NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Beh, Eric J.; Farver, Thomas B. "The box-cox transformation and non-iterative estimation methods for ordinal log-linear models" Australian & New Zealand Journal of Statistics Vol. 54, Issue 4, p. 475-484 (2012)

Available from: <http://dx.doi.org/10.1111/anzs.12007>

This is the accepted version of the following article: Beh, Eric J.; Farver, Thomas B. "The box-cox transformation and non-iterative estimation methods for ordinal log-linear models" Australian & New Zealand Journal of Statistics Vol. 54, Issue 4, p. 475-484 (2012), which has been published in final form at <http://dx.doi.org/10.1111/anzs.12007>

Accessed from: <http://hdl.handle.net/1959.13/1051997>

THE BOX-COX TRANSFORMATION AND NON-ITERATIVE ESTIMATION METHODS FOR ORDINAL LOG-LINEAR MODELS

ERIC J. BEH ^{a,*1}, THOMAS B. FARVER ^b

University of Newcastle, Australia and University of California, Davis, USA

Summary

Recently Beh and Farver investigated and evaluated three non-iterative procedures for estimating the linear-by-linear parameter of an ordinal log-linear model. The study demonstrated that these non-iterative techniques provide estimates that, for most types of contingency tables, are statistically indistinguishable when compared with Newton's unidimensional algorithm. Here we show how two of these techniques are related using the Box-Cox transformation. We also show that by using this transformation, accurate non-iterative estimates are achievable even when a contingency table consists of sampling zeros.

Keywords: Box-Cox transformation; linear-by-linear association; Newton's unidimensional method; ordinal log-linear model.

1. Introduction

Ordinal log-linear models are an extremely useful tool when analysing the association structure of contingency tables. In the very simple case where such a table consists of two categorical variables, the parameter of such a model has a clear and useful interpretation; the parameter estimates the common log-odds ratio of the contingency table. Traditionally, iterative algorithms have been used to estimate this parameter and its standard error; these include iterative proportional fitting and Newton's algorithm and hence estimates derived from such techniques may be considered as the traditional 'benchmark' estimates. One may

* Author to whom correspondence should be addressed.

¹ School of Mathematical & Physical Sciences, University of Newcastle, Callaghan, NSW, 2308, Australia.
email: eric.beh@newcastle.edu.au

² Department of Population Health and Reproduction, School of Veterinary Medicine, 1014 Haring Hall, University of California, Davis, CA 95616 – 8743, USA.

refer to, for example, Galindo – Garre & Vermunt (2004) and Ait-Sidi-Allal, Baccini & Mondat (2004), for recent extensive discussions of these iterative procedures. Non-iterative procedures for estimating the linear-by-linear association parameter may also be considered and were proposed by Beh & Davy (2004) and more recently extended by Beh & Farver (2009). The latter study provided an empirical evaluation of three direct methods for estimating the parameter and showed that, for nearly all types of two-way contingency tables (of varying dimension and sample size), such non-iterative estimates are statistically indistinguishable when compared with Newton’s unidimensional algorithm. Two such estimation methods are referred to as BDNI and LogNI, where BDNI is a first order approximation of the LogNI using the Taylor series expansion of the natural logarithm function. LogNI provides exceptionally good estimates but only when the contingency table does not consist of zero cell frequencies. Conversely, BDNI can easily take into consideration zero cell values, but it is not as accurate as LogNI or the iterative procedures. In this paper we demonstrate how LogNI and BDNI are related using the Box-Cox transformation (Box & Cox 1964) and show that, by using the transformation, the accuracy of LogNI can be considered even when sampling zeros are present in the contingency table.

To discuss these issues, this paper is divided into four further sections. Section 2 provides a very brief overview of the ordinal log-linear model for a two-way contingency table consisting of ordered row and ordered column categories. We also discuss iterative and non-iterative techniques for estimating the parameter of interest which reflects the linear-by-linear association between the two ordered categorical variables. In particular we focus our attention on the BDNI and LogNI non-iterative estimates of this parameter. Section 3 shows how the Box-Cox transformation can be used for the non-iterative estimation of the parameter, while Section 4 demonstrates the application of this transformation. Some concluding remarks are left for the discussion (Section 5).

2. The ordinal log-linear model

Suppose we consider an $I \times J$ two-way contingency table, N , where the (i, j) th cell entry is given by n_{ij} for $i = 1, \dots, I$ and $j = 1, \dots, J$. Let the grand total of N be n and the matrix of joint relative frequencies be P , so that the (i, j) th cell entry is $p_{ij} = n_{ij} / n$ and

$\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1$. Define the i th row marginal proportion by $p_{i\bullet} = \sum_{j=1}^J p_{ij}$ and define the j th column marginal proportion as $p_{\bullet j} = \sum_{i=1}^I p_{ij}$.

Consider the ordinal log-linear model

$$\ln m_{ij} = \mu + \alpha_i + \beta_j + \phi(u_i - \bar{u})(v_j - \bar{v}) \quad (1)$$

where m_{ij} is the expected cell frequency of the (i, j) th cell of the table, μ is the grand mean, and the α_i , $i=1, \dots, I$, and β_j , $j=1, \dots, J$, are main effects of the i th row and j th column, respectively. They may be estimated such that

$$\begin{aligned} \hat{\mu} &= \ln(n) + \frac{1}{I} \sum_{i=1}^I \ln(p_{i\bullet}) + \frac{1}{J} \sum_{j=1}^J \ln(p_{\bullet j}), \\ \hat{\alpha}_i &= \ln(p_{i\bullet}) - \frac{1}{I} \sum_{i=1}^I \ln(p_{i\bullet}) \quad \text{and} \quad \hat{\beta}_j = \ln(p_{\bullet j}) - \frac{1}{J} \sum_{j=1}^J \ln(p_{\bullet j}). \end{aligned}$$

where the latter two parameter are constrained so that $\sum_{i=1}^I \alpha_i = \sum_{j=1}^J \beta_j = 0$. See, for example, Beh & Davy (2004).

For model (1), u_i is the score assigned to the i th ordered row category and v_j is the score for the j th ordered column category. The quantities $\bar{u}$ and $\bar{v}$ are the weighted mean of the I row scores and J columns scores, respectively, such that $\bar{u} = \sum_{i=1}^I p_{i\bullet} u_i$ and $\bar{v} = \sum_{j=1}^J p_{\bullet j} v_j$ where u_i and v_j is the score associated with the i th row and j th column, respectively, and reflect the ordered structure of the categories. A variety of different scoring techniques may be considered; however, we confine our attention to *natural scores* in the example below. These scores are consecutive integer scores such that $u_i = i$ for $i = 1, 2, \dots, I$ and $v_j = j$ for $j = 1, 2, \dots, J$.

Suppose that one treats μ , α_i and β_j as nuisance parameters and focuses solely on the estimation of ϕ . When unit-distant row and column scores (such as natural scores) are

used to reflect the ordinal structure of each variable so that $u_{i+1} - u_i = 1$ and $v_{j+1} - v_j = 1$ respectively, ϕ is the estimate of the constant log-odds ratio of the contingency table:

$$\phi = \ln \left(\frac{m_{ij} m_{i+1, j+1}}{m_{i, j+1} m_{i+1, j}} \right).$$

There are various algorithmic strategies that can be adopted that provide an accurate means of determining the maximum likelihood estimate of ϕ . One of the more popular approaches is to consider iterative proportional fitting, or Newton's algorithm. For discussions on these, and other approaches, the reader is directed to the contributions of Goodman (1979), Haberman (1974), Agresti (1984, appendix B), Nelder & Wedderburn (1972), Galindo-Garre & Vermunt (2004) and Ait-Sidi-Allal, Baccini & Mondat (2004).

The parameter ϕ may be estimated without the use of iterative algorithms by considering the non-iterative estimates

$$\hat{\phi}_{\text{LogNI}} = \frac{1}{\sigma_I^2 \sigma_J^2} \sum_{i=1}^I \sum_{j=1}^J p_{i\bullet} p_{\bullet j} (u_i - \bar{u})(v_j - \bar{v}) \ln \left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right) \quad (2)$$

and

$$\hat{\phi}_{\text{BDNI}} = \frac{1}{\sigma_I^2 \sigma_J^2} \sum_{i=1}^I \sum_{j=1}^J p_{ij} (u_i - \bar{u})(v_j - \bar{v}) \quad (3)$$

where $\sigma_I^2 = \sum_{i=1}^I p_{i\bullet} u_i^2 - \bar{u}^2$ and $\sigma_J^2 = \sum_{j=1}^J p_{\bullet j} v_j^2 - \bar{v}^2$. The estimate $\hat{\phi}_{\text{LogNI}}$ was proposed by Beh & Farver (2009) while Beh & Davy (2004) originally considered the estimate $\hat{\phi}_{\text{BDNI}}$. The study undertaken by Beh & Farver (2009) empirically showed that $\hat{\phi}_{\text{LogNI}}$ and $\hat{\phi}_{\text{BDNI}}$ produced estimates that were statistically indistinguishable when compared with the MLE (under the Poisson model) of ϕ found using Newton's unidimensional algorithm and denoted by $\hat{\phi}_N$. The study of Beh & Farver (2009) confirmed that for small ϕ values both estimates are statistically equivalent to $\hat{\phi}_N$; however, in general, $\hat{\phi}_{\text{LogNI}}$ is preferable since it provides a more stable estimate of the parameter for the variety of sized contingency tables studied and

ϕ considered. $\hat{\phi}_{\text{LogNI}}$ produces estimates of ϕ that are more comparable to those obtained using Newton's unidimensional algorithm than $\hat{\phi}_{\text{BDNI}}$ if such benchmarking is necessary.

The standard error of these estimates may be calculated such that

$$SE(\hat{\phi}) = \left[\sum_{i=1}^I \sum_{j=1}^J (u_i - \bar{u})^2 (v_j - \bar{v})^2 m_{ij} \right]^{-1/2}$$

where m_{ij} is the estimated value of n_{ij} given the estimate of the parameter ϕ .

3. Box-Cox transformation

The derivation of $\hat{\phi}_{\text{BDNI}}$ involves involves the first order Taylor series approximation of $\hat{\phi}_{\text{LogNI}}$ (see Beh & Farver 2009). An alternative link can be established between the two estimates by considering the Box-Cox transformation. Box & Cox (1964) demonstrated that, for a random sample of n observation, $x_1, x_2, \dots, x_n$, the following transformation may be applied:

$$x_i^{(\alpha)} = \begin{cases} \frac{1}{\alpha} (x_i^\alpha - 1), & \alpha \neq 0 \\ \ln(x_i), & \alpha = 0 \end{cases}$$

where $x_i > 0, \forall i$. Typically such a transformation is considered when x_i has a heavily skewed distribution and is therefore applicable for overcoming various normality assumptions that have been violated. There is a wealth of literature on this issue and the reader is encouraged to consider Broemeling (2006) and Loh (2006) for introductory discussions.

It can be seen that the non-iterative estimate of ϕ using $\hat{\phi}_{\text{LogNI}}$ involves the natural logarithm of

$$x_{ij} = \frac{p_{ij}}{p_{i\bullet} p_{\bullet j}}$$

which is the (i, j) th Pearson ratio of the two-way contingency table – see, for example, Beh (2004) and Goodman (1996). Note that (2) and (3) may be alternatively expressed in terms of these ratio's such that

$$\hat{\phi}_{\text{LogNI}} = \frac{1}{\sigma_I^2 \sigma_J^2} \sum_{i=1}^I \sum_{j=1}^J p_{i\bullet} p_{\bullet j} (u_i - \bar{u})(v_j - \bar{v}) \ln(x_{ij})$$

and

$$\hat{\phi}_{\text{BDNI}} = \frac{1}{\sigma_I^2 \sigma_J^2} \sum_{i=1}^I \sum_{j=1}^J x_{ij} p_{i\bullet} p_{\bullet j} (u_i - \bar{u})(v_j - \bar{v}).$$

When the Pearson ratio is 1, there is complete independence between the row and column variable and, by considering an ordinal log-linear model of the form of (1), the linear-by-linear association parameter ϕ is zero. Any value of $0 \leq x_{ij} < 1$ indicates that the expected cell values (under the hypothesis of independence) have been underestimated while overestimation of the cell frequencies arises when $x_{ij} > 1$. Thus, the histogram of the Pearson ratio's for a particular contingency table will be very heavily positively skewed. By incorporating the Box-Cox transformation into $\hat{\phi}_{\text{LogNI}}$, (2), the non-iterative estimate of the linear-by-linear association parameter can be expressed as a function of α such that

$$\hat{\phi}_{\text{BCNI}}(\alpha) = \frac{1}{\alpha \sigma_I^2 \sigma_J^2} \sum_{i=1}^I \sum_{j=1}^J p_{i\bullet} p_{\bullet j} (u_i - \bar{u})(v_j - \bar{v}) \left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\alpha \quad (4)$$

since $\sum_{i=1}^I \sum_{j=1}^J p_{i\bullet} p_{\bullet j} (u_i - \bar{u})(v_j - \bar{v}) = 0$. The estimate (4) is referred to as the BCNI estimate of ϕ and is valid for any real, but unknown, $\alpha \neq 0$. Thus, using the Box-Cox transformation, $\hat{\phi}_{\text{BCNI}}(0) = \hat{\phi}_{\text{LogNI}}$ while $\hat{\phi}_{\text{BCNI}}(1) = \hat{\phi}_{\text{BDNI}}$. To achieve a non-iterative estimate equivalent to the benchmark value often derived using Newton's unidimensional algorithm (or some other iterative procedure), the BCNI estimate (4) suggests estimating the relevant value of α that provides such equality. Finding the value of α such that the BCNI estimate is equivalent to another estimate of ϕ , $\hat{\phi}$ (such as Newton's estimate), can be made (iteratively) by considering

$$\alpha_{n+1} = \alpha_n \left(1 + \frac{\hat{\phi}_{\text{BCNI}}(\alpha_n) - \hat{\phi}}{\hat{\phi}_{\text{BCNI}}(\alpha_n) - \phi(\alpha_n)} \right) \quad (5)$$

where

$$\phi(\alpha_n) = \frac{1}{\sigma_I^2 \sigma_J^2} \sum_{j=1}^J \sum_{i=1}^I \ln \left(\frac{p_{ij}}{p_{i\cdot} p_{\cdot j}} \right) p_{i\cdot} p_{\cdot j} (u_i - \bar{u})(v_j - \bar{v}) \left(\frac{p_{ij}}{p_{i\cdot} p_{\cdot j}} \right)^{\alpha_n}$$

for $\alpha_0 \neq 0$. Equation (5) may be obtained by solving (4) using the Newton-Raphson algorithm. One must also be sure that the non-iterative estimate obtained that is equivalent to an iterative value is of relevance since the study of Beh & Farver (2009) suggest that the LogNI estimate (2) provides non-iterative estimates of ϕ that are often almost indistinguishable when compared to estimates obtained by iterative procedures.

However, zero cell frequencies regularly arise in two-way contingency tables and these have been described as being either *structural zeros* (where it is theoretically impossible for a positive cell frequency to occur) or *sampling zeros* (where the zero cell frequency randomly occurs and has a positive expected cell frequency under some pre-defined model of association) (see, for example, Agresti 1984, page 54). One may also consider Christensen (1990, chapter 8) who referred to these zero cells as fixed and random zero's, respectively. In the case where a contingency table consists of at least one sampling (random) zero, due to the presence of the logarithm function it is not possible to determine an estimate of ϕ using $\hat{\phi}_{\text{LogNI}}$ (2). Ishii-Kuntz (1994, page 18) describes a variety of ways of dealing with sampling zeros. The simplest solution involves replacing a zero cell frequency with 0.5, or alternatively adding 0.5 to all cell frequencies (even those that are non-zero). Another strategy is to replace a zero cell frequency with a value that is smaller than 0.5. For example Evers & Namboodiri (1978) suggest 0.2, Beh & Farver (2009) consider 0.05 while Clogg & Eliason (1988, page 13) suggest 0.00000001, and suggested that perhaps an even smaller quantity may be considered. As an alternative to these strategies one may also consider estimating ϕ by considering $\hat{\phi}_{\text{BCNI}}(\alpha)$ (4) for some value of α close to (but not equal to) zero. Below we examine by example the impact of replacing sampling zeros with small quantities on the estimation of ϕ and examine the BCNI estimate for various values of α .

4. Association of occupational class amongst British men and their fathers

To demonstrate the behaviour of the non-iterative estimate of ϕ as a function of α , we considered Table 1 which was originally analysed by Glass (1954). It has since been examined in detail by Goodman (1979) and in subsequent studies of association for contingency tables. Table 1 cross-classifies 3498 British men according to their occupational class and the occupational class of their father. These classes are divided into eight categories of increasing status so that the table consists of ordered row categories and ordered column categories. Therefore, the association between the two variables can be quantified by estimating the linear-by-linear association parameter, ϕ , from the ordinal log-linear model (1).

TABLE 1

Cross-classification of 3498 British men according to their occupational status and their father's occupational status

Fathers	Sons Status								Total
Status	1	2	3	4	5	6	7	8	
1	50	19	26	8	7	11	6	2	129
2	16	40	34	18	11	20	8	3	150
3	12	35	65	66	35	88	23	21	345
4	11	20	58	110	40	183	64	32	518
5	2	8	12	23	25	46	28	12	156
6	12	28	102	162	90	554	230	177	1355
7	0	6	19	40	21	158	143	71	458
8	0	3	14	32	15	126	91	106	387
Total	103	159	330	459	244	1186	593	424	3498

By implementing Newton's unidimensional algorithm to estimate ϕ , $\hat{\phi}_N = 0.1638$ and convergence to four decimal places occurs after 23 iterations. The standard error of this estimate is 0.0043. Due to the presence of the two zero cell frequencies in the bottom left hand side of Table 1, it is not possible to determine the LogNI estimate of ϕ . However, $\hat{\phi}_{BDNI} = 0.1404$ and has a standard error of 0.0033.

When there are sampling zeros present in a contingency table, obtaining an approximation to the LogNI estimate of ϕ can be achieved in two ways. Firstly, the zero cell frequencies can be replaced with a reasonably small positive value; here we considered 0.00000001, 0.05 and 0.5. Alternatively, the Box-Cox transformation can be applied to the non-iterative estimation techniques and the estimate of ϕ determined using $\hat{\phi}_{\text{BCNI}}(\alpha)$ (4) by considering various values of α .

Prior to performing the Box-Cox transformation it is instructive to examine the Pearson ratios. Figure 1 provides a histogram of these ratios (a) and the natural logarithm of their values (b). Figure 1 (a) shows that the Pearson ratios are very positively skewed while Figure 1 (b) shows that the natural logarithm of the ratios produces a fairly symmetric distribution. This suggests that the Box-Cox transformation can be considered for estimating ϕ and observing its behaviour for different values of α . Since the BCNI estimate, $\hat{\phi}_{\text{BCNI}}(\alpha)$, is equivalent to the LogNI estimate, $\hat{\phi}_{\text{LogNI}}$, when $\alpha = 0$ and is equal to the BDNI estimate, $\hat{\phi}_{\text{BDNI}}$, when $\alpha = 1$, we confined our attention to the case where $0 \leq \alpha \leq 1$. However, one may also examine the magnitude of the parameter estimate for values of α that lie outside of this interval.

Figure 1: The histograms of the Pearson ratio's (a) and the natural logarithm of the Pearson ratio's (b) of Table 1.

Table 2 provides a summary of the estimates of ϕ for Table 1 using Newton's univariate method, and the BCNI estimate, $\hat{\phi}_{\text{BCNI}}(\alpha)$, when $\alpha = 0$ and 1 (thereby providing the LogNI and BDNI estimates, respectively) and $\alpha = 0.1$. It also considers such estimates when the sampling zeros are replaced by 0.00000001, 0.05 and 0.5.

Table 2 shows that, when substituting sampling zeros with reasonably small quantities, there is very little variation in the estimate of ϕ for Table 1 when using Newton's method. Also, there is no variation (up to the fourth decimal place) in the standard error of these estimates. However, there is considerable variation in the LogNI estimates when substituting zero cell frequencies with 0.00000001, 0.05 and 0.5 due to the presence of the logarithm function in $\hat{\phi}_{\text{LogNI}}$. If a value close (but not equal) to zero is chosen, the BDNI estimates become more consistent and are more comparable with Newton's estimate. Table 2 shows that choosing $\alpha = 0.1$ and not making any substitutions for the sampling zeros, the estimate of ϕ is 0.1829 with a standard error of 0.0025.

TABLE 2

Comparison of estimates of ϕ using Newton's unidimensional algorithm and $\hat{\phi}_{\text{BCNI}}(\alpha)$ for $\alpha = 0, 0.1$ and 1 with 0, 0.00000001, 0.05 and 0.5 substituted for sampling zeros

Estimation Procedure	Substitution of Sampling Zeros			
	0	0.00000001	0.05	0.5
$\hat{\phi}_{\text{N}}$	0.1638	0.1639	0.1638	0.1631
$\text{SE}[\hat{\phi}_{\text{N}}]$	(0.0043)	(0.0043)	(0.0043)	(0.0043)
$\hat{\phi}_{\text{BCNI}}(\alpha = 0) = \hat{\phi}_{\text{LogNI}}$	Does Not Exist	0.2525	0.1574	0.1432
$\text{SE}[\hat{\phi}_{\text{BCNI}}(\alpha = 0)]$	(---)	(0.0016)	(0.0030)	(0.0032)
$\hat{\phi}_{\text{BCNI}}(\alpha = 0.1)$	0.1829	0.1753	0.1473	0.1381
$\text{SE}[\hat{\phi}_{\text{BCNI}}(\alpha = 0.1)]$	(0.0025)	(0.0027)	(0.0032)	(0.0033)
$\hat{\phi}_{\text{BCNI}}(\alpha = 1) = \hat{\phi}_{\text{BDNI}}$	0.1404	0.1404	0.1403	0.1399
$\text{SE}[\hat{\phi}_{\text{BCNI}}(\alpha = 1)]$	(0.0033)	(0.0033)	(0.0033)	(0.0033)

Figure 2 graphically shows the relationship between $\hat{\phi}_{\text{BCNI}}(\alpha)$ and α when taking into consideration our small quantity substitutions for the sampling zeros. Figure 2 (a) shows the change in the non-iterative estimate as α varies from 0 to 1 while Figure 2 (b) shows the change in the standard error of the estimate. When $\alpha = 0$, the non-iterative estimate of ϕ is the LogNI estimate studied by Beh & Farver (2009). We can see here that replacing a zero cell value with either 0.000001, 0.05 or 0.5 has a relatively large impact on the estimate that is calculated; this confirms the results summarised in Table 2 for $\alpha = 0, 0.1$ and 1. As α increases from 0 to 1, Figure 2 (a) shows that the parameter estimate decreases until α reaches approximately 0.6 then increases to its BDNI estimate. However, the change in the parameter estimate becomes almost negligible between $\alpha = 0.4$ and 1. Therefore, if one were to consider the original non-iterative estimate proposed by Beh & Davy (2004), adding a small fraction to the zero cell entries has a negligible effect. Figure 2 (b) shows that this is also the case when considering the standard error of the estimate.

Figure 2. (a) BCNI estimates of ϕ for $\alpha \in [0, 1]$ and (b) its standard error.

For zero cell frequencies, the following lines reflect the substitution made for n_{ij} : — $n_{ij} = 0$, - - - $n_{ij} = 0.00000001$, $n_{ij} = 0.05$, - . - . $n_{ij} = 0.5$.

The solid horizontal line in Figure 2 (a) corresponds to the Newton estimate of ϕ which is 0.1638 while the solid horizontal line of Figure 2 (b) identifies the standard error of

0.0043 for the Newton estimate of ϕ . If one wishes, it is possible to calculate a non-iterative estimate of the parameter that is very similar to its iterative counterpart in one of two ways. Near equivalence of the non-iterative estimate and Newton's estimate can be achieved by replacing a zero cell value with 0.05 and calculating the LogNI estimate; such an estimate is 0.1639. This is indeed the practice that was undertaken in Beh & Farver (2009) and provides very comparable estimates with Newton's estimate. Alternatively, one may also keep in place the zero cell value, but consider the BCNI estimate, (4), with an approximate α value, such as 0.15. Using this approach $\hat{\phi}_{\text{NI}}(0.15) = 0.1615$ with a standard error of 0.00029.

A non-iterative estimate of ϕ that is equivalent to the Newton estimate may be achieved by considering a more refined estimate of α . In fact, by considering (5) when $\alpha_0 < 0.6$, $\alpha = 0.1423$ produces such an estimate and this can be seen verified by considering Figure 2 (a). Other values of α that lie outside of the interval $[0, 1]$ may also be examined and it can be shown that by setting $\alpha_0 > 0.6$ in (5), $\alpha = 1.3074$ provides equivalent BCNI and Newton estimates.

5. Discussion

This paper has explored the application of the Box-Cox transformation for the non-iterative estimation of the ϕ parameter for the ordinal log-linear model (1). While the BDNI estimate is a first order approximation of the LogNI approximation, their link can also be established by considering the BCNI estimate, (4), for specific values of α ; namely $\hat{\phi}_{\text{BCNI}}(0) = \hat{\phi}_{\text{LogNI}}$ and $\hat{\phi}_{\text{BCNI}}(1) = \hat{\phi}_{\text{BDNI}}$.

While Beh & Farver (2009) empirically show that the BDNI estimate (3) produces relatively poor estimates when compared with the LogNI estimate (2) and Newton's estimate, the LogNI estimate produces estimates that are almost indistinguishable when compared with the estimate calculated using Newton's method. However, due to the presence of the natural logarithm function, the LogNI estimate is not appropriate for two-way contingency tables that consist of zero's. This paper has shown that, by treating zero cell frequencies as sampling zero's, the BCNI estimate, (4), can be used as an accurate and reliable estimation method for small α values. For Table 1 the appropriate value of α to choose so that the non-iterative estimate is comparable to Newton's estimate was approximately 0.15. In fact, empirical

evidence based on the application of the BCNI estimate to contingency tables of varying size and ϕ value with sampling zeros suggests that using a α value of between -0.2 and $+0.2$ often produces a non-iterative estimate that is comparable to Newton's unidimensional method. For tables with all cell frequencies that are positive, an α value close to zero will work well and this is verified by considering the accuracy and reliability of the LogNI estimate. Calculating a non-iterative estimate that is equivalent to the MLE found by an iterative procedure can be also done by considering α values that lie outside of the interval $[0, 1]$.

The application of the BCNI estimate to Table 1 shows that the original non-iterative estimation procedure proposed by Beh & Davy (2004) produces an estimate of ϕ that is close to the minimum value it can take. In fact, evidence derived from considering applications of the BDNI estimate to a large variety of contingency tables suggest that such an estimate is indeed close to the global minimum (for positively associated ordinal categorical variables) or the global maximum (for negatively associated variables) that $\hat{\phi}$ can attain. Depending on the concavity of the relationship between $\hat{\phi}_{\text{BCNI}}(\alpha)$ and α given by (4), the accuracy of the BDNI estimate may be quite good, or very poor. This issue needs further investigation. One may further explore, such as via simulation, the behaviour of α and its impact on the estimation of $\hat{\phi}$. This issue will also be left for future consideration.

Caution must be taken not to always seek out estimates of ϕ that are (close to) equivalent to the ML estimate. Due to the nature of iterative algorithms, parameter estimates may not be calculable due to lack of convergence or because the initial value of the parameter of interest was poorly chosen; although, these problems are a rarity when considering Newton's unidimensional algorithm for contingency table analysis. However, the LogNI and BDNI estimates provide a simple, yet accurate and reliable estimate of the parameter, ϕ . As described in this paper using $\hat{\phi}_{\text{BCNI}}(\alpha)$ establishes further the link between these estimates and provides an alternative non-iterative approach that further generalises the estimation of ϕ from an ordinal log-linear model.

References

- AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.
- AIT-SIDI-ALLAL, M.L., BACCINI, A. & MONDOT, A.M. (2004). A new algorithm for estimating the parameters and their asymptotic covariance in correlation and association models. *Comput. Statist. Data Anal.* **45**, 389 – 421.
- BEH, E.J. (2004). Simple correspondence analysis: A bibliographic review. *Int. Stat. Rev.* **72**, 257 – 284.
- BEH, E.J. & DAVY, P.J. (2004). A non-iterative alternative to ordinal log-linear models, *J. Appl. Math. Decis. Sci.* **8**, 67-86.
- BEH, E.J. & FARVER, T.B. (2009). An evaluation of non-iterative methods for estimating the linear-by-linear parameter of ordinal log-linear models. *Aust. N. Z. J. Stat.* **51**, 335 – 352.
- BOX, G.E.P. & COX, D.R. (1964). An analysis of transformations. *J. R. Stat. Soc. Ser. B.* **26**, 211 – 252.
- BROEMELING, L.D. (2006). Box-Cox transformations – I. In *Encyclopedia of Statistical Sciences* (2nd Edn), eds N. Balakrishnan, C. B. Read, B. Vidakovic, pp. 636 – 637. Hoboken, NJ: Wiley.
- CHRISTENSEN, R. (1990). *Log-Linear Models and Logistic Regression*. New York: Springer.
- CLOGG, C. C. & ELIASON, S. R. (1987). Some common problems in log-linear analysis. *Sociol. Methods Res.* **16(8)**, 8 – 44.
- EVERS, M. & NAMBOODIRI, N.K. (1979). On the design matrix strategy in the analysis of categorical data. In *Sociological Methodology 1979*, ed K. F. Schuessler, pp. 86 – 111. San Francisco, CA: Jossey-Bass.
- GALINDO-GARRE, F. & VERMUNT, J.K. (2004). The order-restricted association model: Two estimation algorithms and issues in testing. *Psychometrika* **69**, 641 – 654.
- GLASS, D.V. (1954). *Social Mobility in Britain*. Glencoe, IL: Free Press.
- GOODMAN, L.A. (1979). Simple models of the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74**, 537–552.
- GOODMAN, L.A. (1996). A single general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondence analysis and association analysis. *J. Amer. Statist. Assoc.* **91**, 408–428.

- HABERMAN, S.J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics* **30**, 589 – 600.
- ISHII-KUNTZ, M. (1994). *Ordinal Log-Linear Models*. Sage University Series on Quantitative Applications in the Social Sciences. Newbury Park, CA: Sage.
- LOH, W.Y. (2006). Box-Cox transformations – II. In *Encyclopedia of Statistical Sciences* (2nd Edn), eds N. Balakrishnan, C. B. Read, B. Vidakovic, pp. 637 – 644. Hoboken, NJ: Wiley.
- NELDER, J.A. & WEDDERBURN, R.W.M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A.* **135**, 370 – 384.