STATISTICA, anno LXXV, n. 4, 2015

ON MODIFIED SKEW LOGISTIC REGRESSION MODEL AND ITS APPLICATIONS

C. Satheesh Kumar[1]
Department of Statistics, University of Kerala, Trivandrum-695 581, India

L. Manju
Department of Community Medicine, Sree Gokulam Medical College, Trivandrum-695 607, India

1. Introduction

Regression methods are usually used for studying the relationship between a response variable and one or more explanatory variables. Over the last decade the logistic regression model, also known as logit model, has become the standard method of analysis when the outcome variable is dichotomous in nature and it has been found applications in several areas of scientific studies such as bioassay problems (Finney, 1952), study of income distributions (Fisk, 1961), analysis of survival data (Plackett, 1959) and modelling of the spread of an innovation (Oliver, 1969). The main drawback of a logit model is that it consider variables of only symmetric and unimodal nature. But asymmetry may arise in several practical situations where the logit model is not appropriate. So through this paper we develop certain regression models based on a modified version of the skew-logistic distribution and compare it with the existing logit model as well as a regression model based on the skew-logistic distribution of (Nadarajah, 2009).

The paper is organized as follows. In section 2, we describe some important aspects of the skew-logistic regression model(SLRM) and propose a modified version of the SLRM, which we termed as the "Modified Skew Logistic Regression Model (MSLRM)". In section 3, we obtain some structural properties of the modified skew logistic distribution. In section 4, we consider the estimation of the parameters of the MSLRM and in section 5, two real life medical datasets are considered for illustrating the usefulness of the model compared to both the logit and skew logit models. In section 6, a generalized likelihood ratio test procedure is suggested for testing the significance of the parameters and a simulation study is also conducted to test the efficiency of the maximum likelihood estimators(MLEs) of MSLRM.

[1] Corresponding Author. E-mail: drcskumar02@gmail.com

2. Skew Logistic and Modified Skew Logitic regression models

A random variable X is said to follow the logistic distribution (LD) if its probability density function (p.d.f) is of the following form.

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \tag{1}$$

where $x \in \ R = (-\infty,\ +\infty)$. The logistic regression model(LRM) is given by

$$p = \frac{1}{1+e^{-z}}, \tag{2}$$

in which

$$z = a + \sum_{r=1}^{s} b_r X_r. \tag{3}$$

The importance of the logistic regression model is due to its mathematical flexibility and in several medical applications it provides clinically meaningful interpretations (Hosmer and Lemeshow, 2000). Nadarajah (2009) developed a modified version of the logistic distribution, similar to the skew-normal distribution of Azzalini (1985), namely skew-logistic distribution(SLD), which he defined through the following p.d.f, in which x $\in R, \beta > 0$ and $\lambda \in R$.

$$f(x) = \frac{2e^{\left(-\frac{x}{\beta}\right)}}{\beta\left[1+e^{\left(\frac{-x}{\beta}\right)}\right]^2\left[1+e^{\left(\frac{-\lambda x}{\beta}\right)}\right]} \tag{4}$$

When $\lambda = 0$, (4) reduces to p.d.f of the standard logistic distribution given in (1). The main feature of skew-logistic distribution is that a new parameter λ is included here for controlling the skewness and kurtosis. Now the skew-logistic regression model (SLRM) can be obtained through the following double series representation.

$$p = \begin{cases} 2\sum\limits_{j=0}^{\infty}\sum\limits_{=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{1}{1+\lambda+\lambda j+k}e^{\left(\frac{(1+\lambda+\lambda j+k)z}{\beta}\right)} & if\ z < 0 \\ \\ 1-2\sum\limits_{j=0}^{\infty}\sum\limits_{=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{1}{1+\lambda j+k}e^{-\left(\frac{(1+\lambda j+k)z}{\beta}\right)} & if\ z > 0 \end{cases} \tag{5}$$

where z is as defined in (3). Here we consider a modified version of the skew-logistic distribution namely "the modified skew-logistic distribution(MSLD)" through the following p.d.f, in which $x \in R$, $\alpha \geq -1$ and $\beta > 0$.

$$f(x;\alpha,\beta) = \frac{2}{\alpha+2}\frac{e^{-x}}{(1+e^{-x})^2}\left[1+\frac{\alpha e^{-\beta x}}{1+e^{-\beta x}}\right] \tag{6}$$

Clearly when $\alpha = 0$ and/or $\beta = 0$, the p.d.f reduces to the p.d.f of the logistic distribution and when α = -1 the p.d.f reduces to the p.d.f of the skew-logistic density given in (4). The p.d.f $f(x;\alpha,\beta)$ of the $MSLD(\alpha,\beta)$ can also be expressed interms of the following single series as well as double series representations,

$$f(x;\alpha,\beta)=\begin{cases}\frac{2}{\alpha+2}\left[\frac{e^{-x}}{(1+e^{-x})^2}+\frac{\alpha\sum\limits_{j=0}^{\infty}\binom{-1}{j}e^{(-1+\beta j)x}}{(1+e^{-x})^2}\right], & if\ \ x<0\\ \frac{2}{\alpha+2}\left[\frac{e^{-x}}{(1+e^{-x})^2}+\frac{\alpha\sum\limits_{j=0}^{\infty}\binom{-1}{j}e^{-(1+\beta+\beta j)x}}{(1+e^{-x})^2}\right], & if\ \ x>0\end{cases} \tag{7}$$

$$f(x;\alpha,\beta)=\begin{cases}\frac{2}{\alpha+2}\left[\sum\limits_{k=0}^{\infty}\binom{-2}{k}e^{(1+k)x}+\alpha\sum\limits_{j=0}^{\infty}\sum\limits_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}e^{(1+\beta j+k)x}\right], & if\ x<0\\ \frac{2}{\alpha+2}\left[\sum\limits_{k=0}^{\infty}\binom{-2}{k}e^{-(1+k)x}+\alpha\sum\limits_{j=0}^{\infty}\sum\limits_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}e^{-(1+\beta+\beta j+k)x}\right], & if\ x>0\end{cases} \tag{8}$$

The modified skew-logistic regression model(MLRM) is given by the following double series representation based on (8).

$$p=\begin{cases}\frac{2}{\alpha+2}\left[\sum\limits_{k=0}^{\infty}\binom{-2}{k}\frac{e^{(1+k)z}}{(1+k)}+\alpha\sum\limits_{j=0}^{\infty}\sum\limits_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{(1+\beta j+k)z}}{(1+\beta j+k)}\right], & if\ \ \ z<0\\ 1-\frac{2}{\alpha+2}\left[\sum\limits_{k=0}^{\infty}\binom{-2}{k}\frac{e^{-(1+k)z}}{(1+k)}+\alpha\sum\limits_{j=0}^{\infty}\sum\limits_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{-(1+\beta+\beta j+k)z}}{(1+\beta+\beta j+k)}\right], & if\ z\geq 0\end{cases} \tag{9}$$

where z is as given in (3). For the derivation of (9), see Appendix I. A graphical representation of MSLRM for particular values of α and $\beta=2$ is given in Figure 1.

3. Some structural properties of modified skew logistic model

Proposition 1. *If X follows $MSLD(\alpha,\beta)$, then $Y=-X$ follows a convex mixture of standard logistic and skew-logistic distributions.*

Proof. *The p.d.f $f(y)$ of Y is the following , for $y\in R$, $\beta\in R$ and $\alpha\geq -1$.*

$$\begin{aligned}f(y)&=f(-y;\alpha,\beta)\left|\frac{dx}{dy}\right|\\ &=\frac{2}{\alpha+2}\frac{e^{y}}{(1+e^{y})^2}\left[1+\alpha\frac{e^{\beta y}}{1+e^{\beta y}}\right]\\ &=\frac{2}{\alpha+2}\frac{e^{-y}}{(1+e^{-y})^2}\left[1+\frac{\alpha}{1+e^{-\beta y}}\right]\\ &=\frac{2}{\alpha+2}\frac{e^{-y}}{(1+e^{-y})^2}+\frac{2\alpha}{\alpha+2}\frac{e^{-y}}{(1+e^{-y})^2}\frac{1}{1+e^{-\beta y}},\end{aligned}$$

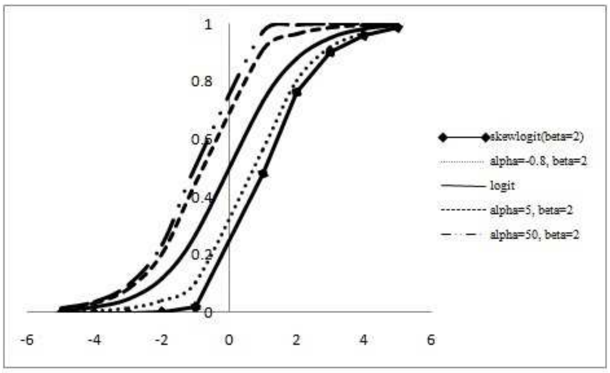

Figure 1 – Plots of regression function of $MSLD(\alpha, 2)$ for different values of α .

which shows that the p.d.f of Y can be considered as a convex mixture of the p.d.f of the standard logistic and and skew-logistic distributions.

PROPOSITION 2. *If X follows $MSLD(\alpha,\beta)$, then $Y = |X|$ follows a half logistic distribution.*

PROOF. *For $y > 0$, the p.d.f $f(y)$ of Y is*

$$f(y) = f(y;\alpha,\beta)\left|\tfrac{dx}{dy}\right| + f(-y;\alpha,\beta)\left|\tfrac{dx}{dy}\right|$$

$$= \tfrac{2}{\alpha+2}\tfrac{e^{-y}}{(1+e^{-y})^2}\left[1+\alpha\tfrac{e^{-\beta y}}{1+e^{-\beta y}}\right] + \tfrac{2}{\alpha+2}\tfrac{e^{-y}}{(1+e^{-y})^2}\left[1+\tfrac{\alpha}{1+e^{-\beta y}}\right]$$

$$= \tfrac{2e^{-y}}{(1+e^{-y})^2},$$

which is the p.d.f of the half logistic distribution.

PROPOSITION 3. *If X follows $MSLD(\alpha,\beta)$, then $Y = X^{1/c}, c \in (0,1]$ follows a distribution with p.d.f*

$$f(y) = \frac{2c}{\alpha+2}\frac{y^{c-1}e^{-y^c}}{(1+e^{-y^c})^2}\left[1+\alpha\frac{e^{-\beta y^c}}{1+e^{-\beta y^c}}\right]$$

PROOF. *For any $y > 0$, the p.d.f $f(y)$ of $Y = X^{1/c}$ is given by*

$$f(y) = \tfrac{2}{\alpha+2}\tfrac{y^{c-1}e^{-y^c}}{(1+e^{-y^c})^2}\left[1+\alpha\tfrac{e^{-\beta y^c}}{1+e^{-\beta y^c}}\right]\left|\tfrac{dx}{dy}\right|$$

$$= \tfrac{2c}{\alpha+2}\tfrac{y^{c-1}e^{-y^c}}{(1+e^{-y^c})^2}\left[1+\alpha\tfrac{e^{-\beta y^c}}{1+e^{-\beta y^c}}\right]$$

PROPOSITION 4. *The first four raw moments of $MSLD(\alpha,\beta)$ are given by*

$$\mu_1' = \frac{2\alpha}{\alpha+2}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\left[\frac{1}{(1+\beta+\beta j+k)^2}+\frac{(-1)}{(1+\beta j+k)^2}\right], \quad (10)$$

$$\mu_2' = \frac{4}{\alpha+2}\left\{2\sum_{k=0}^{\infty}\frac{\binom{-2}{k}}{(1+k)^3}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\left[\frac{1}{(1+\beta+\beta j+k)^3}+\frac{1}{(1+\beta j+k)^3}\right]\right\}, \quad (11)$$

$$\mu_3' = \frac{12\alpha}{\alpha+2}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\left[\frac{1}{(1+\beta+\beta j+k)^4}+\frac{(-1)}{(1+\beta j+k)^4}\right] \quad (12)$$

and

$$\mu_4' = \frac{48}{\alpha+2}\left\{2\sum_{k=0}^{\infty}\frac{\binom{-2}{k}}{(1+k)^5}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\left[\frac{1}{(1+\beta+\beta j+k)^5}+\frac{1}{(1+\beta j+k)^5}\right]\right\}. \quad (13)$$

PROOF. *By using the double series representation of the p.d.f of the $MSLD(\alpha,\beta)$ as given in (8) we obtain its first raw moment as*

$$\begin{aligned}\mu_1' &= \int_{-\infty}^{\infty} x f(x)dx \\ &= \tfrac{2}{\alpha+2}\Bigg[\sum_{k=0}^{\infty}\tbinom{-2}{k}\int_{-\infty}^{0} xe^{(1+k)x}dx + \alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\tbinom{-1}{j}\tbinom{-2}{k}\int_{-\infty}^{0} xe^{(1+\beta j+k)x}dx + \\ &\qquad \sum_{k=0}^{\infty}\tbinom{-2}{k}\int_{0}^{\infty} xe^{-(1+k)x}dx + \alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\tbinom{-1}{j}\tbinom{-2}{k}\int_{0}^{\infty} xe^{-(1+\beta+\beta j+k)}dx\Bigg],\end{aligned}$$

which gives (10), by using the standard results of integration. In a similar way one can obtain (11),(12) and (13).

Now by using proposition 4 we can evaluate the mean, variance, skewness and kurtosis of the $MSLD(\alpha,\beta)$ with the help of R softwares as obtained in Table 1.

TABLE 1
Mean, Variance, Skewness and Kurtosis of the MSLD for particular choice of α and β.

(α,β)	Mean	Variance	Skewness	Kurtosis
(-0.5,2)	0.4113	3.1209	0.0045	1.4013
(2, 10)	-0.6896	2.8145	0.0247	1.9544
(5,20)	-0.9901	2.3098	-0.0009	2.8710
(10, 50)	-1.1584	1.9482	-0.0645	3.5265
(20, 100)	-1.2659	1.6875	-0.4065	3.8782
(50, 150)	-1.3401	1.4941	-1.1858	3.9307
(-200, 200)	-1.3806	1.3839	-2.0794	3.8041

PROPOSITION 5. *The median of $MSLD(\alpha,\beta)$ is given by the following equations*

$$\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{(1+k)x_m}}{(1+k)}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{(1+\beta j+k)x_m}}{(1+\beta j+k)}\right]=\frac{1}{2}\quad if\ x_m<0$$

$$\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{\left(2-e^{-(1+k)x_m}\right)}{(1+k)}+\right.$$
$$\left.\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\left(\frac{1}{(1+\beta j+k)}+\frac{\left(1-e^{-(1+\beta+\beta j+k)x_m}\right)}{(1+\beta+\beta j+k)}\right)\right]=\frac{1}{2}\quad if\ x_m>0$$

PROOF. *The median of a probability density function f(x) is a point x_m on the real line which satisfies the equation $\int_{-\infty}^{x_m} f(x)dx=\frac{1}{2}$. Using the double series expansion of the p.d.f we get,*

Case i: If $x_m<0$

$$\int_{-\infty}^{x_m}\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}e^{(1+k)x_m}+\alpha\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\binom{-1}{j}\binom{-2}{k}e^{(1+\beta j+k)x_m}\right]dx=\frac{1}{2}$$

$$\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{(1+k)x_m}}{(1+k)}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{(1+\beta j+k)x_m}}{(1+\beta j+k)}\right]=\frac{1}{2}\qquad(14)$$

Case ii: If $x_m>0$

$$\int_{-\infty}^{0}\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{(1+k)x_m}}{(1+k)}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{(1+\beta j+k)x_m}}{(1+\beta j+k)}\right]dx+$$
$$\int_{0}^{x_m}\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{-(1+k)x_m}}{(1+k)}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{-(1+\beta+\beta j+k)x_m}}{(1+\beta+\beta j+k)}\right]dx=\frac{1}{2}$$

which on simplification yields

$$\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{\left(2-e^{-(1+k)x_m}\right)}{(1+k)}+\right.$$
$$\left.\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\left(\frac{1}{(1+\beta j+k)}+\frac{\left(1-e^{-(1+\beta+\beta j+k)x_m}\right)}{(1+\beta+\beta j+k)}\right)\right]=\frac{1}{2} \tag{15}$$

Closed form for x_m is not obtainable, on solving equations (14) or (15) using mathematical softwares MATHEMATICA OR MATHCAD one can obtain the median.

PROPOSITION 6. *The mode of $MSLD(\alpha,\beta)$ is given by the following equation*

$$\frac{-2e^{-x}\left[1-e^{-x}+(2+\alpha+\alpha\beta)e^{-\beta x}-(2+\alpha-\alpha\beta)e^{-(1+\beta)x}+(1+\alpha)e^{-2\beta x}-(1+\alpha)e^{-(1+2\beta)x}\right]}{(2+\alpha)(1+e^{-x})(1+e^{-\beta x})^2}=0 \tag{16}$$

PROOF. *The mode of a probability density function is obtained by equating the derivative of the density function to zero and solving for the variable. Thus, differentiating $f(x;\alpha,\beta)$ with respect to x yields (16). Using the mathematical softwares MATHCAD OR MATHEMATICA one can solve (16) and obtain the mode.*

PROPOSITION 7. *The mean deviation about the average A denoted by $\delta_1(X)$ is given by the following*

$$\delta_1(X)=\begin{cases}\delta_{11}(A), & if A<0\\ \delta_{12}(A), & if A\geq 0\end{cases}$$

where,

$$\delta_{11}(A)=\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{(1+k)A}}{(1+k)^2}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{(1+\beta j+k)A}}{(1+\beta j+k)^2}\right]-A,$$

$$\delta_{12}(A)=\frac{2}{\alpha+2}\left[2\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{-(1+k)A}}{(1+k)^2}+\right.$$
$$\left.\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\left(\frac{\left(2e^{-(1+\beta+\beta j+k)A}-1\right)}{(1+\beta+\beta j+k)^2}+\frac{1}{(1+\beta j+k)^2}\right)\right]+A$$

PROOF. *The mean deviation about A is defined by*

$$\delta_1(X)=\int_{-\infty}^{\infty}|x-A|f(x)\,dx$$

$\delta_1(X)$ on simplification yields

$$\delta_1(X)=2AF(A)-A-\int_{-\infty}^{A}xf(x)\,dx+\int_{A}^{\infty}xf(x)\,dx$$

Case (i). $A < 0$

$$\delta_1(X) = 2AF(A) - A - \int_{-\infty}^{A} xf(x)\,dx + \int_{A}^{0} xf(x)\,dx + \int_{0}^{\infty} xf(x)\,dx \tag{17}$$

Using the double series expansion of p.d.f we get,

$$\int_{-\infty}^{A} xf(x)\,dx = \frac{2}{\alpha+2}\left\{\sum_{k=0}^{\infty}\frac{\binom{-2}{k}e^{(1+k)A}}{(1+k)^2}\left[A(1+k)-1\right] + \alpha\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{-1}{j}\binom{-2}{k}}{(1+\beta j+k)^2}\left[A(1+\beta j+k)-1\right]\right\} \tag{18}$$

$$\int_{A}^{0} xf(x)\,dx = \frac{2}{\alpha+2}\left\{\sum_{k=0}^{\infty}\frac{\binom{-2}{k}}{(1+k)^2}\left[e^{(1+k)A}\left(1-A(1+k)\right)-1\right] + \alpha\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{-1}{j}\binom{-2}{k}}{(1+\beta j+k)^2}\left[e^{(1+\beta j+k)A}\left(1-A(1+\beta j+k)\right)-1\right]\right\} \tag{19}$$

$$\int_{0}^{\infty} xf(x)\,dx = \frac{2}{\alpha+2}\left\{\sum_{k=0}^{\infty}\frac{\binom{-2}{k}}{(1+k)^2} + \alpha\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{-1}{j}\binom{-2}{k}}{(1+\beta+\beta j+k)^2}\right\} \tag{20}$$

Substituting (18)-(20) in (17) we get the equation of $\delta_{11}(A)$.
Case (ii). $A \geq 0$

$$\delta_1(X) = 2AF(A) - A - \int_{-\infty}^{0} xf(x)\,dx - \int_{0}^{A} xf(x)\,dx + \int_{A}^{\infty} xf(x)\,dx \tag{21}$$

$$\int_{-\infty}^{0} xf(x)\,dx = \frac{-2}{\alpha+2}\left\{\sum_{k=0}^{\infty}\frac{\binom{-2}{k}}{(1+k)^2} + \alpha\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{-1}{j}\binom{-2}{k}}{(1+\beta j+k)^2}\right\} \tag{22}$$

$$\int_{0}^{A} xf(x)\,dx = \frac{-2}{\alpha+2}\left\{\sum_{k=0}^{\infty}\frac{\binom{-2}{k}}{(1+k)^2}\left[e^{-(1+k)A}\left(1+A(1+k)\right)-1\right] - \alpha\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{-1}{j}\binom{-2}{k}}{(1+\beta+\beta j+k)^2}\left[e^{-(1+\beta+\beta j+k)A}\left(1+A(1+\beta j+k)\right)-1\right]\right\} \tag{23}$$

$$\int_{A}^{\infty} xf(x)\,dx = \frac{2}{\alpha+2}\left\{\sum_{k=0}^{\infty}\frac{\binom{-2}{k}}{(1+k)^2}e^{-(1+k)A}\left[1+A(1+k)\right] + \alpha\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{-1}{j}\binom{-2}{k}}{(1+\beta+\beta j+k)^2}e^{-(1+\beta+\beta j+k)A}\left[1+A(1+\beta j+k)\right]\right\} \tag{24}$$

Substituting (22) - (24) in (21) we get the equation of $\delta_{12}(A)$.

4. Estimation

This section deals with the maximum likelihood estimation of the parameters of the MSLRM. Suppose we have a sample of n independent observations of the pair (x_i, y_i) , i=1,2,...,n, where y_i denotes the value of the dichotomous outcome variable and x_i is the value of the independent variable for the i^{th} subject. Let $p_i = P(Y_i = 1|X_i)$, so that $P(Y_i = 0|X_i) = 1 - p_i$. The probability of observing the outcome Y_i whether it is 0 or 1 is given by $P(Y_i|X_i) = p_i^{y_i}(1-p_i)^{1-y_i}$. If there are n sets of values of X_i, say $\boldsymbol{X}$, the probability of observing a particular sample of n values of Y, say $\boldsymbol{Y}$ is given by the product of n probabilities, since the observations are independent. That is,

$$P(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} \tag{25}$$

Let $z = a + \sum_{r=1}^{s} b_r X_r$ and $\Theta = (\alpha, \beta, a, b_1, b_2, ..., b_s)$ be the vector of parameters of the MSLD regression model and let $\widehat{\Theta} = \left(\widehat{\alpha}, \widehat{\beta}, \widehat{a}, \widehat{b_1}, \widehat{b_2}, ..., \widehat{b_s}\right)$ be the maximum likelihood estimator(MLE) of Θ. The log-likelihood function of MSLRM is given by

$$l = \log\ L(y|z,\theta) = \sum_{i=1}^{n} y_i \log p_i + \sum_{i=1}^{n}(1-y_i)\log(1-p_i) \tag{26}$$

The MLE of the parameters are obtained by solving the following set of likelihood equations, in which, $\delta(p,y) = \sum_{i=1}^{n}\frac{y_i}{p_i} - \sum_{i=1}^{n}\frac{(1-y_i)}{(1-p_i)}$, with p_i is as defined in (9).

Case 1 : For z $\geq$ 0,

$$\frac{\partial l}{\partial \alpha} = 0$$

or equivalently

$$\delta(p,y)\left[\frac{2}{(\alpha+2)^2}\left(\sum_{k=0}^{\infty}(-1)^k e^{-(1+k)z} - 2\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{(1+k)\,e^{-(1+\beta+\beta j+k)z}}{(1+\beta+\beta j+k)}\right)\right] = 0, \tag{27}$$

$$\frac{\partial l}{\partial \beta} = 0$$

or equivalently

$$\delta(p,y)\left[\frac{2\alpha}{(\alpha+2)}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{(1+k)(1+j)\,e^{-(1+\beta+\beta j+k)z}}{(1+\beta+\beta j+k)^2}\left[(1+\beta+\beta j+k)z+1\right]\right] = 0, \tag{28}$$

$$\frac{\partial l}{\partial a} = 0$$

or equivalently

$$\delta\left(p,y\right)\left[\frac{2}{\left(\alpha+2\right)}\left(\sum_{k=0}^{\infty}\left(-1\right)^{k}\left(1+k\right)e^{-\left(1+k\right)z}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(-1\right)^{j+k}\left(1+k\right)e^{-\left(1+\beta+\beta j+k\right)z}\right)\right]=0 \tag{29}$$

and for r = 1, 2,..., s,

$$\frac{\partial l}{\partial b_r}=0$$

or equivalently

$$\delta\left(p,y\right)\left[\frac{2x_r}{\left(\alpha+2\right)}\left(\sum_{k=0}^{\infty}\left(-1\right)^{k}\left(1+k\right)e^{-\left(1+k\right)z}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(-1\right)^{j+k}\left(1+k\right)e^{-\left(1+\beta+\beta j+k\right)z}\right)\right]=0. \tag{30}$$

Case 2 : For z <0

$$\frac{\partial l}{\partial \alpha}=0$$

or equivalently

$$\delta\left(p,y\right)\left[\frac{2}{\left(\alpha+2\right)^2}\left(2\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(-1\right)^{j+k}\left(1+k\right)\frac{e^{\left(1+\beta j+k\right)z}}{\left(1+\beta j+k\right)}-\sum_{k=0}^{\infty}\left(-1\right)^{k}e^{-\left(1+k\right)z}\right)\right]=0, \tag{31}$$

$$\frac{\partial l}{\partial \beta}=0$$

or equivalently

$$\delta\left(p,y\right)\left[\frac{2\alpha}{\left(\alpha+2\right)}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(-1\right)^{j+k}\left(1+k\right)\frac{je^{\left(1+\beta j+k\right)z}\left[\left(1+\beta j+k\right)z-1\right]}{\left(1+\beta j+k\right)^2}\right)\right]=0, \tag{32}$$

$$\frac{\partial l}{\partial a}=0$$

or equivalently

$$\delta\left(p,y\right)\left[\frac{2}{\left(\alpha+2\right)}\left(\sum_{k=0}^{\infty}\left(-1\right)^{k}\left(1+k\right)e^{\left(1+k\right)z}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(-1\right)^{j+k}\left(1+k\right)e^{\left(1+\beta j+k\right)z}\right)\right]=0, \tag{33}$$

and for r = 1, 2,..., s,

$$\frac{\partial l}{\partial b_r}=0$$

or equivalently

$$\delta\left(p,y\right)\left[\frac{2x_r}{\left(\alpha+2\right)}\left(\sum_{k=0}^{\infty}\left(-1\right)^{k}\left(1+k\right)e^{\left(1+k\right)z}+\alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(-1\right)^{j+k}\left(1+k\right)e^{\left(1+\beta j+k\right)z}\right)\right]=0. \tag{34}$$

Second order partial derivatives of equation (26) with respect to the parameters are observed with the help of MATHEMATICA software and find that the equation gives negative values for all $\alpha \geq -1$, $\beta > 0$, $a, b \in R$. Now we can obtain the $\widehat{\Theta}$ by solving the likelihood equations (27) to (34) with the help of some mathematical softwares such as MATHCAD, MATHEMATICA, R Softwares etc.

5. Applications

Here we illustrate the procedures discussed in section 4 with the help of the following two data sets:
Data Set 1. Shock data set obtained from (Afifi and Azen, 1979).
(see https://www.umass.edu/statdata/statdata/stat-logistic.html). These data were collected at the Shock Research Unit at the University of Southern California, Los Angeles, California. Data were collected on 113 critically ill patients. Here we consider the explanatory variable as the urine output (ml/hr) at the time of admission and the dependent variable y as whether the person survived or not.
Data Set 2.Prostate cancer data set:
(see https://www.umass.edu/statdata/statdata/stat-logistic.html).The data is on 380 subjects of which 153 had tumor that penetrated the prostatic capsule. The variable *capsule* denotes the status of the tumor, whether it is penetrated or not, which we consider as the dichotomous dependent variable(Y) and *Prostatic Specimen Antigen Value(PSA)*in mg/ml as the explanatory (X) variable. These data set is also studied in (Hosmer and Lemeshow, 2000). These data are copyrighted by John Wiley and Sons Inc.
Here we consider the simplest model, $Z = a + bX$. We obtained the MLE of the parameters α, β, a, b with the help of R software 3.0.3 by using the “nlm package”. The estimated parameters of logit and skew logit model are also obtained by the same procedure. The estimated values of the parameters of the LRM, SLRM and MSLRM along with the computed values of the Akaike Information Criterion (AIC), Bayes Information Criterion(BIC) and pseudo R^2 values (Mcfadden, 1973; Cameron and Windmeijer, 1996) are given in Table 2. Also, we have plotted the Empirical cummulative distribution function (ECDF) of the data sets and the three fitted regression models in Figure 2 and Figure 3. The values of AIC, BIC are relatively less in case of the MSLRM compared to other existing models while the values of the of pseudo R^2 (such as McFadden's R^2, McFadden's Adj R^2, Cox Snell R^2, Cragg-Uhler(Nagelkerke)R^2) are more in case of MSLRM. This shows that the MSLRM gives a better fit to the given data sets compared to other existing models. Also Figure 2 and Figure 3 support this conclusion.

TABLE 2

Estimated values of the parameters with the corresponding Pseudo R^2 values and Information criteria values.

		Distribution		
Data set		LRM(a,b)	SLRM(λ,β,a,b)	MSLRM(α,β,a,b)
Data set 1	$\hat{\alpha}$	–	–	12.134
	$\hat{\lambda}$	–	11.400	–
	$\hat{\beta}$	–	8.317	1.793
	$\hat{a}$	0.154	11.774	0.006
	$\hat{b}$	0.024	0.086	0.001
	AIC	150.000	148.980	144.100
	BIC	150.106	149.192	144.202
	McFadden's R^2	0.040	0.073	0.105
	McFadden's Adj R^2	0.014	0.021	0.053
	Cox Snell R^2	0.053	0.094	0.132
	Cragg-Uhler(Nagelkerke)R^2	0.071	0.127	0.179
Dataset 2	$\hat{\alpha}$	–	–	11.080
	$\hat{\lambda}$	–	34.100	–
	$\hat{\beta}$	–	13.238	9.170
	$\hat{a}$	-1.114	5.044	-0.002
	$\hat{b}$	0.057	0.488	0.0002
	AIC	468.240	466.860	464.260
	BIC	469.399	469.179	466.579
	McFadden's R^2	0.094	0.104	0.109
	McFadden's Adj R^2	0.086	0.089	0.097
	Cox Snell R^2	0.119	0.131	0.140
	Cragg-Uhler(Nagelkerke)R^2	0.160	0.177	0.185

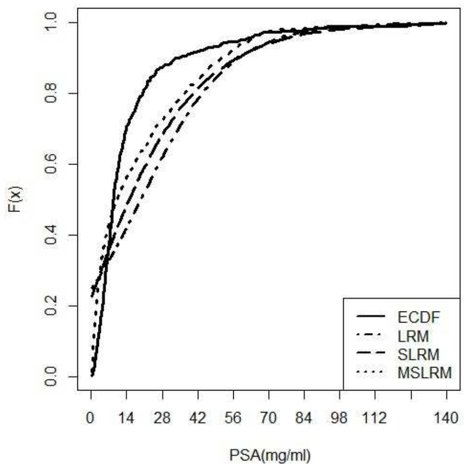

Figure 2 – ECDF of the data set1 and fitted regression models - LRM, SLRM and MSLRM

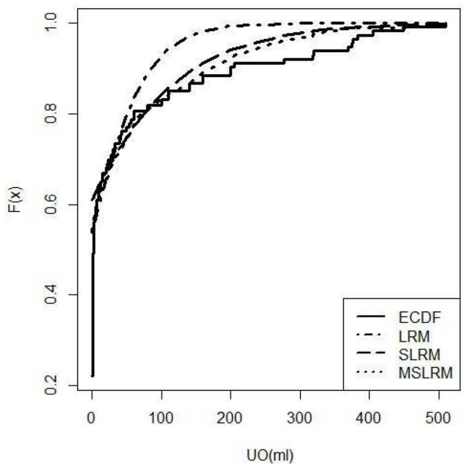

Figure 3 – ECDF of the data set2 and fitted regression models - LRM, SLRM and MSLRM

TABLE 3
Calculated values of the test statistic

Data set	Hypothesis	$\log L\left(\hat{\Omega}; y\|x\right)$	$\log L\left(\hat{\Omega}^*; y\|x\right)$	Test statistic
Data set1	$H_0 : \alpha = 0$	-68.047	-73.001	9.908
	$H_0 : \alpha = -1$	-68.047	-70.490	4.886
Data set2	$H_0 : \alpha = 0$	-228.130	-232.120	7.980
	$H_0 : \alpha = -1$	-228.130	-230.430	4.600

6. Testing of hypothesis and Simulation

In this section we discuss certain test procedures for testing the significance of the additional parameter α of the MSLRM and carry out a brief simulation study for examining the performance of the maximum likelihood estimators. First we discuss the generalized likelihood ratio test procedure for testing the following hypothesis:
Test 1. $H_0 : \alpha = 0$ against the alternative hypothesis $H_1 : \alpha \neq 0$
Test 2. $H_0 : \alpha = -1$ against the alternative hypothesis $H_1 : \alpha \neq -1$.
Here the test statistic is,

$$-2 \log \Lambda = 2 \left[\log L\left(\hat{\Omega}; y|x\right) - \log L\left(\hat{\Omega}^*; y|x\right) \right] \tag{35}$$

where $\hat{\Omega}$ is the maximum likelihood estimator of $\Omega = (\alpha, \beta, a, b)$ with no restriction, and $\hat{\Omega}^*$ is the maximum likelihood estimator of Ω when $\alpha = 0$ in case of Test 1 and $\alpha = -1$ in case of Test 2. The test statistic $-2 \log \Lambda$ given in (35) is asymptotically distributed as χ^2 with one degree of freedom (Rao, 1973). The computed values of $logL\left(\hat{\Omega}; y|x\right)$, $logL\left(\hat{\Omega}^*; y|x\right)$ and test statistic in case of both data sets are listed in Table 3.
Since the critical value at the significance level 0.05 and degree of freedom one is 3.84, the null hypothesis is rejected in both the case, which shows the appropriatness of the MSLRM to the data sets.

Next we conduct a simulation study for assessing the performance of the MLEs of the parameters of the MSLRM. We consider the following two sets of parameters.
1. α=12.134, β= 1.793, a= 0.006 , b=0.001
2. α=11.086, β= 9.170, a= -0.002 , b=0.0002

TABLE 4. Bias and Mean Square Error(MSE) within brackets of the simulated data sets

Data Set	sample size:	α	β	a	b
Data Set1	100	4.62E-01 (7.30E-01)	-1.09E-01 (1.23E-02)	5.62E-04 (8.29E-07)	1.44E-03 (4.11E-06)
	200	-4.31E-01 (3.98E-01)	-4.96E-02 (4.24E-03)	3.55E-04 (2.56E-07)	9.87E-04 (9.92E-07)
	300	-1.14E-01 (2.22E-01)	-2.33E-02 (1.44E-03)	1.74E-04 (5.62E-08)	7.65E-04 (7.18E-07)
	500	-6.02E-02 (7.22E-02)	-6.06E-03 (2.54E-04)	9.14E-05 (1.64E-08)	3.29E-04 (2.92E-07)
Data Set2	100	6.49E-01 (1.01E+00)	1.21E-01 (8.24E-01)	-1.10E-03 (1.84E-06)	3.88E-04 (6.89E-07)
	200	5.14E-01 (5.92E-01)	-7.29E-02 (2.22E-02)	-6.07E-04 (4.46E-07)	1.03E-04 (1.10E-08)
	300	-1.34E-01 (4.54E-01)	8.57E-03 (4.55E-03)	-1.34E-04 (6.46E-08)	4.86E-05 (2.94E-09)
	500	9.94E-02 (7.76E-02)	2.60E-03 (5.65E-05)	4.95E-05 (2.45E-08)	2.07E-05 (1.87E-09)

The computed values of the bias and mean square error(MSE) corresponding to sample sizes 100, 200, 300 and 500 respectively are given in Table 4.

From the table it can be seen that both the absolute bias and MSEs in respect of each parameters of the MSLRM are in decreasing order as the sample size increases.

Acknowledgements

The authors would like to express their sincere thanks to the Editor and the anonymous Referee for for their valuable comments on an early version of the paper which greatly improved the quality and presentation of the manuscript.

Appendix

A. Proofs of equation (9)

By definition, using the double series expansion for p.d.f, the c.d.f of the $MSLD(\alpha,\beta)$ takes the following form, for $x < 0$.

$$\begin{aligned} F(x) &= \frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\int_{-\infty}^{x} e^{(1+k)x}dx + \alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\int_{-\infty}^{x} e^{(1+\beta j+k)x}dx\right] \\ &= \frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{(1+k)x}}{(1+k)} + \alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{(1+\beta j+k)x}}{(1+\beta j+k)}\right] \qquad (36) \end{aligned}$$

In a similar way, the c.d.f of the $MSLD(\alpha,\beta)$ can be written as given below, for $x \geq 0$.

$$\begin{aligned} F(x) &= 1-\int_{x}^{\infty} f(x)\,dx \\ &= 1-\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\int_{x}^{\infty} e^{-(1+k)x}dx + \alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\int_{x}^{\infty} e^{-(1+\beta+\beta j+k)x}dx\right] \\ &= 1-\frac{2}{\alpha+2}\left[\sum_{k=0}^{\infty}\binom{-2}{k}\frac{e^{-(1+k)x}}{(1+k)} + \alpha\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\binom{-1}{j}\binom{-2}{k}\frac{e^{-(1+\beta+\beta j+k)x}}{(1+\beta+\beta j+k)}\right] \qquad (37) \end{aligned}$$

Thus (36) and (37) gives (9).

References

A. A. Afifi, S. P. Azen (1979). *Statistical Analysis: A Computer Oriented Approach.* Academic Pres, New York.

A. Azzalini (1985). *A class of distributions which includes the normal ones.* Scandinavian Journal of Statistics, 12, pp. 171–178.

A. C. Cameron, F. A. Windmeijer (1996). *R-squared measures for count data regression models with applications to health-care utilization.* Journal of Business and Economic Statistics, 14, no. 2, pp. 209–220.

D. J. Finney (1952). *Statistical methods in Biological Assay.* Hafner Publications, NewYork.

P. R. Fisk (1961). *The graduation of income distributions.* Econometrica, 29, pp. 171–185.

D. W. Hosmer, S. Lemeshow (2000). *Applied Logistic Regression.* JohnWiley and Sons, NewYork.

D. Mcfadden (1973). *Conditional logit analysis of qualitative choice behavior in zarembka,p. (ed.).* In *Frontiers in Econometrics*, Academic Press, NewYork, pp. 105–142.

S. Nadarajah (2009). *The skew logistic distribution.* AStA. Adv. Stat. Anal, 93, pp. 187–203.

F. R. Oliver (1969). *Another generalisation of the logistic growth function.* Econometrica, 37, pp. 144–147.

R. L. Plackett (1959). *The analysis of life test data.* Technometrics, 1, pp. 9–19.

C. R. Rao (1973). *Linear statistical inference and its applications.* NewYork, Wiley.

Summary

Here we consider a modified form of the logistic regression model useful for situations where the dependent variable is dichotomous in nature and the explanatory variables exhibit asymmetric behaviour. Certain structrual properties of the modified skew logistic model is discussed and the proposed regression model has been fitted to some real life data sets by using the method of maximum likelihood estimation. Further, two data illustrations are given for highlighting the usefulness of the model in certain medical applications and a simulation study is conducted for assessing the performance of the estimators.

Keywords: Logistic regression; Maximum likelihood estimation; Asymmetric distributions; Simulation.