

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática

Adequacy of the Asymptotic Variance-Covariance Matrix  
using Bootstrap and Jackknife Techniques  
in Latent Trait Analysis of  
Binary Data

Maria T. Albanese

Cadernos de Matemática e Estatística

Série A, nº 16, ABR/91  
Porto Alegre, abril de 1991

## SUMMARY

	Page
1- <i>Introduction</i>	3
2- <i>The Model and its Estimation</i>	3
3- <i>Jackknife</i>	4
4- <i>Bootstrap</i>	6
5- <i>Applications</i>	10
5.1- Attitudes toward the U.S. Army	11
5.2- Attitudes toward Situations of Conflict	14
5.3- Cancer Knowledge	19
5.4- Arithmetic Reasoning Test on Black Women	22
6- <i>Comparison of Bootstrap, Normal Bootstrap and ML Estimation</i>	27
7- <i>Conclusions</i>	31
8- <i>References</i>	34

## 1- Introduction

When interpreting the asymptotic variance-covariance matrix of the parameter estimates it is assumed that the model is appropriate for the data. Since this assumption may be false in practice, or the sample size is not large enough for the number of parameters which have been estimated or even the standard asymptotic theory does not apply, the standard deviation and covariances obtained asymptotically will probably represent lower limits for the actual ones, and they must be analysed carefully.

A common way to investigate questions of this type is to generate bootstrap or jackknife samples, and use these as substitute for the sampling distributions of the estimators.

The aim of this paper is to investigate the adequacy of the variance-covariance matrix in a logit-probit model for binary data through the jackknife and bootstrap techniques. All results presented here were taken from Albanese(1990, Chapter 3). We shall present some examples used to carried out the study and report the main results. The results about the adequacy of the asymptotic covariances will be given only in the conclusions.

## 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  $x_{ij}$ . 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

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

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

or

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

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  $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})$  for individual  $j$  with latent variable value  $z$  is

$$\begin{aligned} g(x_j|z) &= \prod_{i=1}^p g_i(x_{ij}|z) \\ &= \prod_{i=1}^p (\pi_i(z))^{x_{ij}} (1-\pi_i(z))^{1-x_{ij}} \end{aligned}$$

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  $\alpha_{i,1}$ ,  $i=1,2,\dots,p$  are estimated by marginal maximum likelihood method, using a modified E-M algorithm (see Albanese(1990) or Bartholomew(1987)) available as Fortran programs FACONE (Shea (1984)) and TWOMISS (Albanese and Knott (1991)).

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

### 3- Jackknife

Jackknifing is a statistical technique first proposed by Quenouille (1956), which is used for reducing bias in the estimation of parameters and for estimating the variance-covariance matrix of the estimates. Miller (1974) gives an review of the subject.

In the basic jackknife the observations are randomly divided into  $g$  groups of size  $h$  each. We shall consider the number of groups equal to the sample size  $n$  we shall have  $n$  groups of size one.

Let  $X_1, X_2, \dots, X_p$  be a sample of independent and identically distributed (iid) random variables and  $\hat{\alpha}$  be an estimator of the parameter vector  $\alpha$  based on the sample size  $n$ .

Let  $\hat{\alpha}_{-i}$  be the corresponding estimator based on the sample of size  $n-1$ , where the  $i^{\text{th}}$  group (observation) has been deleted.

Then jackknife pseudovalues are defined by

$$\tilde{\alpha}_i = n \hat{\alpha} - (n-1) \hat{\alpha}_{-i} \quad (2)$$

for  $i=1, 2, \dots, n$ .

The jackknife estimates  $\tilde{\alpha}$  and its estimated variance-covariance matrix are obtained from the  $n$  pseudovalues by treating them as independently identically distributed observations from a multivariate normal distribution (Tukey, 1958). These estimates are given by

$$\tilde{\alpha} = \frac{\sum \tilde{\alpha}_i}{n} \quad (3)$$

$$\tilde{\alpha} = g \hat{\alpha} - \frac{n-1}{n} \sum \hat{\alpha}_{-i}$$

and

$$\sum (\tilde{\alpha}) = \frac{\sum (\tilde{\alpha}_i - \tilde{\alpha}) (\tilde{\alpha}_i - \tilde{\alpha})^t}{n(n-1)} \quad (4)$$

Since it often happens that  $\tilde{\alpha}$  and  $\hat{\alpha}$  are asymptotically equivalent,  $\sum (\tilde{\alpha})$  is sometimes used to estimate the variance-covariance matrix of  $\hat{\alpha}$ .

The jackknife estimate of bias is the difference between the parameter estimate  $\hat{\alpha}$  and  $\tilde{\alpha}$  multiplied by the correction factor

$n/(n-1)$ , i.e.,

$$\text{bias} = \frac{n}{n-1} (\hat{\alpha} - \tilde{\alpha}) \quad (5)$$

The jackknife technique has been applied in many areas, including factor analysis. Pennell (1972) demonstrated how the method can be used to find confidence intervals for the factor loadings, while Clarkson (1979) discussed the results of simulation studies using jackknife techniques and proposed modifications.

Clarkson's studies do not include the jackknife samples which provide Heywood cases. He argue that in these cases the jackknife estimates of the factor loadings are not representative of the 'usual' jackknife results because they are too large in absolute value.

Jorgensen (1987) gave a modification of the jackknife method for estimating the dispersion of the parameter estimates that are obtained as limits of iterative processes. He also gave examples to show how the method can be applied to the E-M algorithm and to iteratively reweighted least-squares.

#### 4- *Bootstrap*

The bootstrap is a general resampling procedure introduced by Efron (1979) to estimate the distribution of statistics based on independent observations. It can be carried out non-parametrically and parametrically, depending on the distribution from which the bootstrap samples are drawn.

We shall first present the non-parametric or empirical bootstrap method.

Suppose  $X_1, X_2, \dots, X_p$  are independent and identically distributed(iid) random variables from a population with unknown distribution function  $F$ , and suppose the goal is to make inferences about the parameter vector  $\alpha$  of the population.

Let  $\hat{\alpha}(x_1, x_2, \dots, x_p)$  be an estimator of  $\alpha$  based on the sample size  $n$  and let  $\hat{F}$  be the empirical distribution, that is, the distribution function that assign mass  $1/n$  to each  $X_i$ .

The bootstrap approximates the sampling distribution of  $\alpha$  under  $F$  by the sampling distribution of  $\hat{\alpha}$  under  $\hat{F}$ . This procedure is carried out using Monte Carlo method as follow:

(1) Construct  $\hat{F}$

(2) Draw a bootstrap sample,  $X_1^*, X_2^*, \dots, X_p^*$  iid with cdf  $\hat{F}$  and calculate

$$\hat{\alpha}^* = \hat{\alpha}(X_1^*, X_2^*, \dots, X_p^*)$$

(3) Independently do  $B$  times the step 2 (for some large  $B$ ), obtaining

$$\hat{\alpha}_b^*, b=1, 2, \dots, B.$$

The distribution function of  $\hat{\alpha}$  is approximated by

$$\hat{F}_B(y) = \#(\hat{\alpha}_b^* \leq y) / B.$$

The bootstrap estimate of  $\alpha$  based on the  $B$  replications is the mean of the  $\hat{\alpha}_b^*$  estimates, i.e.,

$$\hat{\alpha}^* = \sum \hat{\alpha}_b^* / B \tag{6}$$

and the bootstrap variance-covariance matrix estimate of  $\alpha$  based on the  $B$  replications is the variance-covariance matrix of the  $\hat{\alpha}_b^*$  estimates, i.e.,

$$\hat{\Sigma}_B = (B-1)^{-1} \sum (\hat{\alpha}_b^* - \hat{\alpha}^*) (\hat{\alpha}_b^* - \hat{\alpha}^*)^t \tag{7}$$

As the number of replications  $B \rightarrow \infty$ ,  $\hat{\alpha}^*$  will approach the bootstrap estimate of  $\alpha$  and  $\hat{\Sigma}_B$  the corresponding bootstrap estimate of the variance-covariance matrix  $\Sigma$ .

The bootstrap estimate of bias is the difference between the parameter estimate  $\hat{\alpha}$  and the bootstrap estimate  $\hat{\alpha}^*$ , that is,

$$\text{bias} = \hat{\alpha} - \hat{\alpha}^* \tag{8}$$

The basic result of the bootstrap theory is that the empirical distributions of the parameter estimates obtained by this method are asymptotically the same as the sampling distribution of those parameters in sampling from the population from which the original sample was drawn.

There is nothing which says that the bootstrap must be carried out non-parametrically. If we have reason to believe that the true distribution  $F$  is Normal, for example, then we can estimate  $F$  by its parametric ML estimate  $\hat{F}$ . The bootstrap samples at step (1) of the algorithm could then be drawn from  $\hat{F}_{\text{normal}}$  instead of  $\hat{F}$  (empirical distribution) and steps (2) and (3) carried out as before.

Efron (1979) also suggests that Taylor series expansion method can be used to obtain the approximate mean and variance of the bootstrap distribution of  $\hat{\alpha}^*$ , and he shows that it turns out to be the same as Jaeckel's infinitesimal jackknife (Miller, 1974), which differ only in detail from the standard jackknife described before.

Efron and Tibshirani (1986) discuss the number of replications  $B$  necessary to give reasonable results when we are estimating the standard deviation of one parameter. They set out the following approximation

$$CV(\hat{\sigma}_B) = \{ CV(\hat{\sigma})^2 + [ E(\hat{\delta}) + 2 ] / 4B \}^{\frac{1}{2}}$$

where  $CV(\hat{\sigma})$  is the limiting coefficient of variation of  $\sigma$  as  $B \rightarrow \infty$ ,  $\hat{\delta}$  is the kurtosis of the bootstrap distribution of  $\hat{\alpha}^*$ , given the observed data  $\mathbf{x}=(x_1, x_2, \dots, x_n)$ , and  $E(\hat{\delta})$  its expected value average over  $\mathbf{x}$ . For typical situations,  $CV(\hat{\sigma})$  lies between 0.10 and 0.30.

From this approximation and assuming that  $E(\hat{\delta})=0$ , they point out that for values of  $CV(\hat{\sigma}) \gg 0.10$ , there is little improvement when  $B$  is bigger than 100. They affirm that  $B$  as small as 25 gives reasonable results. We suggest that the number of bootstrap samples be determined by the point where stability of the estimates is obtained.

Chatterjee (1984) gives an application of the non-parametric bootstrap method to the problem of estimating the variability of the estimates of factor loadings. The number of bootstrap samples was settled empirically; it appeared that 300 gave reasonable stability.



Combining the bootstrap with graphical techniques he examines the variability of the estimator of the factor loadings. He points out that bootstrap may very well reveal when the asymptotic results are poor approximations.

Grönroos (1985) applies bootstrap methods to confirmatory factor analysis of a LISREL submodel (Jöreskog and Sörbom, 1984) to estimate factor loadings and their standard deviations.

His simulation studies involve artificial data with sample size 100, 150 and 300 and initially 300 replications. However the number of bootstrap samples become smaller, since he deletes from the analysis all those which provide the occurrence of Heywood cases.

Comparing asymptotic theory with bootstrap and Normal bootstrap results, he points out that the difference between the two bootstrap methods is very small, but it is larger, even though not essential significant, when compared with the asymptotic results.

Beran and Srivastava (1985) use bootstrap test and confidence regions for functions of the population covariance matrix, for example, eigenvalues and eigenvectors, which have the desired asymptotic levels if model restrictions, such as multiple eigenvalues, are taken into account in designing the bootstrap algorithm.

Efron and Tibshirani (1986) give a review of bootstrap methods for estimating standard errors and confidence intervals. The bootstrap is also extended to other measures of statistical accuracy such as bias and prediction error, and to complicated data structures such as time series, censored data, and regression models.

Bootstrap confidence intervals have been discussed with new improvements by Efron (1987) and their applications to problems in a wide range of situations is given by Diccio and Tibshirani (1987).

Albanese and Knott(1990) applies empirical and parametric bootstrap methods to find symmetric percentil-t confidence intervals for  $\alpha_{i,1}$  in a logit-probit model, and to test the reliability of the ranking of the respondents on the latent scale.

## 5 - Applications

We shall compare the bootstrap, jackknife and the original ML parameter estimates for  $\hat{\alpha}_{i,1}$ ,  $\hat{\alpha}_{i,0}^*$  and  $\hat{\alpha}_{i,0} = \hat{\alpha}_{i,0} / (1 + \hat{\alpha}_{i,1}^2)$ ,  $i=1, \dots, 4$ , their variability and how close is the bootstrap distribution of these parameter estimates and their jackknife pseudovalues to a Normal distribution. The reparametrization  $\hat{\alpha}_{i,0}^*$  provides a better likelihood behaviour, independent of the size of the parameter estimates (see Albanese(1990, Chapter 2)).

We shall look also at the jackknife and bootstrap bias of the discrimination parameter estimates  $\hat{\alpha}_{i,1}$ .

The standard deviation of the ML parameter estimate  $\hat{\alpha}_{i,0}^*$  is obtained from the following approximation:

$$\begin{aligned} \text{Var}(\alpha_{i,0}^*) \approx & \left[ \frac{\partial \alpha_{i,0}^*}{\partial \alpha_{i,0}} \right]^2 \text{Var}(\alpha_{i,0}) + 2 \frac{\partial \alpha_{i,0}^*}{\partial \alpha_{i,0}} \frac{\partial \alpha_{i,0}^*}{\partial \alpha_{i,1}} \text{Cov}(\alpha_{i,0}, \alpha_{i,1}) + \\ & + \left[ \frac{\partial \alpha_{i,0}^*}{\partial \alpha_{i,1}} \right]^2 \text{Var}(\alpha_{i,1}). \end{aligned}$$

Therefore

$$\begin{aligned} \text{Var}(\alpha_{i,0}^*) \approx & \frac{1}{1 + \alpha_{i,0}^2} \text{Var}(\alpha_{i,0}) - 2 \frac{\alpha_{i,0} \alpha_{i,1}}{(1 + \alpha_{i,1}^2)^2} \text{Cov}(\alpha_{i,0}, \alpha_{i,1}) + \\ & + \frac{(\alpha_{i,0} \alpha_{i,1})^2}{(1 + \alpha_{i,1}^2)^3} \text{Var}(\alpha_{i,1}), \end{aligned}$$

when  $\alpha_{i,1}$  and  $\alpha_{i,0}$  are replaced by their ML parameter estimates  $\hat{\alpha}_{i,1}$  and  $\hat{\alpha}_{i,0}$ ,  $i=1, \dots, p$ .

The results will take into account all bootstrap and jackknife samples, even those when fitted by a logit-probit model provide very large estimates for  $\hat{\alpha}_{i,1}$ .

The decision about what number of bootstrap samples to take in order to check the adequacy of the asymptotic variance-covariance matrix, was based on the stability of the bootstrap parameter estimates, obtained from 50, 75 and 100 replications. Since the bootstrap estimates for all examples showed great stability from 75 to 100 replications, all results are given for 100 bootstrap samples.

The investigation of the normality of the bootstrap distribution of the parameter estimates and the jackknife distribution of the pseudovalues will be done by Normal probability plotting and looking at  $R^2$ , the proportion of variance explained by the fitted straight line.

As we show later that, at least for the data we have worked with, the fitting of the jackknife pseudovalues by a Normal distribution is not associated with the degree of similarity between the jackknife and the original ML parameter estimates, so we only present Normal probability plots of some bootstrap distributions. This apparent non-association may be due to the small number of different jackknife pseudovalues (16, the number of different response patterns), and a larger number of variable would provide satisfactory results.

If the jackknife gives the same information as the bootstrap about the applicability of the asymptotic theory to estimating of the variability of the parameter estimates then it is more practical to use the jackknife, since it is quicker.

#### 5.1- *Attitudes towards the U.S.Army*

The data used here is on the attitudes to the US Army; it was presented by Stouffer et al (1950, p.21-22). The four items were intended to measure attitudes towards the U.S.Army held by 1000 noncommissioned officers in 1945. The questions asked were:

- (i) how well is the Army run
- (ii) whether you will return to civilian life with a favourable attitude towards the Army
- (iii) whether you have got a square deal in the Army and
- (iv) whether the Army has tried its best to look out for the welfare of enlisted men.

Table 1 shows the frequencies of the observed response patterns, where approval is scored as 1 and disapproval as 0. It also shows the fitted frequencies from the logit-probit model and the scoring of the response patterns by the posterior mean value of the latent variable.

Table 1- Frequency distribution and results obtained by fitting a logit-probit model to the Attitudes towards the U.S.Army.

Response pattern	Observed frequency	Expected frequency	Total score	Posterior mean
0000	229	227.39	0	-1.01
0100	52	52.57	1	-0.47
0010	25	27.91	1	-0.34
0001	16	17.78	1	-0.26
1000	199	194.97	1	-0.24
0110	16	13.00	2	0.12
0101	8	9.03	2	0.20
1100	96	100.99	2	0.21
0011	10	5.84	2	0.32
1010	60	65.60	2	0.33
1001	45	47.40	2	0.40
0111	3	5.57	3	0.76
1110	69	63.72	3	0.77
1101	55	50.03	3	0.85
1011	42	39.24	3	0.97
1111	75	78.95	4	1.45
Total	1000	1000.00	-	-

$\chi^2 = 7.39$  with 7 degrees of freedom ( $p \approx 0.40$ ).

It is reasonable to infer from the low  $\chi^2$  value that the data are consistent with a single latent variable measuring attitudes towards the U.S.Army.

Table 2- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,1}$  to the Attitudes towards the U.S.Army.

i	$\hat{\alpha}_{i,1}$	SD( $\hat{\alpha}_{i,1}$ )	CV( $\hat{\alpha}_{i,1}$ )	R $\chi(\hat{\alpha}_{i,1})$
1	1.68 (1.64)	2.14	.25 (.24)	.22
2	1.13 (1.12)	1.08	.15 (.14)	.14
3	1.45 (1.41)	1.50	.20 (.19)	.19
4	1.63 (1.60)	2.24	.20 (.22)	.22

Tables 2 shows an excellent agreement between all the bootstrap results and original ML parameter estimates. This is, perhaps, to be expected since all the bootstrap distributions of the parameter estimates are approximated very well by a normal distribution. The asymptotic theory works well in this example, where the sample size is 1000 and the ML parameter estimates  $\hat{\alpha}_{i,1}$  are nearly equal.

The jackknife parameter estimates  $\hat{\alpha}_{i,1}$ , for  $i=2,3$ , and their standard deviations are very similar to the corresponding original ML, while for items 1 and 4, they are slightly bigger with smaller coefficients of variation.

The relation between the jackknife and the original ML  $\hat{\alpha}_{i,0}^*$  (not presented here) has the same pattern as  $\hat{\alpha}_{i,0}$ , showing similar results, but not as close as those given by the bootstrap.

The bootstrap distribution of the parameter estimates are as well as or better fitted by a normal distribution than the corresponding jackknife pseudovalues.

Figure 1 presents the bootstrap distribution of the parameter estimates  $\hat{\alpha}_{i,1}$  and its fit by a normal distribution.

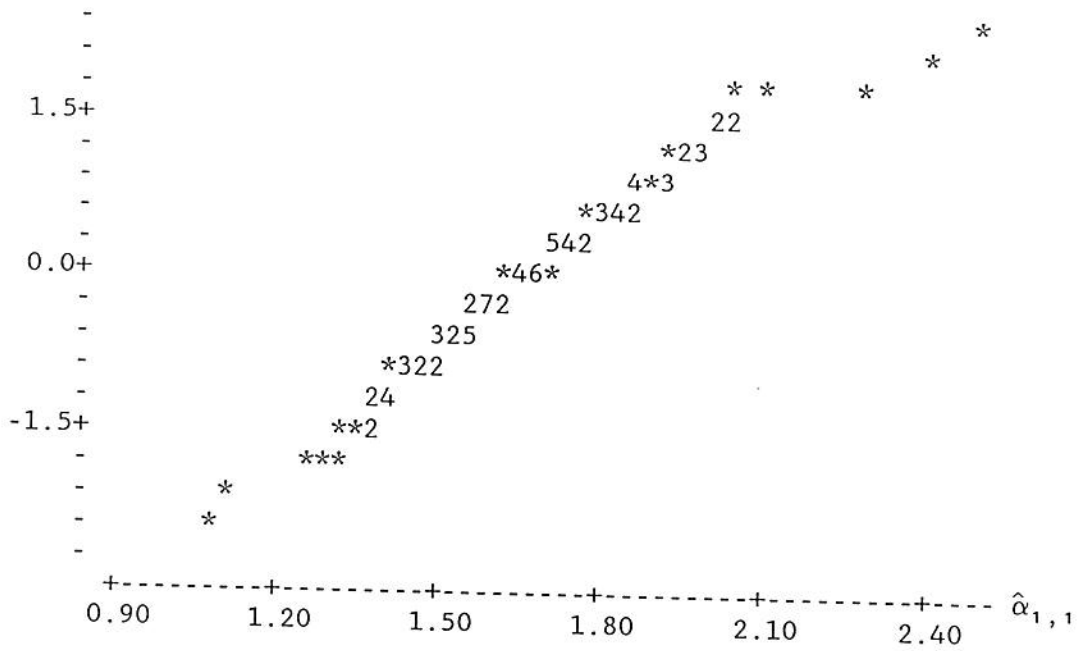


Figure 1- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{1,1}$  to the Attitudes towards the U.S.Army (original ML  $\hat{\alpha}_{1,1} = 1.64$ , bootstrap  $\hat{\alpha}_{1,1} = 1.68$  and  $R^2 = 97.9\%$ ).

Bootstrap biases of  $\alpha_{i,1}$  ( equation 8), for  $i=1, \dots, 4$ , are equal to 0.04, 0.01, 0.04 and 0.03, while the jackknife biases (equation 5), are equal to 0.50, -0.04, 0.09 and 0.64, respectively. Thus, bootstrap has provide estimates  $\hat{\alpha}_{i,1}$  with equal or less bias than the jackknife.

### 5.2- Attitudes towards Situations of Conflict

Stouffer and Toby (1951) report the answers of 216 respondents in 4 situations of conflict. For each situation the respondents can react either by a universalistic attitude (negative response) or particularistic attitude (positive response), which results are given in Table 3.

Table 3- Frequency distribution and results obtained by fitting a logit-probit model for the Stouffer and Toby data.

Response pattern	Observed frequency	Expected frequency	Total score	Posterior mean
0000	20	22.49	0	-1.26
1000	38	38.40	1	-0.72
0010	9	6.92	1	-0.63
0100	6	5.62	1	-0.54
0001	2	1.17	1	-0.33
1010	24	22.93	2	-0.18
1100	25	20.67	2	-0.09
0110	4	4.21	2	-0.02
1001	7	5.39	2	0.09
0011	2	1.14	2	0.16
0101	1	1.11	2	0.25
1110	23	27.38	3	0.39
1011	6	9.18	3	0.58
1101	6	9.86	3	0.66
0111	1	2.34	3	0.74
1111	42	37.19	4	1.20
Total	216	216.00	-	-

$\chi^2=5.85$  with 3 degrees of freedom ( $0.20 < p < 0.10$ ).

Table 3 shows that these data are fitted well by a logit-probit model with one single latent variable as a measure of the attitude of a person when under different situations of conflict.

Table 4- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,1}$  for the Stouffer and Toby data.

i	$\hat{\alpha}_{i,1}$	SD( $\hat{\alpha}_{i,1}$ )	CV( $\hat{\alpha}_{i,1}$ )	R $\chi^2(\hat{\alpha}_{i,1})$
1	1.19 (1.15) 1.33	.38 (.36) .28	.32 (.31) .21	93.5 90.8
2	1.82 (1.58) 1.44	.83 (.44) .42	.46 (.28) .29	66.4 75.5
3	1.44 (1.35) 1.24	.46 (.36) .34	.32 (.27) .27	93.0 77.0
4	2.72 (2.10) 2.18	2.99 (.66) .66	1.10 (.31) .30	35.3 63.5

The similarity between the bootstrap and the original ML results related to  $\hat{\alpha}_{i,1}$ ,  $i=1$  and  $3$ , can be considered very good, since the parameter estimates and their standard deviations are nearly equal. However there are significant differences for the remaining items ( $i=2$  and  $4$ ), especially for item  $4$ , where  $R^2$  is equal to  $35.30\%$ .

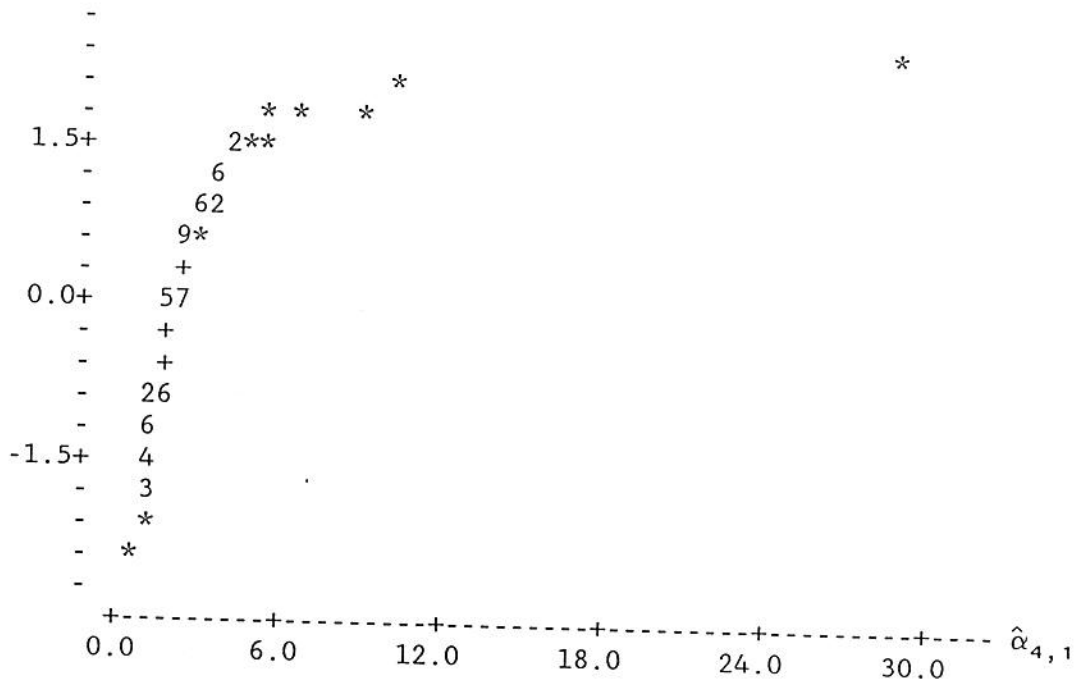


Figure 2- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{4,1}$  for the Stouffer and Toby data (original ML  $\hat{\alpha}_{4,1} = 2.10$ , bootstrap  $\hat{\alpha}_{4,1} = 2.72$  and  $R^2 = 35.3\%$ ).

Looking at the bootstrap distribution of  $\hat{\alpha}_{4,1}$ , we can see that this is due to a sample with  $\hat{\alpha}_{4,1}=28.12$  (Figure 2). On deleting this sample the only changes on the bootstrap results in Tables 4 and 5 are related to item 4, that is,

$$\hat{\alpha}_{4,1} = 2.46 \quad SD(\hat{\alpha}_{4,1}) = 1.41 \quad CV(\hat{\alpha}_{4,1}) = 0.57 \quad \text{and} \quad R^2 = 71.1\%$$

$$\hat{\alpha}_{4,0} = -1.47 \quad SD(\hat{\alpha}_{4,0}) = 0.38 \quad CV(\hat{\alpha}_{4,0}) = 0.26 \quad \text{and} \quad R^2 = 68.4\%$$

The improvement is not significant since we still have a large difference between the bootstrap and the ML parameter estimates  $\hat{\alpha}_{i,1}$ ,  $i=2$  and  $4$  ( $1.82$  compared to  $1.58$  and  $2.46$  compared to  $2.10$ , respectively). Their  $R^2$  are equal to  $66.4\%$  and  $71.1\%$ , respectively, which do not indicate a good approximation of the bootstrap distribution of parameter estimates by a normal distribution, probably responsible



for the values 0.89 and 1.14 for the ratios of bootstrap standard deviations to the corresponding asymptotic standard deviations.

On the other hand, a joint analysis of the bootstrap distribution  $\hat{\alpha}_{2,1}$  and  $\hat{\alpha}_{2,0}$  suggests that both distributions are either very skewed or a mixture of a normal distribution and some infinite values for  $\alpha_{2,1}$ , which are estimated to be only large due to inaccurate computing (see Albanese(1990, Chapter3)).

There is good agreement between jackknife and original ML parameter estimates  $\alpha_{i,1}$  for items 1 and 3. As bootstrap is very close to the original ML parameter estimates for these items, both methods give practically the same information about them. However for items 2 and 4, jackknife results are closer to the original ML than the bootstrap results. This is not desirable, especially because bootstrap is warning that the standard deviations probably are bigger than those given by the asymptotic theory.

Bootstrap biases of  $\alpha_{i,1}$ ,  $i=1,\dots,4$ , are equal to 0.04, 0.24, 0.11 and 0.62, while the jackknife estimates are 0.18, -0.14, -0.11 and 0.08, respectively. The largest biases are yielded by bootstrap and related to  $\alpha_{4,1}$ , followed by  $\alpha_{2,1}$ . hence, in terms of bias of  $\alpha_{i,1}$ , jackknife provides estimates equal or less biased than bootstrap.

Table 5- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,0}$  for the Stouffer and Toby data.

i	$\hat{\alpha}_{i,0}$	SD( $\hat{\alpha}_{i,0}$ )	CV( $\hat{\alpha}_{i,0}$ )	$R^2(\hat{\alpha}_{i,0})$
1	1.70 ( 1.66) 1.81	.28 (.22) .20	.16( .13) .11	94.3 72.6
2	-0.01 ( 0.01) 0.01	.25 (.24) .19	25.00(24.00)14.32	92.5 71.9
3	0.08 ( 0.08) 0.08	.19 (.20) .18	2.38( 2.50) 2.12	98.0 74.2
4	-1.53 (-1.33)-1.41	.90 (.36) .35	.59( .27) 0.24	55.0 70.7

According to the bootstrap results, the asymptotic theory can probably be applied to the estimation of the variability of all parameter estimates  $\hat{\alpha}_{i,0}$  and  $\hat{\alpha}_{i,0}^*$  (not presented here), except for  $\hat{\alpha}_{4,0}$  which is not well fitted by a normal distribution, since  $R^2=55.0\%$ .

The bootstrap and the original ML coefficient of variation of  $\hat{\alpha}_{2,0}^*$  are very large, 50.00 and 26.00, respectively, due to the large  $CV(\hat{\alpha}_{2,0})$  in both cases.

The bootstrap distribution for  $\hat{\alpha}_{1,1}$  and  $\hat{\alpha}_{1,0}$  show a good fit to a normal distribution ( $R^2 \geq 93.5\%$ ) and great similarity in the display of the points. The biggest values for the estimates in these graphs come from the same sample, but it is not the one that has  $\hat{\alpha}_{4,1} = 28.12$ .

Bootstrap results suggest that the application of the asymptotic theory to determine the standard deviation of the parameter estimates may not be adequate when both the sample size is small and at least one of the  $\hat{\alpha}_{i,1}$  is bigger than 2.0 and almost double size of the smallest (1.15).

Table 5 shows that the jackknife parameter estimates  $\hat{\alpha}_{i,0}$  and its coefficient of variation is very similar to the original ML, except for items 2 and 3, for which they are smaller, indicating that the jackknife standard deviations are probably underestimating the true ones.

As in the preceding examples, the pattern of  $R^2$  of the jackknife pseudovalues seems not to be associated to the degree of similarity between its estimates and the original ML as showed by the bootstrap results.

Jackknife parameter estimates are equal or closer to the original ML than the bootstrap ones. Furthermore, while some jackknife estimates of the standard deviation are even smaller than the asymptotic ones, bootstrap results are warning that the true standard deviations are probably bigger.

### 5.3- Cancer Knowledge

Lombard and Doering (1947) give the frequency distribution of the response patterns for a sample from a study on knowledge about cancer. Questions were asked about whether or not the following were sources of general information:

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

As we can see in Albanese(1990) or Bartholomew(1987), 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 ( $\chi^2 = 11.68$  with 6 degrees of freedom ( $0.05 < p < 0.10$ )).

Table 6- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,1}$  for the Lombard and Doering data.

i	$\hat{\alpha}_{i,1}$	SD( $\hat{\alpha}_{i,1}$ )	CV( $\hat{\alpha}_{i,1}$ )	R <sup>2</sup> ( $\hat{\alpha}_{i,1}$ )
1	.73 (.72)	.72 .09 (.09)	.09 .12 (.12)	.12 98.3 87.5
2	4.14 (3.40)	3.01 2.71(1.14)	1.19 1.65 (.34)	.40 67.4 73.9
3	1.39 (1.34)	1.31 .19 (.17)	.17 .14 (.13)	.13 95.4 72.5
4	.82 (.77)	.80 .14 (.14)	.15 .17 (.18)	.19 99.2 65.7

Table 7- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,0}$  for the Lombard and Doering data.

i	$\hat{\alpha}_{i,0}$	SD( $\hat{\alpha}_{i,0}$ )	CV( $\hat{\alpha}_{i,0}$ )	R <sup>2</sup> ( $\hat{\alpha}_{i,0}$ )
1	-1.29 (-1.29)	-1.28 .07 (.06)	.07 .05 (.05)	.05 99.4 79.8
2	0.74 (0.60)	0.55 .50 (.17)	.20 .68 (.28)	.36 66.4 81.3
3	-0.14 (-0.14)	-0.13 .07 (.08)	.06 .50 (.57)	.46 99.2 79.8
4	-2.75 (-2.75)	-2.74 .12 (.18)	.12 .04 (.06)	.04 99.4 48.2

From Tables 6 and 7 we can see that, except for item 2, there is very good agreement between all the bootstrap and the original ML results and the bootstrap distributions of the parameter estimates  $\alpha_{i,1}$  and  $\alpha_{i,0}$  are fitted very well by a Normal distribution ( $R^2 \geq 95.4\%$ ).

For item 2, the differences between the two results are significant and the fittings by a normal distribution are not good, since  $R^2$  equals 67.4% and 66.4% as we can see from Figures 3 and 4 below.

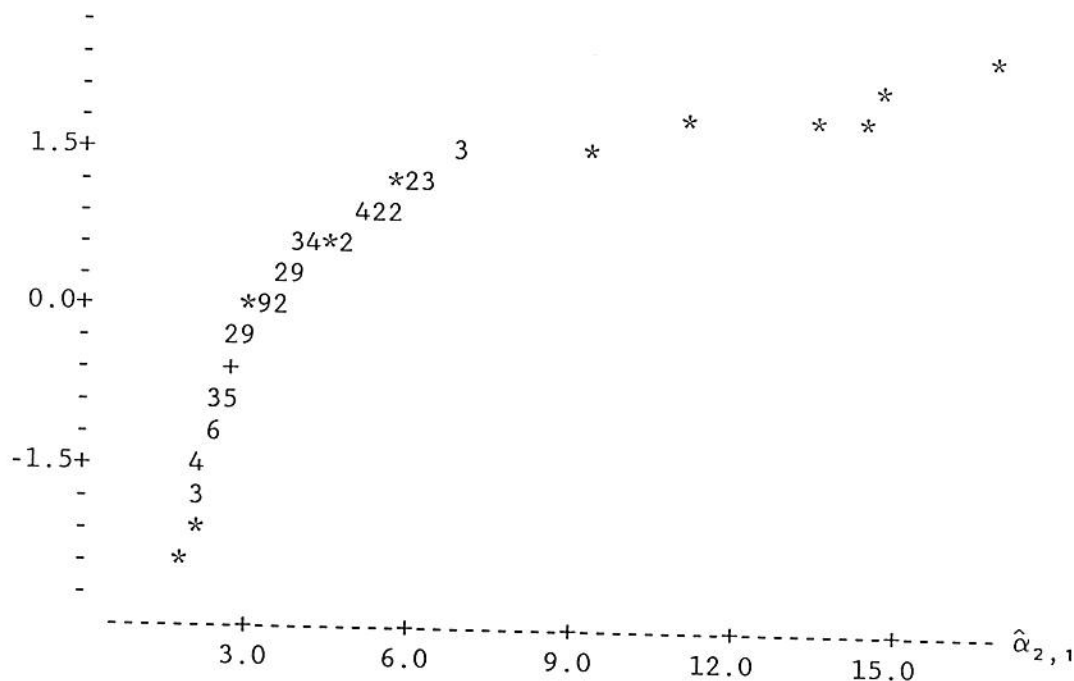


Figure 3- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{2,1}$  for the Lombard and Doering data (original ML  $\hat{\alpha}_{2,1} = 3.40$ , bootstrap  $\hat{\alpha}_{2,1} = 4.14$  and  $R^2 = 67.4\%$ ).

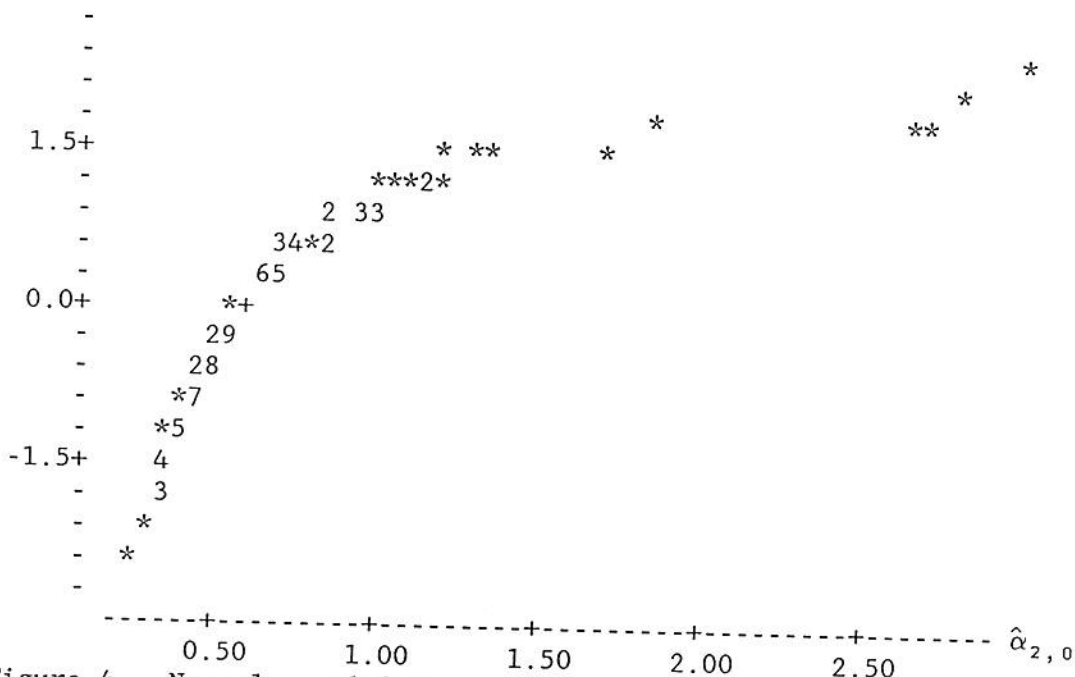


Figure 4- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{2,0}$  for the Lombard and Doering data (original ML  $\hat{\alpha}_{2,0} = 0.60$ , bootstrap  $\hat{\alpha}_{2,0} = 0.74$  and  $R^2 = 66.4\%$ ).

Actually, these two figures suggest that either the bootstrap distributions of  $\hat{\alpha}_{2,1}$  and  $\hat{\alpha}_{2,0}$  are fitted by two distributions, one normal and another with  $\alpha_{2,1}$  equal to infinity) or they are fitted by only one normal distribution extremely skewed. The bootstrap parameter estimates are larger than the original ML so that the bootstrap  $CV(\hat{\alpha}_{2,1})$  and  $CV(\hat{\alpha}_{2,0})$  are 91% and 143% larger than the corresponding original ML ones.

There is very good agreement between all jackknife estimates and their standard deviations and the corresponding original ML ones, except for  $\alpha_{2,1}$ , for which the jackknife estimate is even smaller (3.01 compared to 3.40).

For item 2, jackknife estimates are closer to the original ML than to the bootstrap (4.14), indicating that jackknife tends to be closer to ML estimates than to bootstrap when one of the  $\hat{\alpha}_{i,1}$  is large compared with the other.

Bootstrap and jackknife estimates of  $\alpha_{i,1}$ , for  $i=1,3,4$ , are approximately unbiased, since the bootstrap biases are 0.01, -0.05 and -0.05 and the jackknife biases are 0.00, -0.03 and 0.03, respectively. The jackknife bias of  $\alpha_{2,1}$  (-0.39) is smaller than the corresponding bootstrap one (0.74).

The fit by a normal distribution of the pseudovalues does not give information, as in bootstrap, about the relation between the jackknife results and original ML ones.

Table 8- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,0}^*$  for the Lombard and Doering data.

i	$\hat{\alpha}_{i,0}^*$	$SD(\hat{\alpha}_{i,0}^*)$	$CV(\hat{\alpha}_{i,0}^*)$	$R^2(\hat{\alpha}_{i,0}^*)$
1	-1.04 (-1.04) -1.03	.06 (.06) .05	.06 (.06) .05	98.7 77.6
2	0.17 ( 0.17) 0.18	.03 (.03) .03	.18 (.19) .17	98.8 79.8
3	-0.08 (-0.08) -0.09	.04 (.05) .04	.50 (.62) .44	99.5 70.7
4	-2.12 (-2.18) -2.13	.11 (.12) .11	.05 (.05) .05	98.1 56.7

Table 8 shows that the 3 methods give approximately the same estimates for  $\alpha_{i,0}^*$  and their standard deviations. For item 3, the coefficient of variation of the original ML estimates is slightly bigger

than the corresponding bootstrap and jackknife parameter estimates (0.62 compared to 0.52 and 0.44, respectively).

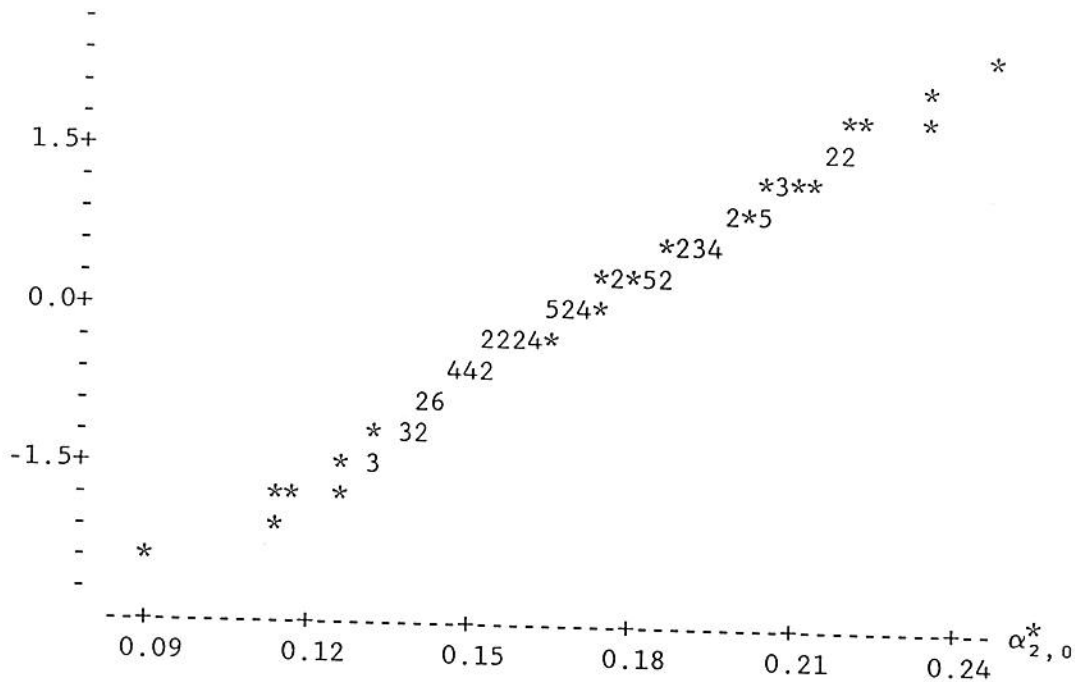


Figure 5- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{2,0}^*$  for the Lombard and Doering data (original ML and bootstrap  $\hat{\alpha}_{2,0}^* = 0.17$ , and  $R^2 = 98.8\%$ ).

As in all preceding examples, in this case too the bootstrap distribution  $\hat{\alpha}_{i,0}^*$ ,  $i=1, \dots, 4$ , has an excellent fit by a normal distribution and most of the  $\hat{\alpha}_{i,0}^*$ 's and their standard deviations are equal to the original ML ones.

According to the bootstrap results, this example seems to indicate that the asymptotic variance matrix may not be trusted when one of the  $\hat{\alpha}_{i,1}$  is large (3.40 or more), even for a large sample size (1729). On the other hand, jackknife does not provide any warning about possible underestimation of the asymptotic standard deviations.

#### 5.4 - Arithmetic Reasoning Test on Black Women

Mislevy(1985) gives frequency distributions of response patterns for samples of American Youth on the Armed Services Vocational Aptitude Battery. The individuals were classified by sex and colour. The results given here are related to a sample of 145 black women. This data set is well fitted by a single later  $2^2$ -variable logit-probit model (see

Bartholomew(1987) or Albanese(1990).

Table 9- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,1}$  for the ART on black women.

i	$\hat{\alpha}_{i,1}$	SD( $\hat{\alpha}_{i,1}$ )	CV( $\hat{\alpha}_{i,1}$ )	R $\chi^2(\hat{\alpha}_{i,1})$
1	6.79(14.39)38.82	7.16(67.78)6.77	1.05 (4.71) .17	83.8 95.4
2	1.63 (0.38) 0.32	3.32 ( .22) .21	2.04 (0.58) .66	49.6 80.8
3	1.56 (0.37) 0.32	3.62 ( .24) .22	2.32 (0.65) .69	47.3 91.4
4	0.14 (0.19) 0.16	2.94 ( .24) .22	21.00 (1.26)1.38	44.0 93.2

Table 10- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,0}$  for the ART on black women.

i	$\hat{\alpha}_{i,0}$	SD( $\hat{\alpha}_{i,0}$ )	CV( $\hat{\alpha}_{i,0}$ )	R $\chi^2(\hat{\alpha}_{i,0})$
1	0.01 (0.24)130.10	1.27(4.63)3.05	127.00(18.52).02	93.2 76.0
2	-0.59(-0.33) -0.32	0.72 (.16) .16	1.22 ( .48).50	54.3 75.8
3	-1.62(-0.96) -0.93	1.89 (.14) .18	1.17 ( .14).19	45.7 70.8
4	-1.56(-1.08) -1.08	1.60 (.16) .18	1.02 ( .15).17	36.7 67.7

This is an example of an extreme case where one  $\hat{\alpha}_{i,1}$  dominates all the other items by its very large value (14.39), and the sample size (145) is very small.

Comparing the bootstrap with the original ML results in Tables 9 and 10 we can see some disagreement for all the results, the bootstrap estimates being larger than the original ML estimates, except for item 1. It seems that the dominating item 1 has affected all the other items, which show an even bigger discrepancy between the bootstrap and the original ML estimates.

The results also show that large values of the bootstrap parameter estimate  $\hat{\alpha}_{i,1}$ , that is,  $\hat{\alpha}_{i,1} \geq 1.42$  for  $i=2,3,4$ , are always associated with small values of  $\hat{\alpha}_{1,1}$ , that is,  $-0.38 \leq \hat{\alpha}_{1,1} \leq 0.77$ .

While the bootstrap  $\hat{\alpha}_{1,1}$ 's are spread from -0.38 to 26.84, at least

90% of the bootstrap  $\hat{\alpha}_{i,1}$  are concentrated between -1.48 and 1.48 for  $i=2,3,4$ . In items 2 and 3, up to 10% of the bootstrap  $\hat{\alpha}_{i,1}$ 's assume values from 1.48 to 17.96 while  $\hat{\alpha}_{4,1}$  varies between -12.75 and 14.98.

The performance of the bootstrap distribution of the parameter estimates  $\hat{\alpha}_{i,1}$  with the corresponding  $\hat{\alpha}_{i,0}$  is very similar, so that large values for  $\hat{\alpha}_{i,1}$  are always associated with large  $\hat{\alpha}_{i,0}$  in absolute value, as we can see in Tables 9 and 10.

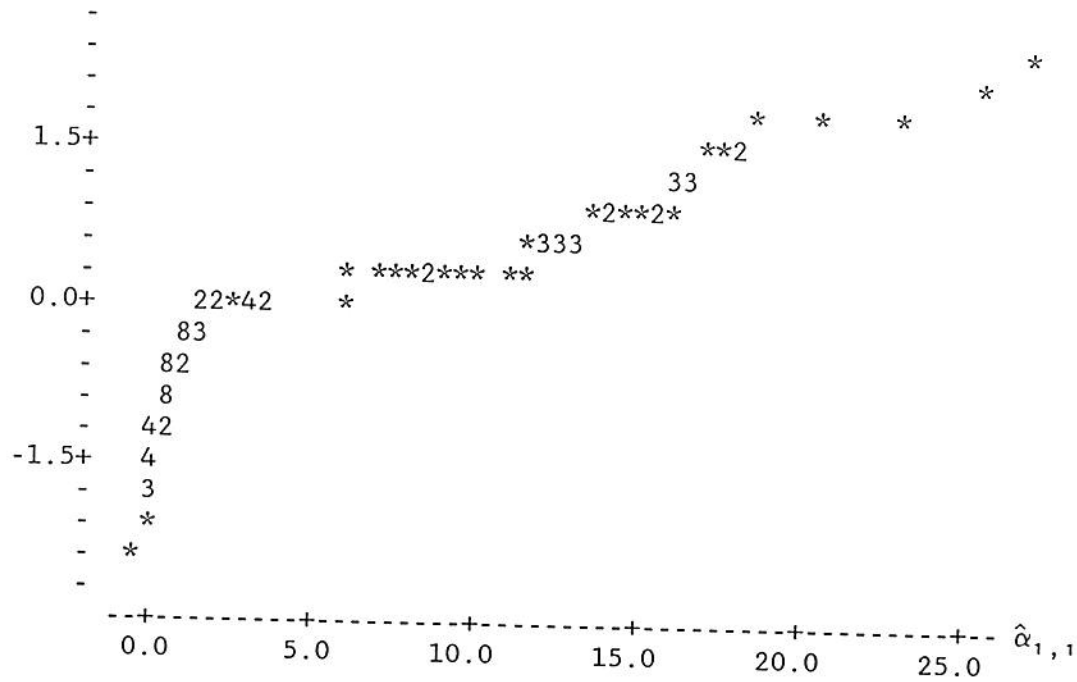


Figure 6- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{1,1}$  to the ART on black women (original ML  $\hat{\alpha}_{1,1} = 14.39$ , bootstrap  $\hat{\alpha}_{1,1} = 6.79$  and  $R^2 = 83.8\%$ ).

Figure 6 shows that the bootstrap distribution of  $\hat{\alpha}_{1,1}$  either could be fitted by a mixture of two Normal distributions or by two different distributions: one normal and another with  $\alpha_{1,1}$  equal to infinity. Although the normal probability plotting for the bootstrap distribution of  $\hat{\alpha}_{1,0}$  provides  $R^2$  equal to 93.2%, we may see a mixed of two normal distributions (Figure 7, next page).

The normal probability plottings for  $\hat{\alpha}_{i,1}$  and  $\hat{\alpha}_{i,0}$  for  $i \neq 1$ , show that most of the bootstrap estimates are fitted by a normal distribution, except for those sample for which  $\hat{\alpha}_{i,1}$ ,  $i \neq 1$ , is very large. In this later case, it could be fitted by a distribution with  $\alpha_{i,1}$ ,  $i \neq 1$ , equal to infinity.



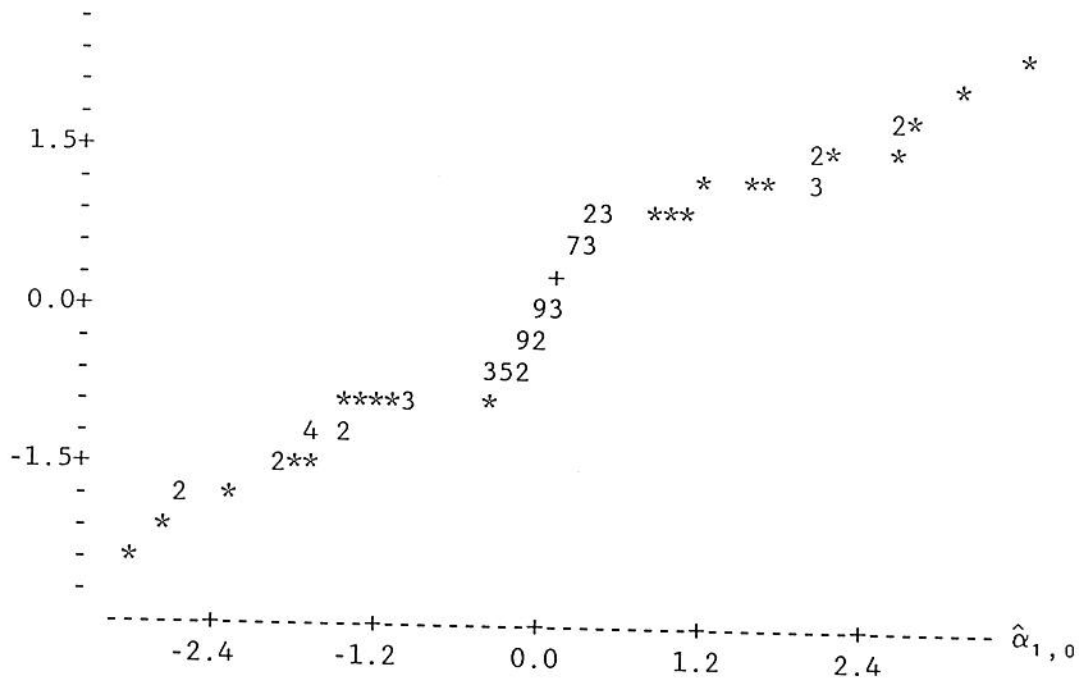


Figure 7- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{1,0}$  to the ART on black women (original ML  $\hat{\alpha}_{1,0} = 0.24$ , bootstrap  $\hat{\alpha}_{1,0} = 0.01$  and  $R^2 = 93.2\%$ ).

The jackknife estimates and standard deviations are very close to the corresponding original ML ones, except for item 1 where jackknife  $\hat{\alpha}_{1,1}$  and  $\hat{\alpha}_{1,0}$  are bigger (38.82 and 130.10 compared to 14.39 and 0.24, respectively) with smaller coefficients of variation (0.17 and 0.20 compared to 4.71 and 18.52), probably underestimating the true standard deviations.

On the other hand, when comparing bootstrap with the original ML results we have seen that they disagree strongly and bootstrap gives a warning that the asymptotic theory probably is not working well.

The bootstrap estimates of bias of  $\alpha_{i,1}$ ,  $i=1, \dots, 4$ , are equal to -7.60, 1.25, 1.19 and -0.05, while the corresponding jackknife estimates are 24.60, -0.06, -0.05 and -0.03, respectively. Therefore jackknife has provided estimates equal or less biased than bootstrap, except for item 1.

The results for all items show great discrepancies between jackknife and bootstrap techniques. These discrepancies are related to the size of the parameter estimates, standard deviations and fit of a normal distribution.

Table 11- Comparison between the bootstrap, original ML(in brackets) and the jackknife parameter estimates  $\hat{\alpha}_{i,0}^*$  for the ART on black women.

i	$\hat{\alpha}_{i,0}^*$	SD( $\hat{\alpha}_{i,0}^*$ )	CV( $\hat{\alpha}_{i,0}^*$ )	$R^2(\hat{\alpha}_{i,0}^*)$
1	0.01 ( 0.02) 0.05	.14 (.24) .20	14.00(12.00)4.00	97.1 72.7
2	-0.31 (-0.31) -0.31	.15 (.13) .15	.48 ( .42) .48	98.7 69.9
3	-0.85 (-0.90) -0.94	.20 (.11) .17	.24 ( .12) .18	99.1 69.6
4	-1.04 (-1.06) -1.13	.20 (.14) .17	.19 ( .13) .15	99.0 64.3

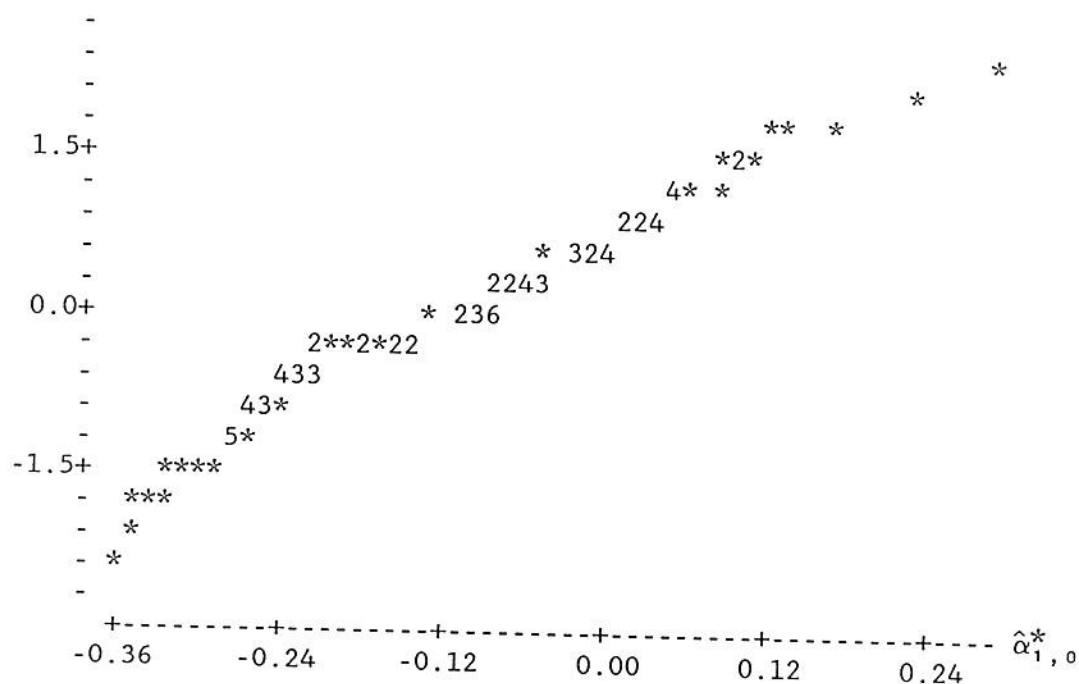


Figure 8- Normal probability plotting of the bootstrap parameter estimate  $\hat{\alpha}_{1,0}^*$  to the ART on black women (original ML  $\hat{\alpha}_{1,0}^* = 0.02$ , bootstrap  $\hat{\alpha}_{1,0}^* = 0.01$  and  $R^2 = 97.1\%$ ).

The bootstrap parameter estimates  $\hat{\alpha}_{i,0}^*$  are not affected by the skewness of  $\hat{\alpha}_{i,1}$  and  $\hat{\alpha}_{i,0}$ ,  $i=1, \dots, 4$ , showing substantial agreement with the original ML ones. This is to be expected, since they are very well approximated by a normal distribution.

Jackknife results for  $\alpha_{1,0}^*$  are very close to the original ML, except for item 1, which coefficient of variation is smaller (4.00 compared to 12.00). This is due to the fact that the reparametrization depends on  $\alpha_{i,1}$  and  $\alpha_{i,0}$ , for with the coefficient of variations are also smaller than those given by the asymptotic theory.

6 - Comparison of Bootstrap, Normal Bootstrap and ML Estimates

The aim of this section is to investigate how close the (empirical) bootstrap parameter estimates  $\hat{\alpha}_{i,1}$ 's are to the corresponding normal bootstrap ones, in order to obtain more evidence which confirm the bootstrap results about the adequacy of the asymptotic standard deviations presented above.

We shall carried out this study considering 100 normal bootstrap samples for each one of the 4 sets of data, which will be drawn from a multinomial distribution with parameters  $\pi_i(z)$ ,  $i=1, \dots, 4$ , where  $\pi_i(z)$  is the response function of a logit-probit model with parameters  $\alpha_{i,1}$  and  $\alpha_{i,0}$  equal to the ML estimates from the real data, and the latent variable  $Z$  is distributed as  $N(0,1)$ .

We shall compare the bootstrap methods (empirical and normal) in relation to the mean, median, interquartile difference  $Q_3-Q_1$  and standard deviation of the corresponding bootstrap distribution of  $\hat{\alpha}_{i,1}$ ,  $i=1, \dots, 4$ . We complement the analysis comparing both bootstrap estimates with the original ML results.

Table 12- Comparison between the bootstrap, original ML (in brackets) the normal bootstrap parameter estimates  $\hat{\alpha}_{i,1}$  to the Attitudes towards the U.S.Army.

$i$	$\hat{\alpha}_{i,1}$	$SD(\hat{\alpha}_{i,1})$	$Q_3 - Q_1$	Median
1	1.68 (1.64) 1.68	.25 (.24) .24	.34 (.32) .32	1.67 1.63
2	1.13 (1.12) 1.11	.15 (.14) .15	.20 (.19) .16	1.13 1.10
3	1.45 (1.41) 1.41	.20 (.19) .18	.31 (.26) .21	1.44 1.39
4	1.63 (1.60) 1.60	.20 (.22) .22	.27 (.30) .28	1.64 1.58

Table 12 shows an excellent agreement between the three procedures. The fitting of both bootstrap parameter estimates by a normal distribution ( $R^2 \geq 92.1\%$ ) is shown by the similarity between mean and median, and as well by  $Q_3-Q_1$ .

Table 13- Comparison between the bootstrap, original ML (in brackets) and the normal bootstrap parameter estimates  $\hat{\alpha}_{i,1}$  for the Stouffer and Toby data.

i	$\hat{\alpha}_{i,1}$	SD( $\hat{\alpha}_{i,1}$ )	$Q_3 - Q_1$	Median
1	1.19 (1.15) 1.24	.38 (.36) .37	.49 (.48) .54	1.12 1.25
2	1.82 (1.58) 1.69	.83 (.44) .64	.62 (.59) .62	1.61 1.56
3	1.44 (1.35) 1.34	.46 (.36) .36	.52 (.48) .48	1.36 1.33
4	2.72 (2.10) 2.90	2.99 (.66) 2.23	.99 (.89) 1.37	2.12 2.32

Items 1 and 3 present estimates nearly equal when comparing the three methods. Items 2 and 4 present some discrepancies, which are stronger for item 4.

The larger bootstrap estimates of  $\alpha_{2,1}$  than the original ML is due to the occurrence of some large values when  $\hat{\alpha}_{4,1}$  was small. This shows some instability of the bootstrap distribution, probably because of the small sample size (216).

Both bootstrap estimates  $\hat{\alpha}_{4,1}$  are closer to each other than to the corresponding ML estimate (2.72 and 2.90 compared to 2.21). The bootstrap medians are closer to the ML estimate than to the means due to the skewness of the bootstrap distribution.

Comparing  $Q_3 - Q_1$ , we can say that the only difference is for item 4, and the (empirical) bootstrap estimate is closer to the ML than the normal bootstrap.

The higher estimates for the normal bootstrap of item 4 is due to the variation in  $\hat{\alpha}_{4,1}$ , which assumes values from 1.14 to 14.75 with 25% of them bigger than 3.15. Although there are some small differences in item 4, we can still say that both bootstrap methods present very similar results.

We can see from Table 14 that all results are nearly equal, except for item 2. Both bootstrap methods present larger estimates  $\hat{\alpha}_{2,1}$  than the original ML, though the normal bootstrap estimate is closer to the latter than to the empirical bootstrap (3.79, 4.14 compared to 3.40). The bootstrap medians for item 2 are closer to the ML estimates than the means.

Table 14- Comparison between the bootstrap, original ML (in brackets) and the normal bootstrap parameter estimates  $\hat{\alpha}_{i,1}$  for the Lombard and Doering data.

i	$\hat{\alpha}_{i,1}$	SD( $\hat{\alpha}_{i,1}$ )	$Q_3 - Q_1$	Median
1	.73 (.72) .73	.09 (.09) .10	.13 (.12) .11	.71 .72
2	4.14 (3.40) 3.79	2.71(1.14)1.92	2.07(1.54)1.28	3.29 3.40
3	1.39 (1.34) 1.38	.19 (.19) .18	.25 (.23) .24	1.37 1.38
4	.82 (.77) .78	.14 (.22) .13	.18 (.19) .18	.81 .77

The differences between both bootstrap methods for  $\hat{\alpha}_{2,1}$  can be better understood if we look at their distributions.

The bootstrap distribution of  $\hat{\alpha}_{2,1}$  assumes values between 1.67 and 16.90 with 25% of them smaller than 2.67 and others 25% bigger than 4.74. On the other hand, in the normal bootstrap  $\hat{\alpha}_{2,1}$  ranges from 2.03 to 13.54, with  $Q_1$  equal to 2.75 and  $Q_3$  equal to 4.03. Therefore in the normal bootstrap the spread of the estimates  $\hat{\alpha}_{2,1}$  is smaller.

Some instability observed in the Stouffer and Toby data reflected by item 4 not occur in this example, though the larger value of ML  $\hat{\alpha}_{2,1}$  (3.40 compared to 2.10), as we can see by the strong similarity among the 3 procedures for items 1, 3 and 4. This is probably due to the larger sample size (1729) of the Lombard and Doering data.

Table 15- Comparison between the bootstrap, original ML (in brackets) and the Normal bootstrap parameter estimates  $\hat{\alpha}_{i,1}$  for the ART on black women.

i	$\hat{\alpha}_{i,1}$	SD( $\hat{\alpha}_{i,1}$ )	$Q_3 - Q_1$	Median
1	6.79(14.39)5.70	7.16(67.78)5.33	12.37(91.43)10.68	2.96 2.53
2	1.63 (.38) .71	3.32 (.14) .15	.20 (.19) .16	1.13 1.10
3	1.56 (.37)1.17	.20 (.19) .18	.31 (.26) .21	1.44 1.39
4	.14 (.19) .24	.20 (.22) .22	.27 (.30) .28	1.64 1.58

The original ML parameter estimates  $\hat{\alpha}_{i,1}$ ,  $i=2,3$ , are smaller than the bootstrap ones, while the ML estimate  $\hat{\alpha}_{4,1}$  is similar, but with much smaller standard deviation.

Medians of both bootstrap methods are closer to the original ML parameter estimates than the corresponding means, except for item 1.

The interquartile differences  $Q_3-Q_1$  are very similar for both bootstrap methods, but they are very different from the asymptotic approximation, specially for  $\hat{\alpha}_{1,1}$ .

The original ML parameter estimate  $\hat{\alpha}_{1,1}$  and its standard deviation are much larger than the bootstrap estimates, 14.39 compared to 6.79 and 5.70, and 67.78 compared to 7.16 and 5.33, respectively.

Albanese(1990, Chapter 2) pointed out when fitting a logit-probit model to the Arithmetic Reasoning Test on black women, the ML parameter estimate  $\hat{\alpha}_{1,1}$  could be equal to any value bigger than 3 or 4, since the likelihood function is flat after this point.

It is worth saying that when carrying out the bootstrap methods we have always considered the same stopping rule for the iterative procedure of the estimation of the parameters. Hence, using the same stop rule, the normal bootstrap estimate of  $\alpha_{1,1}$  is equal to 5.70, though the bootstrap samples are drawn from a distribution with  $\alpha_{1,1}$  equal to 14.39.

The normal bootstrap distribution of  $\hat{\alpha}_{1,1}$  assumes values between 0.09 and 15.79 with  $Q_1$  equal to 0.69 and  $Q_3$  equal to 11.37. The fitting by a normal distribution is the same as for the empirical bootstrap (83.6%). Furthermore, 52% of the parameter estimates are bigger than 3.0 and the median is 2.53. These results seem to indicate that the 'true' parameter could be equal to 3.0.

The disagreement between the bootstrap results for the remaining items are strongly due to the influence of item 1, since  $\hat{\alpha}_{i,1}$ ,  $i=2,3,4$ , assume only large values in the bootstrap samples with  $\hat{\alpha}_{1,1}$  small. That is, 8% of the normal bootstrap samples with  $\hat{\alpha}_{1,1} < 1.0$ , have one of the  $\hat{\alpha}_{i,1}$ 's,  $i \neq 1$ , bigger than 3.0.

The empirical bootstrap distribution of  $\hat{\alpha}_{1,1}$  ranges from -0.39 to 26.84, with  $Q_1$  equal to 0.55 and  $Q_3$  equal to 12.92. Besides, the median is 2.96 and 28% of the bootstrap samples present  $\hat{\alpha}_{1,1} < 3.0$  and  $\hat{\alpha}_{i,1} \geq 3.0$ , for some  $i \neq 1$ . This suggests that as in the normal bootstrap, item 1 is strongly affecting the remaining items, producing skewed distributions and larger estimates.

## 7 - Conclusions

The results from the comparison between the bootstrap, jackknife and ML parameter estimates  $\hat{\alpha}_{i,0}$ ,  $\hat{\alpha}_{i,1}$  and  $\hat{\alpha}_{i,0}^*$  of a logit-probit model and the corresponding variance-covariance matrix suggest that

- (1)- The more closely the bootstrap distribution of the parameter estimates is fitted by a normal distribution, the better is the agreement between the bootstrap and the asymptotic standard deviation.
- (2)- If  $\hat{\alpha}_{i,1}$  is not large, the asymptotic variance matrix can probably be trusted, since the bootstrap estimates and standard deviations are very close to the ML estimates and to the asymptotic standard deviations. Furthermore, this similarity increases as the sample size becomes larger.
- (3)- Large values for  $\hat{\alpha}_{i,1}$  are associated with skewed distributions or a mixture of two distributions, one normal and another with  $\alpha_{i,1}$  equal to infinity. Probably the asymptotic standard deviations of the parameter estimates are smaller than the true ones.
- (4)- If the sample size is small and one of the items has very large  $\hat{\alpha}_{i,1}$  while the remaining ones are small, all with relative large standard deviations then it is likely that most of the estimates can not be trusted.
- (5)- As the bootstrap distribution of the parameter estimates  $\hat{\alpha}_{i,0}^*$  is fitted by a Normal distribution very well, most of the bootstrap results are equal to the corresponding original ML and their asymptotic variance parameter estimates. This shows that  $\hat{\alpha}_{i,0}^*$  is not affected by the skewness of the bootstrap  $\hat{\alpha}_{i,1}$  and  $\hat{\alpha}_{i,0}$ , though their variability is shown through large coefficient of variation of  $\hat{\alpha}_{i,1}^*$ .

(6)- In summary, although the bootstrap distribution must underestimate the variation in the true sampling distribution, there is strong evidence that it gives a better guide than the usual first order normal approximation. Bootstrapping methods seem to be very useful for investigating the adequacy of the normal approximation in doubtful cases. When the discrimination parameters are small the asymptotic theory works well, but when they get large it is inadequate.

(7)- Jackknife parameter estimates and their standard deviations tend to be very similar to the original ML ones, independent of the pattern of the  $\hat{\alpha}_{i,1}$ 's and the sample size. Therefore, jackknife is not as good as bootstrap in warning about possible inadequacy of the asymptotic standard deviations. This undesirable result for the jackknife method may be due to the small number of different jackknife pseudovalues (16 in the case examined), and a larger number of items would provide more satisfactory results.

(8)- Considering the results from the comparison between the bootstrap, the normal bootstrap and the original ML discrimination parameter estimates  $\hat{\alpha}_{i,1}$ , this study suggests that

In general, when there is some difference between the bootstrap, normal bootstrap and the ML results for  $\hat{\alpha}_{i,1}$ , bootstrap estimates are closer to each other than to the ML estimates. The strongest similarity among them is related to the interquartile difference  $Q_3-Q_1$ , which could be expected since most or all of the estimates responsible for the skewness of the distributions are not considered.

The significant agreement between most of the (empirical) bootstrap and the normal bootstrap results is probably because the ML parameter estimates  $\hat{\alpha}_{i,1}$  and  $\hat{\alpha}_{i,0}$  are very close to the 'true' values.

At the same time, it is supporting the evidence that in some situations ( $\hat{\alpha}_{i,1}$  very large) the asymptotic theory is likely underestimating the standard deviations and most of the estimates related to the ART on black women can not be trusted.



(9) Regarding to the comparison between the asymptotic, bootstrap and jackknife estimates of correlations this study suggests that

Except for the Arithmetic Reasoning Test (ART) on black women, there is no difference between correlations based on the observed second derivative matrix and those obtained from the information matrix.

Except for the ART on black women, jackknife estimates of the correlations are equal or closer to the asymptotic ones than the corresponding bootstrap. Actually, the jackknife estimates of the correlations are closer to the asymptotic than to the bootstrap estimates only for the Lombard and Doering data, where the bootstrap largest difference in relation to the asymptotic estimate is 0.41, while for the jackknife it is 0.17. For the first 3 sets of data, the differences in relation to the asymptotic correlations are up to 0.20, and whether they are significant or not it is difficult to say.

For the ART on black women, there are strong discrepancies among all results, whether comparing the asymptotic correlations or those with the corresponding bootstrap or jackknife estimates of the correlations. These results suggest that the asymptotic correlations probably can not be trusted.

8 - *References*

- Albanese, M.T. (1990) Latent variable models for binary response. Ph.D. thesis. University of London.
- Albanese, M.T. and Knott, M. (1990) Bootstrapping latent variable models for binary response. Submitted for publication in *British Journal of Mathematical and Statistical Psychology*.
- 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.
- Beran, R. and Srivastava, M.S. (1985) Bootstrap tests and confidence regions for functions of a covariance matrix. Ann. Statist., 13, 95-115.
- Chatterjee, S. (1984) Variance estimation in factor analysis: An application of the bootstrap. Br. J. Math. Statist. Psychol., 37, 252-262.
- Clarkson, D.B. (1979) Estimating the standard errors of rotated factor loadings by jackknifing. Psychometrika, 44, 297-314.
- Diciccio, T. and Tibshirani (1987) Bootstrap confidence intervals and bootstrap approximations. J. Amer. Statist. Assoc., 82, 163-170.
- Efron, B. (1979) Bootstrap methods: another look at the jackknife. Ann. Statist., 7, 1-26.
- Efron, B. (1984) Better bootstrap confidence intervals. Tech. Rep. Stanford University Dept. Statist.
- Efron, B. and Tibshirani, R. (1986) Bootstrap method for standard errors, confidence intervals and other measures of statistical accuracy. Statist. Science, 1, 54-75.

- Efron, B. (1987) Better bootstrap confidence intervals. J. Amer. Statist. Assoc., 82, 171-185.
- Grönroos, M. (1985) Bootstrapping in Factor Analysis. 45th Biennial Session of the International Statistical Institute, 12-22.
- Jorgensen, M.A. (1987) Jackknifing fixed points of iterations. Biometrika, 74, 207-211.
- Jöreskog, K.G. and Sörbom, D. (1984) LISREL VI user's guide. Mooresville, Indiana: Scientific Software, Inc.
- 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.
- Miller, R.G. (1974) The jackknife - a review. Biometrika, 61, 1-15.
- Mislevy, R.J. (1985) Estimation of latent group effects. J. Am. Statist. Assoc., 80, 993-997.
- Pennell, R. (1972) Routinely computable confidence intervals for factor loadings using the 'jackknife'. Br. J. Math. Statist. Psychol., 25, 107-114.
- Quenouille, M.H. (1956) Notes on bias estimation. Biometrika, 43, 353-360.
- Shea, B.L. (1984) FACONE: A computer program for fitting the logit latent model by maximum likelihood. Department of Statistics, London School of Economics.
- Stouffer, S. and Toby, J. (1951) Role conflict and personality. J. Sociology, 56, 395-406.
- Stouffer, S., Guttman, L., Suchman, E., Lazarsfeld, P., Star, S. and Clausen, J. (1950) Measurement and prediction, Volume 4 of Studies in Social Psychology during World War. Princeton, NJ: Princeton University Press. Reprinted, 1966, New York: John Wiley.
- Tukey, J.W. (1958) Bias and confidence in not quite large samples. Ann. Math. Statist., 29, 614 [61].

Série A: Trabalho de Pesquisa

1. Marcos Sebastiani - Transformation des Singularités - MAR/89.
2. Jaime Bruck Ripoll - On a Theorem of 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 -  $\epsilon$  - 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  $\epsilon$  - Convex Hypersurfaces - NOV/89.
10. Jandyra Maria G. Fachel - Coeficientes de Correlação Tipo -Contingência - JAN/90.
11. Jandyra Maria G. Fachel - The Probability of Occurrence of Heywood Cases - JAN/90.
12. Jandyra Maria G. Fachel - Heywood Cases in Unrestricted Factor Analysis - JAN/90.
13. Julio Cesar R. Claeysen, 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