Journal of the Korean Statistical Society (1998), 27: 4, pp 495 506

The Generalized Logistic Models with Transformations

In-Kwon Yeo[1] and Richard A. Johnson[2]

ABSTRACT

The proposed class of generalized logistic models, indexed by an extra parameter, can be used to model or to examine symmetric or asymmetric discrepancies from the logistic model. When there are a finite number of different design points, we are mainly concerned with maximum likelihood estimation of parameters and in deriving their large sample behavior. A score test and a bootstrap hypothesis test are also considered to check if the standard logistic model is appropriate to fit the data or if a generalization is needed.

Keywords: Bootstrap hypothesis test; Generalized logistic model; Maximum likelihood estimation; Power transformation; Score test

1. INTRODUCTION

The linear logistic regression models are commonly adapted for modeling the dependence of a binary response on several explanatory variables. It is assumed that if $Y_1, \ldots, Y_n$ are independent Bernoulli random variable with mean $\mu_i = P(Y_i = 1)$, then, for $i = 1, \ldots, n$,

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{\beta}'\boldsymbol{x}_i \quad \text{or} \quad \mu_i = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}, \tag{1.1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ is the i-th vector of observations on explanatory variables.

Although the logit link is commonly used to describe how the mean response depends linearly on the predictors, linearity cannot be taken for granted. In practice, the linearity should, if possible, be checked, one check is provided by the chi-square goodness of fit test. Several authors including Pregibon (1980),

[1]Research Institute of Applied Statistics, Sung Kyun Kwan University, Seoul, 110-745, Korea
[2]Department of Statistics, University of Wisconsin - Madison, WI 53706, U.S.A.

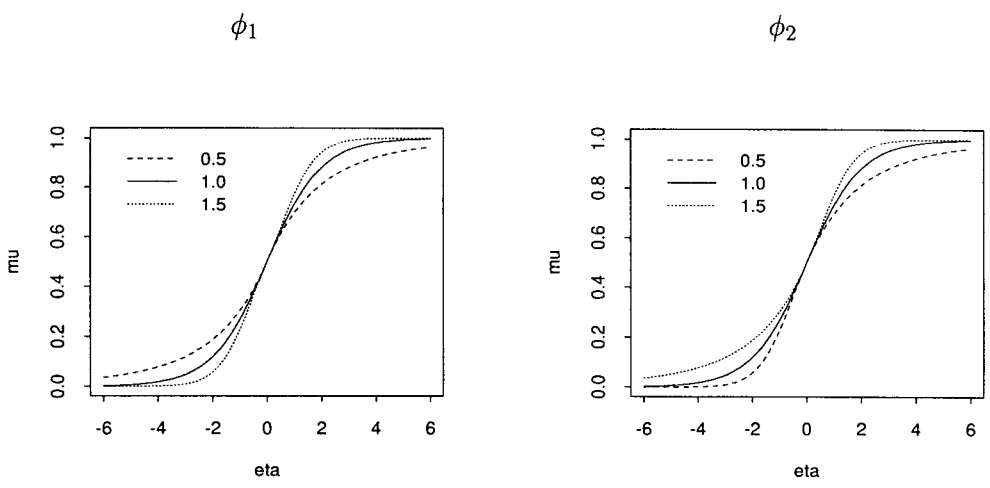

Figure 2.1: Plots of generalized logistic curves using ϕ_1 and ϕ_2 against η.

Aranda-Ordaz (1981), Guerrero and Johnson (1982), and Stukel (1988) have proposed the generalization in which the transformation with additional parameters is applied. They concerned either deriving asymptotics of estimators or testing transformation parameters using the score.

In this article, two families of one-parameter transformations which influence either symmetry or asymmetry of the fitted curve are applied in order to evaluate a certain type of lack of fit and to improve the fit. The selection of transformation parameter is completely based on the data. The determination of plausible values for the transformation parameter allows us to decide whether or not the logistic regression is an appropriate model. We establish the asymptotic properties of maximum likelihood estimators of parameters as well as develop a bootstrap test to examine deviations from the standard logistic model.

2. GENERALIZED LOGISTIC MODELS

As proposed by Stukel (1988), our model takes the form

$$\log\left(\frac{\mu}{1-\mu}\right) = \phi(\lambda,\eta) \quad \text{or} \quad \mu = \frac{\exp(\phi(\lambda,\eta))}{1+\exp(\phi(\lambda,\eta))}, \tag{2.1}$$

where $\eta = \boldsymbol{\beta}'\boldsymbol{x}$ and the $\phi(\lambda,\eta)$ are strictly increasing nonlinear transformations of η which are specified by λ. The two families of power transformations are

considered:

$$\phi_1(\lambda,\eta) = \begin{cases} \big((\eta+1)^\lambda - 1\big)/\lambda & \lambda \neq 0, \eta \geq 0, \\ \log(\eta+1) & \lambda = 0, \eta \geq 0, \\ -\big((-\eta+1)^\lambda - 1\big)/\lambda & \lambda \neq 0, \eta < 0, \\ -\log(-\eta+1) & \lambda = 0, \eta < 0, \end{cases} \tag{2.2}$$

$$\phi_2(\lambda,\eta) = \begin{cases} \big((\eta+1)^\lambda - 1\big)/\lambda & \lambda \neq 0, \eta \geq 0, \\ \log(\eta+1) & \lambda = 0, \eta \geq 0, \\ -\big((-\eta+1)^{2-\lambda} - 1\big)/(2-\lambda) & \lambda \neq 2, \eta < 0, \\ -\log(-\eta+1) & \lambda = 2, \eta < 0. \end{cases} \tag{2.3}$$

Note that for $\lambda = 1$, both ϕ_1 and ϕ_2 reduce to the identical function. In fact, the functions ϕ_1 are the modulus transformations introduced by John and Draper (1980) and ϕ_2 are the power transformations by Yeo and Johnson (1997). The subscript will be suppressed for notational convenience if both transformations are applicable.

The generalized logistic curves, when ϕ_1 and ϕ_2 are applied, are shown in Figure 2.1 for $\lambda = 0.5$ (broken line), 1.0 (solid), and 1.5 (dotted). Since $\phi_1(\lambda,\eta)$ changes from convex to concave as η changes sign, the corresponding curves are symmetric about $\eta = 0$ and $\mu = 0.5$. On the contrary, $\phi_2(\lambda,\eta)$ are concave in η for $\lambda < 1$ and convex for $\lambda > 1$ so the generalized curves are asymmetrically pulled down in the tails for $\lambda < 1$ and pushed up for $\lambda > 1$. Hence, they might be used to select the model in which the response variables are asymmetrically affected by the predictors.

We may also consider the compositions of two transformation, $\phi_1(\lambda_1, \phi_2(\lambda_2,\eta))$ and $\phi_2(\lambda_2, \phi_1(\lambda_1,\eta))$, which provide an extensive class of generalizations and give a flexible improvement in fit. In a future paper, we will discuss the properties of these composition transformations. Here we assume that either model (2.2) or model (2.3) holds throughout.

3. PARAMETER ESTIMATION AND ASYMPTOTIC RESULTS

We assume that there exist some value for λ such that (2.1) holds. Suppose that there are a finite number, K, of distinct design points $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K)'$. Let n_k be the number of replicates and r_k the number of successes at design point $\boldsymbol{x}_k$, $k = 1, \ldots, K$. Under this assumption, maximum likelihood inference is based

on the log-likelihood function for $n = \sum_{k=1}^{K} n_k$ observations,

$$\begin{aligned} l_n(\boldsymbol{\theta}|\boldsymbol{y}) &= \sum_{k=1}^{K} \left\{ r_k \log\left(\frac{\mu_k(\boldsymbol{\theta})}{1-\mu_k(\boldsymbol{\theta})} \right) + n_k \log(1-\mu_k(\boldsymbol{\theta})) \right\} \\ &= \sum_{k=1}^{K} \left\{ r_k \phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k) - n_k \log(1+\exp(\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k))) \right\}, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\theta} = (\lambda, \beta_1, \ldots, \beta_p)'$. The maximum likelihood estimates, $\hat{\boldsymbol{\theta}}$, are obtained by solving the likelihood equations for λ and $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_p)$ by a Newton-Raphson procedure or by using the delta algorithm (Jorgensen, 1984).

Since the small sample distribution of $\hat{\boldsymbol{\theta}}$ is difficult to derive and the inference about $\boldsymbol{\theta}$ is needed, we now investigate the large sample behavior of $\hat{\boldsymbol{\theta}}$. Assuming $\lim_{n\to\infty} n_k/n = \delta_k$ and $0 < \delta_k < 1$ for all $k = 1, \ldots, K$, we conclude, by the strong law of large numbers, that for $\boldsymbol{\theta}$ fixed,

$$\begin{aligned} \frac{1}{n} l_n(\boldsymbol{\theta}|\boldsymbol{Y}) &\xrightarrow{a.s} l_0(\boldsymbol{\theta}) = \sum_{k=1}^{K} \delta_k E\left[l_1^{(k)}(\boldsymbol{\theta}|Y) \right] \\ &= \sum_{k=1}^{K} \delta_k \left\{ \mu_k(\boldsymbol{\theta}) \log\left(\frac{\mu_k(\boldsymbol{\theta})}{1-\mu_k(\boldsymbol{\theta})} \right) + \log(1-\mu_k(\boldsymbol{\theta})) \right\} \\ &= \sum_{k=1}^{K} \delta_k \left\{ \frac{\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k)\exp(\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k))}{1+\exp(\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k))} - \log\left(1+\exp(\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k))\right) \right\}, \end{aligned} \tag{3.2}$$

where $E\left[l_1^{(k)}(\boldsymbol{\theta}|Y) \right]$ is expected value of log-likelihood function for one observation at the k-th design point.

Before stating asymptotic results, we introduce some convenient notations. The column vector $\nabla l_1^{(k)}(\boldsymbol{\theta}_0|Y) = \left(\frac{\partial l_1^{(k)}(\boldsymbol{\theta}|Y)}{\partial \theta_i}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)$ denotes the gradient of the log-likelihood function at design point $\boldsymbol{x}_k$, $k = 1, \ldots, K$, and $\nabla^2 l_1^{(k)}(\boldsymbol{\theta}_0|Y) = \left(\frac{\partial^2 l_1^{(k)}(\boldsymbol{\theta}|Y)}{\partial \theta_i \partial \theta_j}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)$ denotes the Hessian of the log-likelihood function where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+1})' = (\lambda, \beta_1, \ldots, \beta_p)'$.

Theorem 3.1. *Suppose the parameter space Θ and the log-likelihood function $l_n(\boldsymbol{\theta}|\boldsymbol{Y})$, defined in (3.1), satisfy the following conditions;*

(i) the parameter space Θ is a compact subset of R^{p+1},

(ii) $l_0(\boldsymbol{\theta})$, defined in (3.2), has a unique global maximum at $\boldsymbol{\theta}_0 \in \Theta$.

Then,

(A) MLE, $\hat{\boldsymbol{\theta}}$, is a strongly consistent estimator of $\boldsymbol{\theta}_0$.

Furthermore, if

(iii) $\boldsymbol{\theta}_0$ is an interior point of Θ,

(iv) $\bar{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) = \sum_{k=1}^{K} \delta_k \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_0)$ is nonsingular, where $\boldsymbol{\Sigma}_k(\boldsymbol{\theta}_0) = -E[\nabla^2 l_1^{(k)}(\boldsymbol{\theta}_0|Y)]$ $E[\nabla l_1^{(k)}(\boldsymbol{\theta}_0|Y)(\nabla l_1^{(k)}(\boldsymbol{\theta}_0|Y))']$,

then

$$(B) \quad \sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{L}} N_{p+1}\left(0, \bar{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta}_0)\right).$$

The proof is given in Appendix.

Cho et al. (1997) investigate the large sample behavior of the MLE's for transformation parameter as well as regression and variance parameters when the Box-Cox transformation is employed in linear model. These allow the explanatory variables to take a countable distinct points. The additional conditions, although difficult to verify, suggest the kinds of smoothness that needs to be satisfied by the sequence of design matrices and underlying distributions in order for asymptotic normality to hold. Their approach can be also applied to the current families.

4. TESTS CONCERNING THE TRANSFORMATION PARAMETER

4.1. Score tests

By testing the hypothesis $\lambda = 1$, we can check whether the logistic model is appropriate for fitting the data or if a generalization is needed. We propose a score test. The score function for θ_i, $i = 1, \ldots, p+1$, is $S_{(i)}(\boldsymbol{\theta}) = \partial l_n(\boldsymbol{\theta}|\boldsymbol{Y})/\partial\theta_i$. Let $\boldsymbol{S}_2(\boldsymbol{\theta})$ denote the $p \times 1$ vector with i-th entry $S_{(i+1)}(\boldsymbol{\theta})$ and consider the partition of the scores $\boldsymbol{S}(\boldsymbol{\theta}) = (S_{(1)}(\boldsymbol{\theta}), \boldsymbol{S}_2(\boldsymbol{\theta})')'$. Correspondingly, we obtain the partitioned submatrices of the information matrix with

$$I_{11}(\boldsymbol{\theta}) = E[S_{(1)}^2(\boldsymbol{\theta})], \ \boldsymbol{I}_{12}(\boldsymbol{\theta}) = E[S_{(1)}(\boldsymbol{\theta})\boldsymbol{S}_2(\boldsymbol{\theta})'], \text{ and } \mathbf{I}_{22}(\boldsymbol{\theta}) = E[\boldsymbol{S}_2(\boldsymbol{\theta})\boldsymbol{S}_2(\boldsymbol{\theta})'].$$

Letting $\phi_{(i)}(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k) = \partial\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k)/\partial\theta_i$, we evaluate each entry of the submatrices with $E[S_{(i)}(\boldsymbol{\theta})S_{(j)}(\boldsymbol{\theta})] = \sum_{k=1}^{K} n_k \phi_{(i)}(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k)\phi_{(j)}(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k) \exp(\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k)) \ / \ (1+ \exp(\phi(\lambda, \boldsymbol{\beta}'\boldsymbol{x}_k)))^2$.

Let $\hat{\boldsymbol{\beta}}_{H_0}$ be the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ under $H_0 : \lambda = 1$ and define the MLE $\hat{\boldsymbol{\theta}}_{H_0} = (1, \hat{\boldsymbol{\beta}}'_{H_0})'$ of $\boldsymbol{\theta}_{H_0}$. Then, the score statistic for testing $H_0 : \lambda = 1$ with nuisance parameter $\boldsymbol{\beta}$ is

$$X^2 = \frac{S^2_{(1)}(\hat{\boldsymbol{\theta}}_{H_0})}{\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_{H_0})},$$

where $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_{H_0}) = I_{11}(\hat{\boldsymbol{\theta}}_{H_0}) - \boldsymbol{I}_{12}(\hat{\boldsymbol{\theta}}_{H_0})\mathbf{I}^{-1}_{22}(\hat{\boldsymbol{\theta}}_{H_0})\boldsymbol{I}'_{12}(\hat{\boldsymbol{\theta}}_{H_0})$. The asymptotic null distribution of X^2 is χ^2 with one degree of freedom and the test based on X^2 is the asymptotically locally most powerful unbiased test (see Bhat and Nagnur (1965)).

4.2. Bootstrap tests

Although χ^2 plays a role of the limiting distribution of the score statistic, the exact distribution for even moderately large samples may be quite different. Moreover, we have to estimate the variance $\Sigma(\boldsymbol{\theta}_{H_0})$. The bootstrap hypothesis test is a well-established technique to address this kind of problem. Our bootstrap test statistic for H_0 is also based on $S_{(1)}(\hat{\boldsymbol{\theta}}_{H_0})$.

Let $K(\boldsymbol{\theta}, s) = P(S_{(1)}(\boldsymbol{\theta}) \leq s)$. Then, using the continuity of $K(\hat{\boldsymbol{\theta}}_{H_0}, s)$, we obtain the bootstrap estimators of the lower critical value, C_l, and the upper, C_u, at the level α. Here, $\hat{C}_l = K^{-1}(\hat{\boldsymbol{\theta}}_{H_0}, \alpha/2)$ and $\hat{C}_u = K^{-1}(\hat{\boldsymbol{\theta}}_{H_0}, 1 - \alpha/2)$. $\hat{C}_l$ and $\hat{C}_u$ can be approximated by performing the bootstrap sampling.

Let $F_0^{(k)}(\hat{\boldsymbol{\beta}}_{H_0})$ be the estimated distribution at k-th design point under H_0. This is the Bernoulli distribution with probability $\hat{\mu}^{(k)}_{H_0} = \exp(\hat{\boldsymbol{\beta}}'_{H_0}\boldsymbol{x}_k)/(1 + \exp(\hat{\boldsymbol{\beta}}'_{H_0}\boldsymbol{x}_k))$. We generate B independent bootstrap samples $\boldsymbol{Y}^*_b = \left(\boldsymbol{Y}^{(1)*}_b, \ldots, \boldsymbol{Y}^{(K)*}_b\right)$, $b = 1, \ldots, B$, where $\boldsymbol{Y}^{(k)*}_b = \left(Y^{(k)*}_{1b}, \ldots, Y^{(k)*}_{n_k b}\right)$ are randomly generated from $F_0^{(k)}(\hat{\boldsymbol{\beta}}_{H_0})$. To carry out the simple generation, we may sample $r^{(k)*}_b$'s from binomial distributions with $(n_k, \hat{\mu}^{(k)}_{H_0})$ instead of the $\boldsymbol{Y}^{(k)*}_b$'s. Then, the bootstrap analog of $S_{(1)}(\hat{\boldsymbol{\theta}}_{H_0})$ is $S_{(1)}(\hat{\boldsymbol{\theta}}^*_{H_0 b})$ with $\hat{\boldsymbol{\theta}}^*_{H_0 b}$ being the MLE of $\hat{\boldsymbol{\theta}}_{H_0}$ calculated based on $\boldsymbol{Y}^*_b$.

At the α% level of significance, $\hat{C}_l$ and $\hat{C}_u$ are approximated by the $100\alpha/2$-th percentile, $\hat{C}^*_l$, of $S_{(1)}(\hat{\boldsymbol{\theta}}^*_{H_0 b})$ and the $100(1 - \alpha/2)$-th percentile, $\hat{C}^*_u$, respectively. The null hypothesis H_0 is rejected if and only if the rejection regions, $(-\infty, \hat{C}^*_l]$ or $[\hat{C}^*_u, \infty)$, include $S_{(1)}(\hat{\boldsymbol{\theta}}_{H_0})$. If the conditions of Theorem 3.1 are satisfied, then $\hat{\boldsymbol{\beta}}_{H_0}$ is consistent under H_0 and this bootstrap test is asymptotically correct (Beran (1986)).

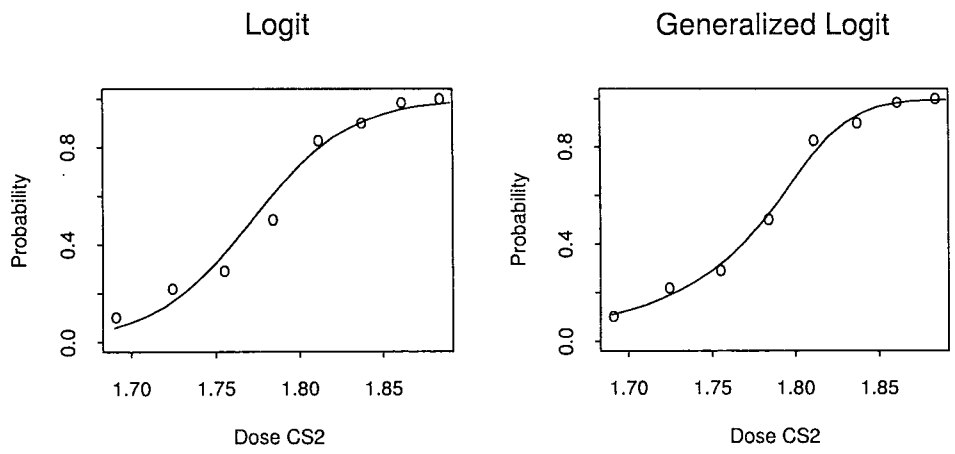

Figure 5.1: Observed Data and Fitted Probability Curves for Beetle Data.

5. EXAMPLES

5.1. Mortality of adult beetle after exposure to CS_2

Prentice (1976), Aranda-Ordaz (1981) and Stukel (1988) who proposed a generalized logistic model analyzed the data of adult beetle mortality after exposure to gaseous carbon disulphide (CS_2) reported by Bliss (1935). The data are given in the first three columns of Table 5.1. According to their studies, the likelihood-ratio χ^2_6 statistic for the logistic model is 11.23 ($p = 0.081$) which gives some evidence of lack of fit. An examination of residuals shows that ϕ_2 tends to make more improvement than ϕ_1.

When the ϕ_2 family is applied, the score function $S_{(1)}(\hat{\boldsymbol{\theta}}_{H_0})$ with nuisances parameters β_0 and β_1 is calculated to be 17.76 The bootstrap critical values based on 1000 samples correspond to -12.35 and 12.31 at the 5% level of significance. The score statistic X^2 for testing $\lambda = 1$ is 8.04 ($p = 0.004$). Neither the score test nor the bootstrap test retains the null hypothesis $\lambda = 1$. Thus, the logit link of linear measurement of CS_2 dosage is not sensible for the present situation.

The parameter estimates for ϕ_2 and the corresponding estimated covariance matrix are

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 1.480 \\ -61.226 \\ 34.366 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0.003 & & \\ 0.006 & 35.224 & \\ -0.004 & -19.570 & 10.878 \end{bmatrix},$$

respectively. Figure 5.1 and Table 5.1 show that the generalized logistic regression with ϕ_2 provides considerable improvement. The likelihood-ratio χ^2_1 test for comparing the ϕ_2 with the logistic model is 8.32 ($p = 0.004$). From an analysis of residuals, we see that the ϕ_2 family with $\lambda = 1.48$ is preferable to the complementary log-log transformation which, according to Aranda-Ordaz (1981), is the most appropriate for these data.

Table 5.1: Adult Beetle Mortality after Exposure to Carbon Disulphide.

Dosage	No. of beetles		Estimated Mortality $n\hat{\mu}$			
CS_2	Exposed	Killed	Logit	C log-log	ϕ_1 family	ϕ_2 family
1.6907	59	6	3.46	5.59	3.28	6.47
1.7242	60	13	9.84	11.28	11.59	11.26
1.7552	62	18	22.45	20.95	24.97	19.66
1.7842	56	28	33.90	30.37	33.15	29.28
1.8113	63	52	50.10	47.78	49.04	48.85
1.8369	59	53	53.29	54.14	53.50	54.89
1.8610	62	61	59.22	61.11	59.95	60.95
1.8839	60	60	58.74	59.95	59.37	59.78

5.2 Age of menarche in Warsaw girls

Milicer and Szczotka (1966) analyzed data determining the age of menarche of a sample of 3918 Warsaw girls in 1965 which are reported in the first three columns of Table 5.2. The likelihood ratio χ^2_{23} test for logistic model is 26.7 ($p = 0.269$) which indicates a non-goodness of fit for logistic model. A glance of residuals suggests that the use of ϕ_2 is also preferable to that of ϕ_1 for these data.

The score function $S_{(1)}(\hat{\boldsymbol{\theta}}_{H_0})$ for ϕ_1 is -39.59 and the corresponding bootstrap critical values based on 1000 samples are -35.16 and 33.81. The score statistic X^2 is 4.28 ($p = 0.039$). These results indicate that an improvement to fit is possible when ϕ_2 is applied. The parameter estimates for ϕ_2 and the corresponding estimated covariance matrix are

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0.881 \\ -21.428 \\ 1.656 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0.003 & & \\ 0.007 & 0.617 & \\ -0.001 & -0.048 & 0.004 \end{bmatrix},$$

Table 5.2: Age of Menarche in 3918 Warsaw Girls.

	No. of		Estimated No. having menstruated			
Age	girls	observed	Logit	Probit(BC)	ϕ_1 family	ϕ_2 family
9.21	376	0	0.76	0.00	0.26	0.28
10.21	200	0	2.06	0.24	1.29	1.15
10.58	93	0	1.74	0.50	1.31	1.12
10.83	120	2	3.34	1.51	2.79	2.36
11.08	90	2	3.72	2.37	3.39	2.86
11.33	88	5	5.36	4.38	5.25	4.46
11.58	105	10	9.33	9.01	9.65	8.34
11.83	111	17	14.19	15.15	15.20	13.50
12.08	100	16	18.06	20.20	19.63	18.05
12.33	93	29	23.15	26.10	25.02	23.97
12.58	100	39	33.26	36.86	35.16	35.12
12.83	108	51	46.27	49.80	47.26	49.06
13.08	99	47	52.46	54.73	51.54	55.16
13.33	106	67	66.67	67.76	64.20	69.07
13.58	105	81	75.42	75.28	72.54	76.99
13.83	117	88	92.79	91.69	89.99	93.61
14.08	98	79	83.51	82.20	81.87	83.55
14.33	97	90	86.97	85.62	86.08	86.57
14.58	120	113	111.45	109.99	111.09	110.67
14.83	102	95	97.05	96.07	97.16	96.29
15.08	122	117	118.00	117.16	118.39	117.10
15.53	111	107	108.55	108.05	109.00	107.80
15.58	94	92	92.61	92.35	92.98	92.06
15.83	114	112	112.87	112.70	113.27	112.31
17.58	1049	1049	1048.40	1048.59	1048.88	1047.40

respectively. Large sample 95% confidence interval for the transformation parameter is $(0.774, 0.988)$ which does not cover the value $\lambda = 1$. The likelihood ratio χ_1^2 for comparing the ϕ_1 family with the logistic model is 4.75 ($p = 0.029$). Table 5.2 shows that both tails are improved when the ϕ_2 family is employed. In fact, Guerrero and Johnson (1982) obtained a remarkable improvement ($\chi_1^2 = 12.2$ ($p = 0.0005$)) using a probit model where the Box-Cox transformation is taken to the explanatory variable, age, and a normal distribution is assumed for transformed variable, see the fifth column of Table 5.2.

The authors gratefully acknowledge the constructive criticism of an anonymous referee.

APPENDIX

Lemma A.1. *Let $\{T_n(\cdot)\}$ be a sequence of random functions defined on a probability space and depend on $\boldsymbol{\theta}$ in a compact set Θ. Suppose that*

(i) there exists a continuous function $T(\boldsymbol{\theta})$ defined on Θ such that $T_n(\boldsymbol{\theta}) \xrightarrow{a.s} T(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta} \in \Theta$,

(ii) $T(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0 \in \Theta$.

Then $\hat{\boldsymbol{\theta}} = \arg\max T_n(\boldsymbol{\theta})$ is a strongly consistent estimator of $\boldsymbol{\theta}_0$.

Since Lemma A.1 is a standard result (*cf.* Yeo and Johnson (1997)), we omit the proof.

Proof of Theorem 3.1 (A) Since $l_1^{(k)}(\boldsymbol{\theta}|y)$, $k = 1, \ldots, K$, are dominated by an integrable function and are equicontinuous in $\boldsymbol{\theta}$ for $y \in \{0,1\}$, where $P(Y \in \{0,1\}) = 1$, the application of Rubin's theorem (1956) allows to conclude that $l_{n_k}^{(k)}(\boldsymbol{\theta}_0|\boldsymbol{Y})/n_k \xrightarrow{a.s} E\left[l_1^{(k)}(\boldsymbol{\theta}|Y)\right]$ uniformly in $\boldsymbol{\theta}$ and the limit function is continuous, where $l_{n_k}^{(k)}(\boldsymbol{\theta}|\boldsymbol{Y})$ are log-likelihood function of n_k observations at the k-th design point. Consequently, $l_n(\boldsymbol{\theta}|\boldsymbol{Y})/n \xrightarrow{a.s} l_0(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}$ and $l_0(\boldsymbol{\theta})$ is continuous. By assumption (iii) of Theorem 3.1, $\boldsymbol{\theta}_0$ maximizes $l_0(\boldsymbol{\theta})$ and is unique so $\hat{\boldsymbol{\theta}} \xrightarrow{a.s} \boldsymbol{\theta}_0$ according to Lemma A.1.

(B) Expanding $n^{-1/2}\nabla l_n(\hat{\boldsymbol{\theta}}|\boldsymbol{Y})$ about $\boldsymbol{\theta}_0$, we obtain that

$$n^{-1/2}\nabla l_n(\hat{\boldsymbol{\theta}}|\boldsymbol{Y}) = n^{-1/2}\nabla l_n(\boldsymbol{\theta}_0|\boldsymbol{Y}) + n^{-1}\nabla^2 l_n(\tilde{\boldsymbol{\theta}}|\boldsymbol{Y})\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \tag{A.1}$$

where $\tilde{\boldsymbol{\theta}} = c_n\hat{\boldsymbol{\theta}} + (1-c_n)\boldsymbol{\theta}_0$, $c_n \in (0,1)$. By assumption (iv), $\boldsymbol{\theta}_0$ is an interior point of Θ and $\hat{\boldsymbol{\theta}}$ is a strongly consistent estimator of $\boldsymbol{\theta}_0$ according to (A) so the left hand side of (A.1) converges to 0 in probability, since $n^{-1/2}\nabla l_n(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = 0$ at the maximum when $\hat{\boldsymbol{\theta}}$ lies in the interior of Θ. Let us decompose $n^{-1/2}\nabla l_n(\boldsymbol{\theta}_0|\boldsymbol{Y}) = \sum_{k=1}^K n_k^{-1/2}\nabla l_{n_k}^{(k)}(\boldsymbol{\theta}_0|\boldsymbol{Y})(n_k/n)^{1/2}$. Then, by the central limit theorem, $n_k^{-1/2}\nabla l_{n_k}^{(k)}(\boldsymbol{\theta}_0|\boldsymbol{Y})$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_k(\boldsymbol{\theta}_0)$. Independence allows us to conclude that $n^{-1/2}\nabla l_n(\boldsymbol{\theta}_0|\boldsymbol{Y}) \xrightarrow{\mathcal{L}} N_{p+1}\left(\mathbf{0}, \bar{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)\right)$. Furthermore, another application of Rubin (1956) and the strong consistency of $\hat{\boldsymbol{\theta}}$ ensure that $n^{-1}\nabla^2 l_n(\tilde{\boldsymbol{\theta}}|\boldsymbol{Y}) = \sum_{k=1}^K n_k^{-1}\nabla^2 l_{n_k}^{(k)}(\tilde{\boldsymbol{\theta}}|\boldsymbol{Y})(n_k/n) \xrightarrow{a.s} \sum_{k=1}^K \delta_k \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_0) = \bar{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$. By Slusky's theorem, we conclude that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0). \xrightarrow{\mathcal{L}} N_{p+1}\left(\mathbf{0}, \bar{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta}_0)\right).$$

REFERENCES

Aranda-Ordaz, F. J. (1981). "On Two Families of Transformations to Additivity for Binary Response Data," *Biometrika* **68**, 357-363.

Beran, R. (1986). "Simulated Power Functions," *Annals of Statistics* **14**, 151-173.

Bhat, B. R. and Nagnur, B. N. (1965). "Locally Asymptotically Most Stringent Tests and Lagrangian Multiplier Tests of Linear Hypotheses," *Biometrika* **52**, 459-468.

Bliss, C. J. (1935). "The Calculation of the Dosage-Mortality Curve," *Annals of Applied Biology* **22**, 134-167.

Cho, K., Yeo, I. K., Johnson, R. A. and Loh, W. Y. (1997). "Asymptotic Theory for Box-Cox Transformation in Linear Models," *Submitted for publication*

Guerrero, V. M. and Johnson, R. A. (1982). "Use of the Box-Cox Transformation with Binary Response Models," *Biometrika* **69**, 309-314.

John, J. A. and Draper, N. R. (1980). "An Alternative Family of Transformations," *Applied Statistics* **29**, 190-197.

Jorgensen, B. (1984). "The Delta Algorithm and GLIM," *International Statistical Review* **52**, 283-300.

Milicer, H. and Szczotka, F. (1966). "Age at Menarche in Warsaw Girls in 1965," *Human Biology* **38**, 199-203.

Pregibon, D. (1980). "Goodness of Link Tests for Generalized Linear Models," *Applied Statistics* **29**, 15-24.

Prentice, R. L. (1976). "Generalization of the Probit and Logit Methods for Dose Response Curves," *Biometrics* **32**, 761-768.

Rubin, H. (1956). "Uniform Convergence of Random Functions with Applications to Statistics," *Annals of Mathematical Statistics* **27**, 200-203.

Stukel, T.A. (1988). "Generalized Logistic Models," *Journal of the American Statistical Association* **83**, 426-431.

Yeo, I. K. and Johnson, R. A. (1997). "A New Family of Power Transformations to Improve Normality or Symmetry," *Submitted for publication*