Integrated likelihood inference in small sample meta-analysis

RUGGERO BELLIO

University of Udine, Italy

and

ANNAMARIA GUOLO

University of Verona, Italy

Abstract

This paper investigates likelihood-based inference in meta-analysis of studies characterized by a small sample size. In such a situation, a reliable approach has to properly account for the uncertainty associated to the estimation of the within-study variances, instead of assuming them to be known, as it is practiced in applications. To this aim, the integrated likelihood is suggested for inference on the mean effect of the meta-analysis. The method is shown to provide more accurate results with respect to standard likelihood assuming known within-study variances, while solving many of the computational problems related to variance components estimation. The proposed methodology is illustrated via simulation and applied to meta-analysis studies in nutritional science and psychological medicine.

Key words: Frequentist inference; Integrated likelihood; Meta-analysis; Small sample inference; Unrelated parameters.

Running title: Integrated likelihood in meta-analysis

1 Introduction

Meta-analysis is a diffuse approach to combine evidence from different studies about the same issue of interest. The usage of meta-analysis pervades almost any area of research, such

as, for example, biological sciences, medicine, epidemiology and, more recently, economics and behavioral investigations (Roberts, 2005; Sutton & Higgins, 2008).

Meta-analysis is typically performed by specifying an appropriate random effects model, with the random component associated to the different studies providing information about the common issue of interest. Inference is then carried out by relying on the procedure by DerSirmonian & Laird (1986), traditionally, or on more recent likelihood approaches developed either from a frequentist or a Bayesian perspective (van Houwelingen et al., 2002). The methods assume that the within-study variances are known and equal to the variance associated to the estimate of the mean effect reported in each study (van Houwelingen et al., 2002, Section 3). The justification is that the sample size of each study is large enough to guarantee a good estimate of the true within-study variance, with little or no impact on the inferential results. Actually, such an assumption can be justifiable in case of large studies, as, for example, those carried out in medical or epidemiological investigations. Conversely, inference performed on small sample studies can provide misleading results, if the uncertainty related to variance estimation is not properly taken into account. This issue may be referred to as *small sample inference*. Several authors pointed out the relevance of the problem, e.g., Hardy & Thompson (1996), Brockwell & Gordon (2001), Sidik & Jonkman (2007), with the suspicion that consequences could affect the variance estimator of the mean effect and related inferential procedures. Nevertheless, no solution to this problem has been provided within the meta-analysis context, to the best of our knowledge.

A similar problem has been recently faced, instead, by Sharma & Mathew (2011) within a different setting, namely, when inferential interest concerns the consensus mean in inter-laboratory studies. Although Sharma & Mathew (2011) never directly refer to meta-analysis and related terminology in their investigation, the framework they focus on can be considered as analogous. For the purpose on investigation on the consensus mean inter-laboratory studies, Sharma & Mathew (2011) take account of the measurements of different within-studies laboratories and propose to improve on likelihood results by applying higher-order

asymptotics via Skovgaard's (1996) second-order likelihood ratio statistic. Nevertheless, their approach may suffer from some computational problems, as illustrated in this paper, requiring a lot of care for its application.

In this paper we consider the problem of small sample in meta-analysis and meta-regression and propose to perform inference through integrated likelihood (Severini, 2000, Section 8.4). Such an approach replaces the elimination of the nuisance parameters represented by the variance components through maximization with their elimination by integration. We show that this method provides a good accuracy of inferential results and, in the meanwhile, it is free of numerical pitfalls.

The plan of the paper is as follows. Section 2 introduces two motivating examples from real meta-analyses in nutritional science and psychological medicine. In Section 3, the model of interest is presented, along with likelihood inferential methods. The integrated likelihood approach for meta-analysis and meta-regression in case of small sample is illustrated in Section 4. Results from simulation studies are summarized in Section 5, while results with empirical examples are given in Section 6. The case of binary data, which requires a specific treatment, is discussed in Section 7. Some final remarks conclude the paper.

2 Examples

This section introduces two motivating examples from meta-analysis studies in nutritional science and psychological medicine, which will be used to illustrate the applicability of the integrated likelihood approach.

2.1 Cocoa intake and blood pressure reduction

Increasing consumption of sources of polyphenols is recommended by physicians as coadjutant therapy to face hypertension and prevent cardiovascular risks. While the protective effects of polyphenols in fruits and vegetables is known, less attention is paid to other

sources, such as, for example, cocoa and tea, although they represent a high proportion of total polyphenol intake in Western countries. Taubert et al. (2007) perform a meta-analysis of randomized controlled studies to evaluate blood pressure-lowering effects of cocoa and tea intake. We focus on a portion of the data about the effectiveness on lowering diastolic blood pressure after two-weeks of cocoa consumption treatment. Data refer to five studies, with sample size ranging from 21 to 41. The estimate of the effect provided by each study is the mean difference in diastolic blood pressure before and after the cocoa consumption treatment. The forest plot of the data is reported in Figure 1, left panel. Information about the estimated mean difference from each meta-analysis study and the associated 95% confidence intervals is reported. The summary estimate obtained from the likelihood analysis is shown as well.

Figure 1 here

2.2 Set-shifting ability in eating disorders

Deficits in executive functions have been associated to eating disorders in psychological medicine. An important executive function is set-shifting, which is defined as the ability to move back and forth between multiple tasks or mental sets. Roberts et al. (2007) perform a meta-analysis to compare set-shifting ability in people with eating disorders and healthy controls. The information provided by each study is the standardized difference in means of the performance on set-shifting tasks between people with eating disorders and healthy controls. We refer to the meta-analysis data examined by Roberts et al. (2007) about the trail making test set-shifting task, provided within the `R` package `metamisc` (Debray, 2013). The data refer to 14 studies with sample size ranging from 20 to 94 and include information about the outcome from each study and the associated standard error. We extend the dataset by including an indicator variable, distinguishing between the type of eating disorders, namely, those referring to anorexia and those also including bulimia nervosa. The forest plot of the data is reported in Figure 2, left panel. Information about the estimated standardized

mean difference from each meta-analysis study and the associated 95% confidence intervals is reported. The summary estimate obtained from the likelihood analysis is shown as well.

Figure 2 here

3 Likelihood inference

Consider a meta-analysis of n independent studies about a common effect β. Let Y_i be the summary measure of the effect β obtained from the i-th study, $i = 1, \ldots, n$, such as, for example, the mean difference. The classical model for meta-analysis is the random effects model (DerSirmonian & Laird, 1986)

$$Y_i = \beta_i + \varepsilon_i,\ \varepsilon_i \sim \text{Normal}(0, \sigma_i^2),$$

where β_i is the random effects component associated to each study,

$$\beta_i = \beta + u_i,\ u_i \sim \text{Normal}(0, \tau^2).$$

Variance components are the within-study variances σ_i^2, $i = 1, \ldots, n$, and the between-study variance τ^2. Thus, marginally, $Y_i \sim \text{Normal}(\beta, \sigma_i^2 + \tau^2)$. The traditional approach to meta-analysis is based on the assumption that each within-study variance σ_i^2 is known and equal to the variance estimate reported in the i-th study. Such an assumption can be justifiable when the sample size of each study included in the meta-analysis is large. Otherwise, inference can provide misleading results, if the uncertainty related to the variance estimation is not properly taken into account. Let S_i^2 denote the measure of the within-study variance σ_i^2 obtained from study i having sample size n_i, with S_i^2 following a scaled chi-square distribution, $S_i^2(n_i - 1)/\sigma_i^2 \sim \chi^2_{n_i - 1}$. According to the specifications above, the log likelihood function for the $(n+2)$-dimensional parameter vector $\boldsymbol{\theta} = (\beta, \tau, \sigma_1, \ldots, \sigma_n)^T$ is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ -\frac{1}{2}\log(\sigma_i^2 + \tau^2) - \frac{1}{2}\frac{(y_i - \beta)^2}{\sigma_i^2 + \tau^2} - \frac{n_i - 1}{2}\log\sigma_i^2 - \frac{(n_i - 1)s_i^2}{2\sigma_i^2} \right\}. \qquad (1)$$

Inferential interest is usually focused on the mean effect β, while variance components are considered as nuisance parameters. Accordingly, we can partition $\boldsymbol{\theta}$ into $\boldsymbol{\theta} = (\beta, \boldsymbol{\lambda})^T$, where $\boldsymbol{\lambda} = (\tau, \sigma_1, \ldots, \sigma_n)^T$. Let $\hat{\boldsymbol{\theta}} = (\hat{\beta}, \hat{\boldsymbol{\lambda}})^T$ denote the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ and let $\hat{\boldsymbol{\lambda}}_\beta$ denote the constrained MLE of $\boldsymbol{\lambda}$ for a given value of β, so that $\tilde{\boldsymbol{\theta}} = (\beta, \hat{\boldsymbol{\lambda}}_\beta)^T$. Let $\ell_{\mathrm{P}}(\beta)$ indicate the corresponding profile log likelihood for β, $\ell_{\mathrm{P}}(\beta) = \ell(\beta, \hat{\boldsymbol{\lambda}}_\beta)$. Given the scalar parameter of interest, inference can be based on the signed profile log likelihood ratio statistic

$$r_{\mathrm{P}}(\beta) = \mathrm{sgn}(\hat{\beta} - \beta)\sqrt{2\left\{\ell_{\mathrm{P}}(\hat{\beta}) - \ell_{\mathrm{P}}(\beta)\right\}},$$

which is asymptotically distributed as a standard normal up to first-order error, under mild regularity conditions (Severini, 2000, Section 4.4).

Despite the feasibility, a serious drawback of first-order asymptotic results is that they can be inaccurate in case of small sample size or large dimension of the nuisance parameter $\boldsymbol{\lambda}$. To face the problem, it is preferable to resort to the theory of higher-order asymptotics (Severini, 2000). Skovgaard (1996) proposes to base inference on a scalar component of interest on statistic

$$r^*_{\mathrm{P}}(\beta) = r_{\mathrm{P}}(\beta) + \frac{1}{r_{\mathrm{P}}(\beta)}\log\frac{u(\beta)}{r_{\mathrm{P}}(\beta)}, \tag{2}$$

which is asymptotically standard normally distributed up to second-order error. The component $u(\beta)$ included in (2) is a function of the observed and the expected information matrices and of covariances of likelihood quantities, evaluated at the MLE and the constrained MLE. From a practical point of view, the computation of $r^*_{\mathrm{P}}(\beta)$ is not involved, since it is comparable to that of the expected information matrix. From a theoretical point of view, Skovgaard's statistic is well defined for a wide class of sufficiently regular parametric models and it is invariant with respect to interest-respecting re-parameterizations. Guolo (2012) investigates the applicability of Skovgaard's statistic in meta-analysis and meta-regression problems, following the convention of assuming known within-study variances. The approach is shown to be satisfactory in improving on the accuracy of standard first-order likelihood analysis when the sample size n is small to moderate. The method is implemented in the `R` (R Core Team,

2014) package `metaLik` (Guolo & Varin, 2012).

Sharma & Mathew (2011) examine the performance of Skovgaard's statistic in inter-laboratory studies where interest relies on the consensus mean, assuming unknown different within-laboratory variances. The simulation studies performed highlight a better accuracy of results based on r_{P}^* with respect to its first-order counterpart. Nevertheless, Skovgaard's approach can suffer from several computational difficulties and numerical instabilities, as Sharma & Mathew (2011) mention. We will return on this issue in Section 5.

Severini (2007) proposes the integrated likelihood approach to perform inference on an interest scalar parameter, when the dimension of the nuisance component is large compared to the sample size. The integrated likelihood eliminates the nuisance parameter by integration of the likelihood function with respect to a prior density, instead of by maximization. With reference to the meta-analysis framework, let $\pi(\boldsymbol{\lambda}|\beta)$ denote a conditional prior density for $\boldsymbol{\lambda}$ given β and let $L(\boldsymbol{\theta}) = \exp\{\ell(\boldsymbol{\theta})\}$. Thus, the integrated log likelihood function for β is

$$\ell_{\mathrm{Int}}(\beta) = \log \int_{\Lambda} L(\boldsymbol{\theta})\pi(\boldsymbol{\lambda}|\beta)d\boldsymbol{\lambda},$$

where integration is with respect to vector $\boldsymbol{\lambda}$, with $\boldsymbol{\lambda} \in \Lambda$. Once the integrated log likelihood is obtained, it can be used as a standard log likelihood function for inference. For example, let $\bar{\beta}$ be the estimate of β obtained from the maximization of $\ell_{\mathrm{Int}}(\beta)$. Then, inference on β can be performed via the signed integrated log likelihood ratio statistic (Severini, 2010)

$$r_{\mathrm{Int}}(\beta) = \mathrm{sgn}(\bar{\beta} - \beta)\sqrt{2\left\{\ell_{\mathrm{Int}}(\bar{\beta}) - \ell_{\mathrm{Int}}(\beta)\right\}}. \tag{3}$$

Advantages of the integrated likelihood approach include better accuracy of the inferential results if compared with those from r_{P} as well as reduced numerical instabilities in case of large dimension of the nuisance parameter (Severini, 2010). The main drawback is the specification of the prior distribution. Severini (2007) provides several suggestions about how to choose the prior distribution in order to make the integrated likelihood share the frequentist properties of a genuine likelihood function and be suitable for non-Bayesian inference. Possible choices are discussed in Section 4.

4 Integrated likelihood in meta-analysis

In our context, we assume that the between-study variance can be nil, whereas the within-study variances are strictly positive. Thus, the parameter space for $\boldsymbol{\lambda}$ is $\Lambda = [0, +\infty) \times (0, +\infty)^n$. The integrated log likelihood for the case of interest has the following form

$$\ell_{\text{Int}}(\beta) = \log \int_0^{+\infty} \underbrace{\int_0^{+\infty} \cdots \int_0^{+\infty}}_{\text{n times}} L(\beta, \tau, \sigma_1, \ldots, \sigma_n) \pi(\tau, \sigma_1, \ldots, \sigma_n | \beta) d\sigma_1 \cdots d\sigma_n d\tau. \quad (4)$$

The evaluation of the integrated log likelihood in (4) requires to overcome two main obstacles. The first one is the choice of the prior distribution for the nuisance parameter vector $\boldsymbol{\lambda} = (\tau, \sigma_1, \ldots, \sigma_n)^T$, conditionally on the parameter of interest β. The second obstacle pertains to the computation of $\ell_{\text{Int}}(\beta)$.

For the choice of the prior distribution for the nuisance parameter vector $\boldsymbol{\lambda}$, we can follow the recommendations by Severini (2007, 2010). He advocates the use of an orthogonal parameterization of the nuisance parameters and the consequent choice of priors for $\boldsymbol{\lambda}$ independent of β. From a frequentist perspective, he shows that the best inferential results are achieved when the model parameterization is expressed so that the nuisance parameter is *strongly unrelated* to β. A nuisance parameter $\boldsymbol{\phi}$ is strongly unrelated to β if

$$E\{\ell_{\boldsymbol{\lambda}}(\beta, \boldsymbol{\lambda}); \beta_0, \boldsymbol{\lambda}_0\}\big|_{(\beta_0, \boldsymbol{\lambda}_0) = (\hat{\beta}, \boldsymbol{\phi})} = 0,$$

where $\ell_{\boldsymbol{\lambda}}$ is the score vector for $\boldsymbol{\lambda}$ and the expected value is computed before the evaluation at $(\beta_0, \boldsymbol{\lambda}_0) = (\hat{\beta}, \boldsymbol{\phi})$. The function $\boldsymbol{\phi} = \boldsymbol{\phi}(\beta, \boldsymbol{\lambda}; \hat{\beta})$ defines a data-dependent parameterization and $\boldsymbol{\phi}$ is called the *zero-score-expectation parameter*. When such a parameterization is employed, the resulting integrated likelihood is a high-order approximation to the modified profile likelihood (Severini, 2000, Section 9.3), which achieves optimal elimination of the nuisance parameters (Severini, 2007). With the zero-score-expectation parameterization the choice of the prior for the nuisance parameter becomes inconsequential.

To derive the zero-score-expectation parameterization for the model corresponding to (1) consider the derivatives of the log likelihood function with respect to the nuisance parameter

vector $\boldsymbol{\lambda} = (\tau, \sigma_1, \ldots, \sigma_n)^T$,

$$\begin{aligned}
\ell_\tau(\beta, \boldsymbol{\lambda}) &= \tau \sum_{i=1}^{n} \left\{ -\frac{1}{(\sigma_i^2 + \tau^2)} + \frac{(y_i - \beta)^2}{(\sigma_i^2 + \tau^2)^2} \right\}, \\
\ell_{\sigma_i}(\beta, \boldsymbol{\lambda}) &= \sigma_i \left\{ -\frac{1}{(\sigma_i^2 + \tau^2)} + \frac{(y_i - \beta)^2}{(\sigma_i^2 + \tau^2)^2} - \frac{f_i}{\sigma_i^2} + \frac{f_i\, s_i^2}{\sigma_i^4} \right\}, \qquad i = 1, \ldots, n.
\end{aligned}$$

The evaluation of the expected value of the above expressions at the parameter value $(\beta_0, \boldsymbol{\lambda}_0)^T$, setting $\beta_0 = \hat{\beta}$ and $\boldsymbol{\lambda}_0 = \boldsymbol{\phi}$, where $\boldsymbol{\phi} = (\zeta, \delta_1, \ldots, \delta_n)^T$ is

$$E\{\ell_\tau(\beta, \boldsymbol{\lambda}); \beta_0, \boldsymbol{\lambda}_0\}\big|_{(\beta_0, \boldsymbol{\lambda}_0) = (\hat{\beta}, \boldsymbol{\phi})} = \tau \sum_{i=1}^{n} \left\{ \frac{(\delta_i^2 - \sigma_i^2) + (\zeta^2 - \tau^2) + (\hat{\beta} - \beta)^2}{(\sigma_i^2 + \tau^2)^2} \right\}$$

and

$$E\{\ell_{\sigma_i}(\beta, \boldsymbol{\lambda}); \beta_0, \boldsymbol{\lambda}_0\}\big|_{(\beta_0, \boldsymbol{\lambda}_0) = (\hat{\beta}, \boldsymbol{\phi})} = \sigma_i \left\{ \frac{(\delta_i^2 - \sigma_i^2) + (\zeta^2 - \tau^2) + (\hat{\beta} - \beta)^2}{(\sigma_i^2 + \tau^2)^2} + \frac{f_i\, (\delta_i^2 - \sigma_i^2)}{\sigma_i^4} \right\}.$$

The zero-score-expectation parameterization is obtained by equating the above expressions to 0 and then solving for $(\zeta, \delta_1, \ldots, \delta_n)^T$. Some simple algebra provides the solution

$$\begin{aligned}
\zeta &= \sqrt{\tau^2 - (\hat{\beta} - \beta)^2}, \\
\delta_i &= \sigma_i, \qquad i = 1, \ldots, n.
\end{aligned} \tag{5}$$

It is not difficult to verify that for the meta-analysis problem we focus on, the nuisance parameter vector $(\tau, \sigma_1, \ldots, \sigma_n)^T$ is orthogonal to β, i.e., the corresponding $\beta\boldsymbol{\lambda}$-block of the expected Fisher information is nil. Moreover, parameters β and σ_i are also strongly unrelated; indeed, $\hat{\sigma}_i^2 \doteq s_i^2$.

The behavior of a strongly unrelated nuisance parameter can be graphically summarized along the lines of Severini (2007, Example 1). For the data of Example 2.1, Figure 3 displays a plot of the scaled constrained estimates $(\hat{\tau}_\beta - \hat{\tau})/\mathrm{SE}(\hat{\tau})$ and $(\hat{\zeta}_\beta - \hat{\zeta})/\mathrm{SE}(\hat{\zeta})$ as a function of β, with $\mathrm{SE}(\cdot)$ denoting the standard error of the estimate. The much weaker dependence of the latter estimates on β compared to the former is apparent.

Figure 3 here

Once a strongly unrelated parameterization for the nuisance parameters is obtained, the prior distributions for ζ and $\sigma_1, \ldots, \sigma_n$ can be chosen with some liberty. A simple choice is given by independent priors for all the components of ϕ, with $\pi(\zeta) \propto 1$ and $\pi(\sigma_i) \propto 1/\sigma_i^k$, for fixed k. Such a choice, coupled with the algebraic form of the score function for β, $\ell_\beta(\beta, \boldsymbol{\lambda})$, implies that the signed integrated log likelihood ratio statistic $r_{\text{Int}}(\beta)$ in (3) is asymptotically standard normally distributed with high accuracy (Severini, 2010, Section 5). The latter property is shared also by the integrated likelihood computed using the original parameterization, provided that similar priors, independent of β, are adopted.

Computation of $\ell_{\text{Int}}(\beta)$ is less demanding than it might seem at first sight. Indeed, due to the assumption of independent meta-analysis information, $L(\beta, \tau, \sigma_1, \ldots, \sigma_n)$ in (4) is the product of n similar terms, which can be readily recovered from formula (1). The aforementioned choice of the prior distribution for $\boldsymbol{\phi}$ with independent components implies that the integral required for $\ell_{\text{Int}}(\beta)$ can be written as

$$\ell_{\text{Int}}(\beta) = \log \int_0^{+\infty} \left\{ \prod_{i=1}^{n} g_i(\beta, \zeta) \right\} \pi(\zeta) d\zeta, \tag{6}$$

where $g_i(\beta, \zeta) = \int_0^{+\infty} L(\beta, \zeta, \sigma_i) \pi(\sigma_i) d\sigma_i$ and $L(\beta, \zeta, \sigma_i)$ is the likelihood term for the i-th study. In other words, each of the n integrals $g_i(\beta, \zeta)$ as well as the main integral in (6) amount to one-dimensional integrals, that can be approximated via standard numerical methods. In our study, the inner integrals for $g_i(\beta, \zeta)$ in (6) is computed by adaptive Gauss-Kronrod quadrature, using the `C` function `Rdqags`, which is the port to the `R` library of `C` functions of the `QUADPACK` routine `dqags` (Piessens et al., 1983). The outer integral instead is computed by a standard Gaussian quadrature. The resulting integrated log likelihood is quite a smooth function of β in all the experiments performed and its maximization by means of a derivative-free optimizer is usually not an issue.

4.1 Meta-regression

Besides the inclusion of the between-study variance component τ^2, the heterogeneity among studies may be partly explained by the inclusion of covariates at the study level. The resulting model is the meta-regression model, specified as follows. Let $\boldsymbol{x_i}$ denote the vector of p covariates available at the aggregated meta-analysis level for each study, with the first value equal to one. Thus, $\boldsymbol{\beta}$ is the associated p-dimensional vector of effects. The meta-regression model extends the meta-analysis model as follows (Thompson & Higgins, 2002),

$$Y_i \sim N(\boldsymbol{x_i}^T\boldsymbol{\beta}, \sigma_i^2 + \tau^2),$$

and the associated log likelihood function for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \tau, \sigma_1, \ldots, \sigma_n)^T$ is a simple modification of (1) by incorporating the study level covariates $\boldsymbol{x_i}$ into the mean of Y_i. Consider $\boldsymbol{\beta}$ as the parameter of interest. The integrated log likelihood for $\boldsymbol{\beta}$ can be obtained similarly to the meta-analysis case described in the previous section. Then, inference on a single element of $\boldsymbol{\beta}$ can be performed by profiling the log likelihood function.

It is still possible to derive the re-parametrized model using the zero-score-expectation parameter $\boldsymbol{\phi}$ for the nuisance component $\boldsymbol{\lambda}$. The approach is the same used for the meta-analysis case and it gives rise to $\boldsymbol{\phi} = (\zeta, \sigma_1, \ldots, \sigma_n)^T$, where ζ is related to τ as follows,

$$\zeta^2 = \tau^2 - \frac{\sum_{i=1}^n (\sigma_i^2 + \tau^2)^{-2}(\boldsymbol{x_i}^T\hat{\boldsymbol{\beta}} - \boldsymbol{x_i}^T\boldsymbol{\beta})^2}{\sum_{i=1}^n (\sigma_i^2 + \tau^2)^{-2}}.$$

Differently from the meta-analysis case, an explicit expression of ζ in terms of τ cannot be obtained apart from the special case of equal within-study variances. In this situation,

$$\zeta = \sqrt{\tau^2 - \sum_{i=1}^{n} (\boldsymbol{x_i}^T\hat{\boldsymbol{\beta}} - x_i^T\boldsymbol{\beta})^2},$$

an expression which is closely related to (5). More generally, instead, the unavailability of a closed form for τ^2 makes the use of the zero-score-expectation parameterization in the integrated likelihood complicated. A viable solution is to adopt an approximate zero-score-expectation parameterization as if within-study variances were equal.

5 Simulation studies

The performance of the integrated likelihood has been investigated via simulation. In particular, the signed integrated log likelihood ratio statistic $r_{\rm Int}$ is compared to the signed profile log likelihood ratio statistic $r_{\rm P}$ and to Skovgaard's statistic $r_{\rm P}^*$, in terms of empirical one-sided rejection rates and empirical coverages of confidence intervals at different nominal levels.

Several specifications of the signed integrated log likelihood ratio are considered, namely,

- $r_{\rm Int}$, based on (4) expressed on the original parameterization, with $\pi(\tau) \propto 1$ and $\pi(\sigma_i) \propto 1/\sigma_i$;
- $\tilde{r}_{\rm Int}$, based on the re-parametrized model using the zero-score-expectation parameter $\boldsymbol{\phi}$ and a prior specification as given in Section 4, with $k = 1$;
- $\bar{r}_{\rm Int}$, based on the re-parametrized model using the zero-score-expectation parameter $\boldsymbol{\phi}$ and a prior specification as given in Section 4, with $k = 1$, and $\hat{\beta}$ in ζ replaced by the maximizer of (4) expressed in the original parameterization.

The latter choice has the virtue of not requiring the MLE of β. This may be an advantage for those data sets where the maximization of $\ell(\boldsymbol{\theta})$ is demanding, as, for example, in the numerical problem considered in Vangel & Rukhin (1999), where the profile likelihood for $(\beta, \tau)^T$ exhibits some local maxima.

Several small-sample scenarios have been taken into account.

a) A meta-analysis of $n = 3$ studies, each of them with dimensions $n_i = 2$, $i = 1, 2, 3$, with parameters values set equal to $(\beta, \tau^2) = (2.0, 2.0)$. The values of the within-study variances σ_i^2, $i = 1, 2, 3$, are initially obtained as independent realizations of a Uniform variable on $[3.0, 10.0]$ and maintained fixed in the simulation.

b) A meta-analysis of $n = 5$ studies, of dimensions $(2, 8, 3, 5, 2)$, with parameters values set equal to $(\beta, \tau^2) = (1.0, 0.5)$. The values of the within-study variances σ_i^2, $i = 1, \ldots, 5$,

are initially obtained as independent realizations of a Uniform variable on $[1.0, 4.0]$ and maintained fixed in the simulation.

c) A meta-analysis of $n = 6$ studies, of dimensions $(2, 15, 2, 15, 2, 15)$, with parameters values set equal to $(\beta, \tau^2) = (1.5, 1.0)$. The values of the within-study variances σ_i^2, $i = 1, \ldots, 6$, are initially obtained as independent realizations of a Uniform variable on $[1.0, 8.0]$ and maintained fixed in the simulation.

The simulation experiment has been repeated 10,000 times for each scenario.

The simulation studies evaluate the empirical one-sided rejection rates and the empirical coverages of confidence intervals for the competing approaches according to different nominal levels. Results are reported in Table 1 and in Table 2, respectively. Figure 4 reports the histograms of the distribution of r_{P}, r_{P}^* and $\bar{r}_{\mathrm{Int}}$, with reference to scenario a).

Table 1 and Table 2 here

Figure 4 here

The standard first-order statistic r_{P} provides empirical one-sided rejection rates which are far from the nominal levels, under all the examined scenarios, and more seriously in the case of small sample size, as in the rather extreme scenario a) in Table 1. As a consequence, empirical coverages of confidence intervals are substantially below the target level, see Table 2. An improvement over first-order results is provided by Skovgaard's statistic, although such an amelioration is not satisfactory. Moreover, from a practical point of view, the evaluation of r_{P}^* suffers from numerical instabilities when estimating the variance components, especially in case of small sample size for the meta-analysis studies, like in scenario a). Sharma & Mathew (2011) apply Skovgaard's statistic in inter-laboratory studies with unknown within-laboratory variances, providing some `R` code for its computation. They suggest a practical strategy to face the computational difficulties related to the evaluation of r_{P}^* that may arise in small samples. For example, the observed information matrix, when not positive definite,

can be substituted with a positive quantity, e.g., the expected information matrix. Moreover, as a supplementary practical measure, they replace r_{P}^{*} by r_{P} whenever the adjustment $u(\beta)$ in (2) fails to be positive. Such a careful computation ensures that the resulting statistic is always well defined. At any rate, it should be noted that the r_{P}^{*} statistic may be unstable when the value under testing is close to $\hat{\beta}$, thus requiring some further adjustments. See, for example, the discussion in Fraser et al. (2003). Though r_{P}^{*} represents a substantial improvement over r_{P}, its finite sample distribution is not standard normal for the chosen settings. This is apparent from Figure 4.

The use of the integrated likelihood approach provides a substantial improvement of the results accuracy with respect to r_{P} and r_{P}^{*}, under all the examined scenarios. Moreover, the application of the approach does not imply any computational inconvenient, especially in case of small sample sizes for the meta-analysis studies. Empirical rejection rates are close to the target levels, with no appreciable difference among the specifications chosen for the integrated likelihood, see Table 1. Correspondingly, empirical coverages of confidence intervals are close to the nominal levels, see Table 2. The high accuracy of the standard normal approximation provided by $\bar{r}_{\mathrm{Int}}$ for scenario a) emerges from Figure 4. Such a behavior is maintained also by r_{Int} and $\tilde{r}_{\mathrm{Int}}$.

Simulation studies for investigating the performance of the integrated likelihood approach has been carried out under specifications of the prior distribution for the nuisance parameters alternative to those chosen here. No appreciable differences in terms of accuracy of the inferential results emerged.

The application of the signed profile log likelihood ratio statistic and of Skovgaard's statistic when the within-study variances are assumed to be known and equal to the values reported by each study has been examined as well. Simulation results showed a severely poor behavior of both the approaches.

Several other simulation studies have been performed in scenarios characterized by a small number of studies n or a small study dimension n_i. Generally speaking, when n or

study dimension n_i increases, then inferential conclusions can be more satisfactory, also when relying on the signed profile log likelihood ratio statistic or on Skovgaard's proposal. Increasing the number of studies n, in fact, ensures inference on the between-study component τ^2 to be more precise, while increasing the study dimension n_i allows a better estimation of the within-study variance σ_i^2. A summary of the main findings is as follows.

- Even for moderate size of study dimension, in settings with small n the signed profile log likelihood ratio statistic may provide inaccurate results, especially for relatively large values of the between-study variance τ^2. In such cases, Skovgaard's statistic is generally recommendable, as for larger study size the numerical problems associated to its computation are only a marginal issue. Instead, in settings with relatively large n and small study dimension n_i, the signed profile log likelihood ratio statistic may still need an adjustment and Skovgaard's proposal may still be affected by numerical problems.

- The signed integrated log likelihood ratio statistic generally provides satisfactory results. A conservative behavior may be experienced for very small n (say, $n < 5$), large n_i (say, n_i around 10 or more) and relatively small τ^2 compared to the within-study variances.

- No major differences were generally found among the various versions of the integrated likelihood. However, the integrated likelihood exploiting the zero-score-expectation parameterization can be slightly more accurate than versions based on the original parameterization in some settings, such as, for example, a modification of scenario a) with $n = 20$ rather than $n = 3$.

6 Real data analysis

This section illustrates the application of the integrated likelihood approach for the analysis of the datasets introduced in Section 2. The integrated likelihood analysis based on $\bar{r}_{\text{Int}}$

is compared to first-order results provided by r_{P}. Results from the application of r_{Int} and $\tilde{r}_{\mathrm{Int}}$ are not reported since they are close to those from $\bar{r}_{\mathrm{Int}}$.

6.1 Cocoa intake and blood pressure reduction

Likelihood analysis provides an estimate of the treatment effect equal to -2.799 (s.e. 1.009), which is found to be significant, given the P-value equal to 0.030 associated to r_{P}. The associated 95% confidence interval for the parameter is $(-5.262, -0.397)$. The integrated likelihood approach based on the zero-score-expectation parameterization suggests a non-significant effect of cocoa consumption on lowering diastolic blood pressure, with the estimate of the treatment effect equal to -2.805 (s.e. 1.270) and the P-value for the effectiveness of the treatment equal to 0.071. The associated 95% confidence interval for the parameter is $(-6.027, 0.349)$. The profile log likelihood function and the integrated log likelihood function are compared in Figure 1, right panel. Note that since the sample size of each study is not too small, then the standard approach assuming known within-study variances provides results very close to those reported above.

6.2 Set-shifting ability in eating disorders

Likelihood analysis provides an estimate of the parameter associated to the eating disorder indicator equal to -0.117 (s.e. 0.216), which is found to be not significant, given the P-value equal to 0.619 associated to r_{P}. The corresponding 95% confidence interval for the parameter is $(-0.566, 0.283)$. A similar conclusion is obtained by the integrated likelihood approach based on the zero-score-expectation parameterization as if within-study variances were equal. The method provides an estimate of the parameter associated to the eating disorder indicator equal to -0.122 (s.e. 0.202) and the P-value equal to 0.561 associated to $\bar{r}_{\mathrm{Int}}$. The corresponding 95% confidence interval for the parameter is $(-0.542, 0.235)$. The profile log likelihood function and the integrated log likelihood function are compared in Figure 2, right panel.

7 Binary data

The case of binary data presents some features making the analysis rather different from what presented in the previous sections for normally distributed data. The most common situation is meta-analysis with summary estimates represented by the log odds ratio. In this case, the within-study variance is a function of the response data yielding the summary estimate. Both the assumption of a scaled chi-square distribution for the within-study variance estimate and the independence assumption between the summary estimate and the within-study variance estimate are rather questionable.

Within this framework, the integrated likelihood approach can still supply useful results. Instead of assuming a model for the estimated log odds ratio and the within-study variance estimate as done for normal data, we consider a model for the original data consisting in the number of successes for the treatment group and the control group. In particular, following Liu & Pierce (1993), we define the following pair of independent observations for the treatment group and the control group within the i-th study,

$$Y_{Ti} \sim Bin\{n_{Ti}, \text{logit}^{-1}(\alpha_i + \beta_i)\}, \qquad Y_{Ci} \sim Bin\{n_{Ci}, \text{logit}^{-1}(\alpha_i)\}, \tag{7}$$

where n_{Ti} and n_{Ci} are the sample sizes of the treatment group and the control group, respectively, α_i is a study-specific intercept and β_i the effect of interest. The usual random effects specification is maintained,

$$\beta_i = \beta + u_i, \;\; u_i \sim \text{Normal}(0, \tau^2),$$

whereas the α_is are treated as fixed nuisance parameters. Similarly to the normal case, the whole parameter vector is $(\beta, \tau, \alpha_1, \ldots, \alpha_n)^T$. Differently from the normal case, instead, it is possible to remove the study-specific intercepts by conditioning on the study total number of successes $y_{\cdot i} = y_{Ti} + y_{Ci}$. The resulting likelihood for study i given the random effect u_i

is the conditional likelihood (Breslow & Day, 1980)

$$L_C(\beta_i) = \frac{\binom{n_{Ci}}{y_{Ci}}\binom{n_{Ti}}{y_{Ti}}\exp(y_{Ti}\,\beta_i)}{\sum_{j\in S_i}\binom{n_{Ci}}{y_{\cdot i}-j}\binom{n_{Ti}}{j}\exp(j\,\beta_i)},$$

where $S_i = \{k : \max(0, y_{\cdot i} - n_{Ci}) \leq k \leq \min(n_{Ti}, y_{\cdot i})\}$. The likelihood function for $\boldsymbol{\theta} = (\beta, \tau)^T$ is then

$$L(\beta, \tau) = \prod_{i=1}^{n} \int_{-\infty}^{+\infty} L_C(\beta + u_i)\,\phi(u_i; 0, \tau) du_i = \prod_{i=1}^{n} h_i(\beta, \tau), \tag{8}$$

where $\phi(u_i; 0, \tau)$ is the probability density function of a Normal$(0, \tau^2)$ distribution and $h_i(\beta, \tau) = \int_{-\infty}^{+\infty} L_C(\beta + u_i)\,\phi(u_i; 0, \tau) du_i$. The integrated log likelihood function is

$$\ell_{\text{Int}}(\beta) = \log \int_0^{+\infty} \left\{ \prod_{i=1}^{n} h_i(\beta, \tau) \right\} \pi(\tau) d\tau, \tag{9}$$

where the prior independence between β and τ has been maintained. There is a close correspondence between the integrated log likelihood (6) for the normal case and the integrated log likelihood (9) for the binary case, since the latter can be obtained from the former by replacing $g_i(\beta, \zeta)$ with $h_i(\beta, \tau)$. The correspondence can be exploited for implementation of (9), paying attention to the approximation of $h_i(\beta, \tau)$ in order to avoid numerical problems when evaluating $L_C(\beta_i)$. Following Liu & Pierce (1993), we found rather effective to approximate the integral required for $h_i(\beta, \tau)$ by the Laplace's approximation; see Severini (2000, Chapter 2) for a gentle treatment.

There is a noteworthy difference between the expressions (6) and (9), despite the close mathematical similarity. Whereas the integration with respect to σ_i in (6) requires the specification of a prior distribution, the distribution assumed for the random effects u_i in (8) can be considered as part of the model specification and the only prior distribution required for obtaining the integrated log likelihood (9) is given by $\pi(\tau)$. Unfortunately, parameters β and τ are not orthogonal, thus making the choice of the prior for τ not inconsequential. Moreover, obtaining a strongly unrelated parameterization for τ by repeating the same steps

illustrated in Section 3 involves rather complicated calculations, discouraging its application for routine use. As a consequence, the specification of a prior distribution for τ has to be based on some practical considerations, supported by numerical experiments. We suggest here two possible specifications for the integrated log likelihood.

i) The first proposal specifies the prior distribution for τ as $\pi(\tau) \propto 1/\tau^k$, for fixed k. An empirical evaluation indicates that $k = 0.5$ is an acceptable choice. Accordingly, inference on β is performed by relying on the signed integrated log likelihood ratio statistic r_{Int} obtained from (9) with $\pi(\tau) \propto 1/\sqrt{\tau}$.

ii) The second proposal considers that achieving a strongly unrelated parameterization is quite involved. Nevertheless, an empirical evaluation of the re-parameterization

$$\zeta = \sqrt{\tau^2 - (\hat{\beta} - \beta)^2},$$

merely by analogy with (5) for the normal case, frequently provides a plot similar to Figure 3. Empirical experiments suggest that $\hat{\zeta}_\beta$ is less dependent on β than $\hat{\tau}_\beta$. Therefore, inference on β can be based on $\tilde{r}_{\text{Int}}$, employing ζ as nuisance parameter and the prior $\pi(\zeta) \propto 1/\sqrt{\zeta}$.

A small-scale simulation study has been carried out to evaluate the performance of the two integrated log likelihood proposals. The simulation considers a scenario involving $n = 6$ studies, with sample sizes given by $\boldsymbol{n_T} = (5, 7, 4, 10, 6, 15)^T$ for the treatment group and $\boldsymbol{n_C} = (5, 8, 4, 10, 6, 15)^T$ for the control group. The simulation experiment generates 10,000 data sets from model (7), with $\beta = 1$ and τ^2 assuming values in $\{0.05, 0.95\}$, representing two different scenarios of effect heterogeneity. The study-specific intercepts are generated as $\alpha_i \sim \text{Normal}(0, 0.25)$, independent of the random effect β_i. Table 3 reports the empirical coverages of confidence intervals at nominal levels 0.90, 0.95, 0.99, based on the signed profile log likelihood ratio statistic and on the two integrated log likelihoods proposed above.

Table 3 here

The profile log likelihood approach and the use of the two integrated log likelihoods provide similar inferential results, none being really off-target. For the profile log likelihood this is not unexpected, as there is a single nuisance parameter. All the solutions are slightly conservative for small effect heterogeneity. For large effect heterogeneity, instead, the use of the profile log likelihood results in empirical coverages of confidence intervals slightly below the target level. The use of the integrated log likelihood provides a moderate correction, especially with reference to $\tilde{r}_{\text{Int}}$.

The computation of the conditional likelihood for large study size may be hampered by numerical problems. In such case, following Liu & Pierce (1993), the conditional likelihood $L_C(\beta_i)$ could be replaced by the modified profile likelihood, with very small loss of accuracy, if any. See also Lee et al. (2010).

8 Concluding remarks

This paper considers inference in meta-analysis with a small dimension of the studies. For normally distributed measures of the effect of interest, the common assumption of known within-study variances is questionable in small sample studies. The proposed integrated likelihood approach allows to account for the uncertainty related to the estimation of the within-study variances and eliminates the variance components via integration with respect to a prior density. The methodology is shown to provide accurate inferential results and it may be preferable to the likelihood analysis performed under the assumption of known within-study variances. In the meanwhile, the method avoids the computational difficulties affecting higher-order solutions when estimating the variance components. An attractive feature of our proposal is that it can be implemented without any recourse to simulation methods, thus yielding fully reproducible results.

The case of binary data has been examined as well. Within this framework, the integrated likelihood approach is shown to maintain an interesting analogy with that under the normal case. Despite the achievement of a prior density for the nuisance components is less

immediate than in the normal case, as the within-study variances are a function of the response data yielding the summary estimate, the integrated likelihood approach still provides accurate inferential results.

From the methodological side, the problem studied in this paper is an instance of two-index asymptotics (Sartori, 2003), meaning that the available sample information grow with both the observations within each study and the number of studies. Although we did not formally cast the study of the available methodology within the two-index setting, it seems worth mentioning that recent results presented in De Bin (2012) substantiate the good properties of the integrated likelihood using the zero-score-expectation parameterization for general statistical models within two-index asymptotics. Findings in De Bin (2012) represent further theoretical support for the results presented here.

Albeit the methodology discussed here is embedded in a frequentist approach, an extension to a full Bayesian formulation is possible. In this case, in fact, the integrated likelihood can be thought of as a likelihood function which results in an effortless implementation, since the prior distribution for the remaining parameter β would be only one-dimensional.

References

Breslow, N. E. & Day, N. E. (1980). *Statistical methods in cancer research: Volume 1 - The analysis of case-control studies*. International Agency for Research on Cancer, Lyon.

Brockwell, S. E. & Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Stat. Med.* **20**, 825–840.

De Bin, R. (2012). Integrated likelihood for the treatment of nuisance parameters. Ph.D. Thesis. University of Padova.

Debray, T. (2013). *metamisc: Diagnostic and prognostic meta analysis (metamisc).* R package version 0.1.1. http://CRAN.R-project.org/package=metamisc.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.

Fraser, D. A. S., Reid, N., Li, R. & Wong, A. (2003). p-value formulas from likelihood asymptotics: Bridging the singularities. *J. Statist. Res.* **37**, 1–15.

Guolo, A. (2012). Higher-order likelihood inference in meta-analysis and meta-regression. *Stat. Med.* **31**, 313–327.

Guolo, A. & Varin, C. (2012). The `R` package `metaLik` for likelihood inference in meta-analysis. *Journal of Statistical Software* **50** (7), 1–14.

Hardy, R. J. & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Stat. Med.* **15**, 619–629.

van Houwelingen, H. C., Arends, L. R. & Stijnen, T. (2002). Tutorial in biostatistics. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat. Med.* **21**, 589–624.

Lee, W., Shi, J. Q. & Lee, Y. (2010). Approximate conditional inference in mixed-effects models with binary data. *Comput. Statist. Data Anal.* **54**, 173–184.

Liu, Q. & Pierce, D. A. (1993). Heterogeneity in Mantel-Haenszel-type models. *Biometrika* **80**, 543–556.

Piessens, R., de Doncker-Kapenger, E., Ueberhuber, C. & Kahaner, D. (1983). *QUADPACK, A subroutine package for automatic integration*. Springer-Verlag.

R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org/`.

Roberts, C. J. (2005). Issues in meta-regression analysis: an overview. *Journal of Economics Surveys* **19**, 295–298.

Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L. & Treasure, J. (2007). A systematic review and meta-analysis of set-shifting ability in eating disorders. *Psychological Medicine* **37**, 1075–1084.

Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* **90**, 533–549.

Severini, T. A. (2000). *Likelihood methods in statistics*. Oxford University Press, Oxford.

Severini, T. A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika* **94**, 529–542.

Severini, T. A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika* **97**, 481–496.

Sharma, G. & Mathew, T. (2011). Higher order inference for the consensus mean in inter-laboratory studies. *Biom. J.* **53**, 128–136.

Sidik, K. & Jonkman, J. N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Stat. Med.* **26**, 1964–1981.

Skovgaard, I. M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli* **2**, 145–165.

Sutton, A. J. & Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Stat. Med.* **27**, 625–650.

Taubert, D., Roesen, R. & Schömig, E. (2007). Effect of cocoa and tea intake on blood pressure: A meta-analysis. *Archives of Internal Medicine* **167**, 626–634.

Thompson, S. G. & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Stat. Med.* **21**, 1559–1573.

Vangel, M.G. & Rukhin, A.L. (1999). Maximum likelihood analysis for heteroscedastic one-way random effects ANOVA in interlaboratory studies. *Biometrics* **55**, 129–136.

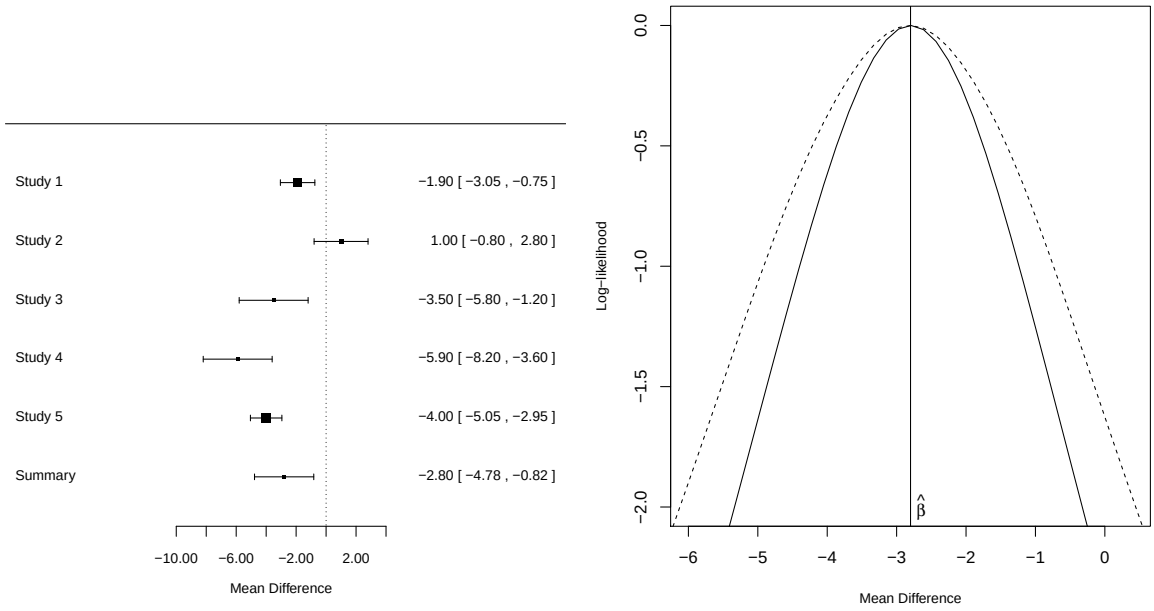

Figure 1: Cocoa data. Left panel: forest plot. Right panel: profile log likelihood function (solid line) and integrated log likelihood function (dashed line).

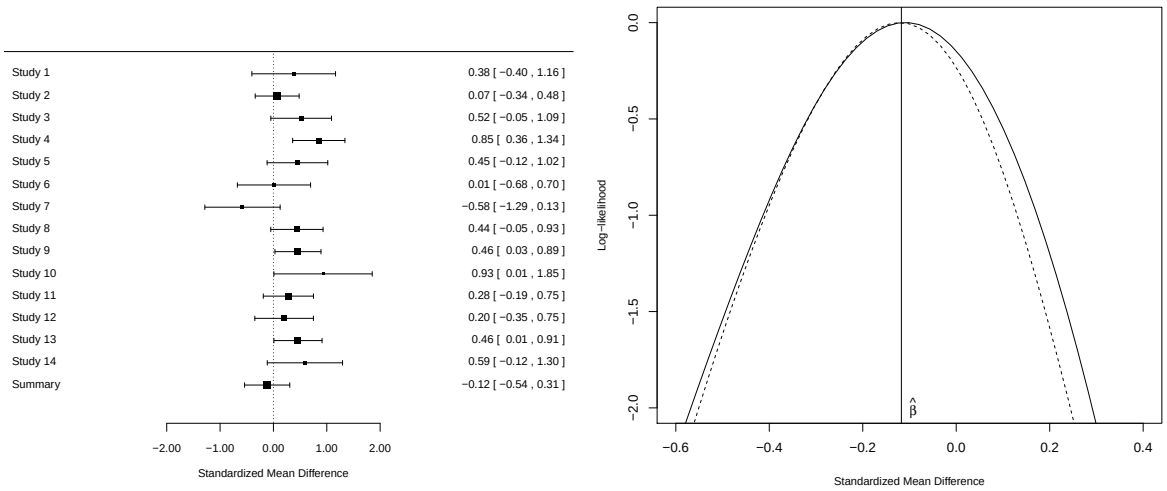

Figure 2: Eating disorders data. Left panel: forest plot. Right panel: profile log likelihood function (solid line) and integrated log likelihood function (dashed line).

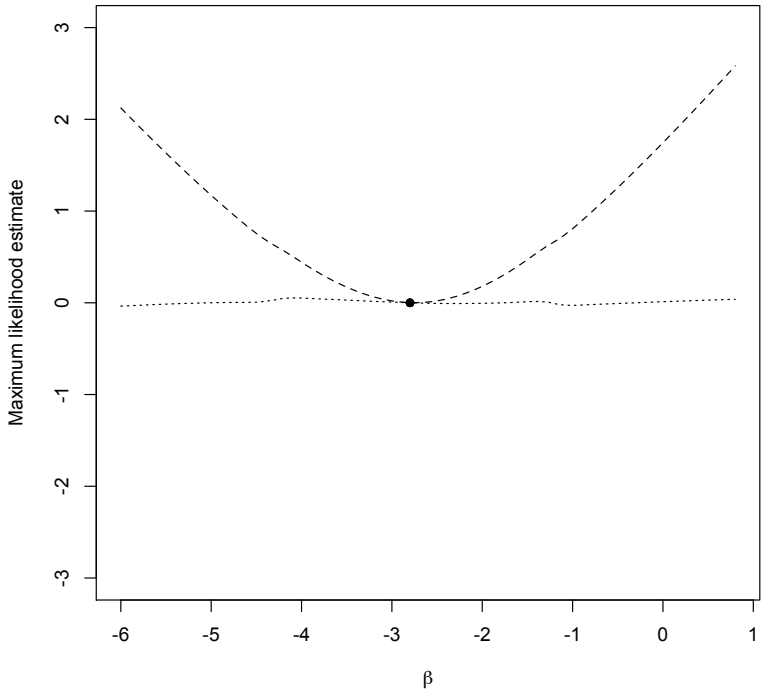

Figure 3: Constrained estimators for Example 2.1. Plot of $(\hat{\tau}_\beta - \hat{\tau})/\mathrm{SE}(\hat{\tau})$ (dashed line) and $(\hat{\zeta}_\beta - \hat{\zeta})/\mathrm{SE}(\hat{\zeta})$ (dotted line) against β.

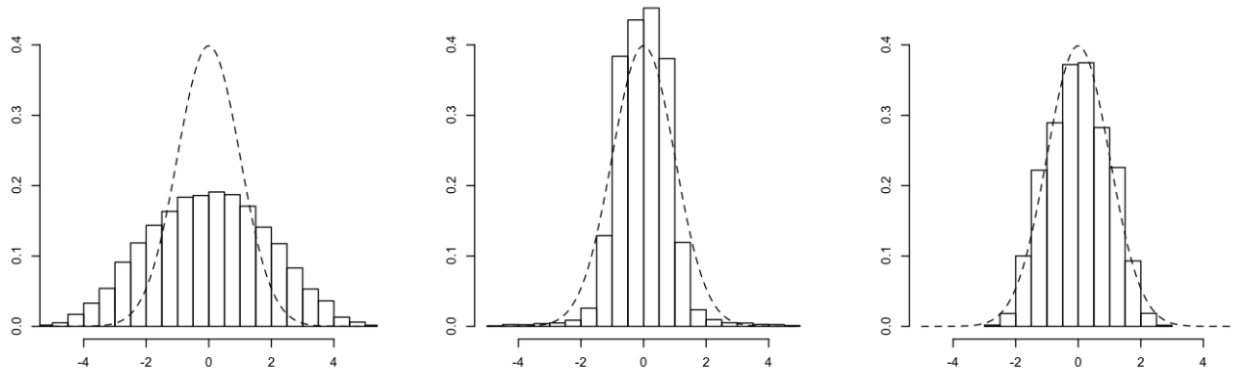

Figure 4: Histograms of the distribution of r_{P} (left panel), r_{P}^{*} (middle panel) and $\bar{r}_{\mathrm{Int}}$ (right panel) for simulation scenario a), based on 10,000 replicates. The standard normal density is superimposed (dashed line).

Table 1: Empirical one-sided rejection rates for the signed profile log likelihood ratio statistic r_{P}, Skovgaard's statistic r_{P}^{*}, and different specifications of the signed integrated log likelihood ratio statistic, r_{Int}, $\tilde{r}_{\mathrm{Int}}$, and $\overline{r}_{\mathrm{Int}}$, based on 10,000 replicates.

	Rejection rates	r_{P}	r_{P}^{*}	r_{Int}	$\tilde{r}_{\mathrm{Int}}$	$\overline{r}_{\mathrm{Int}}$
				Scenario a)		
Lower	0.010	0.121	0.018	0.003	0.002	0.002
	0.025	0.168	0.023	0.013	0.012	0.012
	0.050	0.213	0.029	0.039	0.039	0.037
	0.100	0.270	0.054	0.099	0.106	0.102
Upper	0.900	0.734	0.947	0.906	0.899	0.903
	0.950	0.792	0.970	0.963	0.963	0.964
	0.975	0.837	0.975	0.987	0.989	0.989
	0.990	0.884	0.981	0.997	0.998	0.998
				Scenario b)		
Lower	0.010	0.090	0.028	0.005	0.005	0.005
	0.025	0.125	0.044	0.021	0.022	0.021
	0.050	0.164	0.068	0.049	0.054	0.051
	0.100	0.223	0.116	0.102	0.110	0.105
Upper	0.900	0.782	0.889	0.898	0.890	0.895
	0.950	0.838	0.934	0.953	0.949	0.952
	0.975	0.878	0.961	0.980	0.979	0.980
	0.990	0.917	0.973	0.994	0.993	0.994
				Scenario c)		
Lower	0.010	0.080	0.028	0.008	0.008	0.008
	0.025	0.116	0.043	0.022	0.024	0.023
	0.050	0.151	0.066	0.049	0.052	0.050
	0.100	0.209	0.111	0.102	0.105	0.101
Upper	0.900	0.789	0.879	0.891	0.887	0.890
	0.950	0.844	0.927	0.946	0.942	0.945
	0.975	0.883	0.951	0.975	0.972	0.975
	0.990	0.915	0.968	0.991	0.991	0.991

Table 2: Empirical coverages of confidence intervals at nominal levels 0.90, 0.95, 0.99, for the signed profile log likelihood ratio statistic r_{P}, Skovgaard's statistic r_{P}^{*}, and different specifications of the signed integrated log likelihood ratio statistic, r_{Int}, $\tilde{r}_{\mathrm{Int}}$, and $\overline{r}_{\mathrm{Int}}$, based on 10,000 replicates.

Level	r_{P}	r_{P}^{*}	r_{Int}	$\tilde{r}_{\mathrm{Int}}$	$\overline{r}_{\mathrm{Int}}$
			Scenario a)		
0.90	0.579	0.941	0.924	0.925	0.927
0.95	0.670	0.953	0.974	0.977	0.977
0.99	0.816	0.966	0.998	0.999	0.999
			Scenario b)		
0.90	0.674	0.866	0.904	0.895	0.902
0.95	0.753	0.917	0.960	0.957	0.959
0.99	0.870	0.955	0.996	0.997	0.997
			Scenario c)		
0.90	0.693	0.861	0.896	0.889	0.895
0.95	0.767	0.907	0.953	0.948	0.952
0.99	0.875	0.952	0.993	0.993	0.993

Table 3: Binary data: empirical coverages of confidence intervals at nominal levels 0.90, 0.95, 0.99, for the signed profile log likelihood ratio statistic r_{P} and two different specifications of the signed integrated log likelihood ratio statistic, r_{Int} and $\tilde{r}_{\mathrm{Int}}$, based on 10,000 replicates.

Level	r_{P}	r_{Int}	$\tilde{r}_{\mathrm{Int}}$
		τ^2=0.05	
0.90	0.917	0.935	0.917
0.95	0.966	0.976	0.970
0.99	0.996	0.997	0.998
		τ^2=0.95	
0.90	0.870	0.882	0.888
0.95	0.935	0.942	0.948
0.99	0.989	0.990	0.993

Annamaria Guolo

Department of Economics, University of Verona

Vicolo Campofiore, 2, I-37129, Verona, Italy

annamaria.guolo@univr.it