Smooth Nonparametric Heterogeneity Estimation with an Application to Meta–Analysis

Dankmar Böhning

Southampton Statistical Sciences Research Institute
University of Southampton, UK

Brescia, 19 June 2013

Contents

Contents

scenario

$$\text{random quantity:} \quad Y_{i|\lambda_i,\sigma_i} \sim \frac{1}{\sigma_i}\phi((y-\lambda_i)/\sigma_i)$$

- $\phi(.)$ stand normal density
- observation-specific mean λ_i is **unknown**
- observation-specific variance σ_i^2 is **known**
- $i = 1, \cdots, n$
- distribution is **conditional** on observation-specific mean and variance

this arises in meta–analysis:

study	effect size y_i	σ_i
1	0.38	0.40
2	0.07	0.21
3	0.52	0.29
...	...	...

meta–analysis of set shifting ability

- 14 studies comparing "set shifting" ability (ability to move back and forth between different tasks) in people with eating disorders and healthy controls
- effect size: standardized difference
- positive effect sizes indicate greater deficiency in people with eating disorders
- Roberts *et al.* (2007 Psychol. Med.); Higgins, Thompson, Spiegelhalter (2009 JRSSA)

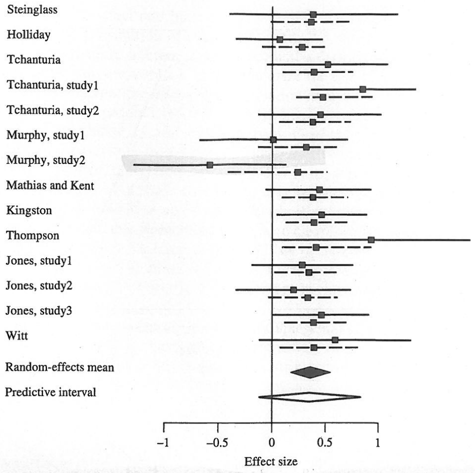


Fig. 2. Bayesian normal random-effects meta-analysis of the set shifting data: for each study the estimated effect size with 95% confidence interval (Table 1) and a posterior median with 95% credible interval are illustrated; 95% credible intervals for μ and for the predicted effect in a new trial, θ_{new}, are given

Contents

the model

- conditionally on i-th observation

$$Y_{i|\lambda_i,\sigma_i} \sim \frac{1}{\sigma_i}\phi((y-\lambda_i)/\sigma_i)$$

- assume that λ_i are realizations of a random quantity Λ_i

$$\Lambda_i \sim Q \text{ unknown}$$

  so that **marginally**

- $$Y_i \sim \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)Q(d\lambda)$$

the model

$$Y_i \sim f(y|Q, \sigma_i^2) = \int \frac{1}{\sigma_i} \phi((y - \lambda)/\sigma_i) Q(d\lambda)$$

- Q is the **mixing distribution**
- $\frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)$ **the component density**
- and $f(y|Q, \sigma_i^2)$ **the mixture distribution**

latent variable Λ with mixing distribution Q

$$Y_i \sim f(y|Q,\sigma_i^2) = \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)Q(d\lambda)$$

- Q is **parametric, continuous** e.g. normal

$$\frac{1}{\tau}\phi((\lambda-\mu)/\tau)$$

$$E(\Lambda)=\mu, Var(\Lambda)=\tau^2$$

- $Var(\Lambda)=\tau^2$ is **heterogeneity variance**

latent variable Λ with mixing distribution Q

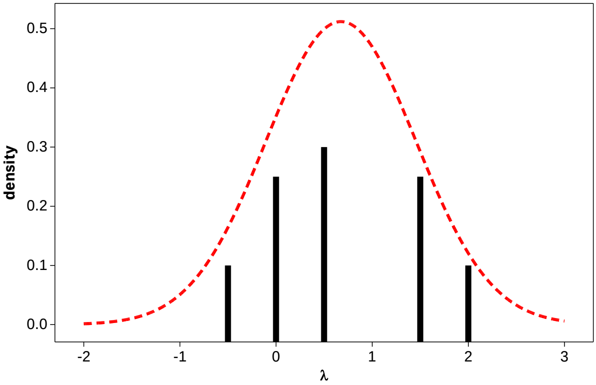
$$Y_i \sim f(y|Q,\sigma_i^2) = \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)Q(d\lambda)$$

- Q is **nonparametric, discrete**

$$\sum_{j=1}^{J} q_j \delta_{\lambda_j}$$

$$E(\Lambda) = \sum_{j=1}^{J} q_j\lambda_j, Var(\Lambda) = \sum_{j=1}^{J} q_j(\lambda_j - E(\Lambda))^2$$

- $Var(\Lambda) = \sum_{j=1}^{J} q_j(\lambda_j - E(\Lambda))^2$ is **heterogeneity variance**


density
0.5
0.4
0.3
0.2
0.1
0.0
-2
-1
0
1
2
3
λ

continuous vs. discrete

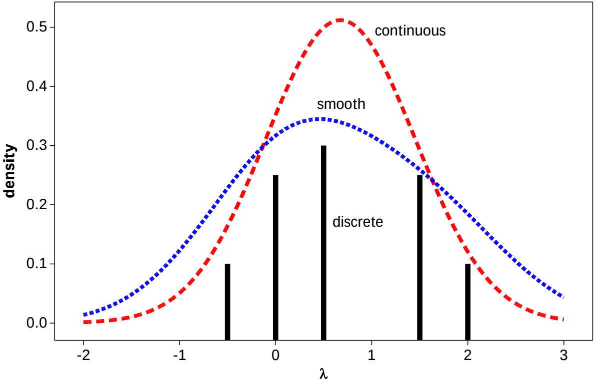
- pros and cons
- heated debates in the 90s
- discrete representing a **cluster-analytic** approach
- continuous (normal) representing a **trait-oriented** approach
- Higgins, Thompson, Spiegelhalter (2009, JRSSA)

compromise: smooth nonparametric mixing

$$Y_i \sim f(y|Q,\sigma_i^2) = \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)Q(d\lambda)$$

- choose mixing distribution Q as **smooth nonparametric**

$$\sum_{j=1}^{J} q_j \frac{1}{\tau}\phi((\lambda-\lambda_j)/\tau)$$


continuous
smooth
discrete
density
λ
0.0
0.1
0.2
0.3
0.4
0.5
-2
-1
0
1
2
3

compromise: smooth nonparametric mixing

$$Y_i \sim f(y|Q,\sigma_i^2) = \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)Q(d\lambda)$$

- choose mixing distribution Q as **smooth nonparametric**

$$\sum_{j=1}^{J} q_j \frac{1}{\tau}\phi((\lambda-\lambda_j)/\tau)$$

-

$$f(y|Q,\sigma_i^2) = \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)\sum_{j=1}^{J} q_j \frac{1}{\tau}\phi((\lambda-\lambda_j)/\tau)d\lambda$$

compromise: smooth nonparametric mixing

$$f(y|Q,\sigma_i^2) = \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)\sum_{j=1}^{J} q_j \frac{1}{\tau}\phi((\lambda-\lambda_j)/\tau)d\lambda$$

$$= \sum_{j=1}^{J} q_j \int \frac{1}{\sigma_i}\phi((y-\lambda)/\sigma_i)\frac{1}{\tau}\phi((\lambda-\lambda_j)/\tau)d\lambda$$

$$= \sum_{j=1}^{J} q_j \frac{1}{(\sigma_i^2+\tau^2)^{1/2}}\phi((y-\lambda_j)/(\sigma_i^2+\tau^2)^{1/2})$$

compromise: smooth nonparametric mixing

$$f(y|Q,\sigma_i^2) = \sum_{j=1}^{J} q_j \varphi(y|\lambda_j, \sigma_i^2 + \tau^2)$$

with $\varphi(y|\lambda, \sigma^2 + \tau^2)$ being the normal with mean λ and variance $\sigma^2 + \tau^2$

- $E(\Lambda) = \sum_{j=1}^{J} q_j \lambda_j$
- $Var(\Lambda) = \sum_{j=1}^{J} q_j [\lambda_j - E(\Lambda)]^2 + \tau^2$
- variance component model with
  - component for **clustering**
  - component for **continuous heterogeneity**

compromise: smooth nonparametric mixing

$$f(y|Q,\sigma_i^2) = \sum_{j=1}^{J} q_j \varphi(y|\lambda_j, \sigma_i^2 + \tau^2)$$

possible models

heterogeneity	parameter J	parameter τ^2
homogenous	1	0
continuous	1	> 0
discrete	≥ 1	0
smooth	≥ 1	> 0

Contents

Likelihood

$$L(Q,\tau^2) = \prod_{i=1}^{n} f(y_i|Q,\sigma_i^2)$$

$$= \sum_{j=1}^{J} q_j \varphi(y|\lambda_j,\sigma_i^2+\tau^2)$$

SNPMLE

$$L(\hat{Q},\hat{\tau}^2) \geq L(Q,\tau^2)$$

for all discrete p.d. Q and all $\tau^2 \geq 0$ is called **smooth nonparametric maximum likelihood estimator** (SNPMLE)

- Magder and Zeger (1996, JASA)
- Tokdar, Martin, Ghosh (2009, Ann. Statist.)

NPMLE

$$L(\hat{Q}, 0) \geq L(Q, 0)$$

for all discrete p.d. Q is called **nonparametric maximum likelihood estimator** (NPMLE)

- Lindsay (1983, Ann. Statist.); Böhning (1982, Ann. Statist.)
- Aitkin (1999, Biometrics)

MLE

$$L(\hat{\mu}, \hat{\tau}^2) \geq L(\mu, \tau^2)$$

for all τ^2 and μ (Q has just one component) is called **maximum likelihood estimator** (MLE)

- Hardy and Thompson (1996, Statist. Med.)
- Böhning *et al.* (2004, Biostatistics); Rukhin (2013, JRSSB)

parametric continuous

$$\log L(\mu, \sigma^2) = \sum_{i=1}^{n} \left(-\frac{1}{2} \log(\sigma_i^2 + \tau^2) - \frac{1}{2} \frac{(y_i - \mu)^2}{\sigma_i^2 + \tau^2} \right)$$

score equations lead to

$$\mu = \sum_i w_i y_i / \sum w_i$$
$$\tau^2 = \sum_i w_i^2 [(y_i - \mu)^2 - \sigma_i^2] / \sum w_i^2$$

where $w_i = 1/(\sigma_i^2 + \tau^2)$

- notice the appealing form since

$$E[(Y_i - \mu)^2 - \sigma_i^2] = \tau^2$$

smooth mixture

incomplete observed likelihood

$$L = \prod_i \sum_{j=1}^{J} f_{ij} q_j$$

complete unobserved likelihood

$$\ell = \prod_i \prod_j (f_{ij} q_j)^{z_{ij}} ,$$

where

$$Z_{ij} = \begin{cases} 1 & \text{if } Y_i \in \text{subpopulation } j \\ 0 & \text{otherwise} \end{cases}$$

$$f_{ij} = \frac{1}{(\sigma_i^2 + \tau^2)^{\frac{1}{2}}} \phi[(y - \lambda_j)/(\sigma_i^2 + \tau^2)^{\frac{1}{2}}]$$

EM for smooth mixture

- step 0. Choose initial values for λ_j, q_j, for $j = 1, \cdots, J$ and τ^2.
- step 1 (E-step). Compute

$$e_{ij} := E(Z_{ij} | Y_i = y_i) = \frac{f_{ij} q_j}{\sum_{j'} f_{ij'} q_{j'}}.$$

EM for smooth mixture

- step 2 (M-step).

$$\hat{q}_j = \sum_i e_{ij}/n$$

Iterate

  - (means.)

$$\hat{\lambda}_j = \frac{\sum_i e_{ij} w_i y_i}{\sum_i e_{ij} w_i}$$

  - (variance of the mixing distribution.)

$$\hat{\tau}^2 = \frac{\sum_i \sum_j e_{ij} w_i^2 [(y_i - \hat{\lambda}_j)^2 - \sigma_i^2]}{\sum_i \sum_j e_{ij} w_i^2}.$$

until convergence

Contents

design of simulation

- $n = 50, 100, 500$
- $\tau^2 = 0, 0.05, 0.15$
- various Q with $J = 1, 2, 3$

J=1

- $\mu = 0, 0.55$

J=2

"overlapping"

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.85 \end{pmatrix}, \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

"well–separated"

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} -1 \\ +1 \end{pmatrix}, \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

J=3

"overlapping"

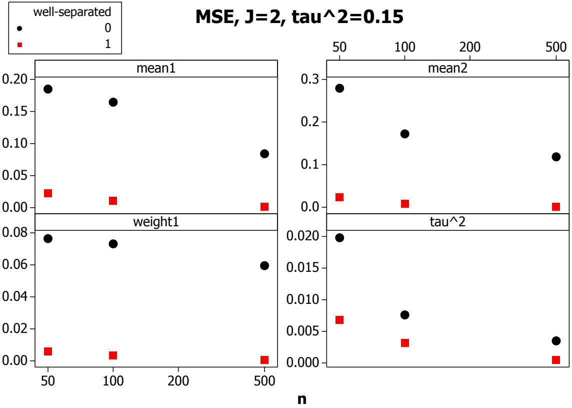
$$\begin{pmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{pmatrix} = \begin{pmatrix}-0.25\\ +0.25\\ +0.85,\end{pmatrix}, \begin{pmatrix}q_1\\ q_2\\ q_3\end{pmatrix} = \begin{pmatrix}0.4\\ 0.2\\ 0.4\end{pmatrix}$$

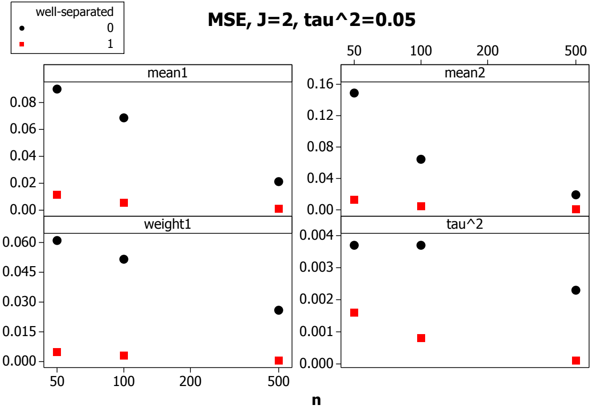
"well–separated"

$$\begin{pmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{pmatrix} = \begin{pmatrix}-1\\ +1\\ +2,\end{pmatrix}, \begin{pmatrix}q_1\\ q_2\\ q_3\end{pmatrix} = \begin{pmatrix}0.4\\ 0.2\\ 0.4\end{pmatrix}$$

results: parameter estimation

- $J = 2$
- $\tau^2 = 0.15, 0.00$
- "well-separated" and "overlapping"
- MSE


MSE, J=2, tau^2=0.15
well-separated
0
1
mean1
mean2
weight1
tau^2
0.20
0.15
0.10
0.05
0.00
0.3
0.2
0.1
0.0
0.08
0.06
0.04
0.02
0.00
0.020
0.015
0.010
0.005
0.000
50
100
200
500
n


well-separated
0
1
MSE, J=2, tau^2=0.05
mean1
mean2
weight1
tau^2
0.08
0.06
0.04
0.02
0.00
0.16
0.12
0.08
0.04
0.00
0.060
0.045
0.030
0.015
0.000
0.004
0.003
0.002
0.001
0.000
50
100
200
500
n

results: model selection, $J = 2$: well-separated components:

Model	AIC	BIC	AIC	BIC	AIC	BIC
n, τ^2	50, 0.15		50, 0.05		50, 0.00	
Smoothed	88.8	79.8	63.2	52.4	6.0	4.4
Discrete	6.4	7.6	36.8	46	94.0	95.6
Continuous	4.8	21.6	0.0	1.6	0.0	0.06
n, τ^2	100, 0.15		100, 0.05		100, 0.00	
Smoothed	99.6	95.2	90.4	84.4	10.8	8.8
Discrete	0.4	0.8	9.6	15.6	88.8	90.8
Continuous	0.0	4.0	0.0	0.0	0.4	0.4
n, τ^2	500, 0.15		500, 0.00		500, 0.00	
Smoothed	100	100	100	100	34.8	32.4
Discrete	0.0	0.0	0.0	0.0	65.2	67.6
Continuous	0.0	0.0	0.0	0.0	0.0	0.0

results: model selection, $J = 2$: overlapping components:

Model	AIC	BIC	AIC	BIC	AIC	BIC
n, τ^2	50, 0.15		50, 0.05		50, 0.00	
Smoothed	8.8	1.2	6.0	0.8	0.8	0.8
Discrete	24.0	15.2	42.8	27.2	78.8	54.8
Continuous	67.2	83.6	51.2	72.0	20.4	44.4
n, τ^2	100, 0.15		100, 0.05		100, 0.00	
Smoothed	12.8	0.4	10.8	0.4	2.0	0.8
Discrete	6.0	4.4	22.0	12.4	87.6	76.4
Continuous	81.2	95.2	67.2	87.2	10.4	22.8
n, τ^2	500, 0.15		500, 0.00		500, 0.00	
Smoothed	20.8	3.0	42.	4.4	4.4	2.8
Discrete	0.0	0.0	0.4	0.4	95.6	97.2
Continuous	79.2	97.0	57.6	95.2	0.0	0.0

results: heterogeneity variance

$$Var(\Lambda) = \sum_{j=1}^{J} q_j [\lambda_j - E(\Lambda)]^2 + \tau^2$$

$J = 2$: well-separated components:

Model	bias	sd	bias	sd	bias	sd
n, τ^2	50, 0.15		50, 0.05		50, 0.00	
Smoothed	-0.0096	0.1938	-0.0127	0.1522	-0.0087	0.1071
Discrete	-0.1833	0.3229	-0.0950	0.2526	-0.0509	0.2138
Continuous	0.0095	0.1977	-0.0116	0.1633	-0.0128	0.1442
n, τ^2	100, 0.15		100, 0.05		100, 0.00	
Smoothed	-0.0072	0.1293	-0.0101	0.0995	-0.0110	0.0669
Discrete	-0.1950	0.3516	-0.1451	0.2981	-0.0949	0.2739
Continuous	-0.0056	0.1338	-0.0078	0.1105	-0.0089	0.0977
n, τ^2	500, 0.15		500, 0.00		500, 0.00	
Smoothed	0.0023	0.0547	0.0013	0.0410	0.0001	0.0225
Discrete	-0.4989	0.4759	-0.3636	0.4576	-0.3150	0.4571
Continuous	0.0029	0.0564	0.0020	0.0460	0.0014	0.0402

$J = 2$: overlapping components:

Model	bias	sd	bias	sd	bias	sd
n, τ^2	50, 0.15		50, 0.05		50, 0.00	
Smoothed	0.0066	0.0939	0.0055	0.0680	0.0038	0.0546
Discrete	-0.0506	0.0977	-0.0177	0.0625	0.0046	0.0375
Continuous	0.0052	0.0927	0.0023	0.0661	0.0036	0.0453
n, τ^2	100, 0.15		100, 0.05		100, 0.00	
Smoothed	0.0071	0.0614	0.0046	0.0412	0.0012	0.0301
Discrete	-0.0629	0.0727	-0.0257	0.458	-0.0002	0.0228
Continuous	0.0064	0.0621	0.0034	0.0414	0.0023	0.0278
n, τ^2	500, 0.15		500, 0.00		500, 0.00	
Smoothed	0.0023	0.0277	0.0018	0.0180	0.0013	0.0084
Discrete	-0.0900	0.0688	-0.0374	0.0404	-0.0019	0.0164
Continuous	0.0019	0.0274	0.0013	0.0180	0.0009	0.0117

Contents

meta–analysis of set shifting ability

- 14 studies comparing ”set shifting” ability (ability to move back and forth between different tasks) in people with eating disorders and healthy controls
- effect size: standardized difference

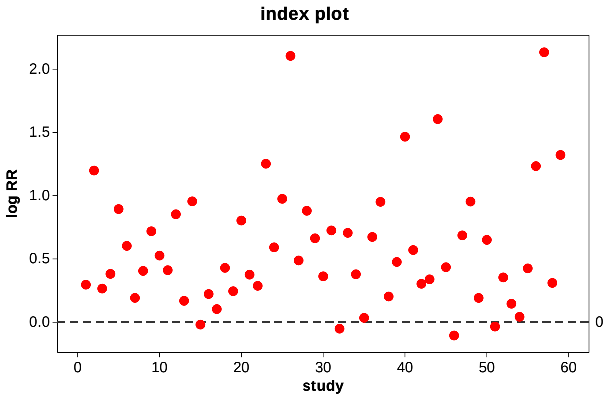
meta–analysis of set shifting ability

model	description	AIC	BIC
$J = 1, \tau^2 = 0$	complete homogeneity	9.9991	10.6382
$J = 1, \tau^2 > 0$	continuous heterogeneity no clustering	11.9652	13.2433
$J > 1, \tau^2 = 0$	no continuous heterogeneity but clustering	13.3226	15.2398
$J > 1, \tau^2 > 0$	continuous heterogeneity and clustering	18.5214	15.9652

- similar conclusion in Higgins, Thompson, Spiegelhalter (2009, JRSSA)

meta–analysis of NRT on quitting smoking

- 59 trials compared the effect of nicotine replacement therapy (NRT) on quitting smoking
- effect size: log-relative risk (risk in treatment group to risk in control group)
- positive effect size indicates success of NRT
- data are from DuMouchel and Normand (2000)


index plot
log RR
2.0
1.5
1.0
0.5
0.0
0
0
10
20
30
40
50
60
study

meta–analysis of NRT on quitting smoking

model	description	AIC	BIC
$J = 1, \tau^2 = 0$	complete homogeneity	59.1428	61.2204
$J = 1, \tau^2 > 0$	continuous heterogeneity no clustering	57.9919	62.1470
$J > 1, \tau^2 = 0$	no continuous heterogeneity but clustering	58.8649	65.0975
$J > 1, \tau^2 > 0$	continuous heterogeneity and clustering	61.9912	70.3014