

On the Existence of an Analytical Solution in Multiple Logistic Regression

Guoping Zeng

4522 Oak Shores Dr, Plano, TX 75024, USA

gzeng314159@gmail.com

ABSTRACT

Fitting a logistic regression model to a given data starts from the likelihood function. Typically, the regression parameters are solved by maximizing the likelihood function. In general, there is no analytical solution since these regression parameters fall into a set of nonlinear equations. So far, only two cases have been known to have an analytical closed-form solution for univariate logistic regression: a dichotomous variable and a categorical variable. In this paper we study the existence of an analytical solution for multiple logistic regression. We discover 3 types of data that yields an analytical solution: Partially balanced data; Perfectly balanced data; Quasi-saturated data. We show that the 2 known cases for univariate logistic regression are special cases to ours with quasi-saturated data.

Keywords: Analytic solution, logistic regression, partially balanced data, perfectly balanced data, quasi-saturated data, overlap.

Mathematics Subject Classification: 62J02, 62-07

1. INTRODUCTION

Logistic regression is a popular statistical and mathematical models for predicting a binary outcome (Siddiqi 2006; Refaat 2011; Zeng 2013; Zeng 2014; Zeng 2015; Wang and Zhu 2015; Habshah et al. 2016; Zeng 2017a; Zeng 2017b; Zeng 2017c; Zeng 2018; Zeng 2019a; Zeng 2019b).

Fitting a logistic regression model to a given data starts from the likelihood function. The parameters (including the intercept) of independent variables are then solved by maximizing the likelihood function. The values of such parameters are called the maximum likelihood estimates (MLEs). For convenience, we also call the MLEs a solution to the logistic regression model.

Albert and Anderson (1984) completely characterized the existence and uniqueness of MLEs using the configuration of the data points: complete separation, quasi-complete separation and overlap. Consequently, the MLEs exist and are finite and unique if and only if there is overlap of the data points. Since the equations to determine the MLEs are nonlinear, in general, there is no explicit solution or closed-form solution to logistic regression (Albert and Anderson, 1984; Agresti, 1996). Rather, a numerical solution can be found by numerical methods such as Newton-Raphson iterative method.

In mathematics, there is no standard definition for explicit solution or closed-form solution. Typically, an explicit solution or closed-form solution solves a given equation with a finite number of standard functions (radical root, exponent, logarithm, trigonometric functions, etc.) and well-known mathematical operations (+, −, ×, ÷, etc.) (Borwein and Crandall, 2013). Nevertheless, the set of operations and

functions may vary with authors, mathematical branches and context. In this paper we extend the definition of an explicit solution or a closed-form solution by including matrix operations.

So far, the literature about the analytical solution of logistic regression has been very limited. Hosmer and Lemeshow (2000) indicated a closed-form formula for the parameter of a single dichotomous variable. Allison (2004 and 2008) gave a closed-form expression for the parameter of a single binary variable with values of 0 and 1. Lipovetsky (2015) improved the result by finding a closed-form formula for both the intercept (often labeled the constant) and the parameter of a single dichotomous variable. He also found a closed-form formula for a single categorical variable. However, Lipovetsky (2015) only covers a single variable.

In this paper we study the existence of an analytical solution to multivariate logistic regression. We discover 3 types of data that yields an analytical solution: (1) Partially balanced data; (2) Perfectly balanced data; and (3) Quasi-saturated-like data.

This paper is structured as follows. Section 2 briefly introduces logistic regression and its maximum likelihood estimates. In Section 3, we show that three types of data in multiple logistic regression have an analytical solution: (1) Partially balanced data; (2) Perfectly balanced data; and (3) Quasi-saturated data. In Section 4, numerical examples are provided for the 2 known cases of Lipovetsky (2005) to demonstrate how to find the analytical closed-form solution by using our result and methods for saturated data. The paper is concluded in Section 5.

2. MAXIMUM LIKELIHOOD ESTIMATES

Let $x = (x_1, x_2, \dots, x_p)$ be the vector of p independent variables and y be the binary dependent variable with values 0 and 1. Note that y cannot be all 1's or all 0's. Let us consider a sample of N independent observations $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$, $i = 1, 2, \dots, N$, where y_i is the value of y and $x_{i1}, x_{i2}, \dots, x_{ip}$ the values of $x_1, x_2, \dots, x_p$ for the i -th observation.

Let X_i be the row vector $(1, x_{i1}, x_{i2}, \dots, x_{ip})$ for $i = 1, 2, \dots, N$, and denote X the $N \times (p + 1)$ matrix, called Design Matrix, with X_i as rows. Let us denote the N -dimensional column vector with all elements of 1 by x_0 . For convenience, we also denote the N -dimensional vectors

$$(x_{11}, x_{21}, \dots, x_{N1})^T, (x_{12}, x_{22}, \dots, x_{N2})^T, (x_{1p}, x_{2p}, \dots, x_{Np})^T$$

by $x_1, x_2, \dots, x_p$, respectively, where superscript T is matrix transport. The following log likelihood is used in logistic regression

$$l(X, \beta) = \sum_{i=1}^N y_i \ln \left(\frac{e^{X_i \beta}}{1 + e^{X_i \beta}} \right) + \sum_{i=1}^N (1 - y_i) \ln \left(\frac{1}{1 + e^{X_i \beta}} \right), \quad (2.1)$$

where β is column vector $(\beta_0, \beta_1, \dots, \beta_p)^T$, and $X_i \beta$ represents matrix multiplication.

The MLEs are the values of the unknown parameters $\beta_0, \beta_1, \dots, \beta_p$ that maximize (2.1). By algebraic manipulation, we see that the MLEs satisfy the following $(p + 1)$ nonlinear equations

$$\sum_{i=1}^N y_i = \sum_{i=1}^N \pi_i \quad (2.2)$$

$$\sum_{i=1}^N x_{ij} y_i = \sum_{i=1}^N x_{ij} \pi_i, j = 1, 2, \dots, p$$

as in Hosmer and Lemeshow (2000), where $\pi_i = \pi(x_{i1}, x_{i2}, \dots, x_{ip}) = \frac{e^{X_i \beta}}{1 + e^{X_i \beta}}$ represents the conditional probability that y is equal to 1 given x for $i = 1, 2, \dots, N$.

Clearly, (2.2) can be written into the matrix form:

$$X^T(y - \pi) = 0 \quad (2.3)$$

where y and π are N -dimensional column vectors with $y = (y_1, y_2, \dots, y_N)^T$ and $\pi = (\pi_1, \pi_2, \dots, \pi_N)^T$.

According to Albert and Anderson (1984), the MLEs exist and are finite and unique if and only if there is overlap of the data points. In this case, all the probabilities $\pi_1, \pi_2, \dots, \pi_N$ will be in the open interval $(0, 1)$.

In general, the $(p + 1)$ nonlinear equations in (2.2) cannot be solved explicitly. Rather, they can be approximately solved numerically by Newton-Raphson algorithm (See Refaat 2011)

$$\beta^{(i+1)} = \beta^{(i)} + \hat{I}^{-1} g, \quad (2.4)$$

where $\hat{I} = X^T V X$, and V is the $N \times N$ diagonal matrix with elements

$$\pi_1(1 - \pi_1), \pi_2(1 - \pi_2), \dots, \pi_N(1 - \pi_N).$$

In addition, $g = X^T(y - \pi)$, and $\hat{I}$ and g are evaluated at $\beta^{(i)}$ in (2.3).

Although (2.2) holds no matter whether $N > p + 1$, the relationship between N and $p + 1$ is ignored in literature. Albert and Anderson (1984) assumed $N > p + 1$ and X has a full column rank $p + 1$, which have since been implicitly assumed in literature, say, Hosmer and Lemeshow (2000) and Refaat (2011).

However, if $N < p + 1$, then $\hat{I}$ is singular and so Newton-Raphson iteration (2.4) is not applicable. Indeed, the $p + 1$ vectors $x_0, x_1, x_2, \dots, x_p$ are linear dependent according to Zeng (2019b). Therefore, one of $x_0, x_1, x_2, \dots, x_p$ is a linear combination of others. Let us assume x_p is a linear combination of $x_0, x_1, x_2, \dots, x_{p-1}$, without loss of generality. Then, there are constants $c_0, c_1, c_2, \dots, c_{p-1}$ such that $x_p = \sum_{i=0}^{p-1} c_i x_i$. Now if $(\beta_0, \beta_1, \dots, \beta_p)^T$ is a solution to (2.2), then

$$\beta_0 x_0 + \beta_1 x_1 + \dots + \beta_p x_p = \sum_{i=0}^{p-1} (\beta_i + \beta_p c_i) x_i. \quad (2.5)$$

That is, $(\beta_0 + \beta_p c_0, \beta_1 + \beta_p c_1, \dots, \beta_{p-1} + \beta_p c_{p-1}, 0)^T$ is also a solution. This is not generally acceptable. In practice, it is rare that $N < p + 1$.

3. THREE TYPES OF DATA WITH AN ANALYTICAL SOLUTION

As in Albert and Anderson (1984), we make the 3 basic assumptions:

- C1: $N > p + 1$.

- C2: X has Full column rank $p + 1$.
- C3: There is overlap in the N data points.

Let us assume all categorical variables have been already converted into continuous variables.

3.1. Partially balanced data

Definition 3.1. The data is partially balanced if

$$\sum_{i=1}^N x_{ij} y_i = \frac{\sum_{i=1}^N y_i}{N} \times \sum_{i=1}^N x_{ij}, j = 1, 2, \dots, p. \quad (3.1)$$

Theorem 3.2. There is a closed-form solution if the data is partially balanced. Specifically, $\beta_1 = \beta_2 = \dots = \beta_p = 0$ and

$$\beta_0 = \ln \frac{\frac{\sum_{i=1}^N y_i}{N}}{1 - \frac{\sum_{i=1}^N y_i}{N}}. \quad (3.2)$$

Proof. It suffices to prove that β_0 defined by (3.2) and $\beta_1 = \beta_2 = \dots = \beta_p = 0$ satisfy (2.2).

Since $\beta_1 = \beta_2 = \dots = \beta_p = 0$, we have

$$\pi_1 = \pi_2 = \dots = \pi_N = \frac{e^{\beta_0}}{1 + e^{\beta_0}} = \frac{\sum_{i=1}^N y_i}{N}.$$

Hence, $\sum_{i=1}^N \pi_i = \sum_{i=1}^N y_i$, which is the first equation of (2.2). Since $\pi_1 = \pi_2 = \dots = \pi_N = \frac{\sum_{i=1}^N y_i}{N}$, the p equations in (3.1) are the second to the last equations of (2.2).

■

In particular, if $p = 1$, then (3.1) reduces to

$$\frac{\sum_{i=1}^N x_{i1} y_i}{N_1} = \frac{\sum_{i=1}^N x_{i1} (1 - y_i)}{N_0}$$

where N_1 and N_0 are number of records with $y_i = 1$ and $y_i = 0$, respectively. That is, the sample average of x_1 for observations with $y_i = 1$ equals the sample average of x_1 for observations with $y_i = 0$. Thus, $\beta_1 = 0$ by Owen and Roediger (2014). Indeed, the right-hand side of (3.1) can be written as

$$\frac{N_1}{N} \sum_{i=1}^N x_{i1} (y_i + 1 - y_i) = \frac{N_1}{N} \sum_{i=1}^N x_{i1} y_i + \frac{N_1}{N} \sum_{i=1}^N x_{i1} (1 - y_i).$$

By some algebra manipulations, we obtain

$$\frac{\sum_{i=1}^N x_{i1} y_i}{N_1} = \frac{\sum_{i=1}^N x_{i1} (1 - y_i)}{N_0}.$$

It should be not confused with intercept-only model in which there are no independent variables.

3.2. Perfectly balanced data

Definition 3.3. If all the N observations $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$ exist in pairs with respect to $y = 0$ and $y = 1$, that is, if $(x_{i1}, x_{i2}, \dots, x_{ip}, 0)$ is one observation, so is $(x_{i1}, x_{i2}, \dots, x_{ip}, 1)$ and vice versa.

Theorem 3.4. There is a closed-form solution if the data is perfectly balanced. Specifically, $\beta_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$.

Proof. It suffices to show that $\beta_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$ satisfy (2.2). If $\beta_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$, then

$$\pi_1 = \pi_2 = \dots = \pi_p = \frac{e^0}{1 + e^0} = \frac{1}{2}.$$

Since the data is perfectly balanced,

$$\sum_{i=1}^N y_i = \frac{N}{2} = \sum_{i=1}^N \pi_i,$$

which is the first equation of (2.2). Next, for $j = 1, 2, \dots, p$, since $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$ exist in pairs with respect to $y = 1$ and $y = 0$, $\forall i = 1, 2, \dots, N$ there exists a $i' \in \{1, 2, \dots, N\}$ such that

$x_{i'j} = x_{ij}$ and $y_{i'} = 1 - y_i$. Hence, $x_{ij}y_i + x_{i'j}y_{i'} = x_{ij}(y_i + y_{i'}) = x_{ij} = \frac{1}{2}x_{ij} + \frac{1}{2}x_{i'j}$. Therefore, for $j = 1, 2, \dots, p$,

$$\sum_{i=1}^N x_{ij}y_i = \sum_{i=1}^N x_{ij} \times \frac{1}{2} = \sum_{i=1}^N x_{ij}\pi_i,$$

which are the second to the last equations of (2.2).

■

There is no need to assume data configuration of overlap for perfectly balanced data. Indeed, if otherwise there is quasi-complete separation in the data points, then there is a vector $b = (b_0, b_1, \dots, b_p)^T$ such that

$$\begin{cases} \sum_{j=0}^p b_j x_{ij} = X_i b = b^T X_i^T \geq 0, & y_i = 1, \\ \sum_{j=0}^p b_j x_{ij} = X_i b = b^T X_i^T \leq 0, & y_i = 0. \end{cases}$$

Since each pair of the N observations has one $y = 0$ and $y = 1$, $\sum_{j=0}^p b_j x_{ij} = 0$. Hence, $x_0, x_1, x_2, \dots, x_p$ are linearly dependent, which is contradictory to the assumption that X has a rank of $p + 1$. We also note that perfectly balanced is partially balance.

3.3. Quasi-Saturated data

If $N = 1 + p$, then it is a saturated model by Hosmer et al. (2000). The design matrix X becomes a square matrix. If the square matrix X is singular, then the solution is not unique as in the case of $N <$

$p + 1$. Again, this is not generally acceptable. Let's assume X is nonsingular. Rather than solving the N nonlinear equations (2.2) for $\beta_0, \beta_1, \dots, \beta_p$, we solve 2 sets of linear equations, first for probabilities $\pi_1, \pi_2, \dots, \pi_N$ and then for regression parameters $\beta_0, \beta_1, \dots, \beta_p$. Let us rewrite (2.2) into a matrix-vector format:

$$X^T \begin{pmatrix} \pi_1 \\ \pi_2 \\ \dots \\ \pi_N \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{i1} y_i \\ \dots \\ \sum_{i=1}^N x_{ip} y_i \end{pmatrix}. \quad (3.3)$$

Since X is nonsingular, X^T is too. Then (3.3) has a unique solution $\pi = (\pi_1, \pi_2, \dots, \pi_N)^T$

$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \dots \\ \pi_N \end{pmatrix} = (X^T)^{-1} \times \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{i1} y_i \\ \dots \\ \sum_{i=1}^N x_{ip} y_i \end{pmatrix} \quad (3.4)$$

Next, if $0 < \pi_i < 1$ for all $i = 1, 2, \dots, N$, then

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = \ln \frac{\pi_i}{1 - \pi_i}, i = 1, 2, \dots, N. \quad (3.5)$$

Rewriting the N linear equations in (3.5) about $\beta_0, \beta_1, \dots, \beta_p$ into the matrix-vector format:

$$X \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \ln \frac{\pi_1}{1 - \pi_1} \\ \ln \frac{\pi_2}{1 - \pi_2} \\ \dots \\ \ln \frac{\pi_N}{1 - \pi_N} \end{pmatrix} \quad (3.6)$$

Solving for $\beta_0, \beta_1, \dots, \beta_p$, we obtain

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix} = X^{-1} \times \begin{pmatrix} \ln \frac{\pi_1}{1 - \pi_1} \\ \ln \frac{\pi_2}{1 - \pi_2} \\ \dots \\ \ln \frac{\pi_N}{1 - \pi_N} \end{pmatrix}. \quad (3.7)$$

However, it follows from (2.3) that $(\pi_1, \pi_2, \dots, \pi_N)^T = (y_1, y_2, \dots, y_N)^T$, when X^T is nonsingular. This is a limit case for $0 < \pi_i < 1$ for all $i = 1, 2, \dots, N$. Therefore, if $N = p + 1$, there is no finite solution, not to mention an analytical closed-form solution unlike stated by Allison (2008).

Although a saturated model does not have a finite solution, it gives us an idea: If we can reduce to square matrices in (2.2) then we can solve 2 set of linear equations, first for probabilities and then for

regression parameters, rather than solving nonlinear equations. This motivates us to seek a saturated kind model.

Definition 3.5. The data is quasi-saturated if the N observations fall into $p + 1$ groups such that $(x_1, x_2, \dots, x_p)$ have the same values in each group.

Theorem 3.6. There is an analytical solution if the data is quasi-saturated.

Proof. Instead of solving a set of nonlinear equations, we solve 2 set of linear equations, first for probabilities and then for regression parameters.

Let J_j denote the set of indices of group j and n_j the cardinal number of set J_j , and U_j the probability π_i for group J_j , for $j = 1, 2, \dots, p + 1$ so that $n_1 + n_2 + \dots + n_{p+1} = N$. Let us rearrange the observations to make n_i observations together in group J_j , for $j = 1, 2, \dots, p + 1$ such that

- $J_1 = \{1, 2, \dots, n_1\}$
- $J_2 = \{n_1 + 1, n_1 + 2, \dots, n_1 + n_2\}$
- ...
- $J_p = \{n_1 + n_2 + \dots + n_{p-1} + 1, n_1 + n_2 + \dots + n_{p-1} + 2, \dots, n_1 + n_2 + \dots + n_p\}$
- $J_{p+1} = \{n_1 + n_2 + \dots + n_p + 1, n_1 + n_2 + \dots + n_p + 2, \dots, n_1 + n_2 + \dots + n_{p+1}\}.$

Rewrite (2.2) as

$$\begin{aligned} \sum_{j=1}^{p+1} n_j U_j &= \sum_{i=1}^N y_i \\ \left(\sum_{i \in J_1} x_{i1} \right) U_1 + \left(\sum_{i \in J_2} x_{i1} \right) U_2 + \dots + \left(\sum_{i \in J_{p+1}} x_{i1} \right) U_{p+1} &= \sum_{i=1}^N x_{i1} y_i \\ \dots & \\ \left(\sum_{i \in J_1} x_{ip} \right) U_1 + \left(\sum_{i \in J_2} x_{ip} \right) U_2 + \dots + \left(\sum_{i \in J_{p+1}} x_{ip} \right) U_{p+1} &= \sum_{i=1}^N x_{ip} y_i. \end{aligned} \quad (3.8)$$

Writing into the matrix-vector form

$$\begin{pmatrix} n_1 & n_2 & \dots & n_{p+1} \\ \sum_{i \in J_1} x_{i1} & \sum_{i \in J_2} x_{i1} & \dots & \sum_{i \in J_{p+1}} x_{i1} \\ \dots & \dots & \dots & \dots \\ \sum_{i \in J_1} x_{ip} & \sum_{i \in J_2} x_{ip} & \dots & \sum_{i \in J_{p+1}} x_{ip} \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_{p+1} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{i1} y_i \\ \dots \\ \sum_{i=1}^N x_{ip} y_i \end{pmatrix} \quad (3.9)$$

The coefficient matrix of (3.9) is

$$\begin{aligned}
 & \begin{pmatrix} \sum_{i \in J_1}^{n_1} x_{i1} & \sum_{i \in J_2}^{n_2} x_{i1} & \dots & \sum_{i \in J_{p+1}}^{n_{p+1}} x_{i1} \\ \dots & \dots & \dots & \dots \\ \sum_{i \in J_1} x_{ip} & \sum_{i \in J_2} x_{ip} & \dots & \sum_{i \in J_{p+1}} x_{ip} \end{pmatrix} = \begin{pmatrix} n_1 & n_2 & \dots & n_{p+1} \\ n_1 x_{11} & n_2 x_{n_1+1,1} & \dots & n_{p+1} x_{n_1+\dots+n_p+1,1} \\ \dots & \dots & \dots & \dots \\ n_1 x_{1p} & n_2 x_{n_1+1,p} & \dots & n_{p+1} x_{n_1+\dots+n_p+1,p} \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{n_1+1,1} & \dots & x_{n_1+\dots+n_p+1,1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{n_1+1,p} & \dots & x_{n_1+\dots+n_p+1,p} \end{pmatrix} \begin{pmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & n_{p+1} \end{pmatrix} \\
 &= \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{n_1+1,1} & \dots & x_{n_1+1,p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n_1+\dots+n_p+1,1} & \dots & x_{n_1+\dots+n_p+1,p} \end{pmatrix}^T \begin{pmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & n_{p+1} \end{pmatrix} \\
 &\triangleq X_0^T \begin{pmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & n_{p+1} \end{pmatrix}.
 \end{aligned}$$

Note that

$$X_0 = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{n_1+1,1} & \dots & x_{n_1+1,p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n_1+\dots+n_p+1,1} & \dots & x_{n_1+\dots+n_p+1,p} \end{pmatrix} \quad (3.10)$$

is part of design matrix X . Indeed, if its first row is repeated n_1 times, the second row is repeated n_2 times, ..., the $(p+1)$ -th row is repeated n_{p+1} times, it becomes X . Since the column rank of X is $p+1$, and the column rank equals the row rank in any matrix (Strang, 2006), the square matrix X_0 is nonsingular. Hence, (3.9) is solved

$$\begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_{p+1} \end{pmatrix} = \begin{pmatrix} 1/n_1 & 0 & 0 & 0 \\ 0 & 1/n_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1/n_{p+1} \end{pmatrix} (X_0^{-1})^T \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{i1} y_i \\ \dots \\ \sum_{i=1}^N x_{ip} y_i \end{pmatrix}. \quad (3.11)$$

Next, we set up a set of linear equations for $\beta_0, \beta_1, \dots, \beta_p$ as in (3.6)

$$X_0 \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \ln \frac{U_1}{1-U_1} \\ \ln \frac{U_2}{1-U_2} \\ \dots \\ \ln \frac{U_{p+1}}{1-U_{p+1}} \end{pmatrix}. \quad (3.12)$$

Again, since the square matrix X_0 is nonsingular, $\beta_0, \beta_1, \dots, \beta_p$ can be solved from (3.12) as

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix} = (X_0)^{-1} \begin{pmatrix} \ln \frac{U_1}{1 - U_1} \\ \ln \frac{U_2}{1 - U_2} \\ \dots \\ \ln \frac{U_{p+1}}{1 - U_{p+1}} \end{pmatrix}. \quad (3.13)$$

Hence, (3.13) combined with (3.11) is the analytic closed-form solution for regression parameters $\beta_0, \beta_1, \dots, \beta_p$. This concludes our proof. ■

Remark 3.7. If there is at least one $y = 0$ and at least one $y = 1$ for each group in the saturated data, then there is overlap in the data points. Indeed, if otherwise there is quasi-complete separation in the data points, then there exists a vector $b = (b_0, b_1, \dots, b_p)^T$ such that

$$\begin{cases} \sum_{j=0}^p b_j x_{ij} = X_i b = b^T X_i^T \geq 0, & y_i = 1, \\ \sum_{j=0}^p b_j x_{ij} = X_i b = b^T X_i^T \leq 0, & y_i = 0. \end{cases}$$

Since each group of the N observations has at least one $y = 0$ and at least one $y = 1$, $\sum_{j=0}^p b_j x_{ij} = 0$. Hence, $x_0, x_1, x_2, \dots, x_p$ are linearly dependent, which is contradictory to the assumption that X has a rank of $p + 1$.

4. NUMERICAL METHODS

Lipovetsky (2015) presented 2 cases in which univariate logistic regression has an analytic closed-form solution for $\beta_0, \beta_1, \dots, \beta_p$:

- Case 1. A single dichotomous variable.
- Case 2. A single categorical variable.

Both cases are related to a single independent variable. In particular, Case 1 is an extension of a single binary continuous variable by Allison (2004 and 2008) and Hosmer and Lemeshow (2000).

4.1. Case 1 - A single dichotomous continuous variable

For Case 1, let us consider data with only a single dichotomous continuous variable x with 2 different values a_0 and a_1 . Let us use a 2×2 contingency table (Table 1) of x and y as in Lipovetsky (2015). Then, the model has an analytical closed-form solution

$$\begin{aligned} \beta_0 &= \frac{\ln \frac{(N_{01}/N_{00})^{a_1}}{(N_{11}/N_{10})^{a_0}}}{a_1 - a_0} \\ \beta_1 &= \frac{\ln \frac{N_{11} N_{00}}{N_{10} N_{01}}}{a_1 - a_0}. \end{aligned} \quad (4.1)$$

4.2. Case 2 - A single categorical variable

Assume a categorical variable x has K levels $a_1, a_2, \dots, a_K$. Consider a $K \times 2$ contingency table of counts of all the combinations of x and y as in Lipovetsky (2015).

To convert x into a continuous variable, $(K - 1)$ dummy (or design) variables, denoted by $X_1, X_2, \dots, X_{K-1}$ are needed: $X_i = 1$ if $x = a_i$ and $X_i = 0$ if $x \neq a_i, i = 1, 2, \dots, K - 1$. Then, there is an analytical closed-form solution:

$$\begin{aligned} \beta_0 &= \ln \frac{N_{K1}}{N_{K0}} \\ \beta_j &= \ln \frac{N_{j1}/N_{j0}}{N_{K1}/N_{K0}} = \ln \frac{N_{j1}N_{K0}}{N_{K1}N_{j0}}, j = 1, 2, \dots, K - 1. \end{aligned} \quad (4.2)$$

Remark 4.1. Both Lipovetsky's cases are quasi-saturated. In Case 1, the N observations can be divided into 2 groups: Group 1 when $x = a_0$ and Group 2 when $x = a_1$. In Case 2, the N observations can be divided into K groups: Group j when $x = a_j$ for $j = 1, 2, \dots, K$.

Remark 4.2. Due to denominators and logarithms, (4.1) may not be defined when any of $N_{01}, N_{11}, N_{10}, N_{00}$ are 0. This can be prevented if we assume overlap in the data points. Indeed, if we assume none of $N_{01}, N_{11}, N_{10}, N_{00}$ are 0, then (4.1) is well defined. In this case, each group of $x = a_0$ and $x = a_1$ has at least one $y = 1$ and at least one $y = 0$. By Remark 3.7, there is overlap in the data points.

Remark 4.3. Similarly, (4.2) may not be defined when $N_{10}, N_{20}, \dots, N_{K0}; N_{11}, N_{21}, \dots, N_{K1}$ are 0. This can be prevented if we assume overlap in the data points. Indeed, if we assume none of $N_{10}, N_{20}, \dots, N_{K0}; N_{11}, N_{21}, \dots, N_{K1}$ are 0, then (4.2) is well defined. In this case, each group of when $x = a_j$ for $j = 1, 2, \dots, K$ has at least one $y = 1$ and at least one $y = 0$. By Remark 3.7, there is overlap in the data points.

In the following we will demonstrate how to find the analytical closed-form solution for the 2 known cases in view of Theorem 3.6 and its proof.

4.3. Numerical examples

Example 4.4. For Case 1, referring to Table 1, we have

$$\begin{aligned} n_1 &= N_{01}, n_2 = N_{11}, \\ \sum_{i=1}^N y_i &= N_{11} \\ \sum_{i=1}^N x_i y_i &= a_0 N_{01} + a_1 N_{11} \\ X_0 &= \begin{pmatrix} 1 & a_0 \\ 1 & a_1 \end{pmatrix}. \end{aligned}$$

Denote the probability when $x = a_0$ by U_1 , and the probability when $x = a_1$ by U_2 . The inverse of X_0 is

$$(X_0)^{-1} = \begin{pmatrix} \frac{a_1}{a_1 - a_0} & -\frac{a_0}{a_1 - a_0} \\ -\frac{1}{a_1 - a_0} & \frac{1}{a_1 - a_0} \end{pmatrix} = \frac{1}{a_1 - a_0} \begin{pmatrix} a_1 & -a_0 \\ -1 & 1 \end{pmatrix}.$$

Since $N_{01} + N_{11} = N_{.1}$ from Table 1, it follows from (3.11) that

$$\begin{aligned} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} &= \frac{1}{a_1 - a_0} \begin{pmatrix} \frac{1}{N_{0.}} & 0 \\ 0 & \frac{1}{N_{1.}} \end{pmatrix} \begin{pmatrix} a_1 & -1 \\ -a_0 & 1 \end{pmatrix} \begin{pmatrix} N_{.1} \\ a_0 N_{01} + a_1 N_{11} \end{pmatrix} = \frac{1}{a_1 - a_0} \begin{pmatrix} \frac{1}{N_{0.}} & 0 \\ 0 & \frac{1}{N_{1.}} \end{pmatrix} \begin{pmatrix} a_1(N_{.1} - N_{11}) - a_0 N_{01} \\ a_0(N_{01} - N_{.1}) + a_1 N_{11} \end{pmatrix} \\ &= \frac{1}{a_1 - a_0} \begin{pmatrix} \frac{1}{N_{0.}} & 0 \\ 0 & \frac{1}{N_{1.}} \end{pmatrix} \begin{pmatrix} a_1 N_{01} - a_0 N_{01} \\ -a_0 N_{11} + a_1 N_{11} \end{pmatrix} = \begin{pmatrix} \frac{N_{01}}{N_{0.}} \\ \frac{N_{11}}{N_{1.}} \end{pmatrix}. \end{aligned}$$

Since $N_{00} + N_{01} = N_{0.}$ and $N_{10} + N_{11} = N_{.1}$ from Table 1, it follows from (3.13) that

$$\begin{aligned} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} &= (X_0)^{-1} \begin{pmatrix} \ln \frac{U_1}{1 - U_1} \\ \ln \frac{U_2}{1 - U_2} \end{pmatrix} = (X_0)^{-1} \begin{pmatrix} \ln \frac{N_{01}}{N_{00}} \\ \ln \frac{N_{11}}{N_{10}} \end{pmatrix} = \frac{1}{a_1 - a_0} \begin{pmatrix} a_1 & -a_0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \ln \frac{N_{01}}{N_{00}} \\ \ln \frac{N_{11}}{N_{10}} \end{pmatrix} \\ &= \frac{1}{a_1 - a_0} \begin{pmatrix} \ln \frac{(N_{01}/N_{00})^{a_1}}{(N_{11}/N_{10})^{a_0}} \\ \ln \frac{N_{11} N_{00}}{N_{10} N_{01}} \end{pmatrix} \end{aligned}$$

whose 2 components β_0 and β_1 are identical to β_0 and β_1 in (4.1).

Example 4.5. For Case 2, referring to Table 2, we have

$$n_1 = N_{1.}, n_2 = N_{2.}, \dots, n_K = N_{K.}$$

$$\sum_{i=1}^N y_i = N_{.1}$$

$$\sum_{i=1}^N x_{ij} y_i = N_{j1}, j = 1, 2, \dots, K - 1$$

$$X_0 = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The submatrix of X_0 resulting from removing the first column and the last row is a unit matrix of dimension $K - 1$.

The inverse of X_0 is

$$(X_0)^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & 0 & \dots & -1 \\ 0 & 1 & 0 & 0 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & -1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -1 \end{pmatrix}.$$

The submatrix of $(X_0)^{-1}$ resulting from removing the first row and the last column is a unit matrix of dimension $K - 1$.

$$(X_0^{-1})^T = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & -1 & -1 & -1 & \dots & -1 \end{pmatrix}$$

The submatrix of $(X_0^{-1})^T$ resulting from removing the last row and the first column is a unit matrix of dimension $K - 1$.

Let's denote the probability when $x = a_j$ by U_j for $j = 1, 2, \dots, K$. It follows from (3.11) that

$$\begin{aligned} \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_{K-1} \\ U_K \end{pmatrix} &= \begin{pmatrix} \frac{1}{N_{1\cdot}} & 0 & \dots & 0 \\ 0 & \frac{1}{N_{2\cdot}} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{N_{K\cdot}} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & -1 & -1 & -1 & \dots & -1 \end{pmatrix} \begin{pmatrix} N_{\cdot 1} \\ N_{11} \\ N_{21} \\ \dots \\ N_{K-1,1} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{N_{1\cdot}} & 0 & \dots & 0 \\ 0 & \frac{1}{N_{2\cdot}} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{N_{K\cdot}} \end{pmatrix} \begin{pmatrix} N_{11} \\ N_{21} \\ \dots \\ N_{K-1,1} \\ N_{\cdot 1} - N_{11} - N_{21} - \dots - N_{K-1,1} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{N_{1\cdot}} & 0 & \dots & 0 \\ 0 & \frac{1}{N_{2\cdot}} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{N_{K\cdot}} \end{pmatrix} \begin{pmatrix} N_{11} \\ N_{21} \\ \dots \\ N_{K-1,1} \\ N_{K1} \end{pmatrix} = \begin{pmatrix} \frac{N_{11}}{N_{1\cdot}} \\ \frac{N_{21}}{N_{2\cdot}} \\ \dots \\ \frac{N_{K-1,1}}{N_{K-1\cdot}} \\ \frac{N_{K1}}{N_{K\cdot}} \end{pmatrix}. \end{aligned}$$

Above we have used the fact that $N_{\cdot 1} - N_{11} - N_{21} - \dots - N_{K-1,1} = N_{K1}$ from Table 2.

Since for $j = 1, 2, \dots, K$, we have

$$\ln \frac{U_j}{1 - U_j} = \ln \frac{N_{j1}/N_{j\cdot}}{(N_{j\cdot} - N_{j1})/N_{j\cdot}} = \ln \frac{N_{j1}}{N_{j\cdot} - N_{j1}} = \ln \frac{N_{j1}}{N_{j0}}$$

from Table 2.

By (3.13), we have

$$\begin{aligned}
 \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{K-1} \end{pmatrix} &= (X_0)^{-1} \begin{pmatrix} \ln \frac{U_1}{1-U_1} \\ \ln \frac{U_2}{1-U_2} \\ \dots \\ \ln \frac{U_K}{1-U_K} \end{pmatrix} = (X_0)^{-1} \begin{pmatrix} \ln \frac{N_{11}}{N_{10}} \\ \ln \frac{N_{21}}{N_{20}} \\ \dots \\ \ln \frac{N_{K1}}{N_{K0}} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & 0 & \dots & -1 \\ 0 & 1 & 0 & 0 & \dots & -1 \\ 0 & 0 & 1 & 0 & \dots & -1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -1 \end{pmatrix} \begin{pmatrix} \ln \frac{N_{11}}{N_{10}} \\ \ln \frac{N_{21}}{N_{20}} \\ \dots \\ \ln \frac{N_{K1}}{N_{K0}} \end{pmatrix} \\
 &= \begin{pmatrix} \ln \frac{N_{K1}}{N_{K0}} \\ \ln \frac{N_{11}}{N_{10}} - \ln \frac{N_{K1}}{N_{K0}} \\ \dots \\ \ln \frac{N_{K-1,1}}{N_{K-1,0}} - \ln \frac{N_{K1}}{N_{K0}} \end{pmatrix} = \begin{pmatrix} \ln \frac{N_{K1}}{N_{K0}} \\ \ln \frac{N_{11}N_{K0}}{N_{K1}N_{10}} \\ \dots \\ \ln \frac{N_{K-1,1}N_{K0}}{N_{K1}N_{K-1,0}} \end{pmatrix}
 \end{aligned}$$

whose components $\beta_0, \beta_1, \dots, \beta_p$ are identical to those in (4.2).

Table 1. Contingency 2×2 table for predictor x and target y

	$y = 0$	$y = 1$	Row Totals
$x = a_0$	N_{00}	N_{01}	$N_{0\cdot}$
$x = a_1$	N_{10}	N_{11}	$N_{1\cdot}$
Column Totals	$N_{\cdot 0}$	$N_{\cdot 1}$	N

Table 2. Contingency $K \times 2$ table for predictor x and target y

	$y = 0$	$y = 1$	Row Totals
$x = a_1$	N_{10}	N_{11}	$N_{1\cdot}$
$x = a_2$	N_{20}	N_{21}	$N_{2\cdot}$
4			
...	...	...	...
$x = a_K$	N_{K0}	N_{K1}	$N_{K\cdot}$
Column Totals	$N_{\cdot 0}$	$N_{\cdot 1}$	N

5. CONCLUSIONS

In this paper, we studied the existence of an analytical solution for multivariate logistic regression. We discovered three types of data that yields an analytical solution: (1) Partially balanced data; (2) Perfectly balanced data; and (3) Quasi-saturated data. We showed that the 2 known cases for univariate logistic regression are each a special case to ours with quasi-saturated data. Future research may focus on finding if there are any other types of data that have an analytical solution.

6. REFERENCES

- Albert, A. & Anderson, J. A., 1984, On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71 (1), 1-10.
- Allison, P., 2004, "Convergence Problems in Logistic Regression." pp. 238-252 in Micah Altman, Jeff Gill and Michael P. McDonald (eds.), *Numerical Issues in Statistical Computing for the Social Scientist*. Hoboken, NJ: John Wiley & Sons.

Allison, P., 2008, Convergence Failures in Logistic Regression, presented at SAS Global Forum., San Antonio, Texas, Paper 360-2008.

Habshah, M., Ariffin, S. B., Imon, A. H. M. R., 2016, Robust Bootstrap Procedure for Estimation of Binary Logistic Regression Model in the Presence of High Leverage Points with Medical Applications. *International Journal of Applied Mathematics & Statistics*. **54** (1), 10-32.

Hosmer, D. W., Lemeshow, 2000, *Applied Logistic Regression* (2nd edition). New York: John Wiley & Sons, Inc.

Jonathan M. Borwein and Richard E. Crandall, January 2013, Closed Forms: What They Are and Why We Care. *Notices of the American Mathematical Society*, **60**(1), 50–65.

Lipovetsky, S., 2015, Analytical closed-form solution for binary logit regression by categorical predictors. *Journal of Applied Statistics* **42**(1), 37-49.

Owen, A. B., and Roediger, P. A., 2014, The Sign of the Logistic Regression Coefficient. *The American Statistician*, **68**(4), 297-301.

Refaat, M., 2011, *Credit Risk Scorecards: Development and Implementation Using SAS*. Raleigh, North Carolina, USA: LULU.COM.

Siddiqi, N., 2006, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Strang, A., 2006, *Linear Algebra and Its Applications* (4th edition). Belmont, CA: Thomson Brooks/Cole.

Wang, L., Zhu, W., 2015, Logistic Regression Model for Cognitive Biases in Financial Analysts Recommendations. *International Journal of Applied Mathematics & Statistics*. **53**(3), 111-119.

Zeng, G., 2013, Metric divergence measures and information value in credit scoring. *Journal of Mathematics*, vol. 2013, Article ID 848271. doi: 10.1155/2013/848271.

Zeng, G., 2014, A rule of thumb for reject inference in credit scoring. *Mathematical Finance Letters*, vol. 2014, article 2, 1-13.

Zeng, G., 2015, A unified definition of mutual information with applications in machine learning. *Mathematical Problems in Engineering*, vol. 2015, Article ID 201874. doi: 10.1155/2015/201874.

Zeng, G., 2017a, A comparison Study of Computational Methods of Kolmogorov–Smirnov Statistic in Credit Scoring. *Communications in Statistics: Simulation and Computation*. **46**(10), 7744-7760. doi: 10.1080/03610918.2016.1249883.

Zeng, G., 2017b, Invariant Properties of Logistic Regression Model in Credit Scoring under Monotonic Transformations. *Communications in Statistics: Theory and Methods*. **46**(17), 8791-8807. doi: 10.1080/03610926.2016.1193200.

Zeng, G., 2017c, On the Existence of Maximum Likelihood Estimates for Weighted Logistic Regression. *Communications in Statistics: Theory and Methods*. **46**(22), 11194-11203. doi: 10.1080/03610926.2016.1260743

Zeng, G., and Zeng, E., 2018, On the Three-way Equivalence of AUC in Credit Scoring with Tied Scores. *Communications in Statistics: Theory and Methods*. **28**(7), 1635-1650. doi: 10.1080/03610926.2018.1435814.

Zeng, G., 2019a, On the Confusion Matrix in Credit Scoring and Its Analytical Properties. *Communications in Statistics: Theory and Method*. doi: 10.1080/03610926.2019.1568485.

Zeng, G. and Zeng, E., 2019b, On the Relationship between Multicollinearity and Separation in Logistic Regression. *Communications in Statistics: Simulation and Computation*. doi: 10.1080/03610918.2019.1589511.