Meta-Analysis of Diagnostic Studies by Means of (S)ROC-Modelling – a Profile-likelihood Approach based upon the Lehmann-family

Dankmar Böhning

Professor and Chair in Applied Statistics, School of Biological Sciences
University of Reading, UK

Vienna, Institute for Medical Statistics, 18 March 2009

Cooperation

Professor Dr. Heinz Holling
Statistics and Quantitative Methods
Faculty of Psychology and Sport Science
University of Münster, Germany

Support

German Research Foundation (DFG)

Often studies are done in medicine or psychology to determine:

discriminatory ability of a diagnostic test to separate people

- **with** a specific disease (or condition)
- from those **without**

Measures of Diagnostic Accuracy

- **Specificity**: $P(T-|D-) = 1-u$
  Probability of a negative test result for a healthy person
- **Sensitivity**: $P(T+|D+) = p$
  Probability of a positive test result for a diseased person

Estimating Diagnostic Accuracy

- **Specificity**: $P(\widehat{T-|D}-) = 1 - \hat{u} = \frac{n-x}{n}$
  where x are the number of false-positives out of n healthy individuals, $n - x$ are the true-negatives

- **Sensitivity**: $P(\widehat{T+|D}+) = \hat{p} = \frac{y}{m}$
  where y are the number of true-positives out of m healthy individuals, $y - m$ are the false-negatives

Frequently available:

- a variety of diagnostic studies
- providing diagnostic measures

$$x_i, n_i \text{ (specificity)}$$

$$y_i, m_i \text{ (sensitivity)}$$

- for $i = 1, \ldots, k$
- leading to the field of **meta-analysis**

An Example: Meta-Analysis of Diagnostic Accuracy of Natriuretic Peptides for Heart Failure

- diagnosis of heart failure is difficult
- overdiagnosis and underdiagnosis is occurring
- natriuretic peptides have been proposed as a diagnostic test
- meta-analysis provided by Doust *et al.* (2004) for brain natriuretic peptide (BNP)
- restriction on studies that use left ventricular ejection fraction of 40% or less as gold standard

Data of Meta-Analysis on Diagnostic Accuracy of BNP for Heart Failure

	diseased		healthy		
study	y(TP)	$m-y$(FN)	$n-x$(TN)	x(FP)	$n+m$
Bettenc. 2000	29	7	46	19	101
Choy 1994	34	6	22	13	75
Valli 2001	49	9	78	17	153
Vasan 2002a	4	6	1612	85	1707
Vasan 2002b	20	40	1339	71	1470
Hutcheon 2002	29	2	102	166	299
Landray 2000	26	14	75	11	126
Smith 2000	11	1	93	50	155

The Cut-off Value Problem

- Why not proceed with the **available armada of meta-analysis methods**?
- continuous or ordered categorical test uses **cut-off value**
- sensitivities and specificities from different studies **not comparable**
- different values for sensitivity and specificity might be due to different **diagnostic accuracy** or **different cut-off value**
- cut-off problem introduces **bias of unknown direction and size**

Illustration of the cut-off value problem for a **single study**:

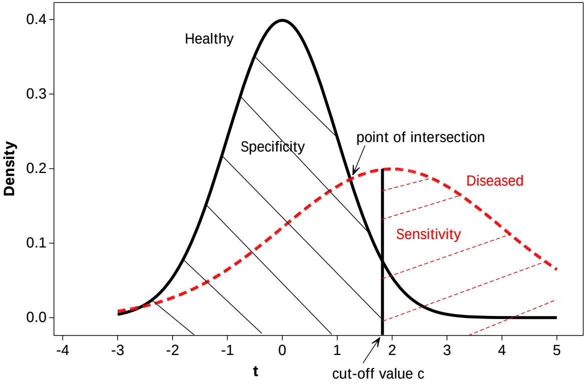

Healthy
Specificity
point of intersection
Diseased
Sensitivity
Density
t
cut-off value c
0.4
0.3
0.2
0.1
0.0
-4
-3
-2
-1
0
1
2
3
4
5

Illustration for a single study on Depression

- Lotrakul *et al.* (2008) seek to determine the diagnostic accuracy for the Thai version of the **Patient Health Questionnaire (PHQ-9)**
- a screening tool for major depression in primary care patients
- sensitivity and specificity were estimated in a diagnostic study involving 279 patients for different cut-off values
- Mini International Neuropsychiatric Interview and the Hamilton Rating Scale for Depressions were used as **gold standards**
- Lotrakul *et al.* (2008) consider different cut-off values and determine associated sensitivities and specificities

Illustration for a single study

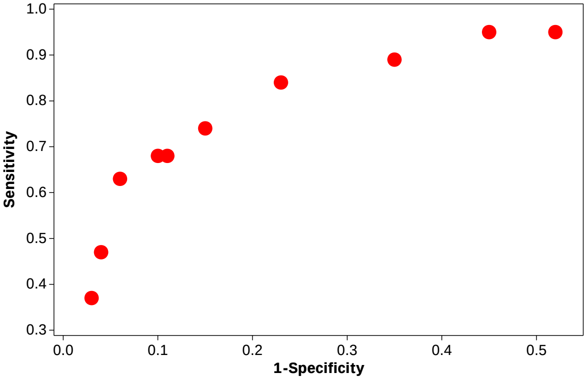
Table: *Performance of various PHQ-9 cut-off scores in detecting major depression (following Lotrakul* et al. *2008)*

Cut-off	Sensitivity	Specificity
6	0.95	0.48
7	0.95	0.55
8	0.89	0.65
9	0.84	0.77
10	0.74	0.85
11	0.68	0.89
12	0.68	0.90
13	0.63	0.94
14	0.47	0.96
15	0.37	0.97

Coping with the cut-off value problem for a single study: The ROC-curve

- let $\hat{p}_i$ and $\hat{u}_i$ be the different values of sensitivity and 1-specificity according to different cut-off values c_i, $i = 1, ..., k$
- construct a diagram with pairs $(\hat{p}_i, \hat{u}_i)$
- called the **Receiver Operating Characteristic (ROC)**
- benefit: **incorporates** the different values of the cut-off

Illustration of the cut-off value problem for a **single study**:


Sensitivity
1-Specificity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.1
0.2
0.3
0.4
0.5

The SROC-diagram for meta-analytic situations

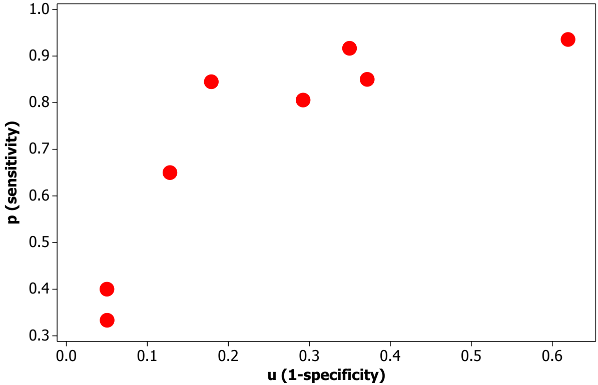
- Consider the pairs (sensitivity, 1-specificity) estimated by

$$(\hat{p}_i, \hat{u}_i) = (y_i/m_i, x_i/n_i)$$

  for $i = 1, ..., k$

- include them in a ROC diagram
- it is called **summary** ROC because the points relate to **different studies** instead of **different cut-off values**

SROC-diagram for MA of BNP and Heart Failure



Modelling of the SROC-diagram

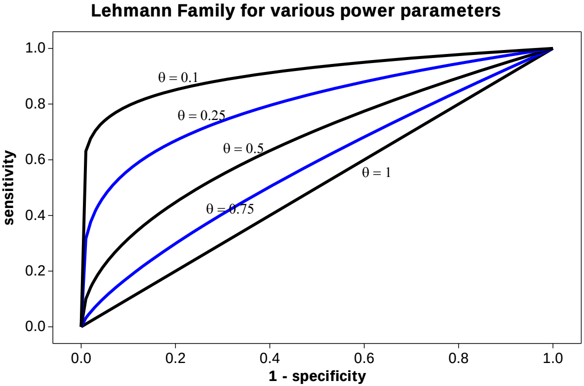
- Consider the **Lehmann family** for $\theta > 0$ and $i = 1, ..., k$ (Le 2006):

$$p_i = u_i^\theta$$

- or as a simple slope-only model

$$\log p_i = \theta \log u_i$$

- note model has **one parameter of interest** θ and k **nuisance parameters** $u_1, ..., u_k$
- note that θ represents the **diagnostic power** whereas the nuisance parameter captures heterogeneity in the specificities


Lehmann Family for various power parameters
θ = 0.1
θ = 0.25
θ = 0.5
θ = 0.75
θ = 1
sensitivity
1 - specificity
0.0
0.2
0.4
0.6
0.8
1.0

Inference

- consider the product-binomial likelihood as the joint distribution of Y_i and X_i for the $i-$th study (index is suppressed for notational convenience)

$$\binom{m}{y} p^y (1-p)^{m-y} \times \binom{n}{x} u^x (1-u)^{n-x}$$

- which we replace by the normal approximation for $\log Y_i$ and $\log X_i$

$$\frac{1}{\sqrt{2\pi s^2}} \exp\{-\frac{1}{2}\frac{(\log y - \log(mp))^2}{s^2}$$

$$\times \frac{1}{\sqrt{2\pi t^2}} \exp\{-\frac{1}{2}\frac{(\log x - \log(nu))^2}{t^2}$$

Inference

- the normal approximation for $\log Y_i$ and $\log X_i$

$$\frac{1}{2\pi st}\exp\{-\frac{1}{2}\frac{(\log y-\log(mp))^2}{s^2}\}\times\exp\{-\frac{1}{2}\frac{(\log x-\log(nu))^2}{t^2}\}$$

- with the Taylor-series variance estimates
  $s^2=\frac{1}{y}-\frac{1}{m}$ and $t^2=\frac{1}{x}-\frac{1}{n}$
- the normal approximation is justified if **the sizes per study are not small** and matches well with the Lehmann family
- consider now the log-likelihood for study i

$$-\frac{1}{2}\frac{(\log y-\log(mp))^2}{s^2}-\frac{1}{2}\frac{(\log x-\log(nu))^2}{t^2}$$

Inference

- and further with setting brackets differently

$$-\frac{1}{2s^2}(\log y - \log m - \log p)^2 - \frac{1}{2t^2}(\log x - \log n - \log u)^2$$

$$= -\frac{1}{2s^2}(\underbrace{\log y - \log m}_{z} - \log p)^2 - \frac{1}{2t^2}(\underbrace{\log x - \log n}_{w} - \log u)^2$$

$$= -\frac{1}{2s^2}(z - \overbrace{\theta \log u}^{\log p})^2 - \frac{1}{2t^2}(w - \log u)^2$$

Inference

- leading to the log-likelihood

$$\ell(\theta, u') = -\frac{1}{2s^2}(z - \theta u')^2 - \frac{1}{2t^2}(w - u')^2$$

- maximizing $\ell(\theta, u')$ in u' for **fixed** θ leads to

$$\hat{u}'_\theta = \frac{\theta t^2 z + s^2 w}{t^2\theta^2 + s^2}$$

- plugging $\hat{u}'_\theta$ in provides the **profile log-likelihood**

$$\ell(\theta) = \ell(\theta, \hat{u}'_\theta) = -\frac{1}{2s^2}(z - \theta\hat{u}'_\theta)^2 - \frac{1}{2t^2}(w - \hat{u}'_\theta)^2$$

Inference

- plugging $\hat{u}'_\theta$ in provides the **profile log-likelihood**

$$\ell(\theta) = \ell(\theta, \hat{u}'_\theta) = -\frac{1}{2s^2}(z - \theta\hat{u}'_\theta)^2 - \frac{1}{2t^2}(w - \hat{u}'_\theta)^2$$

  with $\hat{u}'_\theta = \frac{\theta t^2 z + s^2 w}{t^2\theta^2 + s^2}$

- ... **after some work** ... simplifies to

$$\ell(\theta) = \ell(\theta, \hat{u}'_\theta) = -\frac{1}{2}\frac{(z - w\theta)^2}{t^2\theta^2 + s^2}$$

  a profile log-likelihood of **remarkable simplicity**

Why profile likelihood?

- eliminates nuisance parameter
- **two forms** of the model:

$$\log p = \theta \log u \ \textbf{or} \ \log u = \frac{1}{\theta} \log p$$

- it is invariant if u or p chosen to be the nuisance parameter

$$\ell(\theta, \hat{u}'_\theta) = \ell(\theta, \hat{p}'_\theta)$$

- suitable for **symmetric** regression problems

Profile or Adjusted Profile Likelihood?

- $\ell(\theta)$ is **almost Gaussian**

$$\ell(\theta) = \ell(\theta, \hat{u}'_\theta) = -\frac{1}{2} \frac{(z - w\theta)^2}{\underbrace{t^2\theta^2 + s^2}_{\sigma^2(\theta)}}$$

- it differs **only** from

$$L(\theta) = -\frac{1}{2} \log \sigma^2(\theta) - \frac{1}{2} \frac{(z - w\theta)^2}{\sigma^2(\theta)}$$

by $\log \sigma^2(\theta)$

Profile or Adjusted Profile Likelihood?

- disadvantage of profile likelihood: it is **not** a likelihood
- hence, first and second order properties not necessarily valid
- in particular, it is thought that the curvature of the profile likelihood is **not** correct to give a valid variance estimate
- since the profile likelihood takes the estimated nuisance parameter as a true parameter value it is thought of **underestimating the variance** of the parameter of interest

Profile or Adjusted Profile Likelihood?

- but **adjustment factor** $\hat{I}(\hat{u}_\theta)^{-1/2}$ available (Cox and Reed 1987; Lee, Nelder Pawitan 2006; Murphy and van der Vaart 2000)

$$\hat{I}(\hat{u}_\theta) = -\frac{\partial^2}{\partial u'^2}\ell(\theta, u') = \frac{\partial^2}{\partial u'^2}\left(\frac{1}{2s^2}(z-\theta\hat{u}')^2 + \frac{1}{2t^2}(w-\hat{u}')^2\right)$$

- where, for fixed θ, $\hat{I}(\hat{u}_\theta)$ is the **observed Fisher information** $\hat{I}(u)$ evaluated at $\hat{u}_\theta$

Profile or Adjusted Profile Likelihood?

- as can be seen directly from above

$$\hat{I}(\hat{u}_\theta) = \frac{\partial^2}{\partial u'^2}\left(\frac{1}{2s^2}(z-\theta\hat{u}')^2 + \frac{1}{2t^2}(w-\hat{u}')^2\right) = \frac{t^2\theta^2+s^2}{s^2t^2}$$

- so that

$$-\frac{1}{2}\log[\hat{I}(\theta)] + \ell(\theta) = L(\theta)$$

- providing an **excellent** justification of the adjusted profile likelihood

Full Sample Profile Likelihoods

for a sample of k studies

- we have the **full-sample profile** log-likelihood

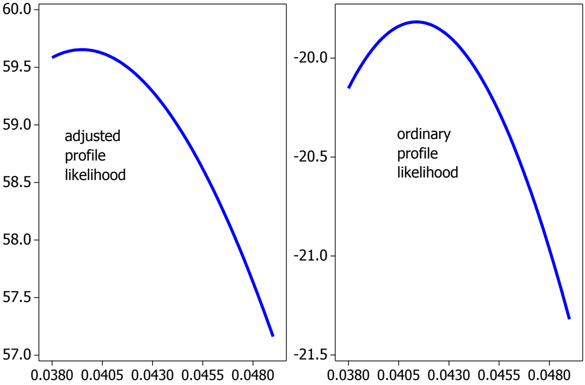
$$\ell(\theta) = -\sum_i \frac{1}{2} \frac{(z_i - w_i\theta)^2}{\sigma_i^2(\theta)}$$

- and the **full-sample adjusted profile** log-likelihood

$$L(\theta) = -\sum_i \frac{1}{2} \log \sigma_i^2(\theta) - \sum_i \frac{1}{2} \frac{(z_i - w_i\theta)^2}{\sigma_i^2(\theta)}$$

where $\sigma_i^2(\theta) = t_i^2\theta^2 + s_i^2$.

Ordinary and Adjusted Profile Likelihoods



Estimation: Maximum Profile Likelihood

- score for the ordinary profile likelihood

$$\frac{d}{d\theta}\ell(\theta) = -\frac{d}{d\theta}\sum_i \frac{1}{2}\frac{(z_i - w_i\theta)^2}{\sigma_i^2(\theta)}$$

$$= \sum_i \frac{(z_i - w_i\theta)w_i}{\sigma_i^2(\theta)} + \frac{1}{2}\frac{(z_i - w_i\theta)^2\sigma_i^2(\theta)'}{\sigma_i^4(\theta)}$$

- and the score for the adjusted profile likelihood

$$\frac{d}{d\theta}L(\theta) = \frac{d}{d\theta}\ell(\theta) - \frac{d}{d\theta}\sum_i \frac{1}{2}\log\sigma_i^2(\theta)$$

$$= \frac{d}{d\theta}\ell(\theta) - \frac{1}{2}\frac{\sigma_i^2(\theta)'}{\sigma_i^2(\theta)}$$

Estimating Equation Approach

- suggestion: fix θ in $\sigma_i^2(\theta)$ and maximize the Gaussian loss in θ:

$$-\sum_i \frac{(z_i - w_i\theta)^2}{\sigma_i^2(\theta)}$$

- or solve the estimating equation

$$\sum_i \frac{(z_i - w_i\theta)w_i}{\sigma_i^2(\theta)} = 0$$

- leading to the **iterative reweighted least-squares** approach:

$$\theta = \frac{\sum_i z_i w_i / \sigma_i^2(\theta)}{\sum_i w_i^2 / \sigma_i^2(\theta)}$$

Estimating Equation Approach

- neither ordinary nor adjusted profile likelihood is equivalent to IWLS
- but ... the latter is close because:
- look at the score for the **adjusted profile likelihood**

$$= \sum_i \frac{(z_i - w_i\theta)w_i}{\sigma_i^2(\theta)} + \frac{1}{2}\frac{\overbrace{(z_i - w_i\theta)^2}^{\sigma_i^2(\theta)}\sigma_i^2(\theta)'}{\sigma_i^4(\theta)} - \frac{1}{2}\frac{\sigma_i^2(\theta)'}{\sigma_i^2(\theta)}$$

$$\approx \sum_i \frac{(z_i - w_i\theta)w_i}{\sigma_i^2(\theta)}$$

- equals **estimating equation** approach

Simulation Study

- previous analysis suggests: profile and adjusted profile likelihood inference differs
- but how much? Look at **Bias and variance**!
- ... and
- how **valid** are the second derivate approximations of the true variances for both likelihoods

Fisher Information

- we developed before:

$$L(\theta) = \sum_i L_i(\theta) = -\sum_i \frac{1}{2} \log \sigma_i^2(\theta) - \sum_i \frac{1}{2} \frac{(z_i - w_i\theta)^2}{\sigma_i^2(\theta)}$$

-

$$\frac{d}{d\theta} L(\theta) = \sum_i \frac{d}{d\theta} L_i(\theta)$$

$$= \sum_i \frac{(z_i - w_i\theta)w_i}{\sigma_i^2(\theta)} + \frac{1}{2} \frac{(z_i - w_i\theta)^2 \sigma_i^2(\theta)'}{\sigma_i^4(\theta)} - \frac{1}{2} \frac{\sigma_i^2(\theta)'}{\sigma_i^2(\theta)}$$

- so that

$$\widehat{var(\hat{\theta})} = 1 / \sum_i \left(\frac{d}{d\theta} L_i(\hat{\theta}) \right)^2$$

Fisher Information

- Recall, if

$$U(\theta) = \frac{d}{d\theta} L(\theta) = \sum_i \frac{d}{d\theta} L_i(\theta) = \sum_i U_i(\theta)$$

- Fisher information

$$I(\theta) = E[U(\theta)^2] = \sum_i E[U_i(\theta)^2]$$

- which leads to the **plug-in estimate**

$$\hat{I}(\hat{\theta}) = \sum_i U_i(\hat{\theta})^2$$

Simulation Study: Design

for $i = 1, ..., k = 10$:

1. $u_i \sim U[0.05, .5]$
2. use model: $p_i = u_i^\theta$ for $\theta = 0.1, 0, 2, 0.3$
3. $n_i, m_i \sim Po(100)$ or $n_i, m_i \sim Po(10)$ (sparsity case)
4. $Y_i \sim Bin(p_i, m_i)$ and $X_i \sim Bin(u_i, n_i)$
5. determine various estimators of θ
6. replicate this process 1,000 times

Simulation Study: Results

Table: *Mean and Variance for Profile (PMLE), Adjust Profile (APMLE) and Iterative Weighted Least Squares (IWLS) Estimator*

estimator for $\theta = 0.1$	$E(\hat{\theta})$	$SE(\hat{\theta})$	$\widehat{SE(\hat{\theta})}$
$E(n_i) = E(m_i) = 100$			
IWLS	0.0961	0.0104	-
PMLE	0.0977	0.0104	0.0119
APMLE	0.0960	0.0101	0.0117
$E(n_i) = E(m_i) = 10$			
IWLS	0.0899	0.0291	-
PMLE	0.0981	0.0313	0.0561
APMLE	0.0812	0.0260	0.0468

Simulation Study: Results

Table: *Mean and Variance for Profile (PMLE), Adjust Profile (APMLE) and Iterative Weighted Least Squares (IWLS) Estimator*

estimator for $\theta = 0.2$	$E(\hat{\theta})$	$SE(\hat{\theta})$	$\widehat{SE}(\hat{\theta})$
$E(n_i) = E(m_i) = 100$			
IWLS	0.1959	0.0153	-
PMLE	0.1988	0.0153	0.0194
APMLE	0.1955	0.0151	0.0191
$E(n_i) = E(m_i) = 10$			
IWLS	0.1722	0.0499	-
PMLE	0.1917	0.0536	0.0838
APMLE	0.1597	0.0442	0.0654

Simulation Study: Results

Table: *Mean and Variance for Profile (PMLE), Adjust Profile (APMLE) and Iterative Weighted Least Squares (IWLS) Estimator*

estimator for $\theta = 0.3$	$E(\hat{\theta})$	$SE(\hat{\theta})$	$\widehat{SE}(\hat{\theta})$
$E(n_i) = E(m_i) = 100$			
IWLS	0.2953	0.0210	-
PMLE	0.3004	0.0211	0.0262
APMLE	0.2953	0.0208	0.0255
$E(n_i) = E(m_i) = 10$			
IWLS	0.2693	0.0694	-
PMLE	0.3011	0.0742	0.1137
APMLE	0.2517	0.0622	0.0869

Simulation Study: Results for small n but large k

Table: *Mean and Variance for Profile (PMLE), Adjust Profile (APMLE) and Iterative Weighted Least Squares (IWLS) Estimator* k*=100*

estimator for $\theta = 0.3$	$E(\hat{\theta})$	$SE(\hat{\theta})$	$\widehat{SE(\hat{\theta})}$
$E(n_i) = E(m_i) = 20$			
IWLS	0.2753	0.0153	-
PMLE	0.2970	0.0156	0.0189
APMLE	0.2718	0.0143	0.0164

Simulation Study: Conclusions

large n_i, m_i

- all three estimators behave similar
- minimal gain in efficiency with APMLE
- Fisher information estimate a bit conservative for variance estimation

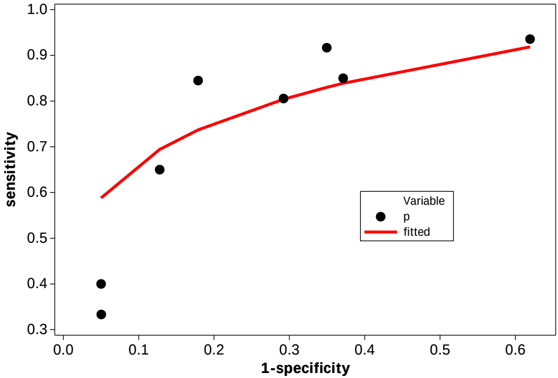
small n_i, m_i

- ordinary PMLE less biased
- APMLE more efficient
- Fisher information estimate overestimates variance of PMLE and APMLE

Application to BNP Meta-Analysis

- APMLE for $L(\theta)$ provides $\hat{\theta} = 0.1774$
- PMLE for $\ell(\theta)$ provides $\hat{\theta} = 0.1802$
- and IWLS gives $\hat{\theta} = 0.1755$

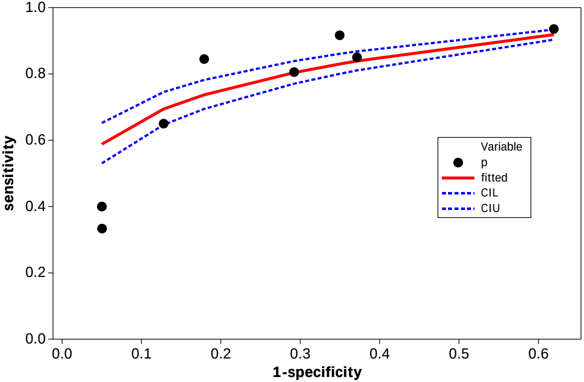
Observed and Fitted Lehmann Model



Fisher Information

- finally
$$\widehat{var(\hat{\theta})} = 1 / \sum_i \left(\frac{d}{d\theta} L_i(\hat{\theta}) \right)^2$$
- hence, 95% CI: $\hat{\theta} \pm 1.96 \sqrt{var(\hat{\theta})}$

Incorporating SE of Estimate



Goodness-of-Fit

- since $E(Z_i) = \theta \log u_i$ and $E(W_i) = \log u_i$
- it follows
$$E(Z_i - \theta W_i) = 0$$
- also $Var(Z_i) = s_i^2$ and $Var(\theta W_i) = \theta^2 t_i^2$
- hence
$$Var(Z_i - \theta W_i) = s_i^2 + \theta^2 t_i^2$$
- so that
$$\frac{Z_i - \theta W_i}{\sqrt{s_i^2 + \theta^2 t_i^2}} \sim N(0,1)$$

Goodness-of-Fit

χ^2- statistic

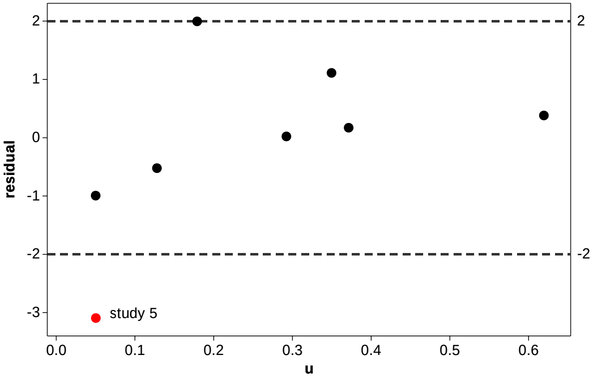
$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(Z_i - \hat{\theta} W_i)^2}{s_i^2 + \hat{\theta}^2 t_i^2}$$

BNP meta-analysis

based upon **all** 8 studies: $\chi^2_7 = 16.23$ and $P = 0.0231$
based upon 7 Studies (without **study 5**): $\chi^2_6 = 6.66$ and $P = 0.4655$ since plot of residuals:

$$\frac{Z_i - \hat{\theta} W_i}{\sqrt{s_i^2 + \hat{\theta}^2 t_i^2}}$$

provides evidence that study 5 is source of heterogeneity


2
1
0
-1
-2
-3
residual
2
-2
study 5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
u

Signposts on the Road Map

- mixed model approach to include residual heterogeneity as a further variance component
- nonparametric mixture approach to model unobserved heterogeneity
- classification of studies into different components of homogeneous diagnostic accuracy