Universidade Federal do Rio Grande do Sul
Instituto de Matemática

Behaviour of the Likelihood Function in Latent Trait Analysis of Binary Data

Maria T. Albanese

Cadernos de Matemática e Estatística
Série A, nº 14, ABR/91
Porto Alegre, abril de 1991

SUMMARY

1- *Introduction*

This paper investigates the adequacy of the maximum likelihood method for estimating the parameters of a single latent logit/probit model for binary response data.

We are interested in checking on whether the likelihood has a smooth unimodal shape, or whether it has multiple relative maxima. The shape of the likelihood around the maximum point will show whether the information matrix will give a good guide to the variability of the estimates. It is a counter-indication to the use of maximum likelihood estimates if there is a flat plateau, or a ridge moving off to infinity.

A badly behaved likelihood function suggests either that a reparametrization is necessary, or that the model is a poor fit for the data, or that the inference is particularly difficult.

We shall study the behaviour of the likelihood function by profiling and an approximate method, using 3 sets of real data. These examples represent a good range of different patterns of parameter estimates and sample sizes.

2- *The Model and its estimation*

We shall suppose that n individuals respond 0 or 1 (no/yes, disagree/agree, for example) to each of p items designed to measure a single latent variable. The response of individual j on item i is written x_{ji}. Individual j has a value z for a latent variable Z, and we assume that Z has a standard normal distribution. Thus the response function of the logit/probit model for individual j on item i may be given by

$$P(X_{ij} = 1 \mid z) = \pi_i(z)$$

where

$$\ln \left[\frac{\pi_i(z)}{1 - \pi_i(z)} \right] = \alpha_{i,0} + \alpha_{i,1} z$$

or

$$\pi_i(z) = \frac{\exp(\alpha_{i,0} + \alpha_{i,1} z)}{1 + \exp(\alpha_{i,0} + \alpha_{i,1} z)} \qquad (1)$$

We assume that the responses to items by an individual are independent given the latent value. This implies that the probability of the response pattern $x_j = (x_{j1}, x_{j2}, \ldots, x_{jp})$ for individual j with latent variable value z is

$$g(x_j | z) = \prod_{i=1}^{p} g_i(x_{ji} | z)$$

$$= \prod_{i=1}^{p} (\pi_i(z))^{x_{ji}} (1-\pi_i(z))^{1-x_{ji}} \qquad (2)$$

This means that the single latent variable Z explains all the association between the responses to different items by an individual.

The difficulty parameter $\alpha_{i,0}$ and the discrimination parameter $\alpha_{i,1}$, $i=1,2,\ldots,p$ are estimated by marginal maximum likelihood method, using a modified E-M algorithm (see Albanese(1990) or Bartholomew(1987)).

Models of this type for binary response were popularised by Bartholomew(1987). Properties of these models were extensively investigated by Albanese(1990).

3- *Comparison between the Profile and an Approximate Method*

Let us consider a single latent variable logit/probit model for fitting binary responses given by (1). Thus the likelihood is a function of $\alpha_{i,0}$ and $\alpha_{i,1}$, $i=1,2,\ldots,p$. A profile likelihood can be obtained for $\alpha_{i,0}$ and $\alpha_{i,1}$ by maximising the likelihood over the remaining variables j, $j=1,2,\ldots,p$ and $j\neq i$. We repeat this procedure to get the profile likelihood at a representative set of values of $(\alpha_{i,0}, \alpha_{i,1})$.

We usually choose to look at the profile likelihood for those parameters for which the likelihood seems to be less satisfactory. One guide to possible poor behaviour is the size of the ML estimate $\hat{\alpha}_{i,1}$. A value of $\hat{\alpha}_{i,1}$ greater than 3.0 may be a sign of a badly behaved likelihood function.

Obtaining the behaviour of the likelihood function using the profile method, described above, takes much computer time, since if we evaluate it for eighty $(\hat{\alpha}_{i,0},\hat{\alpha}_{i,1})$ points we have to maximise the likelihood function that number of times.

Clearly it would be useful to have a quicker method that gives the same information as the profile likelihood.

A simple alternative is to replace the maximisation procedure by some approximation. We have tried using the original marginal ML estimates for $\alpha_{j,0}$ and $\alpha_{j,1}$ for $j \neq i$ instead of maximising again for each new choice of values for $\alpha_{i,0}$ and $\alpha_{i,1}$.

We shall call the latter approach method A, the profile likelihood method B . Put

$L_A(\alpha_{i,0},\alpha_{i,1})$ = loglikelihood value obtained by fixing the remaining parameter at these ML values $\hat{\alpha}_{j,0}$ and $\hat{\alpha}_{j,1}$, $j=1,2,\ldots,p$, $j \neq i$.

$L_B(\alpha_{i,0},\alpha_{i,1})$ = loglikelihood value obtained by maximising over $\hat{\alpha}_{j,0}$ and $\hat{\alpha}_{j,1}$, $j=1,2,\ldots,p$, $j \neq i$.

We apply and compare both methods by contouring the values for L_A, L_B as a function of $\alpha_{i,0}$ and $\alpha_{i,1}$, as defined above, using the subroutine library GINO-SURF. This is done using 3 sets of real data, which represent a good range of different pattern of parameter estimates and sample sizes. The computer program used for fitting the model was FACONE, but also can be done using TWOMISS, using 48 quadrature points. The asymptotic standard deviations of the parameter estimates are obtained by inverting the observed second derivative matrix at the ML solution point.

3.1- Arithmetic Reasoning Test on White Women

The frequency distribution of the response patterns for the first and second examples are samples of the Arithmetic Reasoning Test (ART) from the American Youth on the Armed Services Vocational Aptitude Battery, given by Mislevy (1985). The individuals were classified by sex and colour, but the results given here relate to white and black women.

Table 1- Score distribution and results obtained by fitting a logit/probit model to the Arithmetic Reasoning Test on white women.

Response pattern	Observed frequency	Expected frequency	Total score	Posterior mean
0000	20	26.79	0	-1.20
0010	14	9.83	1	-0.69
1000	23	18.43	1	-0.67
0100	20	15.78	1	-0.58
0001	8	4.86	1	-0.48
1010	9	11.24	2	-0.20
0110	11	10.55	2	-0.11
1100	18	20.21	2	-0.09
0011	2	3.57	2	-0.02
1001	8	6.86	2	0.04
0101	5	6.70	2	0.09
1110	20	21.87	3	0.37
1011	6	8.16	3	0.46
0111	7	8.74	3	0.56
1101	15	17.18	3	0.58
1111	42	37.23	4	1.09
Total	228	228.00	-	-

$\chi^2 = 8.39$ on 6 degrees of freedom ($p = 0.21$)

Thus it is reasonable to infer that the data are consistent with a single latent variable indicating the arithmetic reasoning ability. The scaling given by the posterior mean is consistent with that of the total score because the $\hat{\alpha}_{i,1}$'s are very similar as we can see in Table 2.

Table 2- Parameter estimates and asymptotic standard deviations from fitting a logit/probit model to the Arithmetic Reasoning Test on white women.

Item i	$\hat{\alpha}_{i,1}$	SE($\hat{\alpha}_{i,1}$)	$\hat{\alpha}_{i,0}$	SE($\hat{\alpha}_{i,0}$)	$\hat{\pi}_i$
1	1.04	0.32	0.59	0.17	0.64
2	1.24	0.39	0.56	0.17	0.64
3	1.00	0.30	-0.06	0.16	0.48
4	1.44	0.45	-0.51	0.21	0.38

The parameter estimates show that the items are neither very easy nor too difficult with approximately equal discriminating power.

We apply below methods A and B to discover the behaviour of the likelihood for the data in Table 1 and parameter estimates in Table 2.

Let us choose the first item as our item i. Since all the slope parameters are approximately the same, we would expect to get the same behaviour by choosing any other item.

Figures 1 and 2 have been obtained from 183 pairs $(\hat{\alpha}_{1,0},\hat{\alpha}_{1,1})$, where $\hat{\alpha}_{1,0} \in (-3.50,3.50)$ and $\hat{\alpha}_{1,1} \in (0.10,12.00)$.

According to Table 2, the ML estimates for item 1 are $\hat{\alpha}_{1,1}$=1.04 and $\hat{\alpha}_{1,0}$=0.59. However Figure 1 suggests that the value of the likelihood does not change much along a whole straight line of values for $\hat{\alpha}_{1,1}$ and $\hat{\alpha}_{1,0}$. Close inspection of the input data shows that there is a slight decrease but not enough to show up in the contouring. Figure 2 shows a result much closer to Figure 1 than one might expect, though the peak is slightly better defined. Comparing both graphs this is the only difference between them and it is due to the fact that in method A the likelihood decrease faster than in method B.

The most striking aspect of both figures is the long ridge in the picture going off in a vaguely North Easterly direction. This suggests that there is very little information in the data to choose between $(\hat{\alpha}_{1,0},\hat{\alpha}_{1,1})$ values along that ridge, and casts doubt on the validity of the ML estimates for $(\alpha_{1,0},\alpha_{1,1})$.

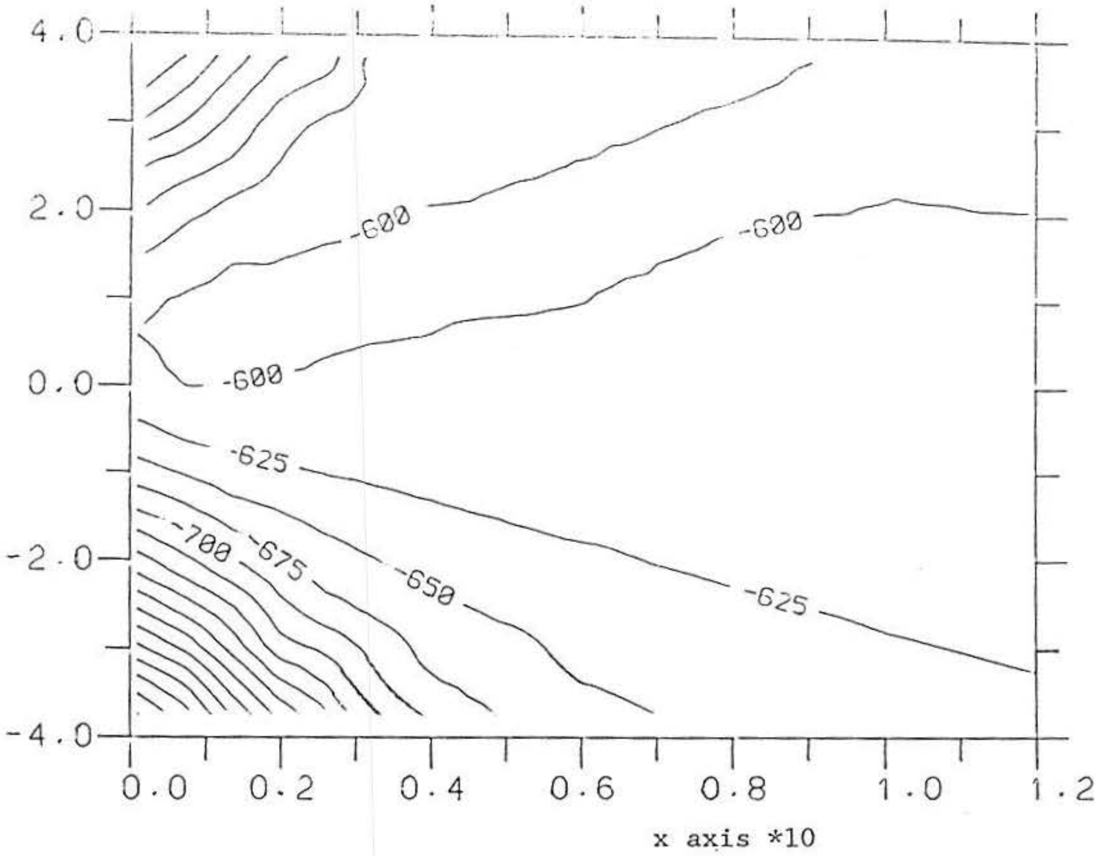

Figure 1- Loglikelihood values as a function of $\hat{\alpha}_{1,1}$ and $\hat{\alpha}_{1,0}$, using method B (profile) to the ART on white women.

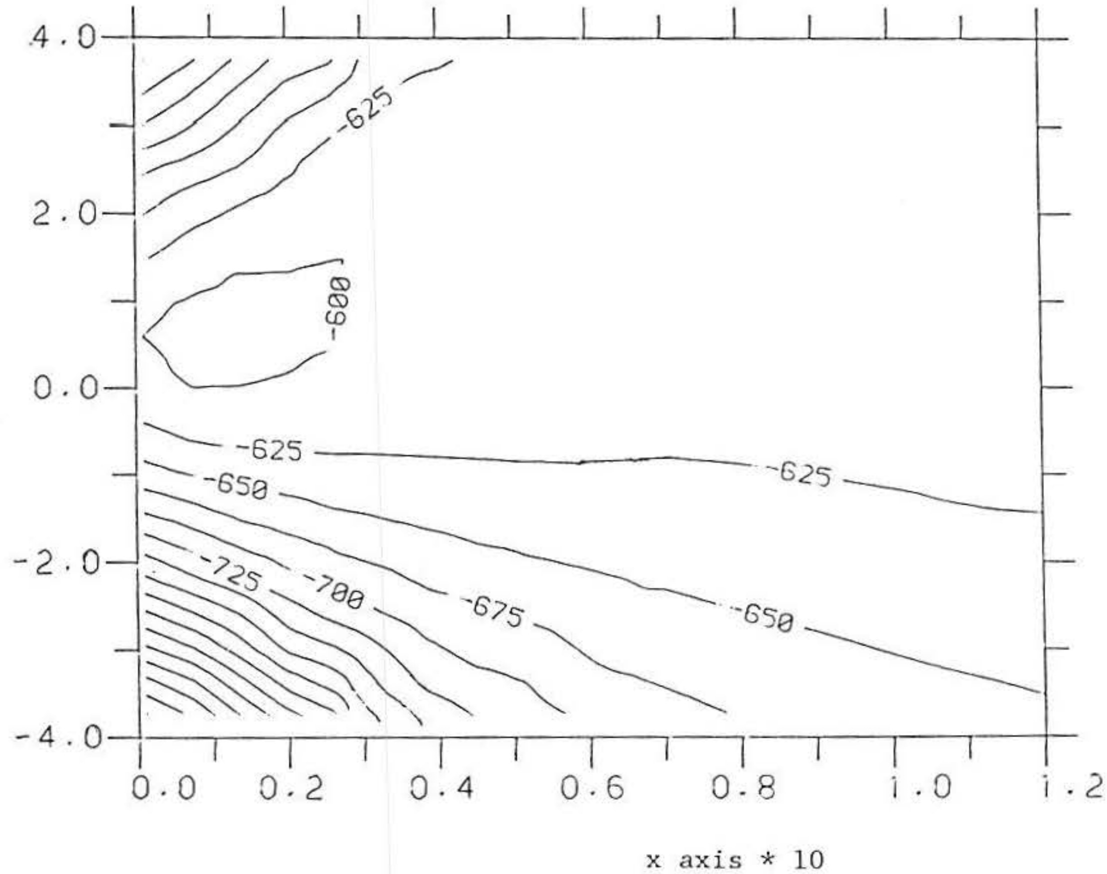

Figure 2- Loglikelihood values as a function of $\hat{\alpha}_{1,1}$ and $\hat{\alpha}_{1,0}$, using approximate method A to the ART on white women.

3.2- Arithmetic Reasoning Test on Black Women

As a second example, we analyse the results of the Arithmetic Reasoning Test on black women.

Table 3- Score distribution and results obtained by fitting a logit/probit model to the Arithmetic Reasoning Test on black women.

Response pattern	Observed frequency	Expected frequency	Total score	Posterior mean
0000	29	28.39	0	-0.84
0001	8	8.19	1	-0.74
0010	7	7.99	1	-0.63
0100	14	14.95	1	-0.63
0011	3	2.36	2	-0.54
0101	5	4.42	2	-0.54
0110	6	4.41	2	-0.43
1000	14	17.74	1	0.49
1001	10	6.88	2	0.58
1010	11	8.90	2	0.69
1100	19	16.77	2	0.70
1011	2	3.54	3	0.79
1101	5	6.66	3	0.79
1110	8	8.84	3	0.92
1111	4	3.62	4	1.02
Total	145	145.00	-	-

$\chi^2 = 6.42$ on 3 degrees of freedom ($p = 0.10$)

As for the test on white women (Table 1) we can also infer that the logit/probit model with one latent variable fits reasonably well.

Note that Table 4 below shows significant differences between the slope parameter estimates($\hat{\alpha}_{i,1}$, $i=1,\ldots,4$).

Table 4- Parameter estimates and asymptotic standard deviations from fitting a logit/probit model to the Arithmetic Reasoning Test on black women.

Item i	$\hat{\alpha}_{i,1}$	$SE(\hat{\alpha}_{i,1})$	$\hat{\alpha}_{i,0}$	$SE(\hat{\alpha}_{i,0})$	$\hat{\pi}_i$
1	14.39	67.78	0.25	4.63	0.56
2	0.38	0.22	-0.33	0.16	0.42
3	0.37	0.24	-0.96	0.20	0.28
4	0.19	0.24	-1.08	0.21	0.25

The results show that item 1, due its large discriminating power, divides the sample into two totally separate groups, those answering the item positively and those who do not. On the other hand, its standard deviation is too large to be trusted. Even for the other $\hat{\alpha}_{1,1}$ the standard deviations may be considered so large that little information is present about them.

Due to the very large slope parameter estimate of item 1 and its strikingly wild standard deviation, it is an obvious choice to look at the behaviour of the likelihood function for 185 pairs $(\hat{\alpha}_{1,0},\hat{\alpha}_{1,1})$.

Since both methods give exactly the same picture, we present just one (Figure 3). There is only a tiny difference between the 185 loglikelihood values from methods A and B, for $\hat{\alpha}_{1,1}$ bigger than 3.0 and any $\hat{\alpha}_{0,1}$.

Figure 3 shows that the likelihood function assumes practically the same values for all $\hat{\alpha}_{1,1}$, and as $\hat{\alpha}_{1,1}$ increases the best values for $\hat{\alpha}_{1,0}$ cover all its interval of variation. Although the subroutine used to draw the graph does not show small differences, analysing the input data we can confirm that the likelihood continues to increase indefinitely, indicating that the actual value for $\hat{\alpha}_{1,1}$ is infinity, which is not sensible.

This is one example where the loglikelihood does not behave appropriately for ML method of estimation.

The broad ridge going from West to East strongly suggests that $\alpha_{1,0}$ is not a meaningful parameter for values of $\alpha_{1,1}$ giving the highest likelihood, since every value of $\hat{\alpha}_{1,0}$ larger than -1.0 will provide the same maximum for loglikelihood function.

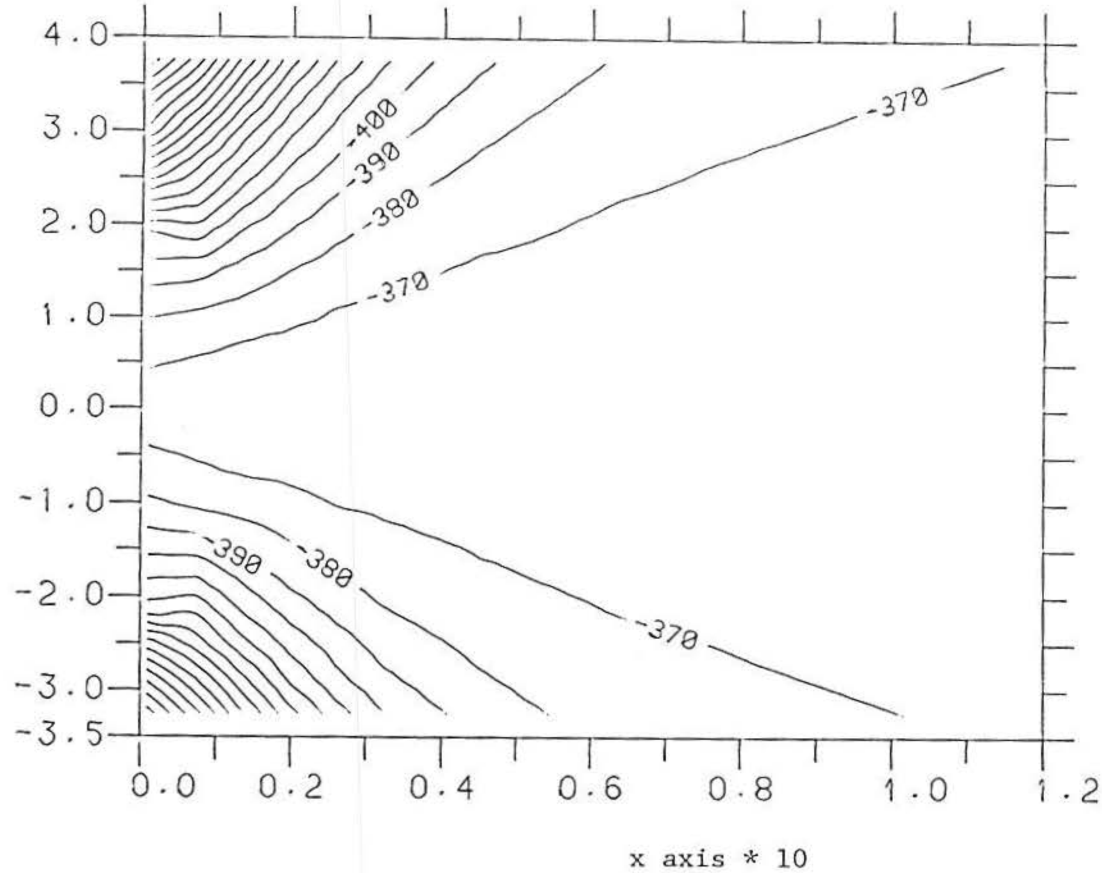

Figure 3- Loglikelihood values as a function of $\hat{\alpha}_{1,1}$ and $\hat{\alpha}_{1,0}$, using methods A or B to the ART on black women.

3.3- Cancer Knowledge

The data in Table 5 comes from a study on knowledge about cancer by Lombard and Doering (1947). Questions were asked about whether or not the following were sources of general information:

(1)radio (2)newspaper (3)solid reading (4)lectures

Table 5 shows that these data are fitted reasonably well by a logit/probit model with one single latent variable as a measure of how well-informed a person is.

Table 5- Score distribution and results obtained by fitting a logit/probit model to the Lombard and Doering's data.

Response pattern	Observed frequency	Expected frequency	Total score	Posterior mean
0000	477	467.37	0	-0.98
1000	63	70.80	1	-0.68
0001	12	16.62	1	-0.66
0010	150	155.93	1	-0.46
1001	7	3.10	2	-0.41
1010	32	33.30	2	-0.22
0011	11	7.98	2	-0.20
1011	4	2.02	3	0.02
0100	231	240.52	1	0.16
1100	94	82.16	2	0.41
0101	13	20.29	2	0.43
0110	378	362.29	2	0.66
1101	12	8.51	3	0.72
1110	169	181.61	3	1.00
0111	45	46.04	3	1.02
1111	31	30.49	4	1.42
Total	1729	1729.00	-	-

$\chi^2 = 11.68$ with 6 degrees of freedom ($0.05<p<0.10$)

The scaling of the sample is not exactly the same when using the total and the component scores. This is due to the large value assumed by $\hat{\alpha}_{2,1}$ as showed in Table 6.

Table 6- Parameter estimates and asymptotic standard deviations from fitting a logit/probit model to the Lombard and Doering data.

Item i	$\hat{\alpha}_{i,1}$	SE($\hat{\alpha}_{i,1}$)	$\hat{\alpha}_{i,0}$	SE($\hat{\alpha}_{i,0}$)	$\hat{\pi}_i$
1	0.72	0.09	-1.29	0.06	0.22
2	3.40	1.14	0.60	0.17	0.64
3	1.34	0.17	-0.14	0.08	0.46
4	0.77	0.14	-2.70	0.18	0.06

The large value for the discriminating power of item 2 indicates that the newspaper has the largest effect on getting information about cancer. Its standard deviation, however, is relatively large. The difficulty parameter estimates range from 'popular source of information' (item 4) to 'not very popular' (item 2).

To carry out the analysis of the behaviour of the likelihood, we used 138 values for $(\hat{\alpha}_{2,0},\hat{\alpha}_{2,1})$, since $\hat{\alpha}_{2,1}$ is very large compared with the other parameter estimates $\hat{\alpha}_{i,1}$.

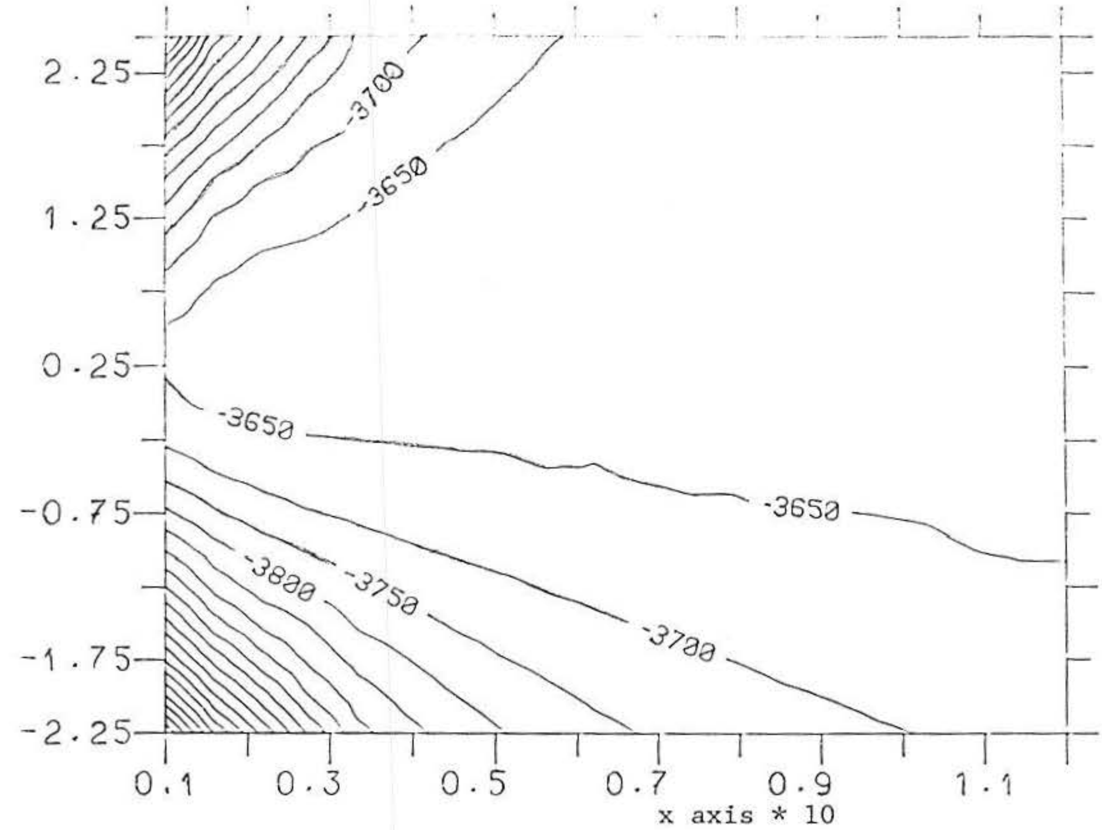

Figure 4- Loglikelihood values as a function of $\hat{\alpha}_{2,1}$ and $\hat{\alpha}_{2,0}$, using method B (profile) for the Lombard and Doering data.

According to Table 6 the likelihood function assumes its maximum value when $\hat{\alpha}_{i,1}$=3.40 and $\hat{\alpha}_{i,0}$=0.60 for item 2. Both Figures 4 and 5 show that $\hat{\alpha}_{2,1}$ could be equal to any number bigger than 1.0 and the range of $\hat{\alpha}_{2,0}$ increases as $\hat{\alpha}_{2,1}$ also increases. As when analysing Figures 1 and 2, this happens because the likelihood values change very little for $\hat{\alpha}_{2,1}$ bigger than 3.40.

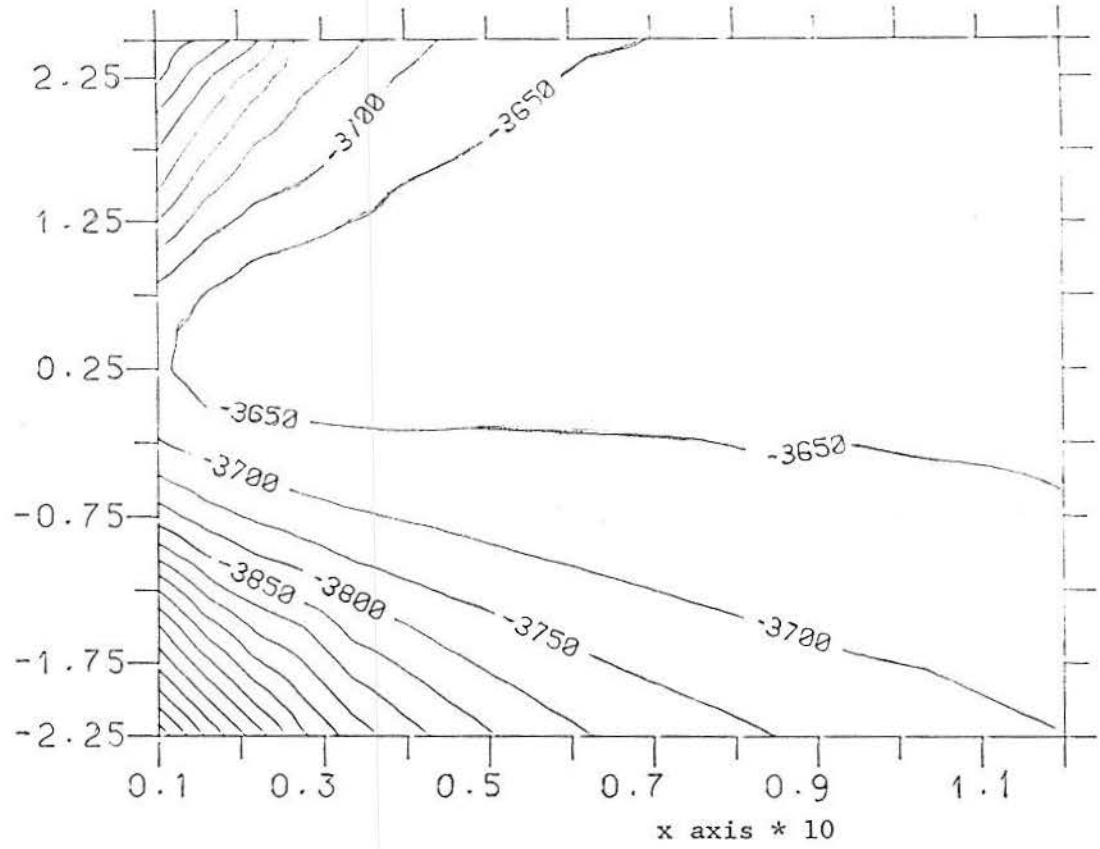

Figure 5- Loglikelihood values as a function of $\hat{\alpha}_{2,1}$ and $\hat{\alpha}_{2,0}$, using approximate method A for the Lombard and Doering data.

In this case too, methods A and B give the same information about the shape of the likelihood function, which does not seem suitable for the ML method.

Conclusion

We have compared 3 sets of data for which a logit/probit model with one latent variable seemed to fit reasonably well.

The results suggest that when one of the $\hat{\alpha}_{i,1}$ is large this probably indicates bad behaviour of the likelihood.

It is difficult to say exactly how large each $\hat{\alpha}_{i,1}$ can be before the ridge in the likelihood appears and the second observed derivatives or the information matrix are not good guides to the variability of this estimates.

There is strong evidence that we can use the approximate method A instead of the profile likelihood, since they give the same information about the behaviour of the likelihood function.

4- Another Look at the Likelihood Function

Working with the contoured likelihood is not always easy, since a lot of points are required and it is hard to see small changes in the likelihood values. It is useful to plot the shape of the likelihood function along the ridge that is evident in Figures 1 to 6. This corresponds to maximising the previously obtained loglikelihood values over $\alpha_{i,0}$. Using the data points $(\hat{\alpha}_{i,1},\hat{\alpha}_{i,0})$ from Figures 1 to 5, results are in the plotting points of Tables 7 to 9 and the likelihood functions in Figures 6 to 8.

Table 7-Maximum loglikelihood value over $\hat{\alpha}_{1,0}$, fixing $\hat{\alpha}_{1,1}$ to the ART on white women.

$\hat{\alpha}_{1,1}$	L_A	L_B
0.0	-601.37	-601.14
1.0	-592.14	-592.12
2.0	-594.59	-594.22
3.0	-598.28	-596.62
4.0	-601.03	-597.87
5.0	-602.99	-598.51
6.0	-604.17	-598.85
7.0	-604.96	-599.06
8.0	-605.50	-599.19
9.0	-605.85	-599.28
10.0	-605.99	-599.33
11.0	-606.13	-599.37

Table 8-Maximum loglikelihood value over $\hat{\alpha}_{1,0}$, fixing $\hat{\alpha}_{1,1}$ to the ART on black women.

$\hat{\alpha}_{1,1}$	L_A	L_B
0.0	-368.08	-367.48
1.0	-365.66	-365.33
2.0	-365.01	-364.90
3.0	-364.83	-364.78
4.0	-364.77	-364.74
5.0	-364.74	-364.72
6.0	-364.72	-364.71
7.0	-364.71	-364.70
8.0	-364.71	-364.69
9.0	-364.70	-364.69
10.0	-364.70	-364.69
11.0	-364.70	-364.69

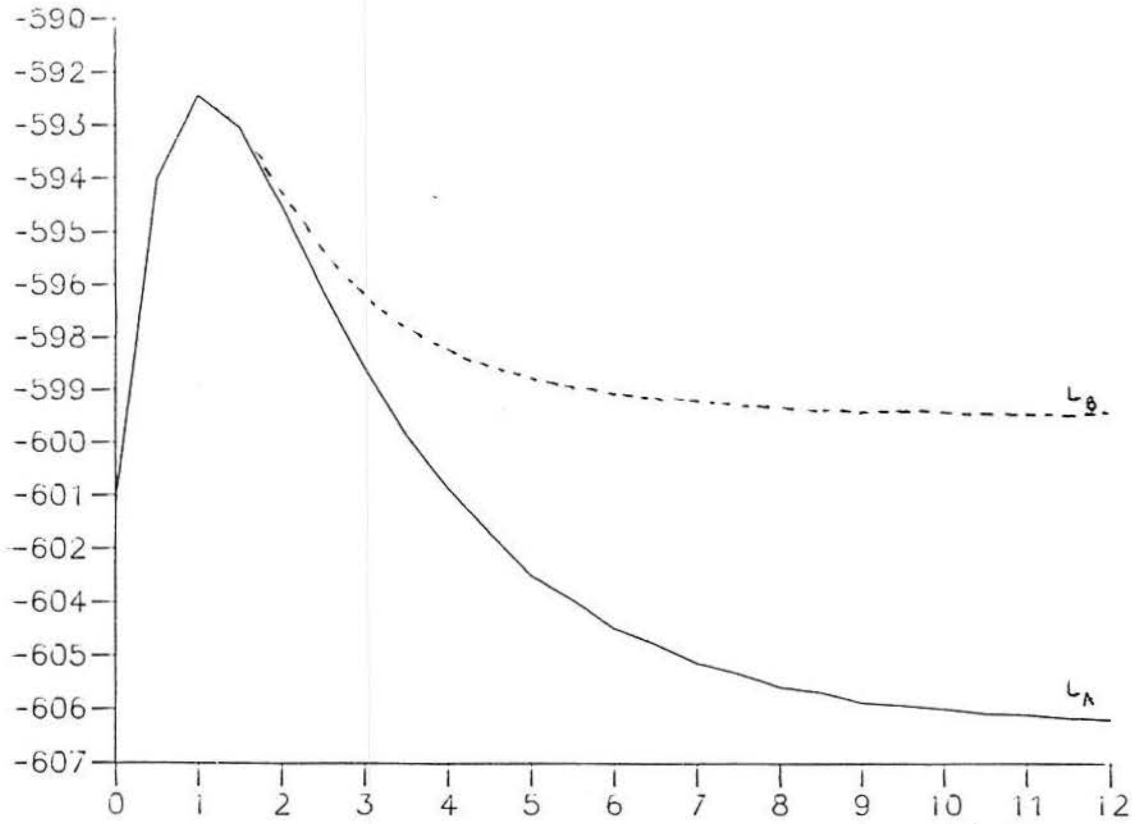

Figure 6- Maximum loglikelihood value over $\hat{\alpha}_{1,0}$ for each $\hat{\alpha}_{1,1}$ fixed, for the ART on white women presented in Table 7.

Table 9- Maximum loglikelihood value over $\hat{\alpha}_{2,0}$, fixing $\hat{\alpha}_{2,1}$ to the Lombard and Doering data.

$\hat{\alpha}_{2,1}$	L_A	L_B	$\hat{\alpha}_{2,1}$	L_A	L_B
0.1	-3758.59	-3755.13	6.0	-3624.05	-3622.90
1.0	-3656.02	-3645.49	7.0	-3624.71	-3623.17
2.0	-3637.84	-3625.53	8.0	-3625.10	-3623.24
3.0	-3622.71	-3622.47	9.0	-3625.29	-3623.27
4.0	-3622.68	-3622.52	10.0	-3625.47	-3623.31
5.0	-3623.54	-3622.80	11.0	-3625.62	-3623.39

Figure 6 shows that both methods give approximately the same loglikelihood values for $\hat{\alpha}_{1,1}$ smaller than 2, increasing up to $\hat{\alpha}_{1,1}$=1.04 (ML estimate) and decreasing faster when using method A than the profile likelihood. This result agrees with our analysis of Figures 1 and 2.

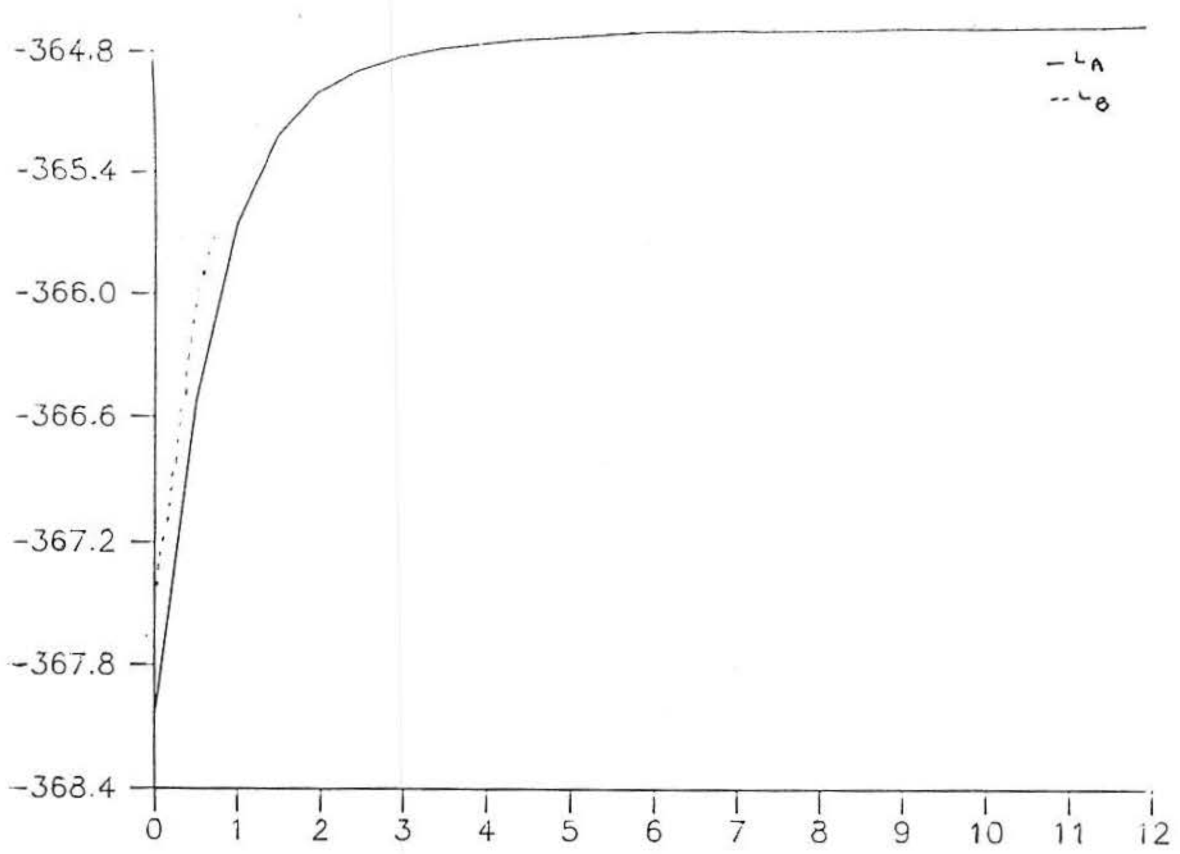

Figure 7- Maximum loglikelihood value over $\hat{\alpha}_{1,0}$ for each $\hat{\alpha}_{1,1}$ fixed, to the ART on black women, presented in Table 8.

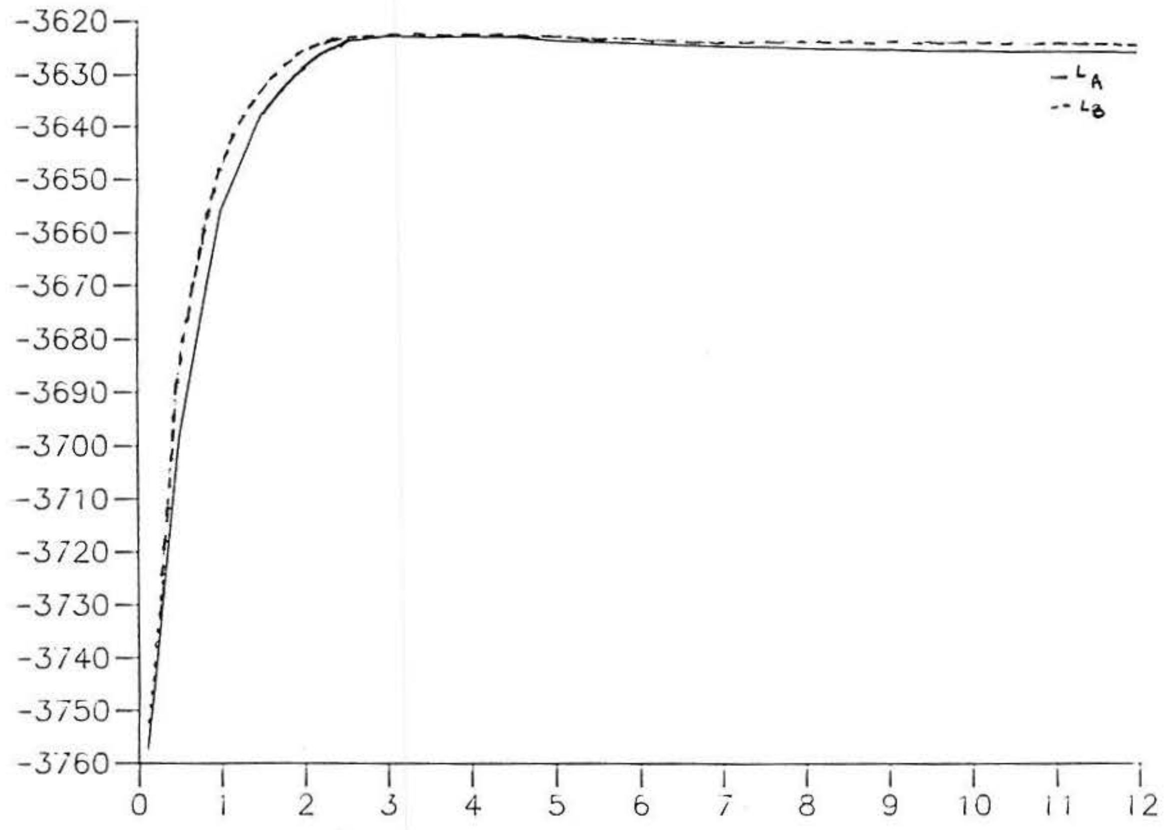

Figure 8- Maximum loglikelihood value over $\hat{\alpha}_{2,0}$ for each $\hat{\alpha}_{2,1}$ fixed, to the Lombard and Doering data, presented in Table 9.

As in the three dimensional graph, Figures 7 and 8 confirm that both methods give roughly the same information about the behaviour of the likelihood. Inspection of the data in Table 8 shows that the likelihood continues increasing, while in Table 9 the likelihood assumes a maximum value, but after that decreases so slightly that the change is insignificant when plotting the data.

<u>Plotting the results for all items</u>

Since approximate method A followed by a simple plot is easy to apply, we shall look at the shape of the likelihood for all items, instead of only one, for the ART on white and black women, and the Lombard and Doering data.

Table 10- Maximum loglikelihood value, $L_A(i)$, over $\hat{\alpha}_{i,0}$, fixing $\hat{\alpha}_{i,1}$, i=1,2..,4, to the ART on white women, using approximate method A.

$\hat{\alpha}_{i,1}$	$L_A(1)$	$L_A(2)$	$L_A(3)$	$L_A(4)$
0.0	-601.37	-603.57	-601.63	-606.09
1.0	-592.14	-592.22	-592.05	-592.74
2.0	-595.06	-593.84	-595.18	-592.68
3.0	-598.28	-595.98	-599.43	-595.06
4.0	-601.22	-598.18	-602.58	-596.31
5.0	-602.88	-599.46	-604.71	-597.47
6.0	-604.17	-600.42	-606.19	-598.33
7.0	-604.90	-600.99	-607.26	-598.71
8.0	-605.50	-601.37	-608.07	-599.13
9.0	-605.77	-601.66	-608.71	-599.27
10.0	-606.10	-601.79	-609.22	-599.44
11.0	-606.13	-601.93	-609.63	-599.51

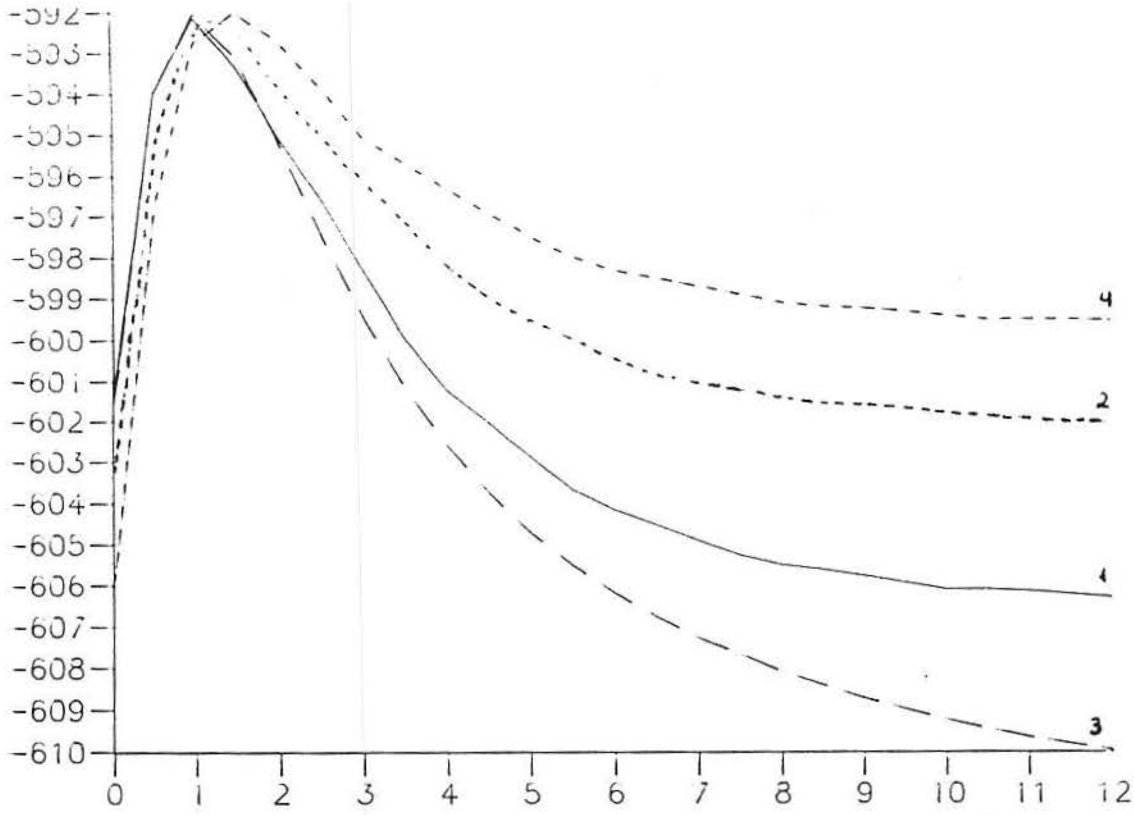

Figure 9- Maximum loglikelihood value over $\hat{\alpha}_{i,0}$ for each $\hat{\alpha}_{i,1}$ fixed, i=1,...,4, to the ART on white women, presented in Table 10, using approximate method A.

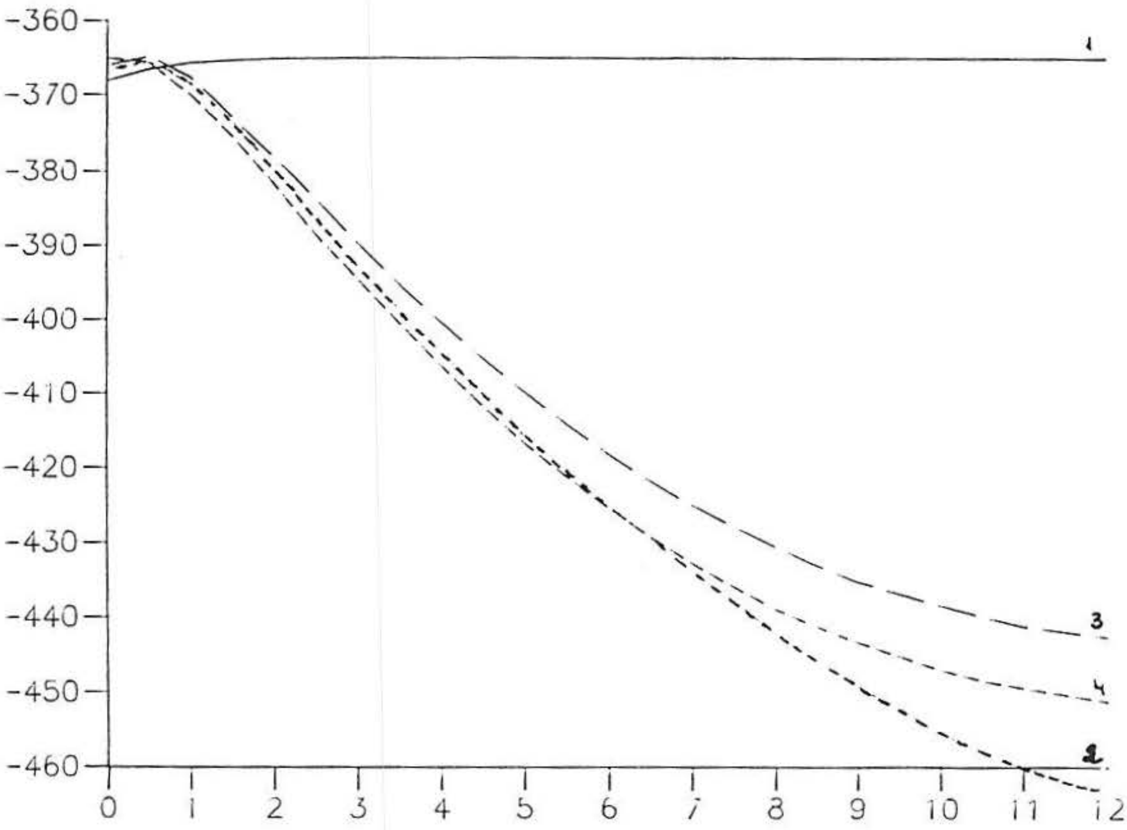

Figure 10- Maximum loglikelihood value over $\hat{\alpha}_{i,0}$ for each $\hat{\alpha}_{i,1}$ fixed, i=1,...,4, to the ART on black women, presented in Table 11, using approximate method A.

Table 11- Maximum loglikelihood values, $L_A(i)$, over $\hat{\alpha}_{i,0}$ fixing $\hat{\alpha}_{i,1}$, i=1,2,..,4, to the ART on black women, using approximate method A.

$\hat{\alpha}_{i,1}$	$L_A(1)$	$L_A(2)$	$L_A(3)$	$L_A(4)$
0.0	-368.08	-366.79	-366.01	-365.07
1.0	-365.66	-368.07	-367.59	-369.97
2.0	-365.01	-379.56	-377.88	-381.54
3.0	-364.83	-392.35	-389.50	-394.39
4.0	-364.77	-404.22	-400.20	-406.07
5.0	-364.74	-415.00	-409.66	-416.58
6.0	-364.72	-424.78	-418.04	-425.29
7.0	-364.71	-433.57	-424.92	-432.68
8.0	-364.71	-441.96	-430.51	-438.91
9.0	-364.70	-449.16	-435.21	-443.32
10.0	-364.70	-455.29	-438.39	-446.98
11.0	-364.70	-460.03	-441.16	-449.43

Table 12- Maximum loglikelihood values, $L_A(i)$, over $\hat{\alpha}_{i,0}$ fixing $\hat{\alpha}_{i,1}$, i=1,2,..,4, to the Lombard and Doering data, using the approximate method A.

$\hat{\alpha}_{i,1}$	$L_A(1)$	$L_A(2)$	$L_A(3)$	$L_A(4)$
0.0	-3666.35	-3790.44	-3755.44	-3640.81
1.0	-3626.13	-3660.79	-3630.09	-3624.84
2.0	-3680.93	-3627.58	-3635.98	-3650.53
3.0	-3739.91	-3622.71	-3666.38	-3682.70
4.0	-3785.70	-3623.91	-3693.62	-3710.37
5.0	-3818.87	-3623.95	-3714.12	-3730.04
6.0	-3842.86	-3624.05	-3728.73	-3743.71
7.0	-3859.40	-3624.97	-3740.07	-3752.89
8.0	-3871.61	-3625.30	-3749.21	-3759.02
9.0	-3881.08	-3625.29	-3756.80	-3763.18
10.0	-3886.42	-3625.97	-3762.16	-3765.29
11.0	-3890.69	-3625.62	-3766.36	-3766.72

Figure 9 shows that whichever item we choose, all items are well-behaved. However it is interesting to point out that the order of the curves from the top to the botton of the page is inversely related to size of the $\hat{\alpha}_{i,1}$,i=1,...,4, since here they all have the same coefficient of variation (0.31).

As we can see in Figure 10, the bad behaviour of the likelihood is indicated by item 1, with very large $\hat{\alpha}_{1,1}$ and its large standard deviation. Item 2 and 3 present very similar $\hat{\alpha}_{i,1}$ (0.38 and 0.37), with coefficient of variation 0.58 and 0.65, respectively, but the latter loglikelihood decreases slowly. The value of $\hat{\alpha}_{4,1}$ is half the size of item 2 or 3, but with a large coefficient of variation (1.26).

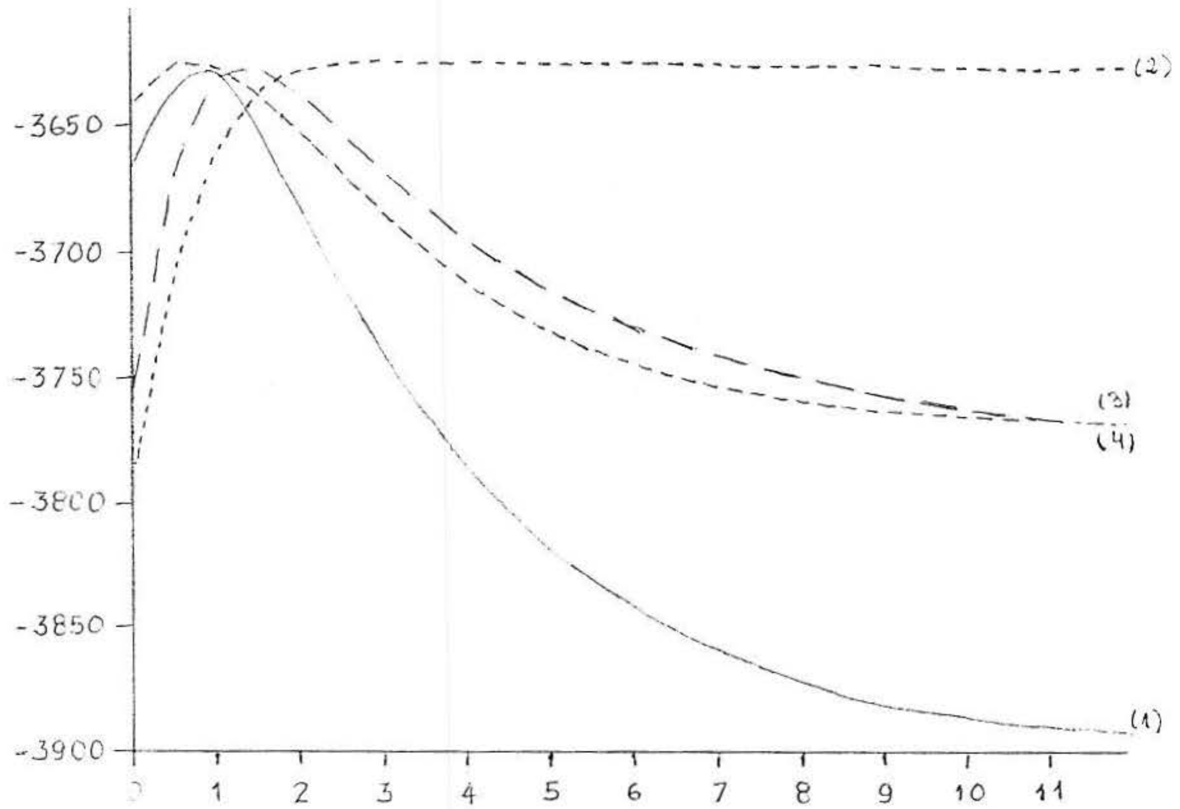

Figure 11- Maximum loglikelihood value over $\hat{\alpha}_{i,0}$ for each $\hat{\alpha}_{i,1}$ fixed, i=1,...,4, for the Lombard and Doering data, presented in Table 12, using approximate method A.

Working through the values of $\hat{\alpha}_{i,1}$ and Figure 11 we see that $\hat{\alpha}_{3,1}$ is bigger than $\hat{\alpha}_{4,1}$ (1.34 and 0.77), but the former has a smaller coefficient of variation, and both items give approximately the same likelihood shape. Item 1 has the smallest $\hat{\alpha}_{1,1}$ and the smallest coefficient of variation (0.12) and the biggest likelihood function decrease, while item 2 has a large value for $\hat{\alpha}_{i,1}$ and large coefficient of variation (0.34) and effectively its likelihood function never decreases.

Conclusions

These results suggest that there is strong evidence that we can look at the behaviour of the likelihood function by the approximate method A, using a graph like those in Figures 6 to 8. However, we should remember that the likelihood values from this method are equal or smaller than the real values and small decreases in the likelihood function should actually be still smaller.

Finally we can conclude that large discriminating parameter estimate ($\hat{\alpha}_{i,1}$) and large standard deviation point to bad likelihood behaviour. The results also indicate that for the same test the shapes of the approximate profile likelihoods obtained for different items i are related to the size of $\hat{\alpha}_{i,1}$ and its coefficient of variation.

5- *Reparametrization*

The investigation of the behaviour of the likelihood function that has been carried out suggests that, at least for the ART on black women (Table 3) and the Lombard and Doering data (Table 5) a reparametrization is necessary.

We have worked through many reparametrizations, as for example,

$$\hat{\alpha}^{*}_{i,1} = \hat{\alpha}_{i,1} / (1 + \exp(\hat{\alpha}_{i,1}))$$

$$\hat{\alpha}^{*}_{i,1} = 1 / \hat{\alpha}_{i,1}$$

$$\hat{\alpha}^{*}_{i,1} = \hat{\alpha}_{i,1} / (1 + \hat{\alpha}^{2}_{i,1})^{\frac{1}{2}}$$

$$\hat{\alpha}^{*}_{i,0} = \hat{\alpha}_{i,0} / (1 + \hat{\alpha}^{2}_{i,1})^{\frac{1}{2}}$$

$$\hat{\alpha}^{*}_{i,0} = - \hat{\alpha}_{i,0} / \hat{\alpha}_{i,1}$$

where i=1 for the ART on black women data and i=2 for the Lombard and Doering data.

We shall present the results just for the reparametrizations that gave useful results, in the sense that it showed better behaviour of the likelihood function, that is, for

$$\hat{\alpha}_{i,0} = \hat{\alpha}^{*}_{i,0} / (1 + \hat{\alpha}^{2}_{i,1})^{\frac{1}{2}} \quad \text{and} \quad \hat{\alpha}^{*}_{i,1} = \hat{\alpha}_{i,1} / (1 + \hat{\alpha}^{2}_{i,1})^{\frac{1}{2}}$$

using the profile and the approximate methods (B and A, respectively).

5.1- Arithmetic Reasoning Test on White Women

The data related to this example are presented in Table 1.

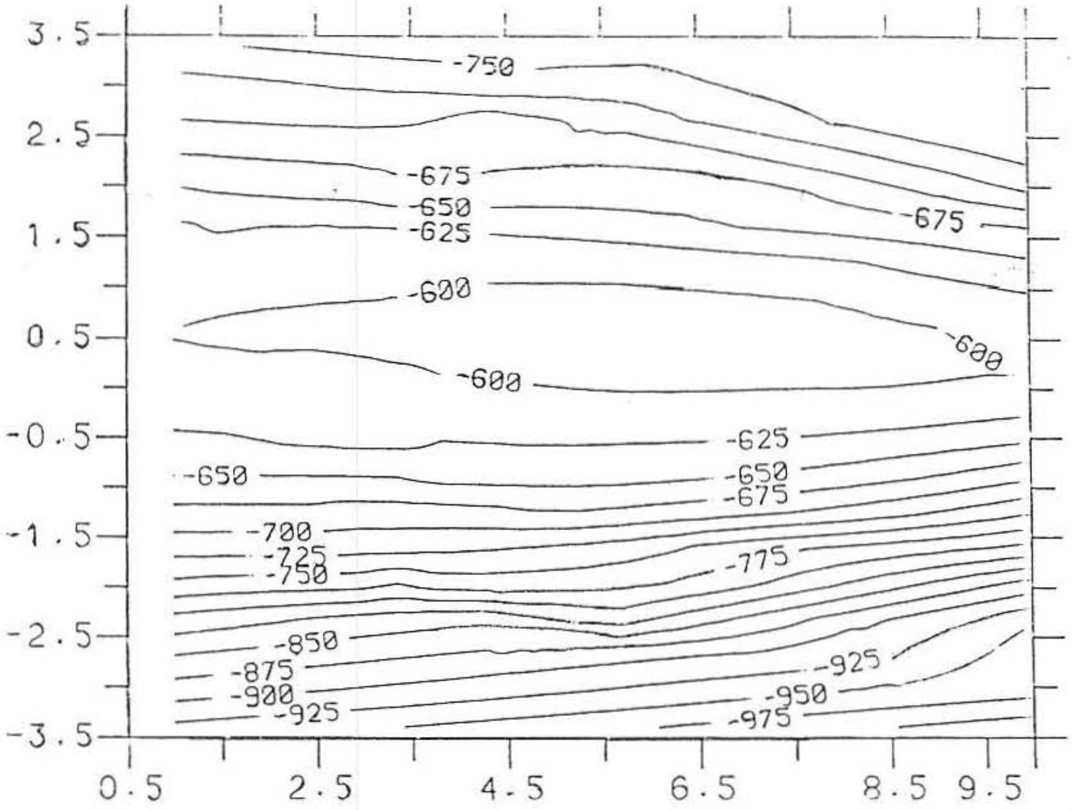

Figure 12- Loglikelihood values as a function of $\hat{\alpha}^{*}_{1,1}$ and $\hat{\alpha}^{*}_{1,0}$, using method B (profile) to the ART on white women.

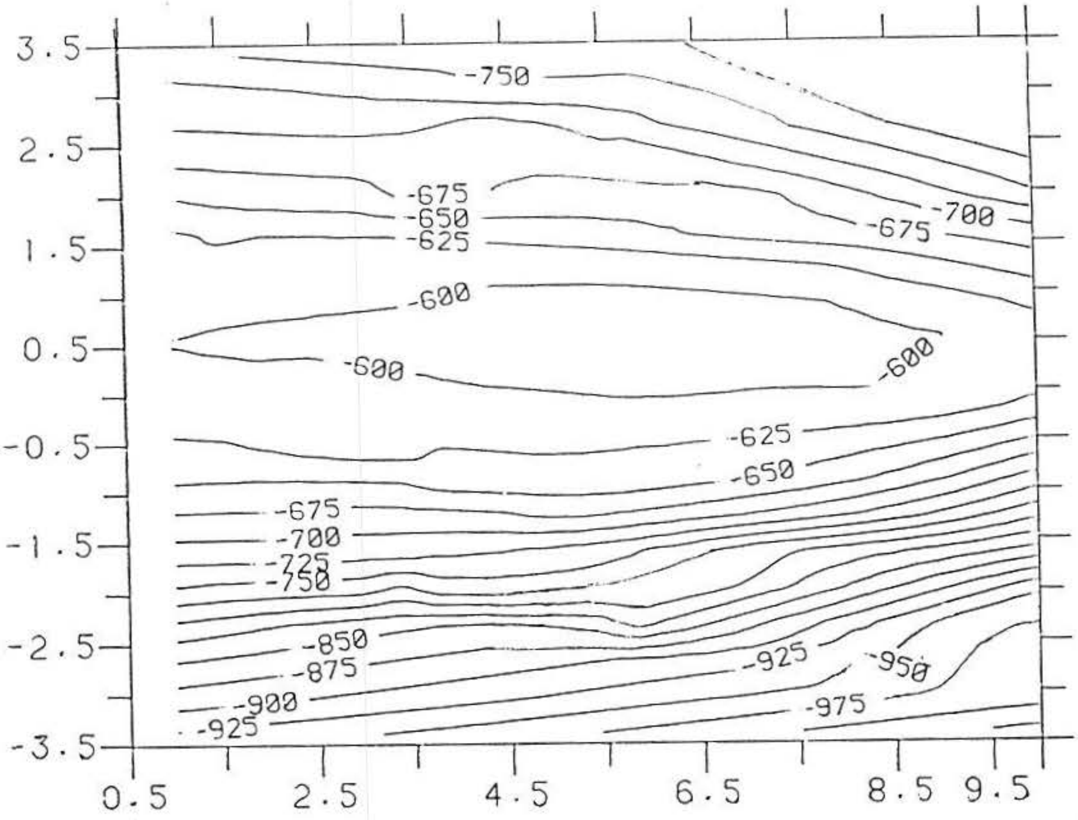

Figure 13- Loglikelihood values as a function of $\hat{\alpha}^*_{1,1}$ and $\hat{\alpha}^*_{1,0}$, using approximate method A to the ART on white women.

Both Figures 12 and 13 show the same shape of the likelihood function and their parallel and almost horizontal lines indicate that the values of the loglikelihood almost do not change for a fixed $\hat{\alpha}^*_{1,0}$ over all range of $\hat{\alpha}^*_{1,1}$. There is a peak inside the ellipse, although the contouring does not show the small differences in the loglikelihood values. We can see it in Figure 15, where we have the maximum loglikelihood values over $\hat{\alpha}^*_{1,0}$ for each $\hat{\alpha}^*_{1,1}$ fixed.

That only one line represents the behaviour of the likelihood function in Figure 14 is due to the fact that methods A and B give the same results for all values assumed by $\hat{\alpha}^*_{1,0}$.

From Figures 14 and 15 we can see that the loglikelihood function behaves well in both reparametrizations and that the maximum loglikelihood values for $\hat{\alpha}^*_{1,1}$ range in a larger interval than for $\hat{\alpha}^*_{1,0}$, since maximum loglikelihood $\hat{\alpha}^*_{1,1} \in (-916.11;-592.27)$ while the maximum loglikelihood $\hat{\alpha}^*_{1,0} \in (-606.23;-592.14)$.

Comparing Figures 1, 2 and 6 with 12 to 15 we can conclude that the reparametrization $\hat{\alpha}^*_{1,0}$ and $\hat{\alpha}^*_{1,1}$ give a likelihood function with better behaviour than $\hat{\alpha}_{1,0}$ and $\hat{\alpha}_{1,1}$.

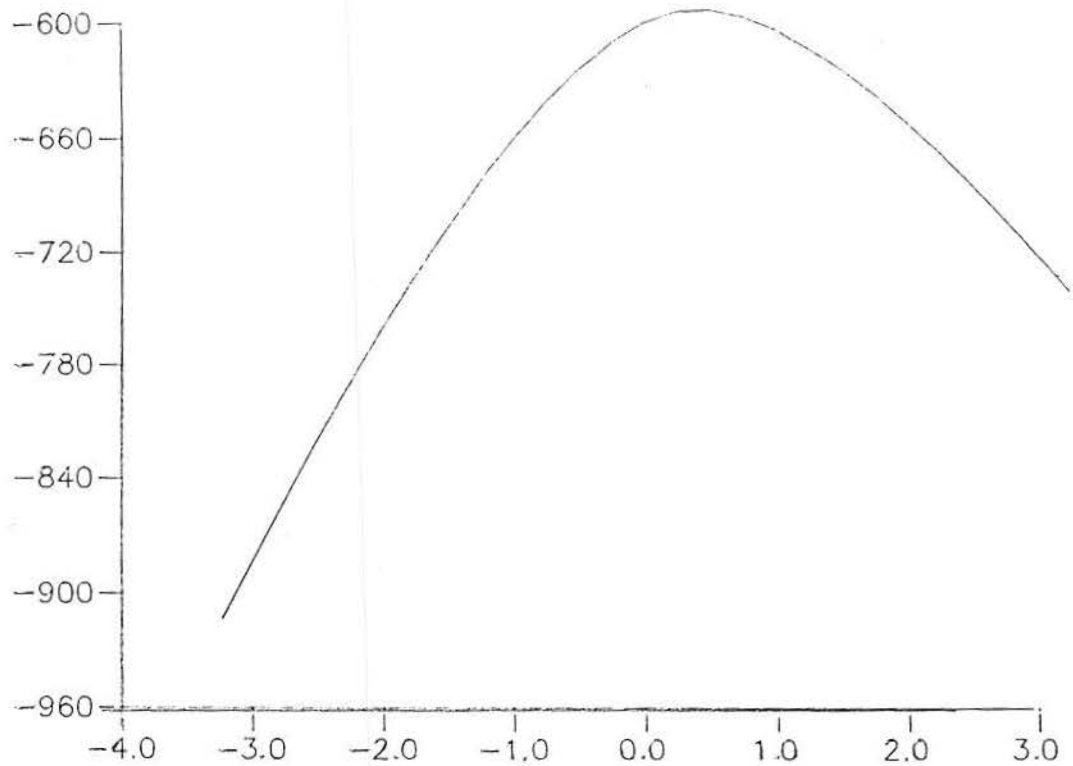

Figure 14- Maximum loglikelihood value over $\hat{\alpha}_{1,1}$ for each $\hat{\alpha}^*_{1,0}$ fixed to the ART on white women, using methods A and B.

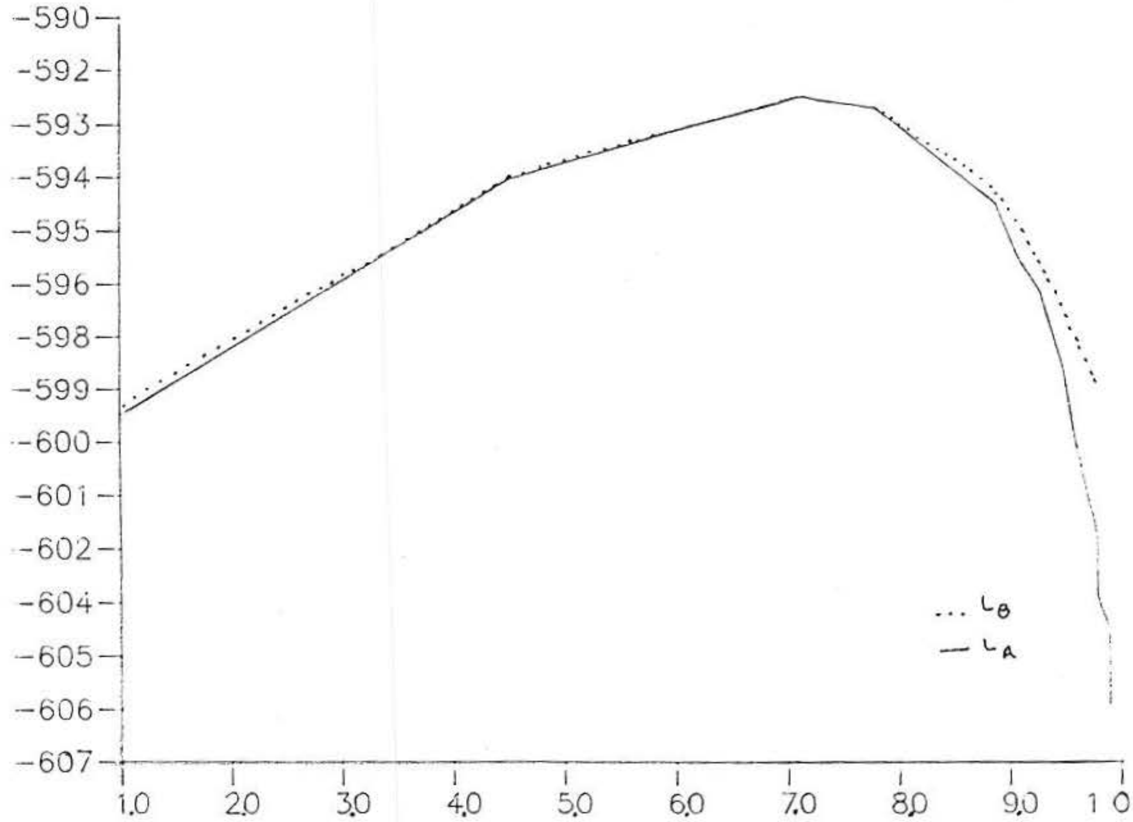

Figure 15- Maximum loglikelihood value over $\hat{\alpha}_{1,0}$ for each $\hat{\alpha}^*_{1,1}$ fixed to the ART on white women, using methods A and B.

5.2- Arithmetic Reasoning Test on Black Women

The following graphs refer to the data in Table 3.

As in the first example, the Figures 16 and 17, 18 and 19 shows the same shape for the likelihood function, whether using profile (method B) or approximate method A.

Figures 16 and 17 show that the parallel lines are becoming horizontal as the loglikelihood function approximates to the maximum value, where we can see a broad bridge going from West to East and $\hat{\alpha}^*_{1,1}$ assuming all values while $\hat{\alpha}^*_{1,0}$ ranges from -0.35 to 0.35.

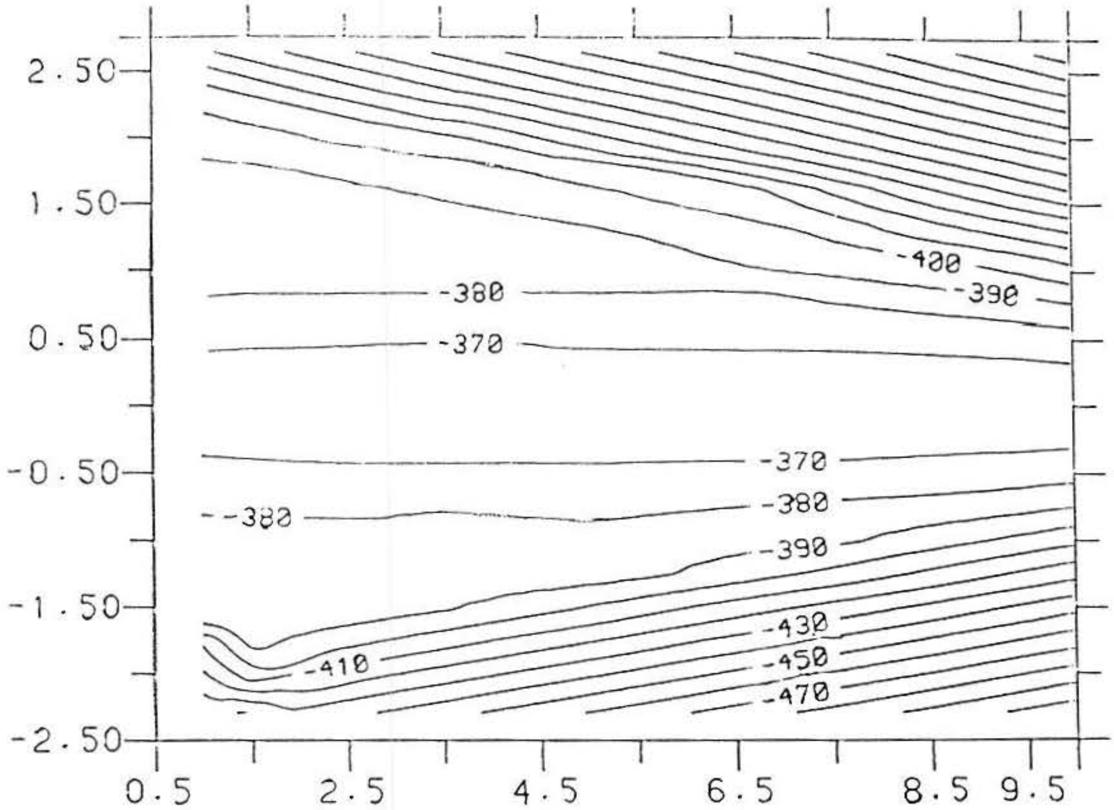

Figure 16- Loglikelihood values as a function of $\hat{\alpha}^*_{1,1}$ and $\hat{\alpha}^*_{1,0}$, using profile method for the ART on black women.

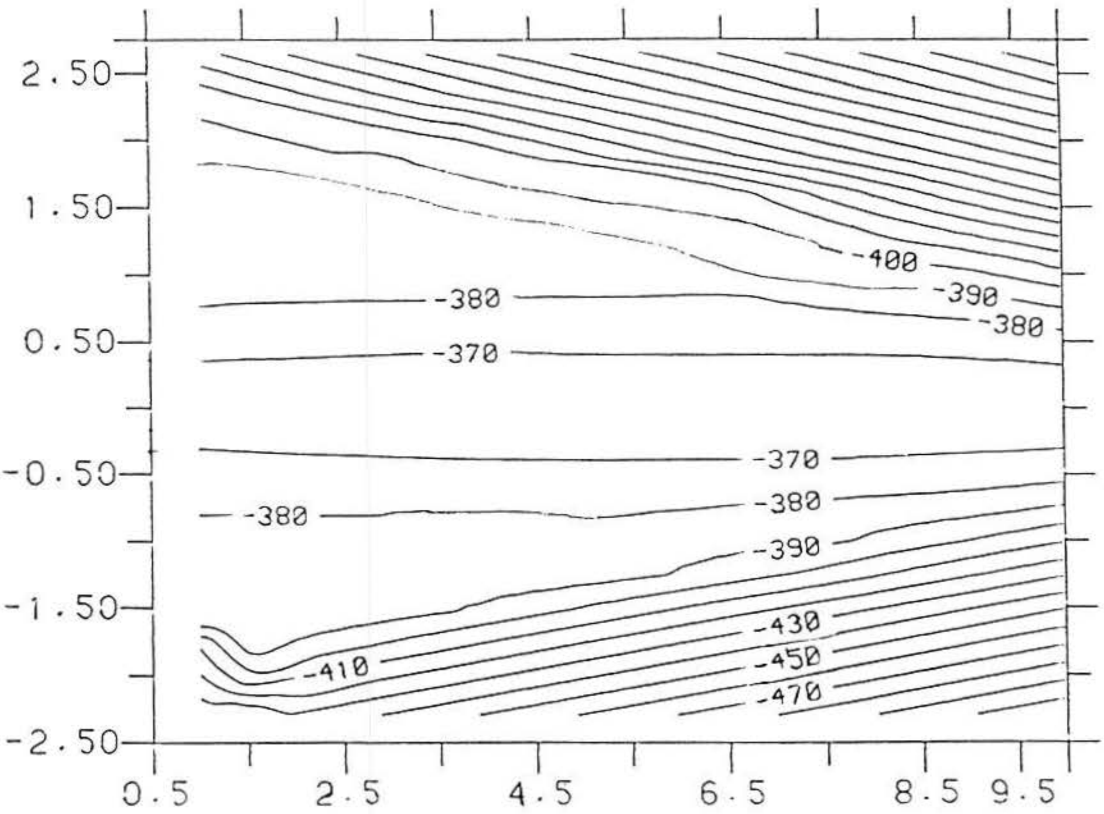

Figure 17- Loglikelihood values as a function of $\hat{\alpha}^*_{1,1}$ and $\hat{\alpha}^*_{1,0}$, using approximate method A for the ART on black women.

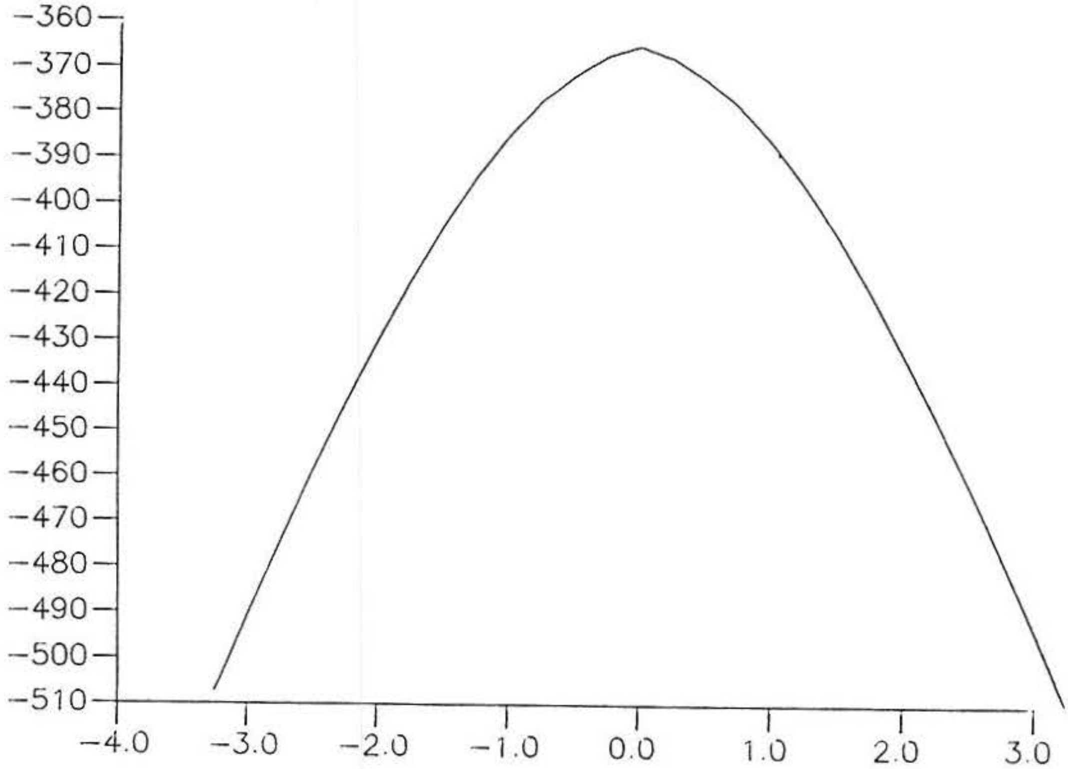

Figure 18- Maximum loglikelihood value over $\hat{\alpha}_{1,1}$ for each $\hat{\alpha}^*_{1,0}$ fixed for the ART on black women, using methods A and B.

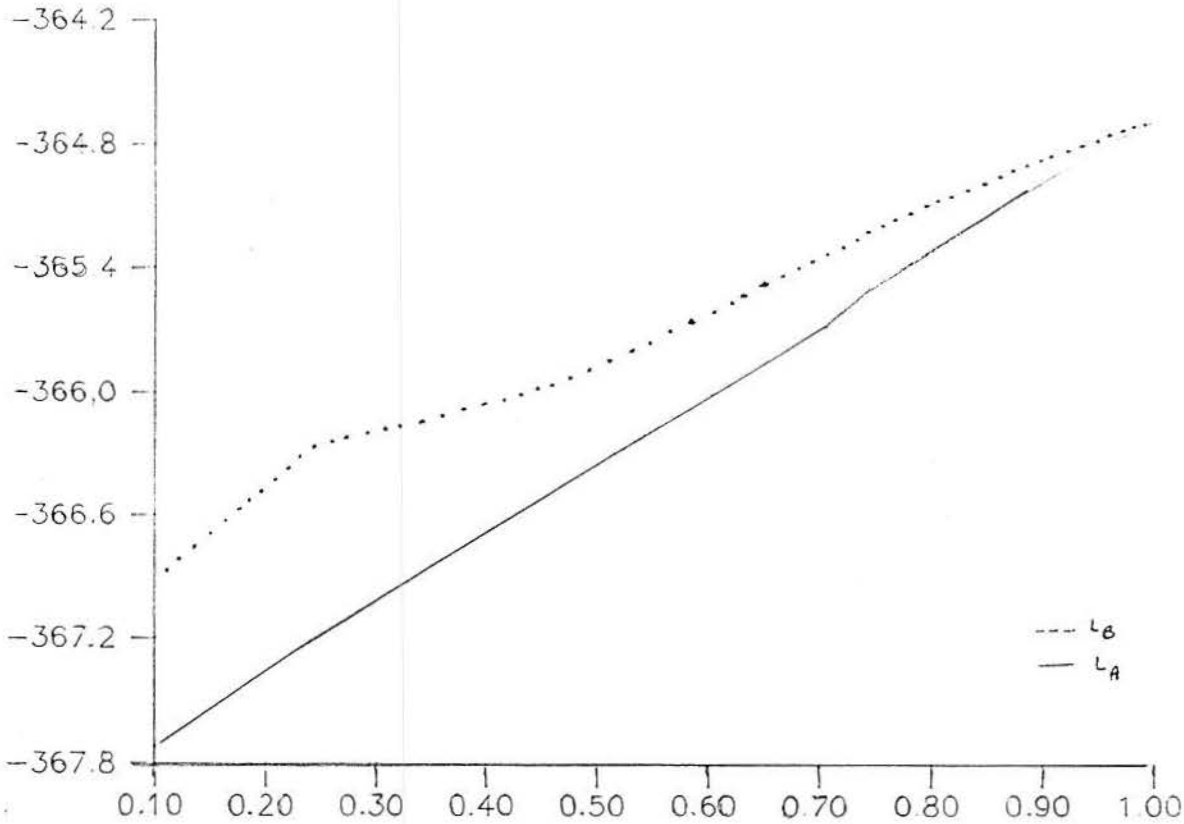

Figure 19- Maximum loglikelihood value over $\hat{\alpha}_{1,0}$ for each $\hat{\alpha}^*_{1,1}$ fixed for the ART on black women, using methods A and B.

The apparent increased likelihood function shown in Figure 19 is,actually, almost constant since it assumes values in a small interval (in the profile method from -367.48 to -364.68 and in the approximate method from -368.08 to -364.69), corresponding to an increase of 0.9%. Thus the reparametrization $\hat{\alpha}^*_{1,1}$ provides a likelihood function that is monotone increasing.

On the other hand, Figure 18 indicates that the reparametrization $\hat{\alpha}^*_{1,0} = \hat{\alpha}_{1,0}/(1+\hat{\alpha}^2_{1,1})^{\frac{1}{2}}$ works very well, since the likelihood function is unimodal, assuming values from -510.35 to -364.68 in both methods (profile and approximate).

5.3- Cancer Knowledge

This example corresponds to the Lombard and Doering data(Table 5).

The small difference between methods A and B (Figures 20 and 21) is because the loglikelihood function for $\hat{\alpha}_{2,1} < 1.1$ in the profile method is bigger than in the approximate method.

The behaviour of the likelihood function after reparametrization in these example is very similar to the former one.

Although Figure 23 seems to show an increased loglikelihood function for $\hat{\alpha}^{*}_{2,1}$, it is almost constant, since it ranges from -3758.59 to -3755.13 which represents a small increase of 3.8%. Therefore the reparametrization $\hat{\alpha}^{*}_{2,1}$ provides a likelihood function that is monotone increasing.

As in the preceding example, the only useful reparametrization is given by $\hat{\alpha}^{*}_{2,0}$, as we can see in Figure 22 an unimodal likelihood function that assumes values between -5813.38 and -3625.14.

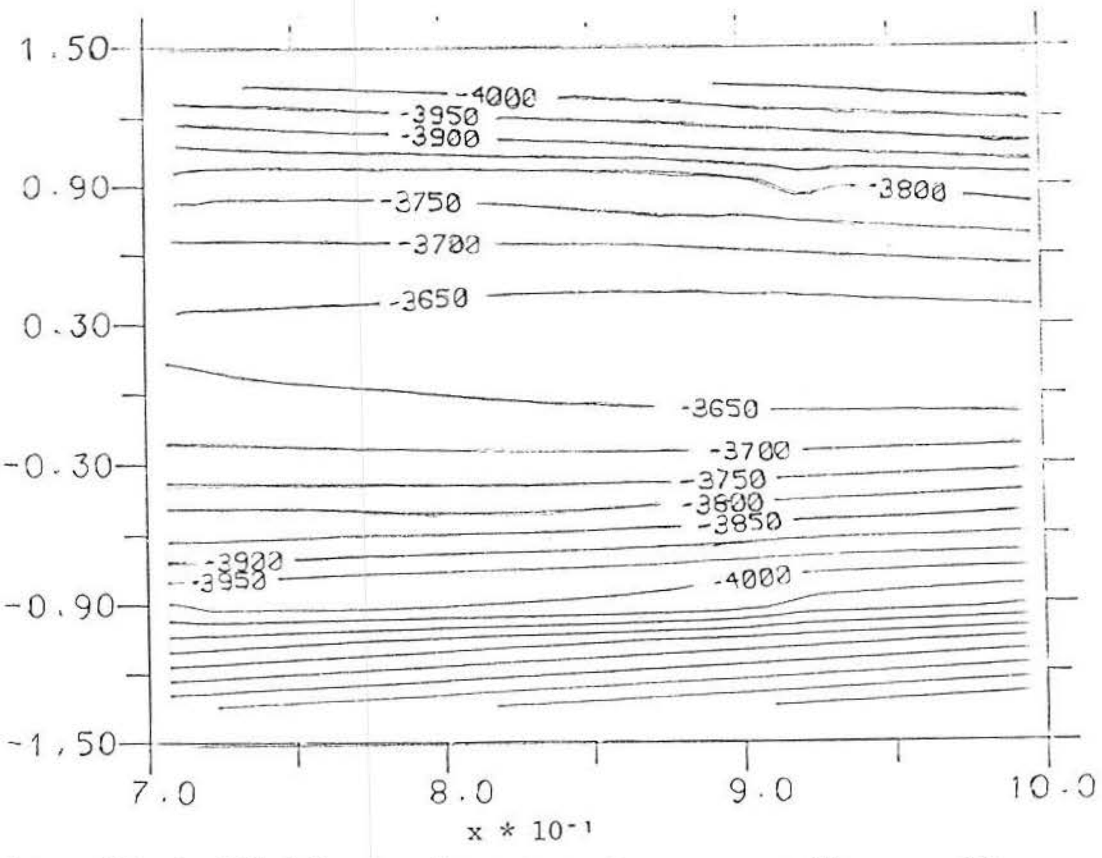

Figure 20- Loglikelihood values as a function of $\hat{\alpha}^{*}_{2,1}$ and $\hat{\alpha}^{*}_{2,0}$, using profile method for the Lombard and Doering data.

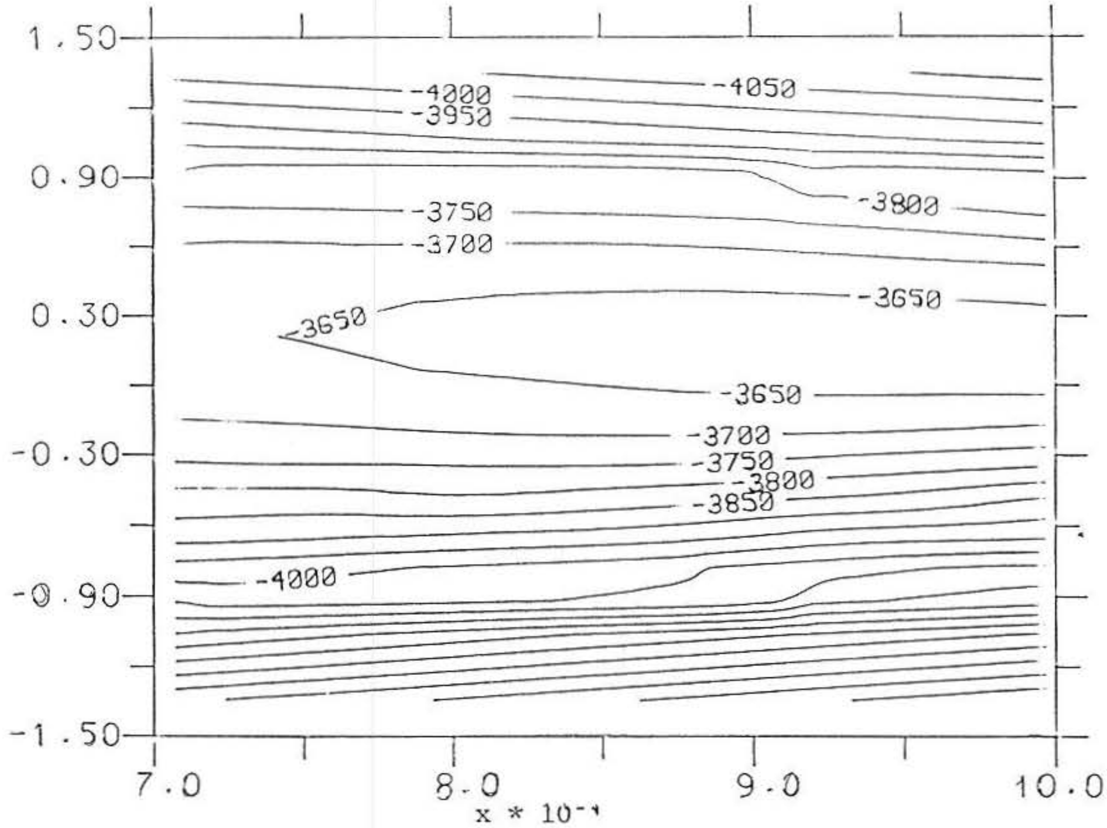

Figure 21- Loglikelihood values as a function of $\hat{\alpha}^*_{2,1}$ and $\hat{\alpha}^*_{2,0}$, using approximate method A for the Lombard and Doering data.

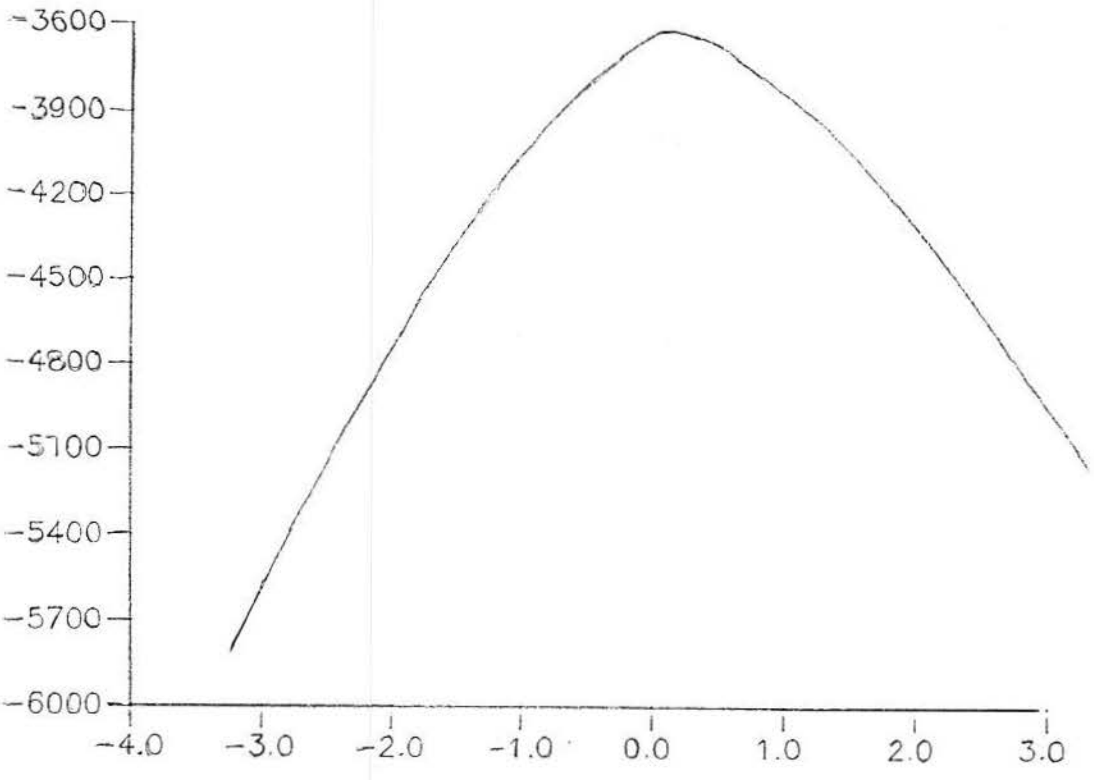

Figure 22- Maximum loglikelihood value over $\hat{\alpha}_{2,1}$ for each $\hat{\alpha}^*_{2,0}$ fixed to the Lombard and Doering data, using methods A and B.

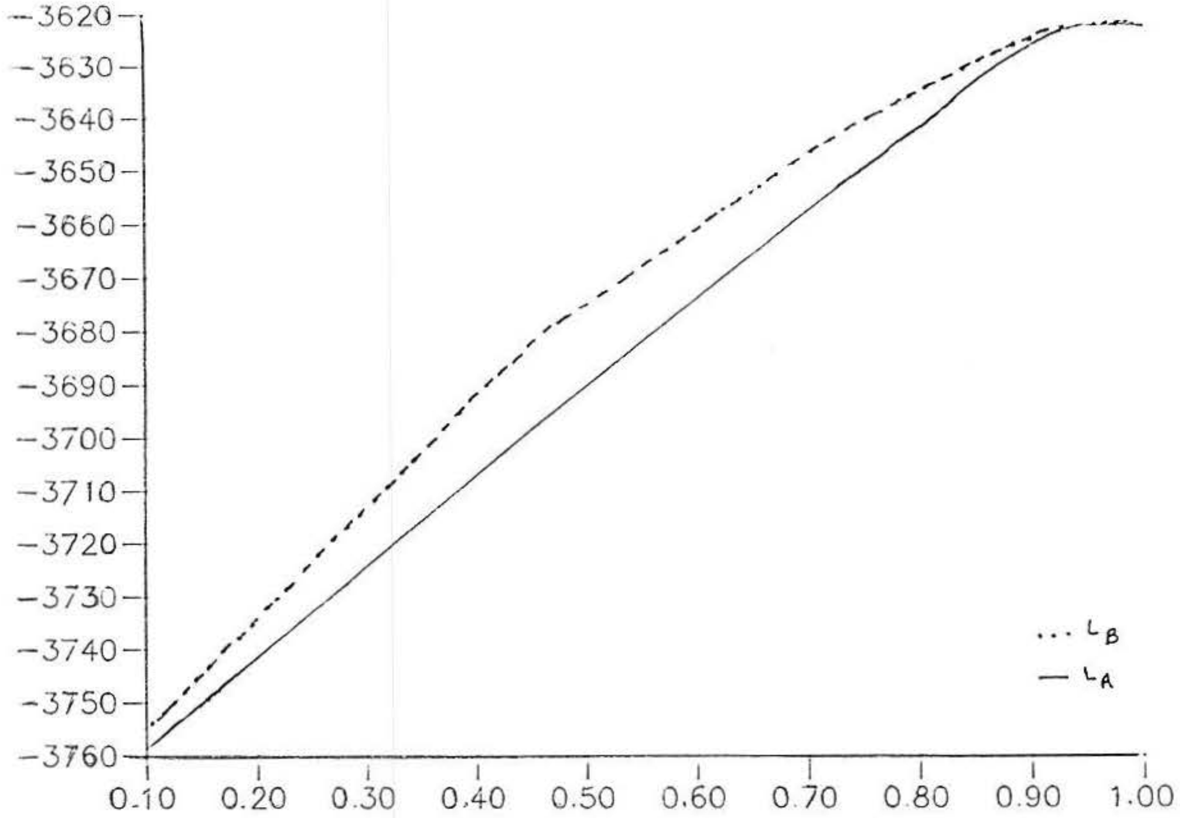

Figure 2.23- Maximum loglikelihood value over $\hat{\alpha}_{2,0}$ for each $\hat{\alpha}^*_{2,1}$ fixed to the Lombard and Doering data, using methods A and B.

Interpretation of $\alpha^*_{i,0}$

The reparametrization $\alpha^*_{i,0} = \hat{\alpha}_{i,0} \,/\, (1 + \hat{\alpha}^2_{i,1})^{\frac{1}{2}}$ corresponds to the probit of the expected value of $\Phi(\alpha_{i,0}+\alpha_{i,1}z)$, the response function of a probit model, i.e.,

$$\alpha^*_{i,0} = \Phi^{-1} \{ E (\Phi(\alpha_{i,0} + \alpha_{i,1}z)) \}.$$

For convenience, let us consider

$\alpha_{i,0} = a$ and $\alpha_{i,1} = b$

Then

$$E(\Phi(a+bz)) = \int_{-\infty}^{\infty} \Phi(a+bz)\ (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2} z^2)\ dz$$

If we take

$$bz = u \text{ and } bdz = du$$

Then

$$E(\Phi(a+bz)) = \int_{-\infty}^{\infty} \Phi(a+u)\ (2\pi)^{-\frac{1}{2}}\ b^{-1} \exp(-\tfrac{1}{2}\ u^2\ b^{-2})du$$

$$= \int_{-\infty}^{\infty} P(Z-u \leqslant a)\ (\text{density for } W \sim N(0,b^2) \text{ at } u)du$$

$$= P(Z+W \leqslant a),\quad Z+W \sim N(0,1+b^2)$$

and therefore

$$E(\Phi(a+bz)) = \Phi(\ a/(1+b^2)^{\frac{1}{2}}\ )$$

or

$$a/(1+b^2)^{\frac{1}{2}} = \Phi^{-1}\{E(\Phi(a+bz))\}.$$

6- *Conclusions*

The results about the investigation of the behaviour of the likelihood function give evidencie that

(1)- The approximate method provides results equivalent to the profile method. Both suggest that large $\hat{\alpha}_{i,1}$ ($\geqslant 3/\sigma$, σ is the standard deviation of the latent distribution) probably indicates bad behaviour of the likelihood, which will be shown by the presence of a long ridge. In this case the second derivative matrix or the information matrix are not good guides to the variability of these estimates.

(2)- If $\hat{\alpha}_{i,1}$ is not large, the likelihood funtion behaves well and thus the first order asymptotic theory is appropriate.

(3)- A badly behaved likelihood function suggests either that a reparametrization is necessary, or that the model is a poor fit for the data, or that the inference is particulary difficult.

(4)- Among the several reparametrizations we tried only the one given by

$$\hat{\alpha}^*_{i,0} = \hat{\alpha}_{i,0} / (1 + \hat{\alpha}^2_{i,1})^{\frac{1}{2}}$$

provided a better behaviour of the likelihood, independent of the size of the parameter estimates.

This reparametrization corresponds to the probit of the expected value of the response function of a probit model, that is,

$$\hat{\alpha}^*_{i,0} = \Phi^{-1}(E(\alpha_{i,0} + \alpha_{i,1} z)) =$$

$$= \Phi^{-1}(E(P(X_i=1 | z)) = \Phi^{-1} (P(X_i=1))$$

7- *References*

Albanese,M.T. (1990) Latent Variable Models for Binary Response Data. University of London, London. Ph.D. thesis.

Albanese,M.T. and Knott,M.(1991) TWOMISS: a computer program for fitting a two-factor logit/probit model to binary response data when observations may be missing.

Bartholomew,D.J. (1987) Latent Variable Models and Factor Analysis. London: Charles Griffin & Company Ltd.

Lombard,H.L. and Doering,C.R. (1947) Treatment of the fourfold table by partial association and partial correlation as it relates to public health problems. Biometrics, 3, 123-128.

Mislevy,R.J. (1984) Estimating latent distributions. Psychometrika, 49, 359-381.

Shea,B.L. (1984) FACONE: A computer program for fitting the logit latent model by maximum likelihood. Department of Statistics, London School of Economics.

Publicações do Instituto de Matemática da UFRGS
Cadernos de Matemática e Estatística

Série A: Trabalho de Pesquisa

1. Marcos Sebastiani - Transformation des Singularités - MAR/89.
2. Jaime Bruck Ripoll - On a Theoremof R. Langevin About Curvature and Complex Singularities - MAR/89.
3. Eduardo Cisneros, Miguel Ferrero e Maria Inés Gonzales - Prime Ideals of Skew Polynomial Rings and Skew Laurent Polynomial Rings - ABR/89.
4. Oclide José Dotto - ϵ - Dilations - JUN/89.
5. Jaime Bruck Ripoll - A Characterization of Helicoids - JUN/89.
6. Mark Thompson, V. B. Moscatelli - Asymptotic Distribution of Liusternik-Schnirelman Eigenvalues for Elliptic Nonlinear Operators -JUL/89.
7. Mark Thompson - The Formula of Weyl for Regions with a Self- Similar Fractal Boundary - JUL/89.
8. Jaime Bruck Ripoll - A Note on Compact Surfaces with Non Zero Constant Mean Curvature - OUT/89.
9. Jaime Bruck Ripoll - Compact ϵ - Convex Hypersurfaces - NOV/89.
10. Jandyra Maria G. Fachel - Coeficientes de Correlação Tipo -Contingência - JAN/90.
11. Jandyra Maria G. Fachel - The Probability of Ocurrence of Heywood Cases - JAN/90.
12. Jandyra Maria G. Fachel - Heywood Cases in Unrestricted Factor Analysis - JAN/90.
13. Julio Cesar R. Claeyssen, Tereza Tsukazan - Dynamical Solutions of Linear Matrix Differential Equations - JUL/90.
14. Maria T. Albanese - Behaviour of the Likelihood Function in Latent Trait Analysis of Binary Data - ABR/91.
15. Maria T. Albanese - Measurement of the Latente Variable in Latent Trait Analysis of Binary Data - ABR/91.
16. Maria T. Albanese - Adequacy of the Asymptotic Variance-Covariance Matrix using Bootstrap and Jackknife Techniques in Latent Trait Analysis of Binary Data - ABR/91.
17. Maria T. Albanese - Latent Variable Models for Binary Response - ABR/91.

Universidade Federal do Rio Grande Sul
Reitor: Professor Tuiskon Dick

Instituto de Matemática
Diretor: Professor Aron Taitelbaum
Núcleo de Atividades Extra Curriculares
Coordenador: Professora Jandyra G. Fachel
Secretária: Rosaura Monteiro Pinheiro

Os Cadernos de Matemática e Estatística publicam as seguintes séries:
Série A: Trabalho de Pesquisa
Série B: Trabalho de Apoio Didático
Série C: Colóquio de Matemática SBM/UFRGS
Série D: Trabalho de Graduação
Série E: Dissertações de Mestrado
Série F: Trabalho de Divulgação
Série G: Textos para Discussão

Toda correspondência com solicitação de números publicados e demais informações deverá ser enviada para:

NAEC - Núcleo de Atividades Extra Curriculares
Instituto de Matemática - UFRGS
Av. Bento Gonçalves, 9500
91.500 - Agronomia - POA/RS
Telefone: 36.11.59 ou 36.17.85 Ramal: 252