A NEW FAMILY OF PARAMETRIC LINKS FOR BINOMIAL GENERALIZED LINEAR MODELS

Nobuhiro Taneichi*, Yuri Sekiya** and Jun Toyama***

In a generalized linear model with binary response, the role of a link function is important to find a model that fits data well. Aranda-Ordaz (1981) proposed a family of link functions that includes a logistic link function and a complementary log-log function. In this paper, we propose a new family of models on the basis of a family of link functions by extending the family proposed by Aranda-Ordaz (1981). We also consider tests to determine whether the new model fits data well. Examples of artificial and real data showing that our new model is more appropriate than the Aranda-Ordaz model are presented.

Key words and phrases: Binary response, cumulative distribution function, generalized linear model, link function.

1. Introduction

Generalized linear models (Nelder and Wedderburn (1972), McCullagh and Nelder (1989)) are composed of random components, linear predictors and link functions. The role of a link function is important when searching for a model that fits data well. In this paper, we consider construction of link functions of generalized linear models with random components that have a binomial distribution $B(n, p)$. In this case, the most popular link function is a logit function:

$$\text{logit}(p) = \log \left( \frac{p}{1 - p} \right), \quad (0 < p < 1),$$

which is a canonical link function. A logistic regression model is defined by

$$\text{logit}(p) = \sum_{m=1}^{k} \beta_m x_m,$$

where $\beta_m, (m = 1, \ldots, k)$ are parameters, $x_m, (m = 1, \ldots, k)$ are covariates and $x_1 = 1$. On the other hand, if we use the inverse of the cumulative distribution function (CDF) $\Phi(\cdot)$ of a standard normal distribution as a link function, we obtain the probit model

$$\Phi^{-1}(p) = \sum_{m=1}^{k} \beta_m x_m.$$
Let $L^*$ be a link function. The links for the logistic regression model and probit model are symmetric about $p = 0.5$ in the sense that

$$L^*(p) = -L^*(1 - p).$$

Therefore, logit and probit models are not appropriate when observations do not fit this symmetric characteristic. The complementary log-log model

$$\log\{-\log(1 - p)\} = \sum_{m=1}^{k} \beta_m x_m$$

is effective for such observations. Two approaches have been proposed to ensure a variety of link functions. Prentice (1976), Pregibon (1980), Aranda-Ordaz (1981), Kaiser (1997), and Czado and Munk (2000) considered parametric approaches, and Weisberg and Welsh (1994) and Muggeo and Ferrara (2008) considered non-parametric approaches.

As a parametric approach, Aranda-Ordaz (1981) considered an asymmetric family of link functions that depend on parameter $c$ for a binary response as follows:

$$L^A_c(p) = \log\left\{(1 - p)^{-c} - 1\right\}/c.$$

$L^A_1$ is the link for the logistic regression model and $L^A_0$ is the link for the complementary log-log model as a limit. This family of links can thus treat the two links inclusively.

In this paper, we propose a new parametric family of link functions by extending the family of link functions (1.1). The proposed link function is given by using an additional power parameter on the cumulative distribution function corresponding to asymmetric link function (1.1). A link function by using an additional power parameter was proposed by Samejima (2000) and Jiang et al. (2013). The link function proposed by Samejima (2000) is derived by using an additional power parameter on the cumulative distribution function corresponding to a logit link function. Since the asymmetric family of link functions (1.1) includes a logit link function as a special case, the family of link functions that is proposed in this paper is a generalization of the link function proposed by Samejima. The link function proposed by Jiang et al. (2013) is derived by using an additional power parameter on the c.d.f. corresponding to a symmetric baseline link function and its mirror reflection. This link function has flexibility to negative skewness as well as positive skewness. The difference between their link and our proposed link is that one uses a symmetric baseline link function and its mirror reflection, while the other uses an asymmetric link (1.1). In order to increase flexibility to skewness of a link function, two different approaches based on the same kind of methods are considered. By using the family of link functions (1.1), we construct generalized linear models for a binary response that can be fitted to various data according to their characteristics. In Section 2, a
new family of link functions is proposed. The necessity of the models based on the new family of link functions is considered. In Section 3, tests to determine whether models based on the family of link functions fit data are described. In Section 4, we show examples of artificial and real data sets for which our model is more appropriate than the model proposed by Aranda-Ordaz.

2. Construction of a new family of link functions

A link function is ordinarily defined by a cumulative distribution function. A link function defined as the inverse of a cumulative distribution function is called an inverse CDF link. Links for a logistic regression model, probit model and complementary log-log model are all inverse CDF links. From this point of view, the inverse function of the link (1.1), that is,

\[ F_c(x) = 1 - \left\{ 1 + c \exp(x) \right\}^{-1/c}, \]

satisfies the following conditions for a cumulative distribution function.

(i) \( F(-\infty) = 0, \quad F(\infty) = 1. \)
(ii) \( x < y \Rightarrow F(x) < F(y). \)
(iii) \( F(x) \) is right continuous.

Therefore, link function (1.1) is an inverse CDF link. \( F_1(x) \) is the inverse function of the link function of a logistic regression model and \( F_0(x) \) is the inverse function of the link function of a complementary log-log model as a limit.

In order to find a model that fits data well, it is important to prepare a family of cumulative distribution functions that have graphs of various shapes. The shape of a graph of a cumulative distribution function depends on the mean, variance, skewness and kurtosis. Skewness indicates symmetry of a distribution and kurtosis indicates sharpness of a density function. Let \( F \) be a cumulative distribution function of a distribution and let \( \beta_1 \) be an arbitrary real number and \( \beta_2 \) be an arbitrary positive real number. We consider values of the mean, variance, skewness and kurtosis of a distribution that has the cumulative distribution function \( F(\beta_1 + \beta_2 x) \). The mean varies depending on the values of \( \beta_1 \) and \( \beta_2 \), and variance varies depending on the value of \( \beta_2 \). However, skewness and kurtosis do not vary with variation in the values of \( \beta_1 \) and \( \beta_2 \). Therefore, if we consider a cumulative distribution function as the inverse function of a link function, we are interested in values of skewness and kurtosis. We investigated the values of skewness and kurtosis of a distribution for which the distribution function \( F_c \) is the inverse function of the link function \( L_A^c(p) \) of Aranda-Ordaz (1981). Horizontal and vertical axes in Fig. 1 denote skewness and kurtosis, respectively. The real curve (curve of \( \gamma = 1 \)) in Fig. 1 shows the locus of the skewness and kurtosis for \( F_c \) in the plane when \(-1.0 \leq c \leq 10.0\). (Details of Fig. 1 will be given later.) From Fig. 1, we find that the distribution defined by \( F_c \) has skewness and kurtosis that are restricted to a special functional relation.

In order to obtain a cumulative distribution function to fit various data, it is necessary to select a model from a family of distributions for which the existence area of skewness and kurtosis is more extensive than that in the family \( F_c \). With
regard to such a family of distributions, put \( G_{c,\gamma}(x) = \{F_c(x)\}^\gamma \), \( \gamma > 0 \) and consider \( G_{c,\gamma}(x) \). From the definition of \( G_{c,\gamma}(x) \), it is easily shown that \( G_{c,\gamma}(x) \) satisfies conditions (i)–(iii). Then, by using the inverse function of \( G_{c,\gamma}(x) \), we define the following new family of inverse CDF links \( L_{c,\gamma}(p) \).

\[
(2.1) \quad L_{c,\gamma}(p) = \log \left\{ \frac{(1 - p^{1/\gamma})^{-c} - 1}{c} \right\}, \quad (\gamma > 0),
\]

and \( L_{0,\gamma}(p) = \lim_{c \to 0} L_{c,\gamma}(p) = \log\left\{ -\log(1 - p^{1/\gamma}) \right\}, \quad (\gamma > 0) \). The family includes \( L_c^A(p) \) defined by (1.1) as a special case of \( \gamma = 1 \). Let skewness and kurtosis of the distribution be defined by \( G_{c,\gamma} \) be \( S(c, \gamma) \) and \( K(c, \gamma) \), respectively, then

\[
(2.2) \quad S(c, \gamma) = \frac{Q_3(c, \gamma) - 3\gamma Q_2(c, \gamma)Q_1(c, \gamma) + 2\gamma^2\{Q_1(c, \gamma)\}^3}{\gamma^{1/2}\{Q_2(c, \gamma) - \gamma\{Q_1(c, \gamma)\}^2\}^{3/2}}
\]

and

\[
(2.3) \quad K(c, \gamma) = \frac{Q_4(c, \gamma) - 4\gamma Q_3(c, \gamma)Q_1(c, \gamma) + 6\gamma^2 Q_2(c, \gamma)\{Q_1(c, \gamma)\}^2 - 3\gamma^3 \{Q_1(c, \gamma)\}^4}{\gamma\{Q_2(c, \gamma) - \gamma\{Q_1(c, \gamma)\}^2\}^2},
\]

where

\[
(2.4) \quad Q_i(c, \gamma) = \begin{cases} \int_0^1 (1 - z)^{\gamma - 1} \left\{ \log \left( \frac{z^{-c} - 1}{c} \right) \right\}^i dz, \quad (c \neq 0), \\ \int_0^1 (1 - z)^{\gamma - 1} \{ -\log z \}^i dz, \quad (c = 0), \end{cases}
\]

for \( i = 1, 2, 3, 4 \). The proof of (2.2) and (2.3) is given in the Appendix. From (2.2) and (2.3), it can be shown that \( S(-1, \gamma) = -2 \) and \( K(-1, \gamma) = 9 \) hold for any positive \( \gamma \). The fact is displayed in Fig. 1.

We consider the points \( (S(c, \gamma), K(c, \gamma)) \). Figure 1 shows the locus of points \( (S(c, \gamma), K(c, \gamma)) \), \((-1.0 \leq c \leq 10.0)\) for \( \gamma = 0.3, 0.5, 1.0, 3.0 \) and \( 20.0 \). From Fig. 1, we find the following. For fixed \( \gamma = 1 \), when we change the value of \( c \), point \( (S(c, 1), K(c, 1)) \) moves on the curve similar to a quadratic function. The curve of \( \gamma = 1 \) therefore does not contain the points of \( K \leq 4 \). As the value of \( \gamma_0 \) becomes small, the graph of the curve of \( \gamma = \gamma_0 \) moves in the positive direction of \( K \). Conversely, as the value of \( \gamma_0 \) becomes large, the graph of the curve of \( \gamma = \gamma_0 \) moves in the negative direction of \( K \). Then, the curve of \( \gamma = 1 \) does not contain the point \( (S, K) = (0, 3) \), which corresponds to a normal distribution, that is, a probit link. However, by choosing an appropriate value \( \gamma_0 \), the point \( (S, K) = (0, 3) \) can be included in the curve of \( \gamma = \gamma_0 \). In this manner, from Fig. 1, we find that there exists a plane domain of skewness and kurtosis that cannot be obtained by a link when \( \gamma = 1 \), that is, the link of Aranda-Ordaz. As seen above, the distribution given by CDF which is defined for a family of links of Aranda-Ordaz is restricted to that for which the skewness and kurtosis have a special functional relation. However, by extension to a new family of inverse CDF links, the existence area of skewness and kurtosis of the distribution given
by the CDF has been greatly extended. This means that even for a data set that does not fit the Aranda-Ordaz model, it is possible to find a model appropriate for the data set in the proposed family of models.

3. Tests of models

In this section, we consider tests to determine whether the models based on the family of link functions fit data. We also consider a test to determine whether a specific model fits data. We consider a family of models based on link functions $L_{c,\gamma}$, $(\gamma > 0)$ that is defined by (2.1). We call the family of models the $L_{c,\gamma}$ model. We also consider a family of models based on link functions $L_{c}^A$, given by Aranda-Ordaz, which is defined by (1.1). We call the family of models the $L_{c}^A$ model. We call the model based on the link function $L_{c,\gamma}$ when $c = c_0$ and $\gamma = \gamma_0$ the $L_{c_0,\gamma_0}$ model.

Let random variable $Y_j$, $(j = 1, \ldots, N)$, which corresponds to numbers of responses in $N$ different subgroups, be independently distributed according to a binomial distribution $B(n_j, p_j)$, $(j = 1, \ldots, N)$. If we consider $L_{c,\gamma}$ given by (2.1) as a link function, then

$$p_j = p_j(\beta, c, \gamma), \quad (j = 1, \ldots, N),$$
where

\[
p_j(\beta, c, \gamma) = \begin{cases} 
1 - \left\{1 + c \exp \left(\sum_{m=1}^{k} \beta_m x_{jm}\right)\right\}^{-1/c} \gamma, & (c \neq 0), \\
1 - \exp \left\{- \exp \left(\sum_{m=1}^{k} \beta_m x_{jm}\right)\right\} \gamma, & (c = 0),
\end{cases}
\]

c, \gamma and \( \beta = (\beta_1, \ldots, \beta_k)' \) are parameters. The log likelihood function is

\[
l(c, \gamma, \beta' | y) = \sum_{j=1}^{N} y_j \log p_j(\beta, c, \gamma) + \sum_{j=1}^{N} (n_j - y_j) \log \left\{1 - p_j(\beta, c, \gamma)\right\} \\
+ \sum_{j=1}^{N} \log \left\{\frac{n_j!}{y_j!(n_j - y_j)!}\right\},
\]

where \( y = (y_1, \ldots, y_N)' \) is the observed value of \( Y = (Y_1, \ldots, Y_N)' \).

We consider the following three tests of hypotheses.

[I] Test of the \( L_{c,\gamma} \) model.

Let the null hypothesis be

\[
H_{0}^{(1)}: p_j = p_j(\beta, c, \gamma), \quad (j = 1, \ldots, N),
\]

where \( c, \gamma(>0) \) and \( \beta = (\beta_1, \ldots, \beta_k)' \) are parameters, \( x_{jm}, (j = 1, \ldots, N; m = 1, \ldots, k) \) are covariates and \( k + 2 < N \). The log likelihood ratio statistic for testing the null hypothesis \( H_{0}^{(1)} \) is given by

\[
T_{1} = 2 \sum_{j=1}^{N} Y_j \log \left(\frac{Y_j}{n_j \hat{p}_j}\right) + 2 \sum_{j=1}^{N} (n_j - Y_j) \log \left(\frac{1 - Y_j}{n_j \left(1 - \hat{p}_j\right)}\right),
\]

where \( \hat{p}_j = p_j(\hat{\beta}, \hat{c}, \hat{\gamma}), \quad (j = 1, \ldots, N), \)

and \( \hat{c}, \hat{\gamma} \) and \( \hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_k) \) are maximum likelihood estimators of parameters under \( H_{0}^{(1)} \), that is, values of \( c, \gamma \) and \( \beta_m, (m = 1, \ldots, k) \) maximize \( l(c, \gamma, \beta' | Y) \).

We consider the following condition:

\[(3.1) \quad \frac{n_j}{n} \to \mu_j, \quad (j = 1, \ldots, N) \quad \text{as} \quad n \to \infty,
\]

where \( 0 < \mu_j < 1, (j = 1, \ldots, N), \sum_{j=1}^{N} \mu_j = 1 \) and \( n = \sum_{j=1}^{N} n_j \). Under condition (3.1) and null hypothesis \( H_{0}^{(1)} \), \( T_{1} \) is asymptotically distributed as a chi-square distribution with \( N - k - 2 \) degrees of freedom. Therefore, for large \( n \), if the observed value of the statistic \( T_{1} \) is greater than the upper \( \alpha \)-point
\( \chi^2_{N-k-2}(\alpha) \) of a chi-square distribution with \( N - k - 2 \) degrees of freedom, we reject the null hypothesis \( H_0^{(1)} \) at a significance level of \( \alpha \).

**[II] Test of the \( L^A_c \) model.**

Let the null hypothesis be \( H_0^{(1)} \) and \( \gamma = 1 \), that is,

\[
H_0^{(2)}: \ p_j = p_j(\beta, c, 1), \quad (j = 1, \ldots, N),
\]

where \( c \) and \( \beta = (\beta_1, \ldots, \beta_k)' \) are parameters, \( x_{jm}, (j = 1, \ldots, N; m = 1, \ldots, k) \) are covariates and \( k + 1 < N \). The log likelihood ratio statistic for testing the null hypothesis \( H_0^{(2)} \) is given by

\[
T_2 = 2 \sum_{j=1}^{N} Y_j \log (Y_j n_j \hat{p}_j) + 2 \sum_{j=1}^{N} (n_j - Y_j) \log \left( \frac{1 - Y_j}{1 - \hat{p}_j} \right),
\]

where

\[
\hat{p}_j = p_j(\tilde{\beta}, \hat{c}, 1), \quad (j = 1, \ldots, N),
\]

\( \hat{c} \) and \( \tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_k)' \) are maximum likelihood estimators of parameters under \( H_0^{(2)} \), that is, values of \( c \) and \( \beta_m, (m = 1, \ldots, k) \) maximize \( l(c, 1, \beta' | Y) \).

Under condition (3.1) and null hypothesis \( H_0^{(2)} \), \( T_2 \) is asymptotically distributed as a chi-square distribution with \( N - k - 1 \) degrees of freedom. Therefore, for large \( n \), if the observed value of the statistic \( T_2 \) is greater than the upper \( \alpha \)-point \( \chi^2_{N-k-1}(\alpha) \) of a chi-square distribution with \( N - k - 1 \) degrees of freedom, we reject the null hypothesis \( H_0^{(2)} \) at a significance level of \( \alpha \).

**[III] Test of the \( L_{c_0,\gamma_{c_0}} \) model.**

Let \( c_0 \) be a real constant and \( \gamma_{c_0} \) be a real positive constant. We consider the null hypothesis

\[
H_0^{(3)}: \ p_j = p_j(\beta, c_0, \gamma_{c_0}), \quad (j = 1, \ldots, N),
\]

where \( \beta = (\beta_1, \ldots, \beta_k)' \) are parameters, \( x_{jm}, (j = 1, \ldots, N; m = 1, \ldots, k) \) are covariates and \( k < N \). The log likelihood ratio statistic for testing the null hypothesis \( H_0^{(3)} \) is given by

\[
T_3 = 2 \sum_{j=1}^{N} Y_j \log \left( \frac{Y_j}{n_j \hat{p}_j^*} \right) + 2 \sum_{j=1}^{N} (n_j - Y_j) \log \left( \frac{1 - Y_j}{1 - \hat{p}_j^*} \right),
\]

where

\[
\hat{p}_j^* = p_j(\beta^*, c_0, \gamma_{c_0}), \quad (j = 1, \ldots, N),
\]
and $\beta^* = (\beta^*_1, \ldots, \beta^*_k)'$ are maximum likelihood estimators of parameters under $H_0^{(3)}$, that is, values of $\beta_m, (m = 1, \ldots, k)$ maximize $l(c_0, \gamma_0, \beta^' | Y)$.

Under condition (3.1) and null hypothesis $H_0^{(3)}$, $T_3$ is asymptotically distributed as a chi-square distribution with $N - k$ degrees of freedom. Therefore, for large $n$, if the observed value of the statistic $T_3$ is greater than the upper $\alpha$-point $\chi^2_{N-k}(\alpha)$ of a chi-square distribution with $N - k$ degrees of freedom, we reject the null hypothesis $H_0^{(3)}$ at a significance level of $\alpha$. When $c_0 = 1$ and $\gamma_0 = 1$, the test of $H_0^{(3)}$ implies a test of the logistic regression model.

For the asymptotic theory of [I] to [III], see section 30.3 of Cramér (1946), Neyman (1949), and section 4.6 of Serfling (1980). With respect to test [III], Taneichi et al. (2014) considered a test for which the link function and test statistic are more general. On the basis of asymptotic expansion, they derived an approximation of the distribution of the test statistic under $H_0^{(3)}$.

We consider recommendation of a model corresponding to results of tests. First, we consider the case in which the null hypothesis $H_0^{(2)}$ is accepted. In this case, if we put $c_0 = \hat{c}$ and $\gamma_0 = 1$, then values of $\beta^*_m, (m = 1, \ldots, k)$ are equal to values of $\hat{\beta}_m, (m = 1, \ldots, k)$, respectively. Then, if we let the observed value of test statistic $T_2$ be $t_2$, the observed value of test statistic $T_3$ is equal to $t_2$. On the other hand, $t_2 \leq \chi^2_{N-k-1}(\alpha)$ implies $t_2 \leq \chi^2_{N-k}(\alpha)$. Therefore, if null hypothesis $H_0^{(2)}$ is accepted at a significance level of $\alpha$, null hypothesis $H_0^{(3)}$ in the case of $c_0 = \hat{c}$ and $\gamma_0 = 1$ is also accepted at a significance level of $\alpha$. So, in the case in which $H_0^{(2)}$ is accepted, we recommend the use of Aranda-Ordaz models $L^A$ with $c = \hat{c}$. Second, we consider the case in which the null hypothesis $H_0^{(2)}$ is rejected and null hypothesis $H_0^{(1)}$ is accepted. In this case, we need a model that is different from the family of Aranda-Ordaz models. If we put $c_0 = \hat{c}$ and $\gamma_0 = \hat{\gamma}$, then values of $\beta^*_m, (m = 1, \ldots, k)$ are equal to values of $\hat{\beta}_m, (m = 1, \ldots, k)$, respectively. Then, if we let the observed value of test statistic $T_1$ be $t_1$, the observed value of test statistic $T_3$ is equal to $t_1$. On the other hand, $t_1 \leq \chi^2_{N-k-2}(\alpha)$ implies $t_1 \leq \chi^2_{N-k}(\alpha)$. Therefore, if $H_0^{(1)}$ is accepted at a significance level of $\alpha$, null hypothesis $H_0^{(3)}$ in the case of $c_0 = \hat{c}$ and $\gamma_0 = \hat{\gamma}$ is also accepted at a significance level of $\alpha$. So, in the case in which null hypothesis $H_0^{(2)}$ is rejected and null hypothesis $H_0^{(1)}$ is accepted, we recommend the use of model $L_{c,\gamma}$ with $c = \hat{c}$ and $\gamma = \hat{\gamma}$. Third, we consider the case in which null hypothesis $H_0^{(2)}$ is rejected and $H_0^{(1)}$ is rejected. In this case, it is concluded that there is no model that represents the data well in the proposed family of models.

4. Application to data

In this section, we show examples of data for which our proposed model is more appropriate than the Aranda-Ordaz model. For the Aranda-Ordaz model, we experimentally found that it is difficult for the model to correspond to data that have the characteristic of relative frequency $y_{ij}/n_{ij}, (j = 1, \ldots, N)$ rapidly approaching 1 as covariate pattern $x_{ij2}, (j = 1, \ldots, N)$ changes. We produce two
artificial data sets that have these characteristics.

Table 1 shows an artificial data set of \( x_j, x_j, n_j, y_j, (j = 1, \ldots, N) \), when \( k = 2, x_j = 1, (j = 1, \ldots, N) \) and \( N = 10 \). First, we test whether the Aranda-Ordaz model fits the data. In order to apply test [II] to the data set, we obtain \( \hat{c} = 3.311, \hat{\beta}_1 = -5.823, \) and \( \hat{\beta}_2 = 1.952 \) as maximum likelihood estimates of \( c, \beta_1 \) and \( \beta_2 \) under \( H_0^{(2)} \), respectively. By using them, the observed value of test statistic \( T_2 \) becomes 20.827. Since \( \chi^2_7(0.05) = 14.067 \), null hypothesis \( H_0^{(2)} \) is rejected at a significance level of 0.05. (By the way, since \( \chi^2_7(0.01) = 18.475 \), null hypothesis \( H_0^{(2)} \) is even rejected at a significance level of 0.01.) Next, we test whether the proposed model fits the data. In order to apply test [I] to the data set, we obtain \( \hat{c} = 3.134, \hat{\gamma} = 0.118, \hat{\beta}_1 = -38.069, \) and \( \hat{\beta}_2 = 2.103 \) as maximum likelihood estimates of \( c, \gamma, \beta_1 \), and \( \beta_2 \) under \( H_0^{(1)} \), respectively. By using them, the observed value of test statistic \( T_1 \) becomes 4.053. Since \( \chi^2_6(0.05) = 12.592 \), null hypothesis \( H_0^{(1)} \) is accepted at a significance level of 0.05. That is, in this data set, null hypothesis \( H_0^{(2)} \) is rejected and null hypothesis \( H_0^{(1)} \) is accepted, and a model different from the model of Aranda-Ordaz is therefore needed to represent the data. Furthermore, it is clear that null hypothesis \( H_0^{(3)} \) with \( c_0 = 37.134 \) and \( \gamma_0 = 0.118 \) is accepted at a significance level of 0.05 since null hypothesis \( H_0^{(1)} \) is accepted at a significance level of 0.05. In Fig. 2, the points of relative frequency \( y_j/n_j, (j = 1, \ldots, 10) \) corresponding to covariate pattern \( x_j, (j = 1, \ldots, 10) \) calculated from the data in Table 1 are plotted. The figure shows a curve of the Aranda-Ordaz model when \( c = 3.311 \) and a curve of the proposed model when \( c = 37.134 \) and \( \gamma = 0.118 \). From the results of tests and results shown in Fig. 2, we recommend the use of the \( L_{37.134,0.118} \) model rather than the \( L_{3,311} \) model for the data set in Table 1.

Table 2 shows an artificial data set of \( x_j, x_j, n_j, y_j, (j = 1, \ldots, N) \), when \( k = 2, x_j = 1, (j = 1, \ldots, N) \), and \( N = 10 \). First, we test whether the Aranda-Ordaz model fits the data. In order to apply test [II] to the data set, we obtain \( \hat{c} = 2.750, \hat{\beta}_1 = 10.673 \) and \( \hat{\beta}_2 = -1.254 \) as maximum likelihood estimates of \( c, \beta_1 \) and \( \beta_2 \) under \( H_0^{(2)} \), respectively. By using them, the observed value of test
Figure 2. Artificial data 1, Aranda-Ordaz model with $c = 3.11$ and proposed model with $c = 37.134$ and $\gamma = 0.118$.

Table 2. Artificial data 2.

<table>
<thead>
<tr>
<th>$j$</th>
<th>$x_{j1}$</th>
<th>$x_{j2}$</th>
<th>$n_j$</th>
<th>$y_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>100</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6</td>
<td>100</td>
<td>88</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>7</td>
<td>100</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>8</td>
<td>100</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>9</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>100</td>
<td>16</td>
</tr>
</tbody>
</table>

Statistic $T_2$ becomes 18.824. Null hypothesis $H_0^{(2)}$ is rejected at a significance level of 0.05. (Null hypothesis $H_0^{(2)}$ is even rejected at a significance level of 0.01 in this data set also.) Next, we test whether the proposed model fits the data. In order to apply test [I] to the data set, we obtain $\hat{c} = 73.000$, $\hat{\gamma} = 0.066$, $\hat{\beta}_1 = 52.210$, and $\hat{\beta}_2 = -8.203$ as maximum likelihood estimates of $c$, $\gamma$, $\beta_1$, and
A NEW FAMILY OF PARAMETRIC LINKS 129

Figure 3. Artificial data 2, Aranda-Ordaz model with $c = 2.750$ and proposed model with $c = 73.000$ and $\gamma = 0.066$.

$\beta_2$ under $H_0^{(1)}$, respectively. By using them, the observed value of test statistic $T_1$ becomes 1.130. Null hypothesis $H_0^{(1)}$ is accepted at a significance level of 0.05. Therefore, we need a model that is different from the model of Aranda-Ordaz to represent the data. Null hypothesis $H_0^{(3)}$ with $c_0 = 73.000$ and $\gamma_0 = 0.066$ is accepted at a significance level of 0.05. In Fig. 3, the points of relative frequency $y_j/n_j$, $(j = 1, \ldots, 10)$ corresponding to covariate pattern $x_j$, $(j = 1, \ldots, 10)$ calculated from data in Table 2 are plotted. The figure shows a curve of the Aranda-Ordaz model when $c = 2.750$ and a curve of the proposed model when $c = 73.000$ and $\gamma = 0.066$. From the results of tests and results shown in Fig. 3, we recommend the use of the $L_{73.000,0.066}$ model rather than the $L_{2.750}^{A}$ model for the data set in Table 2.

Next, we consider a real data application. Table 3 shows a data set analyzed by Thompson and Funderlic (1981). Cox (1987) introduced it as a famous data set and called it BRSH2. The data set represents summarized results obtained for tumor incidence of mice injected with dibenzanthracene taken from Bryan and Shimkin (1943). In order to apply test [II] to the data set, we obtain $\hat{c} = 1.793$, $\hat{\beta}_1 = 8.115$ and $\hat{\beta}_2 = 4.518$ as maximum likelihood estimates of $c$, $\beta_1$ and $\beta_2$ under $H_0^{(2)}$, respectively. Then the observed value of test statistic $T_2$ becomes 12.284. Since $\chi^2_7(0.1) = 12.017$, null hypothesis $H_0^{(2)}$ is rejected at a significance
Table 3. Real data.

<table>
<thead>
<tr>
<th>Log-dose ($x_{j2}$)</th>
<th>Dibenzanthracene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mice injected ($n_j$)</td>
</tr>
<tr>
<td>−2.710</td>
<td>79</td>
</tr>
<tr>
<td>−2.108</td>
<td>40</td>
</tr>
<tr>
<td>−1.807</td>
<td>19</td>
</tr>
<tr>
<td>−1.509</td>
<td>21</td>
</tr>
<tr>
<td>−1.208</td>
<td>20</td>
</tr>
<tr>
<td>−0.903</td>
<td>23</td>
</tr>
<tr>
<td>−0.602</td>
<td>21</td>
</tr>
<tr>
<td>−0.301</td>
<td>21</td>
</tr>
<tr>
<td>0.000</td>
<td>22</td>
</tr>
<tr>
<td>0.301</td>
<td>19</td>
</tr>
</tbody>
</table>

Figure 4. Real data, Aranda-Ordaz model with $c = 1.793$ and proposed model with $c = 10.500$ and $\gamma = 0.177$.

level of 0.1. Next, we test whether the proposed model fits the data. In order to apply test [I] to the data set, we obtain $\hat{c} = 10.500$, $\hat{\gamma} = 0.177$, $\hat{\beta}_1 = 25.597$, and $\hat{\beta}_2 = 17.242$ as maximum likelihood estimates of $c$, $\gamma$, $\beta_1$, and $\beta_2$ under $H_0^{(1)}$, respectively. By using them, the observed value of test statistic $T_1$ becomes 8.408. Since $\chi^2_6(0.1) = 10.645$, null hypothesis $H_0^{(1)}$ is accepted at a significance level of
0.1. In the real data set, null hypothesis $H_0^{(2)}$ is rejected and null hypothesis $H_0^{(1)}$ is accepted, and a model different from the model of Aranda-Ordaz is therefore needed to represent the data. Furthermore, it is clear that null hypothesis $H_0^{(3)}$ with $c_0 = 10.500$ and $\gamma_0 = 0.177$ is accepted at a significance level of 0.1 since null hypothesis $H_0^{(1)}$ is accepted at a significance level of 0.1. In Fig. 4, the points of relative frequency $y_j/n_j$, $\left(\begin{array}{c} j \\ 1, \ldots, 10 \end{array}\right)$ corresponding to covariate pattern $x_{j1} = 1$, $\left(\begin{array}{c} j \\ 1, \ldots, 10 \end{array}\right)$ and $x_{j2}$, $\left(\begin{array}{c} j \\ 1, \ldots, 10 \end{array}\right)$ calculated from data in Table 3 are plotted. The figure shows a curve of the Aranda-Ordaz model when $c = 1.793$ and a curve of the proposed model when $c = 10.500$ and $\gamma = 0.177$. From the results of tests and results shown in Fig. 4, we recommend the use of the $L_{10.500,0.177}$ model rather than the $L_{1.793}$ model for the real data set in Table 3.

In this experiment, since the Hessian matrix sometimes becomes singular, it is not appropriate to use the Newton-Raphson method for estimating parameters. Therefore, maximum likelihood estimates of parameters are searched in the direction which is parallel to each axis.

5. Concluding remarks

We proposed a family of parametric link functions for generalized linear models with a binary response including link functions of Aranda-Ordaz. We considered a test procedure to determine whether we need a model that is constructed by a family of the proposed link functions for data or whether the Aranda-Ordaz model is sufficient for the data. We showed that our model is more appropriate than the Aranda-Ordaz model for some data sets.

Appendix

Proof of (2.2) and (2.3). Let $g_{c,\gamma}(x)$, ($\gamma > 0$) be the derivative of $G_{c,\gamma}(x)$ defined in Section 2. Then, when $c > 0$,

$$g_{c,\gamma}(x) = \gamma e^x (1 + ce^x)^{-1/c - 1} \left\{1 - (1 + ce^x)^{-1/c}\right\}^{\gamma - 1}, \quad (-\infty < x < \infty),$$

when $c < 0$,

$$g_{c,\gamma}(x) = \begin{cases} 
\gamma e^x (1 + ce^x)^{-1/c - 1} \left\{1 - (1 + ce^x)^{-1/c}\right\}^{\gamma - 1}, \\
-\infty < x < \log \left(\frac{-1}{c}\right), \\
0, \quad \log \left(\frac{-1}{c}\right) \leq x < \infty,
\end{cases}$$

and $g_{0,\gamma}(x) = \lim_{c \to +0} g_{c,\gamma}(x)$. Let $X$ be a random variable with a probability density function $g_{c,\gamma}(x)$. Let $M_X(t)$ be the moment generating function of $X$ and let $\kappa_X(t) = \log M_X(t)$ be the cumulant generating function of $X$. Then

$$M_X(t) = \gamma J_{c,\gamma}(t),$$
and
\[ \kappa_X(t) = \log \gamma \log J_{c, \gamma}(t), \]
where
\[ J_{c, \gamma}(t) = \begin{cases} 
\int_0^1 (1 - z)^{\gamma - 1} \left( \frac{z - c}{c} \right)^t dz, & (c \neq 0), \\
\int_0^1 (1 - z)^{\gamma - 1} \log z^t dz, & (c = 0). \end{cases} \]

Then \( J_{c, \gamma}(0) = \gamma^{-1} \) and \( J_{c, \gamma}^{(i)}(0) = Q_i(c, \gamma) \), where \( Q_i(c, \gamma) \) is given by (2.4). The \( i \)-th cumulant \( \kappa_i = \kappa_X^{(i)}(0), \) \( (i = 2, 3, 4) \) of \( X \) are represented as follows:

(A.1) \[ \kappa_2 = \gamma Q_2(c, \gamma) - \gamma^2 \{ Q_1(c, \gamma) \}^2, \]
(A.2) \[ \kappa_3 = \gamma Q_3(c, \gamma) - 3\gamma^2 Q_2(c, \gamma) Q_1(c, \gamma) + 2\gamma^3 \{ Q_1(c, \gamma) \}^3, \]
and

(A.3) \[ \kappa_4 = \gamma Q_4(c, \gamma) - 4\gamma^2 Q_3(c, \gamma) Q_1(c, \gamma) - 3\gamma^2 \{ Q_2(c, \gamma) \}^2 \\
+ 12\gamma^3 Q_2(c, \gamma) \{ Q_1(c, \gamma) \}^2 - 6\gamma^4 \{ Q_1(c, \gamma) \}^4. \]

Since \( S(c, \gamma) \) is the skewness of \( X \) and \( K(c, \gamma) \) is the kurtosis of \( X \), it holds that

(A.4) \[ S(c, \gamma) = \frac{\kappa_3}{\kappa_2^{3/2}}, \]
and

(A.5) \[ K(c, \gamma) = \frac{\kappa_4}{\kappa_2^2} + 3. \]

Therefore, substituting (A.1)–(A.3) for (A.4) and (A.5), we have (2.2) and (2.3).

Acknowledgements

The authors are very grateful to the reviewers for their valuable comments.

References


