Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 1605–1614
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Small sample distribution of the likelihood ratio test in the random effects model

Helle Sørensen

Department of Natural Sciences, University of Copenhagen (LIFE), Thorvaldsensvej 41, DK-1871 Frederiksberg C, Denmark

Received 9 October 2006; received in revised form 8 May 2007; accepted 8 May 2007
Available online 15 August 2007

Abstract

In this paper we examine the small sample distribution of the likelihood ratio test in the random effects model which is often recommended for meta-analyses. We find that this distribution depends strongly on the true value of the heterogeneity parameter (between-study variance) of the model, and that the correct p-value may be quite different from its large sample approximation. We recommend that the dependence of the heterogeneity parameter be examined for the data at hand and suggest a (simulation) method for this. Our setup allows for explanatory variables on the study level (meta-regression) and we discuss other possible applications, too. Two data sets are analyzed and two simulation studies are performed for illustration.

Keywords: Likelihood ratio test; Meta-analysis; Meta-regression; Overdispersion model; Small sample distribution; Random effects model

1. Introduction

In meta-analysis, estimates from independent studies on some (treatment) effect are combined in order to get an estimate across studies and in order to increase the statistical power. Sometimes one takes explanatory variables on the study level into account, then the analysis is referred to as meta-regression. The number of observations is often quite small (around 10), so the usual large sample approximations for computing p-values and confidence limits are not necessarily appropriate. In this paper we investigate the small sample properties of the likelihood ratio (LR) test statistic in the so-called random effects model.

The *random effects model* was introduced for meta-analysis of randomized clinical trials by DerSimonian and Laird (1986). Allowing for explanatory variables, the model is the following:

$$y_i = \mu_i + \varepsilon_i, \quad \mu_i = A_i\theta + e_i \quad (i = 1, \ldots, n). \tag{1}$$

Here, the observation y_i is the estimated treatment effect, for example an estimated log-odds ratio, from the ith study. The true treatment effect under the circumstances of the study is μ_i. Due to the *within-study variation* μ_i is estimated with a standard error, σ_i. This source of variation is described by $\varepsilon_i \sim \mathrm{N}(0, \sigma_i^2)$. The variances σ_i^2 are assumed to be known, although in reality they are estimated in the ith study.

E-mail address: helle@dina.kvl.dk
URL: http://www.dina.kvl.dk/∼ helle.

0378-3758/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2007.05.047

The study-specific mean, μ_i, is modelled as a sum of a systematic part and a random part. The systematic part is $A_i\theta$, where A_i is the ith row of the known design matrix, A, of dimension $n \times p$ and θ is a p-dimensional parameter describing the (hypothetical) population of all possible studies. In the simple case without explanatory variables, θ is the average treatment effect over all possible studies and all $A_i = 1$. Due to the specific circumstances about the ith study (the nature of individuals in the study, the specific design of the study, etc.), μ_i deviates from the population value $A_i\theta$. This across- or *between-study variation* is modelled with $e_i \sim \mathrm{N}(0, \tau^2)$ where the between-study variance, or heterogeneity parameter, τ^2 is unknown. In other words, study enters the model as a random effect. We point out that this random effects model is now recommended for meta-analysis over the fixed effects model, assuming $\tau^2 = 0$; see the fine tutorials by Normand (1999) and van Houwelingen et al. (2002).

We shall mainly be interested in testing hypotheses about the parameter θ. We investigate the *small sample distribution* of the LR test statistic and find that it depends strongly on the true value of the between-study variance τ^2.

If all the within-study variances σ_i^2 are the same, then the random effects model is equivalent to the overdispersion model, $y_i \sim \mathrm{N}(A_i\theta, \omega^2\sigma_i^2)$ where $\omega^2 \geqslant 1$. For this model we prove that the probabilities $P_{\omega^2}(\mathrm{LR} > x)$ are increasing in ω^2 for every x and that there exists a lower bound (for $\omega^2 = 1$) and an upper bound (for $\tau^2 \to \infty$). Here, LR is the likelihood ratio test statistic and P_{ω^2} is the notation for the distribution of LR when the true value of the overdispersion parameter is ω^2. If the within-study variances are not too different, we expect the same properties to hold approximately for the random effects model (with notation similar to the one above): $P_{\tau^2}(\mathrm{LR} > x)$ increasing in τ^2 with a lower bound for $\tau^2 = 0$ and an upper bound for $\tau^2 \to \infty$. If the within-study variances are very different, however, as is often the case, then the monotonicity property does *not* hold, and the limit for $\tau^2 \to \infty$ is *not* an upper bound.

In any case, we recommend that the dependence of τ^2 on the p-value $P_{\tau^2}(\mathrm{LR} > \mathrm{LR}_{\mathrm{obs}})$ be investigated whenever a test is performed. Specifically we suggest to simulate the p-values as a function of τ^2 for a whole range of τ^2-values. Since the number of observations is small for the applications we have in mind, computation time is not a problem.

The rest of the paper is organized as follows. The random effects model and the overdispersion model are described in greater detail in Section 2. The properties of the distribution of LR in the random effects model are illustrated by simulation in Section 3. In Section 4 we analyze two real data sets from the meta-analysis literature. Finally some concluding remarks are given in Section 5.

2. The distribution of the LR test statistic

In this section we go into details about the small sample distribution of the LR test statistic in the random effects model as well as in the related overdispersion model.

2.1. The random effects model: preliminaries

Consider again model (1), and recall that all ε_i's and e_i's are independent with mean zero and variances σ_i^2 (known) and τ^2 (unknown), respectively. The vector of means $\mu = (\mu_1, \ldots, \mu_n)$ belongs to the subspace L_1 spanned by the design matrix A which we assume has full rank, p. In other words $y_1, \ldots, y_n$ are independent with $y_i \sim \mathrm{N}(\mu_i, \sigma_i^2 + \tau^2)$, where $\mu \in L_1$ and $\tau^2 \geqslant 0$ are the unknown parameters. Our main interest lies in tests of hypotheses about μ (or, equivalently, θ).

Various estimation strategies have been discussed in the meta-analysis literature. Thompson and Sharp (1999) compare a moment estimator similar to the heterogeneity estimator from DerSimonian and Laird (1986), the maximum likelihood (ML) estimator, a restricted maximum likelihood (REML) estimator and an empirical Bayes estimator for τ^2. They recommend the REML estimator, but since our main interest is the LR test, we are mainly concerned with ML estimation.

The log-likelihood function is given by

$$l(\mu, \tau^2) = -\frac{1}{2}\sum_{i=1}^{n}\log(\sigma_i^2 + \tau^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{\sigma_i^2 + \tau^2}, \quad \mu \in L_1, \quad \tau^2 \geqslant 0.$$

Due to the additive variance structure, μ cannot be estimated by ML independently from τ^2, and the ML estimator $(\hat{\mu}, \hat{\tau}^2)$ must be found numerically. It is usually fruitful to "profile out" μ and maximize $l(\hat{\mu}(\tau^2), \tau^2)$ wrt. τ^2 where