Composite marginal likelihoods to the normal Bradley-Terry model

Mohand Larbi Feddag

HAL Id: hal-00741503
https://hal.science/hal-00741503

Submitted on 13 Oct 2012

Composite marginal likelihoods to the normal Bradley-Terry model

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

Feddag_CSSC_2010_F.tex

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Response: To the first referee

Composite marginal likelihoods to the normal Bradley-Terry model
Authors: M-L. Feddag
Manuscript: $LSSP - 2010 - 0324$

Many thanks for your valuable and interesting comments on our submitted paper.

Please find in red color in the new version of the document the response to all your comments.

Compared with original submitted paper, in this revised version, the author added some necessary illustrations which make the paper more understandable and clear. Exept for some minor remarks, which I have listed in the following, I have no further comments for the major part of this paper.

1. **On page 3, for the Bradley-Terry model (4). According to the author's response, I know that Y_{ij} is not $logit(P(y_i > y_j))$, but another new continuous random variable. In addition, $Y_{ij} \neq Y_{ji}$. I suggest adding one sentence here in the paper to clarify the above two points in case that the similar notations in model (3) and model (4) make confusion.**

   PLease find the following comments in page 3, just after the model (4).

   Up to our knowledge, there is no work on this defined model where Y_{ij} is continuous random variable and is different from Y_{ji}. This model is obviously different from the model (3) which is defined only for binary random variable.

2. **On page 3, for the pairwise likelihood (6). I suggest adding one or two sentences after the proposed likelihood in order to demonstrate why the pairwise likelihood is constructed in this format. I think the author's responses to the two referees' related questions are good enough.**

   PLease find the following comments in page 4, just after the pairwise likelihood (6).

1

The first order log-likelihood ℓ_1 as defined by Cox and Reid [3] and Varin [11] considers all the $N(N-1)$ univariate marginal distributions of the matrix $Y = (y_{ij})_{1 < i \neq j < N}$. In the second order loglikelihood ℓ_2 defined by Eq. (6), the first part involves all the pairs $(\frac{N(N-1)(N-2)}{2})$ sharing the random effects U_i and the second part involves all the pairs $(\frac{N(N-1)(N-2)}{2})$ sharing the random effects U_j. In total $N(N-1)(N-2)$ pairs of observations.

3. **On page 4, in the middle. The pairwise likelihood $\ell_2(\sigma^2; y)$ is re-written in the form of SS_W and SS_B. I didn't check the rigorous proof for $N >= 3$, but I believe your derivation is correct.**

   It is the generalization of the one given in Cox and Reid [3]

4. **On page 8, for Figure 1. The words in the figure title is upsidedown.**

   The problem deals with the style I am using which is different from the style of this journal, so it will be solved in the last step

Response: To the second referee

Composite marginal likelihoods to the normal Bradley-Terry model
Authors: M-L. Feddag
Manuscript: $LSSP - 2010 - 0324$

Many thanks for your valuable and interesting comments on our submitted paper.

   Please find in red color in the new version of the document the response to all your comments.

   You have given exhaustive explanations to my remarks, but you did not added them in your paper. Therefore, I suggest to add these explanations directly in it. More precisely:

1. you should underline that the normal version of Bradley-Terry model is introduced for the first time in your contribution

2

Please find the following comments in page 3, after the model (4).

Up to our knowledge, there is no work on this defined model where Y_{ij} is continuous random variable and is different from Y_{ji}. This model is obviously different from the model (3) which is defined only for binary random variable.

2. **insert your remarks on points 2, 3 and 4 (second paragraph from "For the logit" to the end) in the paper.**

   - Point 2

     Please find the following comments given in the end of page 3.

     In fact, let write the likelihood for $N = 3$. We denote by $y = (y_{12}, y_{13}, y_{23}, y_{21}, y_{31}, y_{32})'$, then we can write the following transformation:

     $$y = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{pmatrix} e_{12} \\ e_{13} \\ e_{23} \\ e_{21} \\ e_{31} \\ e_{32} \end{pmatrix}$$

     $$= AU + e$$

     Hence the density of y is a normal with mean 0 and variance $\sigma^2 AA' + I_6$. For this case, the likelihood has an explicite expression. It's clear that for an arbitray N, it is not easy to calculate the transformation matrix A associated to the random variable y, so the likelihood associated to the model is complex to evaluate.

   - Point 3

     Please find the following comments given in page 4, after the pairwise likelihood (6).

     The first order log-likelihood ℓ_1 as defined by Cox and Reid [3] and Varin [11] considers all the $N(N-1)$ univariate marginal distributions of the matrix $Y = (y_{ij})_{1 < i \neq j < N}$. In the second order loglikelihood ℓ_2 defined by Eq. (6), the first part involves all the pairs $(\frac{N(N-1)(N-2)}{2})$ sharing the random effects U_i and the second part involves all the pairs $(\frac{N(N-1)(N-2)}{2})$ sharing the random effects U_j. In total $N(N-1)(N-2)$ pairs of observations.

   - Point 4

3

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Please find the following comments given in the end of the discussion.

Finally, this approach could be generalized to the logit version of the model including covariates, which is given by (3). One should evaluate the pairs of probabilities $(y_{ij}, y_{ik}) = (1, 1)$, $(y_{ij}, y_{ik}) = (1, 0)$, $(y_{ij}, y_{ik}) = (0, 1)$ and $(y_{ij}, y_{ik}) = (0, 0)$, which has an analytical expression for the probit version of the model. For the logit version, the scale mixture approximation of Monahan and Stefanski [7] are needed.

4

1

1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Composite marginal likelihoods to the normal Bradley-Terry model

M-L. Feddag

EA 4275 "Biostatistique, Recherche Clinique et Mesures Subjectives en Sante", Faculte de Pharmacie, Universite de Nantes, France

Abstract

Inference in Generalized linear mixed models with crossed random effects is often made cumbersome by the high-dimensional intractable integrals involved in the marginal likelihood. This article presents two inferential approaches based on the marginal composite likelihood for the normal Bradley-Terry model. The two approaches are illustrated by a simulation study to evaluate their performance. Thereafter, the asymptotic variances of the estimated variance component are compared.

Keywords: Asymptotic variance; Bradley-Terry model; Marginal composite likelihood; Monte Carlo; Variance component.

1. Introduction

In paired comparisons, one considers a set of N treatments or players which are presented in pairs, where $N \geq 3$. It is assumed that the responses to the treatments may be described in terms of an underlying continuum on which the worth or ability of the treatments can be relatively located. Let π_i denotes the worth, an index of relative preference of the ith treatment such that $\pi_i > 0$, and $\sum_{i=1}^{N} \pi_i = 1$. Let denote by Y the binary random variable and by y_i and y_j its response values. The Bradley-Terry model [2] postulates that, if y_i and y_j are the response to treatments i and j respectively, then

$$P(y_i > y_j) = \frac{\pi_i}{\pi_i + \pi_j}, \tag{1}$$

in the comparison of treatments i and j. One interprets $y_i > y_j$ as indicative of preference for treatment i over treatment j. If i and j are players, then this event is replaced by player i beats player j. The model can alternatively be expressed in the logit-linear form

$$\text{logit}\,(P(y_i > y_j)) = \lambda_i - \lambda_j, \tag{2}$$

where $\lambda_i = \ln(\pi_i)$ for all i.

The more general model which includes the covariates and random effects is given by $\lambda_i = \sum_{r=1}^{q} x_{ir}\beta_r + U_i$, where $\beta = (\beta_1, \ldots, \beta_q)$ are the regression parameters, and U_i are the random effects supposed independent and identically distributed with $N(0, \sigma^2)$. Hence the model can be expressed as

$$logit\left(P(y_i > y_j)\right) = \sum_{r=1}^{q} (x_{ir} - x_{jr})\beta_r + U_i - U_j. \tag{3}$$

This model without random effects has been widely studied, see for example Firth [5].

Thurstone [9, 10] introduced a model for paired comparisons for continuous preference with several stimuli. It is defined via the threshold latent variables, with applications for example in marketing or psychometry. This model has been extended by Takane [8], where random error is added in the linear predictor.

From now on, we will focus on the normal version of the Bradley-Terry model without covariates and given by

$$Y_{ij} = U_i - U_j + e_{ij}, \quad 1 \le i \ne j \le N \tag{4}$$

where Y_{ij} is the results between the treatments (or players) i and j, and e_{ij} are the residual errors supposed independent and identically distributed as normal with mean 0 and variance 1. They are also supposed independent from the random effects U_i, $i = 1, \ldots, N$. Let lower cases letters y_{ij} denotes realization of Y_{ij}. Up to our knowledge, there is no work on this defined model where Y_{ij} is continuous random variable and different from Y_{ji}. This model is obviously different from the model (3) which is defined only for binary random variable.

This model belongs to the linear mixed model with complex random effects, thus the statistical inference on the variance components σ^2 by the classical likelihood is not straightforward and cumbersome. In fact, let write the likelihood for $N = 3$. We denote by $y = (y_{12}, y_{13}, y_{23}, y_{21}, y_{31}, y_{32})'$, then we can write the following transformation:

$$
\begin{aligned}
y &= \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + \begin{pmatrix} e_{12} \\ e_{13} \\ e_{23} \\ e_{21} \\ e_{31} \\ e_{32} \end{pmatrix} \\
&= AU + e
\end{aligned}
$$

Hence the density of y is a normal with mean 0 and variance $\sigma^2 AA' + I_6$. For this case, the likelihood has an explicit expression.

It's clear that for an arbitray N, it is not easy to calculate the transformation matrix A associated to the random variable y, so the likelihood associated to the model is complex to evaluate.

As an alternative, we propose the marginal composite likelihood proposed by Cox and Reid [3].

3

This paper is structured as follows. In Section 2, we define the two composite marginal likelihood approaches. Section 3 is devoted to the simulation study for four different sample sizes. The asymptotic variances of the estimated parameter provided by the two approaches are compared in Section 3. We conclude in Section 4.

2. Composite marginal likelihood

We are interested by the first and second order loglikelihoods defined respectively

$$\ell_1(\sigma^2; y) = \sum_{i=1}^{N} \sum_{j \neq i} \ln f(y_{ij}; \sigma^2) \tag{5}$$

$$\ell_2(\sigma^2; y) = \sum_{i=1}^{N} \sum_{1 \leq j < k \leq N} \ln f(y_{ij}, y_{ik}; \sigma^2) + \sum_{j=1}^{N} \sum_{1 \leq i < k \leq N} \ln f(y_{ij}, y_{kj}; \sigma^2) \tag{6}$$

The first order log-likelihood ℓ_1 as defined by Cox and Reid [3] and Varin [11] considers all the $N(N-1)$ univariate marginal distributions of the matrix $Y = (y_{ij})_{1 < i \neq j < N}$. In the second order loglikelihood ℓ_2 defined by Eq. (6), the first part involves all the pairs $(\frac{N(N-1)(N-2)}{2})$ sharing the random effects U_i and the second part involves all the pairs $(\frac{N(N-1)(N-2)}{2})$ sharing the random effects U_j. In total $N(N-1)(N-2)$ pairs of observations.
The functions ℓ_1 and ℓ_2 are marginal composite likelihood of order one and two, which are example of the general composite likelihood defined by Lindsay [6]. The pseudo likelihood ℓ_2 is called also pairwise loglikelihood (see Cox and Reid [3], Bellio and Varin [1], Feddag and Bacci [4] and Varin [11]).

These loglikelihoods are given respectively by

$$\ell_1(\sigma^2; y) = -\frac{N(N-1)}{2} \ln(2\sigma^2 + 1) - \frac{1}{2(2\sigma^2 + 1)} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} y_{ij}^2,$$

$$\ell_2(\sigma^2; y) = -\frac{N(N-1)(N-2)}{2} \ln(3\sigma^4 + 4\sigma^2 + 1)$$
$$-\frac{1}{2(3\sigma^4 + 4\sigma^2 + 1)} \left[\sum_{i=1}^{N} \sum_{1 \leq j < k \leq N} \left\{ (1 + 2\sigma^2)y_{ij}^2 - 2\sigma^2 y_{ij} y_{ik} + (1 + 2\sigma^2)y_{ik}^2 \right\} \right]$$
$$-\frac{1}{2(3\sigma^4 + 4\sigma^2 + 1)} \left[\sum_{j=1}^{N} \sum_{1 \leq i < k \leq N} \left\{ (1 + 2\sigma^2)y_{ij}^2 - 2\sigma^2 y_{ij} y_{kj} + (1 + 2\sigma^2)y_{kj}^2 \right\} \right]$$

4

Using the paper of Cox and Reid [3], this pairwise likelihood could be expressed as follows:

$$\ell_2(\sigma^2; y) = -\frac{N(N-1)(N-2)}{2} \ln(3\sigma^4 + 4\sigma^2 + 1)$$
$$-\frac{1}{2} \frac{N - 2 + (2N-3)\sigma^2}{3\sigma^4 + 4\sigma^2 + 1} (SS_W + SS_W^\star)$$
$$-\frac{1}{2} \frac{N-2}{N-1} \frac{1 + \sigma^2}{3\sigma^4 + 4\sigma^2 + 1} (SS_B + SS_B^\star),$$

where $SS_W = \sum_{i=1}^{N} \sum_{j \neq i} (y_{ij} - \bar{y}^i)^2$, $SS_B = \sum_{i=1}^{N} \left(\sum_{j \neq i} y_{ij} \right)^2$, $\bar{y}^i = \frac{1}{N-1} \sum_{j \neq i} y_{ij}$, $SS_W^\star = \sum_{j=1}^{N} \sum_{i \neq j} (y_{ij} - \bar{y}^j)^2$, $SS_B^\star = \sum_{j=1}^{N} \left(\sum_{i \neq j} y_{ij} \right)^2$, $\bar{y}^j = \frac{1}{N-1} \sum_{i \neq j} y_{ij}$

We define pseudo-score functions by loglikelihood derivatives in the usual way

$$U_1(\sigma^2; y) = \frac{\partial \ell_1(\sigma^2; y)}{\partial \sigma^2} \tag{7}$$

$$U_2(\sigma^2; y) = \frac{\partial \ell_2(\sigma^2; y)}{\partial \sigma^2} \tag{8}$$

After classical derivations, we obtain these pseudo-score given by

$$U_1(\sigma^2; y) = -\frac{N(N-1)}{2\sigma^2 + 1} + \frac{1}{(2\sigma^2 + 1)^2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} y_{ij}^2, \tag{9}$$

$$U_2(\sigma^2; y) = -N(N-1)(N-2) \frac{3\sigma^2 + 2}{3\sigma^4 + 4\sigma^2 + 1}$$
$$+ \frac{1}{2} \frac{3(2N-3)\sigma^4 + 6(N-2)\sigma^2 + 2N - 5}{(3\sigma^4 + 4\sigma^2 + 1)^2} (SS_W + SS_W^\star)$$
$$+ \frac{3}{2} \frac{N-2}{N-1} \frac{1}{(3\sigma^2 + 1)^2} (SS_B + SS_B^\star) \tag{10}$$

The estimating equations $U_i(\hat{\sigma}^2; y) = 0$, $i = 1, 2$ are under usual regularity conditions, unbiased. The resulting estimator is for large N asymptotically normal with mean σ^2 and variance

$$\frac{E\left(U_i^2(\sigma^2)\right)}{\left[E(-U_i^{(1)}(\sigma^2))\right]^2} \tag{11}$$

The first derivative of these pseudo-scores with respect to the parameter σ^2 are given by

5

$$U_1^{(1)}(\sigma^2; y) = \frac{2N(N-1)}{(2\sigma^2+1)^2} - \frac{4}{(2\sigma^2+1)^3} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} y_{ij}^2 \qquad (12)$$

$$
\begin{aligned}
U_2^{(1)}(\sigma^2; y) = {} & N(N-1)(N-2)\frac{9\sigma^4 + 12\sigma^2 + 5}{(3\sigma^4 + 4\sigma^2 + 1)^2} \\
& - \frac{1}{2}\frac{18(2N-3)\sigma^6 + 54(N-2)\sigma^4 + (36N-90)\sigma^2 + 10N - 28}{(3\sigma^4 + 4\sigma^2 + 1)^3} \times \\
& (SS_W + SS_W^\star) - 9\frac{N-2}{N-1}\frac{1}{(3\sigma^2+1)^3} \ (SS_B + SS_B^\star) \qquad (13)
\end{aligned}
$$

3. Simulation study

The maximum marginal composite likelihood estimation is performed for two pseudo-score methods where we have considered four sizes $N = 10, 20, 30$ and 50 and four values of σ^2: 0.5, 1, 2 and 4.

We evaluate the performance of the two approaches with a simulation study which is based on 500 data sets. The different results are given in Table 1 for the four sizes considered. For both approaches, these two tables show that there is a small bias for all the value of σ^2 for the three first sizes. The bias is negligible for the size $N = 50$. As expected, the standard deviation decreases when the size N increases. We note that for the value of the variance component equal to 4, the standard deviation is more considerable: greater than 1 for the three first sizes and equal to 0.805 for $N = 50$. According to the estimates and to the standard deviation, it is clearly shown that there is no significant difference between the two approaches.

4. Asymptotic variance

In this section we compare the asymptotic variance provided by U_1 and U_2 and given by (11). It needs the calculation of the expectation $\mathbb{E}\left(-U_i^{(1)}(\sigma^2)\right)$ and $\mathbb{E}\left(U_i^2(\sigma^2)\right),\ i = 1, 2.$

If the evaluation of the first quantity is straightforward, however the second quantity is more tedious and complex. In fact, It is easier to prove that we have the following expectations:

$$
\begin{aligned}
\mathbb{E}(SS_W) &= \mathbb{E}(SS_W^\star) = N(N-2)(\sigma^2 + 1), \\
\mathbb{E}(SS_B) &= \mathbb{E}(SS_B^\star) = N(N-1)(N\sigma^2 + 1).
\end{aligned}
$$

By replacing $\mathbb{E}(Y_{ij}^2) = 2\sigma^2 + 1$ in (12) and the two previous expectations in the expression (13), we derive the following quantities

6

Table 1: Parameter estimate (mean) for σ^2 and its standard deviation (sd) for $N = 10,\ 20,\ 30$ and 50.

	ℓ_1		ℓ_2		ℓ_1		ℓ_2	
		$N = 10$				$N = 20$		
σ^2	mean	sd	mean	sd	mean	sd	mean	sd
0.5	0.510	0.282	0.511	0.281	0.509	0.181	0.510	0.181
1	1.014	0.514	1.015	0.514	1.017	0.344	1.017	0.344
2	2.022	0.983	2.022	0.983	2.031	0.673	2.031	0.673
4	4.035	1.925	4.034	1.924	4.058	1.329	4.057	1.328
		$N = 30$				$N = 50$		
σ^2	mean	sd	mean	sd	mean	sd	mean	sd
0.5	0.512	0.139	0.512	0.139	0.500	0.102	0.500	0.102
1	1.024	0.269	1.024	0.269	0.999	0.202	0.999	0.202
2	2.048	0.531	2.048	0.531	1.996	0.403	1.996	0.403
4	4.098	1.053	4.098	1.052	3.989	0.805	3.991	0.805

$$
\begin{aligned}
\mathbb{E}\left(-U_1^{(1)}(\sigma^2)\right) &= \frac{2N(N-1)}{(2\sigma^2+1)^2}, \\
\mathbb{E}\left(-U_2^{(1)}(\sigma^2)\right) &= N(N-1)(N-2)\frac{9\sigma^4+12\sigma^2+5}{(3\sigma^4+4\sigma^2+1)^2} \\
&\quad - N(N-2)\frac{18(2N-3)\sigma^6+54(N-2)\sigma^4+(36N-90)\sigma^2+10N-28}{(3\sigma^4+4\sigma^2+1)^2(3\sigma^2+1)} \\
&\quad - 18N(N-2)\frac{1}{(3\sigma^2+1)^3}\left(N\sigma^2+1\right).
\end{aligned}
\tag{14}
$$

The two quantities $\mathbb{E}\left(U_i^2(\sigma^2)\right)$, $i = 1, 2$, of the asymptotic variance given by the expression (11), are approximated by Monte Carlo, with the following formula:

Let write the two expectations as follows:

$$
\mathbb{E}\left(U_i^2(\sigma^2)\right) = \int_{-\infty}^{+\infty} U_i^2(\sigma^2;y)f_\sigma(y)dy; \ i = 1, 2,
$$

where $f_\sigma(y)$ is the density distribution of the variable y obtained from the model defined by (4).

Thus the Monte Carlo approximation of these two integrals are given by

$$
\frac{1}{M}\sum_{j=1}^{M} U_i^2(\sigma^2;y^j); \ i = 1, 2,
$$

7

where $y^1, \ldots, y^M$ are random sample from the density distribution $f_\sigma(.)$ and M fixed to 100000.

Figure 1, gives the asymptotic variance of $\hat{\sigma}^2$ as function of σ^2 provided by the two approaches ℓ_1 and ℓ_2 for the four sizes considered in the simulation study: $N = 10$, 20, 30 and 50.

Figure 1 shows that the four graphics has the same shape: the two asymptotic variances are the same for approximatively $\sigma^2 \leq 1.4$ and for $\sigma^2 > 1.4$, the one provided by ℓ_2 is slightly greater than the asymptotic variance provided by ℓ_1. Both curves are increasing from the value of σ^2 equal 0 to almost 2 and slightly decreasing from the value 2.

5. Discussion

The aim of this paper was to present the first order and a pairwise marginal likelihoods estimation in the normal Beadley-Terry model. These two approaches belonging to the broad class of composite likelihood provide simple estimation of the variance component comparatively to the classical maximum likelihood which is more complex.

The conducted simulation study indicates that the two proposed approaches can estimate the variance component of the model even for moderate sizes. We find little bias in the estimation of the parameter and its standard deviation decreases when the sample size increases.

In terms of asymptotic variance, the two approaches provide the same variance for moderate variance component whereas for large value, the first order likelihood method is slightly better than the pairwise likelihood one.

It would be intersting to find a data for applications to this model with the two proposed approaches. Possible applications could be as for the Thurstone's model in marketing or psychometry for continuous preference with different stimuli.

Finally, this approach could be generalized to the logit version of the model including covariates, which is given by (3). One should evaluate the pairs of probabilities $(y_{ij}, y_{ik}) = (1, 1)$, $(y_{ij}, y_{ik}) = (1, 0)$, $(y_{ij}, y_{ik}) = (0, 1)$ and $(y_{ij}, y_{ik}) = (0, 0)$, which has an analytical expression for the probit version of the model. For the logit version, the scale mixture approximation of Monahan and Stefanski [7] are needed.
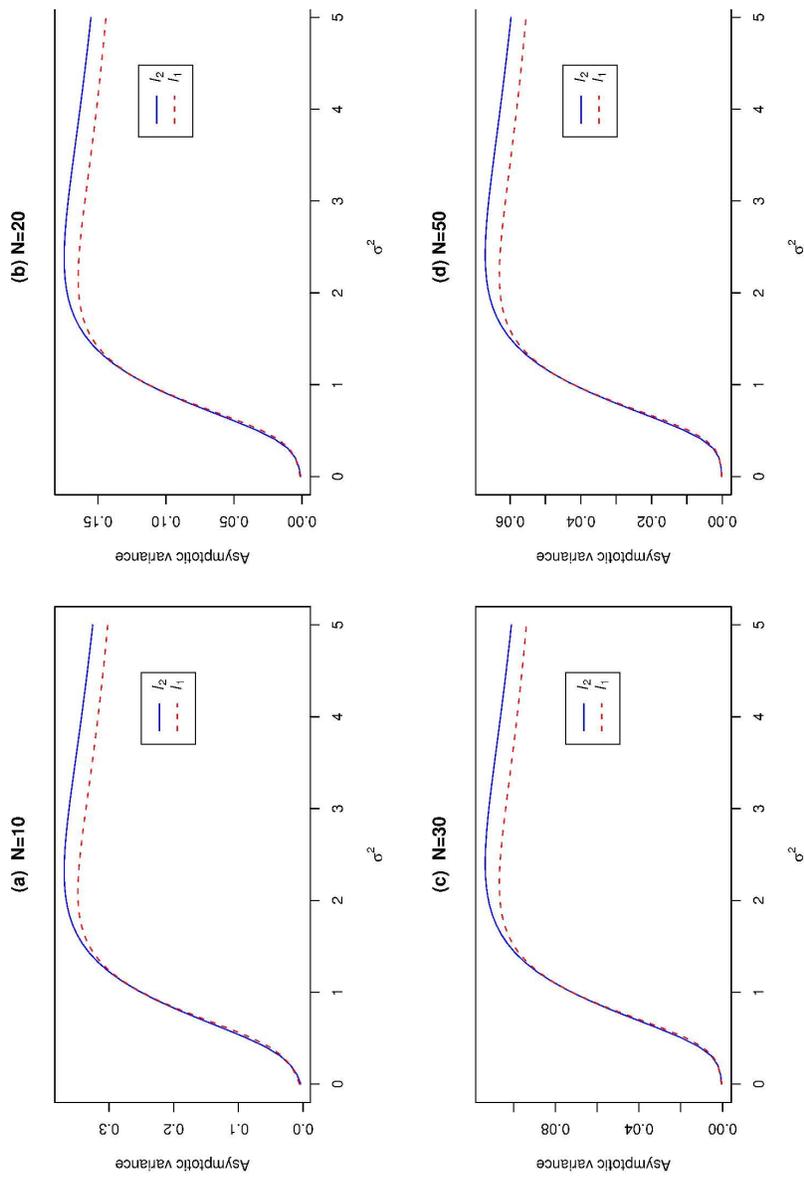
References

[1] R. Bellio and C. Varin, A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling*, **5**, 217–227 (2005).

[2] R. A. Bradley and M. E. Terry, Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika*, **39**, 324–345 (1952).

[3] D. R. Cox and N. Reid, A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737 (2004).

8

Figure 1: Asymptotic variance of $\hat{\sigma}^2$ as function of σ^2 for the two approaches: ℓ_2 (solid curve), ℓ_1 (dashed curve). (a) N=10, (b) N=20, (c) N=30 and (d) N=50.
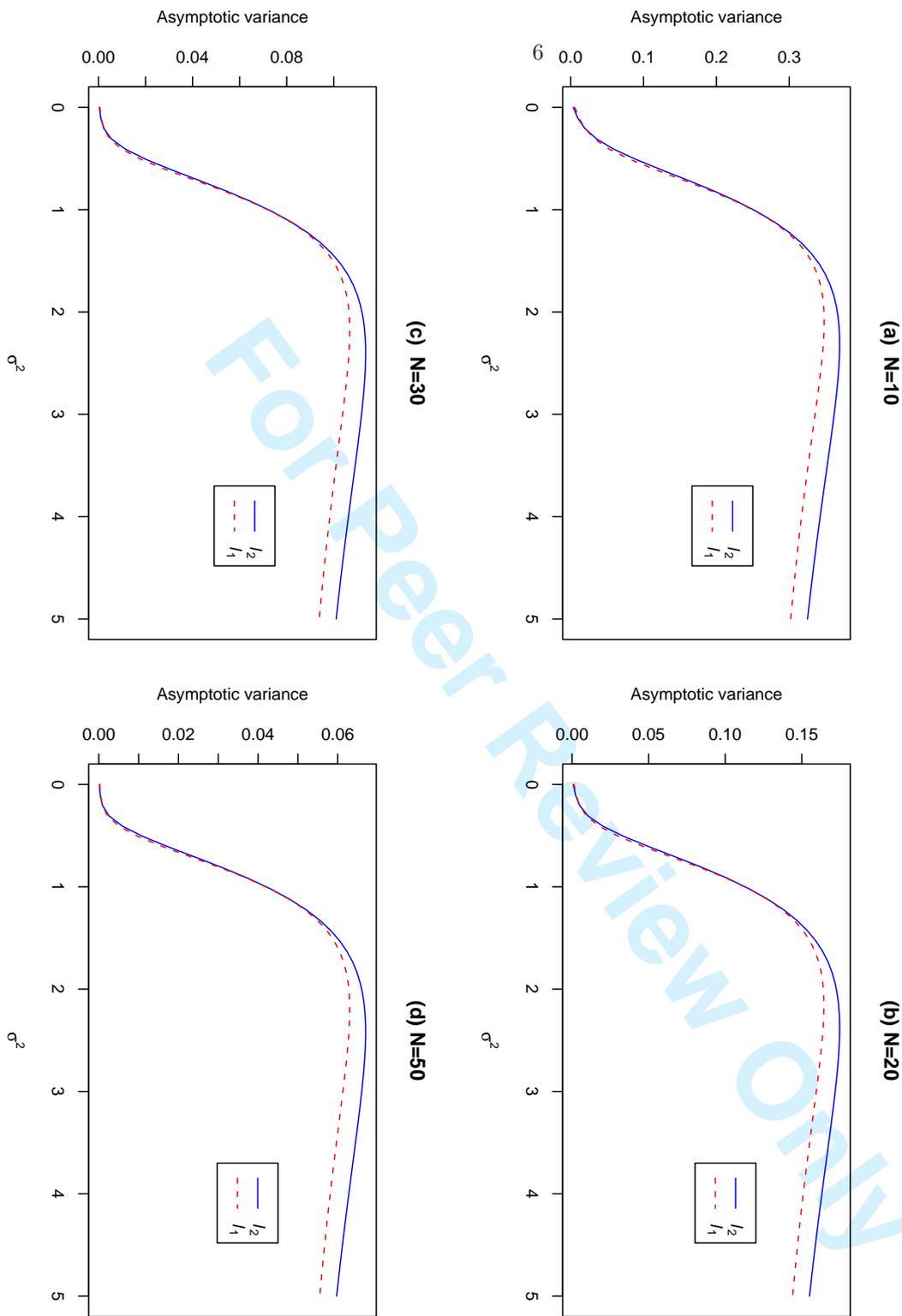
[4] M. L. Feddag and S. Bacci, Pairwise likelihood for the longitudinal mixed Rasch model. *Computational Statistics and Data Analysis*, **53**, 1027-1037 (2009).

[5] D. Firth, Bradley-Terry model in R. *Journal of Statistical software*, **12(1)**, 1-12 (2005).

[6] B. G. Lindsay, Composite likelihood methods. *In Statistical Inference from Stochastic Processes Comtemporary Mathematics 80, Ed. N. U. Prablu. Providence: American Mathematical Society*, 221-39 (1988).

[7] J. F. Monahan and L. A. Stefanski, Normal scale mixture approximations to $F^{\star}(z)$ and computation for the logistic-normal integral. *In Handbook of the Logistic Distribution, N. Balakrishnan (ed.)*, 529-540 (1989). New York: Marcel Dekker.

[8] Y. Takane, Analysis of covariance structure and probabilistic binary choice data. *Communication and Cognition*, **20**, 45-62 (1987).

[9] L. L. Thurstone, A law of comparative judgment. *Psychological Review*, **79**, 281-299 (1927).

[10] L. L. Thurstone, Rank order as psychological method. *Journal of Experimental Psychology*, **14**, 187-201 (1931).

[11] C. Varin, On Composite Marginal Likelihoods. *Advances in Statistical Analysis*, **92(1)**, 1-28 (2008).

10

215x279mm (600 x 600 DPI)