latitude, data=vaccine, sigma2=vaccine$sigma2)
Est. heterogeneity component tau^2: 0.1676 (std.err. 0.1025)
Fixed effects:
          estimate  std.err. signed logLRT  p-value  Skovgaard  p-value
(Intercept)-0.3051    0.2241        -1.3380   0.1809    -1.2246   0.2207
latitude   -0.0154    0.0064        -2.1202   0.0340    -1.8163   0.0693

Log-likelihood: 1.1212
```

test.metaLik function

Hypothesis testing on a scalar component of the fixed-effects vector, using $r_P(\cdot)$ and $\overline{r}_P(\cdot)$

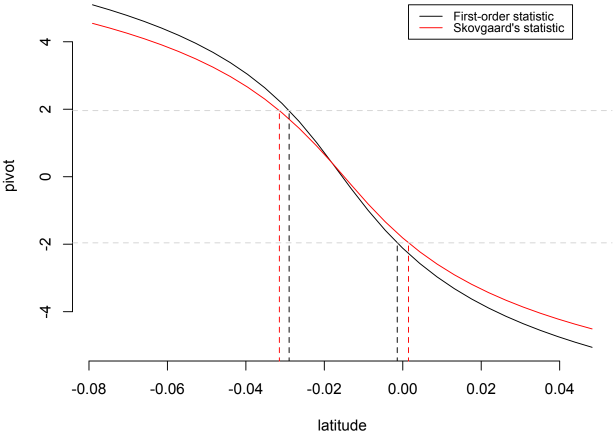
```
##Test on latitude coefficient
test.metaLik(m, param=2, value=0, alternative='less')

Signed profile log-likelihood ratio test for parameter latitude

First-order statistic
r:-2.12, p-value:0.01699
Skovgaard's statistic
rSkov:-1.816, p-value:0.03466
alternative hypothesis: parameter is less than 0
```

profile.metaLik function

```
##95% confidence interval for the latitude coefficient
profile.metaLik(m, param=2, level=0.95, plot=TRUE)
                    2.5%    97.5%
signed logLRT  -0.02897 -0.00141
Skovgaard      -0.03146  0.00146
```



Conclusions and open problems

`metaLik`:

- likelihood approach for meta-analysis and meta-regression
- inference on a scalar parameter of interest
- signed profile log-likelihood ratio statistic and Skovgaard's second-order statistic

Higher-order asymptotics:

- advantages over standard meta-analysis techniques
- superior to first-order results in terms of accuracy
- especially for small sample sizes

Open problems:

- between-study heterogeneity problems
- extension to generalized linear and to nonlinear mixed models
- empirical studies suggest Skovgaard's proposal improves on first-order solutions, but theoretical investigation is needed

Guolo, A. (2011). Higher-order likelihood inference in meta-analysis and meta-regression. *Submitted.*