Commentary

Logistic regression without intercept

Abstract

In this paper, for the first time we study logistic regression without intercept. We first derive the maximum likelihood estimate for logistic regression without intercept and its numerical solution. We then extend the definition of multicollinearity for logistic regression with intercept to the case of without intercept. Next, we extend three types of data configuration (complete separation, quasi-complete separation and overlap) for logistic regression with intercept to the case of without intercept, and characterize the existence and uniqueness of maximum likelihood estimate. In addition, we discuss linear transformations and monotonic transformations of independent variables in logistic regression without intercept. Our numerical examples further compare logistic regression with intercept and without intercept.

1.   Introduction

Logistic regression and linear regression are two most commonly used regressions for modeling the relationship between the dependent variable and one or more independent variables. By default, both linear regression and logistic regression have an intercept which is a constant term with regard to the coefficients of independent variables. Albeit controversial, linear regression without intercept has been studied over the past 2 decades (Kozak & Kozak, 1995; Eisenhauer, 2003; Cincotta, 2011; Othman 2014). However, logistic regression without intercept is rarely studied in the literature, although it has been used to study ridge logistic regression (Lee and Silvapulle, 1988; Duffy and Santner, 1989; Le Cessie & Van Houwelingen, 1992; Tian, et al., 2009; and Zhang, Zhang & Qiu, 2015). In practice, there are circumstances in which logistic regression without intercept is appropriate or even necessary. For instance, logistic regression without intercept can be used in conditional logistic regression with matched pairs data ("SAS User Guide", 2013).

In this paper, we study logistic regression without intercept when it is appropriate. Specifically, we derive maximum likelihood estimate (MLE) and its numerical solution, multicollinearity and

its detection, characterization of the existence and uniqueness of MLE by data configuration, and linear transformations of independent variables.

The rest of the paper is organized as follows. In Section 2, we introduce logistic regression without intercept, its MLE and numerical solution. In Section 3, multicollinearity is discussed for logistic regression without intercept. In Section 4, three types of data configuration are defined for logistic regression without intercept and used to characterize the existence and uniqueness of MLE by data configuration. Linear transformations and monotonic transformations of independent variables are also discussed. Numerical examples are provided in Section 6. In particular, we show an example which has multicollinearity for logistic regression with intercept but not for the case of without intercept. We also show an example which has separation data configuration for logistic regression with intercept but not for the case of without intercept. Finally, the paper is concluded in Section 7.

Throughout the paper, by logistic regression, we mean logistic regression with intercept unless otherwise specified. We first present the results for logistic regression with intercept and then extend the results to logistic regression without intercept.

2. Maximum likelihood estimate

Let us first introduce maximum likelihood estimate (MLE) in logistic regression with intercept. Let $x = (x_1, x_2, \ldots, x_p)$ be the vector of p independent variables and y the binary dependent variable (also called response or target) with values of 0 and 1. Without loss of generality, we assume $x_1, x_2, \ldots, x_p$ do not have missing values. Assume we have a sample of N independent observations $(x_{i1}, x_{i2}, \ldots, x_{ip}, y_i), i = 1, 2, \ldots, N,$ where y_i is the value of y and $x_{i1}, x_{i2}, \ldots, x_{ip}$ are the values of functions $x_1, x_2, \ldots, x_p$ for the i-th observation, respectively.

Let X_i be the row vector $(1, x_{i1}, x_{i2}, \ldots, x_{ip})$ for $i = 1, 2, \ldots, N$, and denote X the $N \times (p + 1)$ matrix, called Design Matrix, with X_i as rows. To accommodate the intercept, let us denote the N-dimensional column vector with all elements of 1 by x_0. For simplicity of notation, we also let $x_1, x_2, \ldots, x_p$ stand for the N-dimensional vectors

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

respectively, when no confusion can arise.

We use $\pi(x) = P(y = 1|x)$ to represent the conditional probability that y is equal to 1 given x.

Then the logistic regression model is given by the equation

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}. \tag{2.1}$$

Note that (2.1) is in the function form. It can be understood in the vector form (dimension N) with i-th component

$$\pi(X_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}}}$$

for $i = 1, 2, \ldots, N$.

Equivalently, the logit transformation of $\pi(x)$ in (2.1) is

$$g(x) = \ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p. \tag{2.2}$$

The log likelihood of logistic regression is defined by

$$l(\beta) = \sum_{i=1}^{N} y_i \ln\left(\frac{e^{X_i \beta}}{1 + e^{X_i \beta}}\right) + \sum_{i=1}^{N} (1 - y_i) \ln\left(\frac{1}{1 + e^{X_i \beta}}\right). \tag{2.3}$$

The value of β that maximizes $l(\beta)$ in (2.3) is called the maximum likelihood estimate (MLE). In particular, if $\beta_1 = \beta_2 = \cdots \beta_p = 0$, then it is called the intercept-only model or null model or base model. In this case, no independent variables are considered and $\beta_0 = \ln\left(\frac{\bar{y}}{1 - \bar{y}}\right)$, where $\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$.

To maximizing function $l(\beta)$ in (2.3), we differentiate it with respect to β, set the derivative equal to 0. We obtain

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

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

(2.4)

as in Hosmer and Lemeshow (2000), where $\pi_i = \pi(x_{i1}, x_{i2}, \dots, x_{ip}) = \frac{e^{X_i\beta}}{1+e^{X_i\beta}}$ for $i = 1, 2, \dots, N$.

In general, (2.4) cannot be solved analytically (Zeng, 2021). Rather, they can be solved numerically by Newton-Raphson algorithm (See Refaat 2011) as follows

$$\beta^{(i+1)} = \beta^{(i)} + (X^TVX)^{-1}g,$$

(2.5)

where V is the $N \times N$ diagonal matrix with its diagonal elements $\pi_1(1-\pi_1), \pi_2(1-\pi_2), \dots, \pi_N(1-\pi_N)$. In addition, $g = X^T(y-\pi)$ with $\pi = (\pi_1, \pi_2, \dots, \pi_N)^T$. Both V and g are evaluated at $\beta^{(i)}$ in (2.5).

2.1. Without intercept

If the intercept is forced out from logistic regression, the parameter vector β will become $(\beta_1, \dots \beta_p)^T$. Vector $x_0 = (1,1, \dots 1)^T$ for the intercept is not needed. Accordingly, row vectors $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ for $i = 1, 2, \dots, N$. Moreover, the dimension of design matrix X will be $N \times p$. Equations (2.1) , (2.2) and (2.3) hold after β_0 is taken out. Again the value of β that maximizes the log likelihood $l(\beta)$ is called the maximum likelihood estimate (MLE).

To maximizing function $l(\beta)$, we differentiate it with respect to β, set the derivative equal to 0, and then solve the resulting set of equations. We obtain

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

(2.6)

.

Clearly, if the original logistic regression model with intercept has a zero intercept $\beta_0 = 0$, logistic regression without intercept will have the same results as logistic regression with intercept (Refer to Example 6.3 in Section 6).

To get a numerical solution to the p simultaneous nonlinear equations of β in (2.6), Newton-Raphson in the form of (2.5) can be used.

3. Multicollinearity

Perfect Multicollinearity or Complete Multicollinearity or Multicollinearity in short refers to a situation in logistic regression in which two or more independent variables are linearly related (Farrar & Glauber, 1967; Zeng, 2019b; Voss, n.d.). Mathematically, multicollinearity means there exist constant $a_0, a_1, \dots, a_p$ such that

$$a_0 x_0 + \sum_{i=1}^{p} a_i x_i = 0 \tag{3.1}$$

where at least two of $a_1, \dots, a_p$ are nonzero. If we treat x_0 as an independent variable, then we just require at least one of $a_1, \dots, a_p$ is nonzero.

In particular, if two independent variables are linearly related, then it is called collinearity. Some authors use near multicollinarity (Chatelain, Jean-Bernard and Ralf, 2014) and partial multicollinearity (Voss, n.d.) when there is an approximate linear relationship among two or more independent variables.

Some authors define multi collinearity to be high correlation between independent variables (Vatcheva et al.; 2016 and references therein; Cincotta, 2011; Dohoo, et al., 1997). Correlation indicates the linear relationship between a pair of independent variables. It can be measured by the well-known Pearson correlation coefficient (Bewick, et al., 2003). The perfect linear relationship is indicated by a correlation coefficient of -1 or 1.

If there is multicollinearity, the design matrix X will not have a full column rank of $p + 1$. Hence, the $(p + 1) \times (p + 1)$ matrix $\hat{I} = X^T V X$ in (2.5) will have a rank less than $p + 1$. Thus, the inverse matrix $\hat{I}^{-1}$ in (2.5) does not exist, which make the iteration in (2.5) impossible.

If there is near multicollinearity and there is no separation of the data points, theoretically $\hat{I} = X^T V X$ in (2.5) has an inverse and the iteration in (2.5) can be proceeded. Yet, iteration (2.5) may not find an approximate inverse $\hat{I} = X^T V X$ and hence may cause unstable estimates and inaccurate variances (Shen and Gao, 2008).

Multicollinearity can be easily detected by using variance inflation factor (VIF). Specifically, the VIF is calculated for each independent variable by doing a linear regression of that independent variable on all the other independent variables, and then obtaining the R^2 (the coefficient of determination) from the regression. The VIF for this independent variable is just $1/(1-R^2)$. The larger the VIF of an independent variable, the larger correlation between this independent variable and others. There is no standard for acceptable levels of VIF.

Multicollinearity can be also detected by correlation coefficient. However, correlation coefficient is used between two independent variables, whereas VIF can be used to check the linear relationship of one independent variable with all other independent variables. In this sense, VIF is preferred.

3.1. Without Intercept

For logistic regression without intercept, we modify the definition of multicollinearity in (3.1) by removing constant a_0. Clearly, if there exists multicollinearity in logistic regression without intercept, there exists multicollinearity in logistic regression with intercept too. Example 6.4 in Section 6 shows that we cannot keep constant a_0 in (3.1) in the case of without intercept.

To calculate VIF, we do a linear regression without intercept for each independent variable on all the other independent variables. We then adjust R^2 (Eisenhauer, 2003) as

$$R^2(\beta_j) = \frac{\sum_{i=1}^{N} \hat{x}_{ij}^2}{\sum_{i=1}^{N} x_{ij}^2}, j = 1,2, \dots, p$$

where $\hat{x}_{ij}$ is the is the i-th fitted value. Finally, the VIF for x_j will be $\frac{1}{\left(1-R^2(\beta_j)\right)}$. Note that correlation coefficient does not necessarily indicate multicollinearity (Cincotta, 2011) in the case of without intercept.

4. Data configuration

Albert and Anderson (1984) first assumed design matrix X to have a full column rank, that is, no multicollinearity. He then introduced the concept of separation (including complete separation and quasi-complete separation) and overlap in logistic regression with intercept. He showed that separation leads to nonexistence of (finite) MLE and that overlap leads to finite and unique MLE.

With slight modifications to the definition of pseudo-complete separation, the concept of separation can be applied to and design matrices. Of course, the assumption of full column rank will need to be added to overlap in order for MLE to exist and to be unique.

Definition 4.1. There is *quasi-complete separation* if the data are not complete separable, but there exists a non-zero vector $b = (b_0, b_1, \ldots, b_p)^T$ such that

$$
\begin{cases}
\sum_{j=0}^{p} b_j x_{ij} = X_i b = b^T X_i^T \geq 0, & y_i = 1, \\
\sum_{j=0}^{p} b_j x_{ij} = X_i b = b^T X_i^T \leq 0, & y_i = 0
\end{cases}
\tag{4.1}
$$

and equality holds for at least one subject in each response group but not for all $i = 1, 2, \ldots, N$.

Remark 4.2. Zeng (2017b) pointed out variants of complete separation and quasi-complete separation.

4.1. Without Intercept

In the case of without intercept, we adjust the definition of complete separation in Albert and Anderson (1984) and Definition 4.1 above by removing constant b_0 from $b = (b_0, b_1, \ldots, b_p)^T$ and changing the lower limit from $j = 0$ to $j = 1$ in the summations. The definition of overlap remains the same.

Clearly, if there is a separation (complete or pseudo-complete) in logistic regression without intercept, there will be a separation in logistic regression with intercept.

Similar to Albert and Anderson (1984), we can completely characterize the existence and uniqueness of MLE in logistic regression without intercept as follows.

Theorem 4.3. For logistic regression without intercept,

    (i)   if there is complete separation, then MLE does not exist;
    (ii)   if there is quasi-complete separation, then MLE does not exist;
    (iii)   If there is overlap and X has a full column rank of p, then MLE exists and is unique.

In the following, we will prove (ii). As to (i) and (iii), they can be proved similarly to those in Zeng (2017c) for weighted logistic regression by taking all weights = 1.

Proof of Theorem 4.3. (ii). Assume $b = (b_1, b_2, \dots, b_p)^T$ satisfies (4.1). Let $\beta(k) = kb$ for $k > 0$. Then

$$
\begin{aligned}
l(\beta(k)) \\
&= \sum_{i=1}^{r} ln\left(\frac{e^{kX_i b}}{1 + e^{kX_i b}}\right) + \sum_{i=r+1}^{n} ln\left(\frac{1}{1 + e^{kX_i b}}\right) \\
&= \sum_{i=1}^{r} ln\left(1 - \frac{1}{1 + e^{kX_i b}}\right) + \sum_{i=r+1}^{n} ln\left(\frac{1}{1 + e^{kX_i b}}\right).
\end{aligned}
$$

Since there is at least one i with $1 \leq i \leq r$ such that $X_i b > 0$ or one i with $r + 1 \leq i \leq N$ such that $X_i b < 0$. If $X_i b = 0$ for $1 \leq i \leq N$, both $ln\left(1 - \frac{1}{1 + e^{kX_i b}}\right)$ and $ln\left(\frac{1}{1 + e^{kX_i b}}\right)$ have a constant value $- ln(2)$. If $X_i b > 0$ for $1 \leq i \leq r$, then $ln\left(1 - \frac{1}{1 + e^{kX_i b}}\right)$ is strictly increasing in k. If $X_i b > 0$ for $r + 1 \leq i \leq N$, then $ln\left(\frac{1}{1 + e^{kX_i b}}\right)$ is strictly increasing in k.

Therefore,

$$l(\beta) = l(\beta(k = 1)) < L(\beta(2)) < \cdots \leq \lim_{k \to \infty} l(\beta(k)).$$

The maximum likelihood estimate & is thus at infinity on the boundary of R^p. ∎

5.   Transformations

Sometimes, linear transformation is used to transfer one independent variable to a new independent variable. For instance, linear transformation can be used for the purpose of scale. Note that we don't consider the trivia case – the constant transformation, as a variable with a constant value will not provide any information (Zeng, 2013). If coefficient β for variable x is too small, we may make the coefficient larger by making a linear transformation $x' = \frac{x}{\alpha}$ with $\alpha > 1$, without changing the model. For instance, if $\alpha = 100$, then the coefficient with transformed variable x' will be 100 times as large as β.

The following result shows that linear transformations do not change logistic regression model.

Theorem 5.1. Logistic regression with intercept is invariant under non-constant linear transformations of independent variables.

Proof. Assume we make linear transformations for k independent variables, where $k \leq p$. Without loss of generality, we assume $x_1, x_2, \ldots, x_k$ are such k independent variables. Let the k linear transformations be $x_i' = a_i x_i + b_i$, where $a_i \neq 0$ for $i = 1, 2, \ldots, k$. Then

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

$$= \beta_0 + \beta_1 \left(\frac{1}{a_1} x_1' - \frac{1}{b_1}\right) + \beta_2 \left(\frac{1}{a_2} x_2' - \frac{1}{b_2}\right) + \cdots + \beta_k \left(\frac{1}{a_k} x_k' - \frac{1}{b_k}\right) + \beta_{k+1} x_{k+1} + \cdots + \beta_p x_p$$

$$= \beta_0 - \left(\frac{\beta_1}{b_1} + \frac{\beta_2}{b_2} + \cdots + \frac{\beta_k}{b_k}\right) + \frac{\beta_1}{a_1} x_1' + \frac{\beta_2}{a_2} x_2' + \cdots + \frac{\beta_k}{a_k} x_k' + \beta_{k+1} x_{k+1} + \cdots + \beta_p x_p.$$

By Albert (1984), the MLE is unique if it exists. If $\beta_0, \beta_1, \ldots \beta_p$ is the MLE with independent variables $x_1, x_2, \ldots, x_p$, then $\beta_0 - \left(\frac{\beta_1}{b_1} + \frac{\beta_2}{b_2} + \cdots + \frac{\beta_k}{b_k}\right), \frac{\beta_1}{a_1}, \frac{\beta_2}{a_2}, \ldots, \frac{\beta_k}{a_k}, \beta_{k+1}, \ldots, \beta_p$ is the MLE with independent variables $x_1', x_2', \ldots, x_k', x_{k+1}, \ldots, x_p$. Moreover, the probabilities $\pi(X_i), i = 1, 2, \ldots, N$ are identical with independent variables $x_1', x_2', \ldots, x_k', x_{k+1}, \ldots, x_p$ are identical to those with independent variables $x_1, x_2, \ldots, x_p$. ∎

Similarly, logistic regression without intercept is invariant under linear transformations without constant.

Theorem 5.2. Logistic regression without intercept is invariant under linear transformations without constant of independent variables.

From the proof of Theorem 5.1, we see that logistic regression without intercept is not invariant under linear transformations with constant unless $\beta_0 = \left(\frac{\beta_1}{b_1} + \frac{\beta_2}{b_2} + \cdots + \frac{\beta_k}{b_k}\right)$. See Example 6.2. in Section 6.

5.1. Monotonic Transformations

Zeng (2017b) showed that MLE, information value (Zeng, 2013), mutual information (Zeng, 2015), Kolmogorov–Smirnov (KS) (2017a) statistics regarding a univariate logistic regression are all invariant under monotonic transformations. Following the same arguments, we can prove that these properties hold for univariate logistic regression without intercept.

6. Numerical examples

We use the statistical software package R (version 3.4.2) and its RStudio for our numerical examples.

We use the German Credit Data from a German bank. They are hosted by the UCI Machine Learning Repository (Lichman, 2013). The Germany data contains 1000 observations for 1000 loan applicants. The dependent variable is credit_status: 1 for good loans and 2 for bad loans. For the purpose of logistic regression, we define a new variable called default as default = credit_status – 1. With the new variable default, 0 is for good loans and 1 is for bad loans.

We will use the 3 following continuous independent variables to run logistic regression:

- duration: Duration in month
- age: Age in years
- num_credits: Nmber of existing credits at this bank.

As we have only 1000 observations, we don't split the data into training and test but use all the 1000 observations. Let's name the Germany Credit Data germany_credit. We italicize the text for the R code and use '>' as prompt and '#' for comments. Since we are not concentrating on model development, we shall not pay attention to the significance (P-value) of the independent variables.

Example 6.1.

Let's call the original logistic regression model with intercept by Model 1.

> *model1 <- glm (default ~ age + duration + num_credits, data = german_credit, family = binomial)*

The coefficients can be found by coefficients(model1) as follows:

```
(Intercept)          age      duration   num_credits
-0.87562523  -0.01738049  0.03747538  -0.12979943
```

Now let's predict the probabilities of the first 5 rows.

> *predict(model1, newdata = german_credit[1:5,], type='response')*

```
        1          2          3          4          5
0.1115639  0.6013331  0.1966501  0.4468069  0.2392163
```

Now let's force no intercept. This can be done by subtracting 1 from or adding 0 to the list of independent variables as follows. Let's call the logistic regression model without intercept by Model 2.

> *model2 <- glm (default ~ age + duration + num_credits - 1, data = german_credit, family = binomial)*

The coefficients are as follows:

```
        age      duration   num_credits
-0.03034242   0.03126543  -0.31047647
```

Now let's predict the probabilities for the first 5 records.

> *predict(model2, newdata = german_credit[1:5,], type='response')*

```
          1            2            3            4            5
0.07825452   0.62778616   0.19433730   0.41029713   0.18561920
```

We see that none of the first 5 records have the same probability as logistic regression with intercept. Moreover, the order of the probabilities has changed. The 5th record has the 4th largest probability in the model without intercept, whereas the 3rd record has the 3rd largest probability in the model with intercept. Indeed, KS and ACU of these 2 models are different as we can see below.

We apply a user-defined R function called KS_AUC (see the appendix).
> *predicted1<- predict(model1, newdata = german_credit, type='response')*

> predicted1 = cbind(german_credit, predicted1)

> predicted1$score = round（1000 * (1 - predicted1$predicted))

> KS_AUC(predicted1, score, default)

KS = 0.207142857142857

Area under the curve: 0.642

> predicted2 <- predict(model2, newdata = german_credit, type='response')

> predicted2 = cbind(german_credit, predicted2)

> predicted2$score = round（1000 * (1 - predicted2$predicted))

> KS_AUC(predicted2, score, default)

KS = 0.192857142857143

Area under the curve: 0.633

We conclude that the logistic regression model without intercept is different from the logistic regression model with intercept.

Example 6.2

Let's make a linear transformation for age, force no intercept and call the logistic regression model by Model 3.

> german_credit$age1 = 2 * german_credit$age - 1

> model3 <- glm (default ~ age1 + duration + num_credits - 1, data = german_credit, family = binomial)

The coefficients are

```
      age1      duration   num_credits
-0.01522484   0.03110889   -0.31561646
```

Now let's predict the probabilities of the first 5 rows.

```
> predict(model3, newdata = german_credit[1:5, ], type='response')
         1            2            3            4            5
0.07802539  0.62783546   0.19479968  0.40997857  0.18494037
```

They are different from but closed to those in model2.

Example 6.3.

We can also eliminate intercept from the original logistic regression model with intercept by absorbing it by making a transformation. Let's call this model by Model 4.

$$age1 \; = \; age \; + \; \beta_0/\beta_1 \; = \; age \; + (-0.87562523)/(-0.01738049).$$

> german_credit$age1 = german_credit$age + (−0.87562523)/(−0.01738049)

> model4 <- glm (default ~ age1 + duration + num_credits, data = german_credit, family = binomial)

To eliminate scientific notation.
> options(scipen=999)

We see that the intercept is almost 0. It is not exactly 0 due to calculation error or precision.

```
(Intercept)                      age1         duration        num_credits
 0.0000001190512  -0.0173804922925   0.0374753807599  -0.1297994307111
```

Let's do logistic regression without intercept and call it by Model 5.

> model5 <- glm (default ~ age1 + duration + num_credits - 1, data = german_credit, family = binomial)

The coefficients are as follows:

```
       age1       duration   num_credits
-0.01738049   0.03747538  -0.12979943
```

We see that coefficients are the same as Model 4. This is because Model 4 has an intercept of 0.

Example 6.4.

> german_credit$age1 = 2 * german_credit$age - 1
> model6 <- glm (default ~ age1 + age + duration + num_credits, data = german_credit, family = binomial)

The coefficients are
```
(Intercept)                age1   age        duration    num_credits
-0.884315473  -0.008690246  NA   0.037475381  -0.129799431
```

We see that the coefficient for age is missing. R automatically keeps only one of age1 and age as they are collinear.

> model7 <- glm (default ~ age1 + age + duration + num_credits - 1, data = german_credit, family = binomial)

The coefficients are

```
        age1           age      duration  num_credits
 0.87562523  -1.76863095   0.03747538  -0.12979943
```

Both age and age1 are included in the model. They are not collinear as age1 and age have a linear relationship with constant. Hence, we have an example which has multicollinearity in logistic regression with intercept but not in logistic regression without intercept.

Example 6.5.

Let's define a new variable x as follows: For $i = 1, 2, \dots, 1000$,

$$x_i = \begin{cases} \min\limits_{y_i=1}(age_i) + 50, & if\ y_i = 1 \\ \max\limits_{y_i=0}(age_i) + 50 & if\ y_i = 0. \end{cases}$$

Then for $i = 1, 2, \dots, 1000$,

$$\begin{cases} 50 + age_i - x_i \geq 0,\ y_i = 1; \\ 50 + age_i - x_i \leq 0,\ \ y_i = 0. \end{cases}$$

Moreover, $50 + age_i - x_i = 0$ when age_i attains the minimum value among those i's such that $y_i = 1$, and $50 + age_i - x_i = 0$ when age_i attains the maximum value among those i's such that $y_i = 0$. Hence, quasi-complete separation exists for both logistic regression with intercept and without intercept as we can see below.

> german_credit$x<- ifelse(german_credit$default == 1,min(german_credit$age[german_credit$default==1] + 50),max(german_credit$age[german_credit$default == 0] + 51))

> *model9 <- glm (default ~ x + age + duration + num_credits, data = german_credit, family = binomial)*

Warning message:
glm.fit: algorithm did not converge

However, overlap exists for logistic regression without intercept.

> model10 <- glm (default ~ x + age + duration + num_credits - 1, data = german_credit, family = binomial)

Hence, we have an example which has data configuration of separation in logistic regression with intercept but not in logistic regression without intercept. The reason is that quasi-complete separation does not exist for logistic regression without intercept as constant 50 is included.

7. Conclusions

In this paper, we have studied logistic regression without intercept. We first derived the maximum likelihood estimate (MLE) for logistic regression without intercept and its numerical solution. We then extended the definition of multicollinearity for logistic regression with intercept to the case of without intercept. Next, we extended three types of data configuration (complete separation, quasi-complete separation and overlap) for logistic regression with intercept to the case of without intercept, and characterized the existence and uniqueness of MLE. In addition, we discussed linear transformations and monotonic transformations of independent variables in logistic regression without intercept. Our numerical examples further compared logistic regression with intercept and without intercept.

Appendix

```
KS_AUC <- function(df, score, target) {
  score <- deparse(substitute(score))
  target <- deparse(substitute(target))
  sample1 <- df[[score]][which(df[[target]] == 0)]
  sample2 <- df[[score]][which(df[[target]] == 1)]
  cdf1 <- ecdf(sample1)
  cdf2 <- ecdf(sample2)
  minMax <- seq(min(sample1, sample2), max(sample1, sample2), length.out = length(sample2))
  x0 <- minMax[which(abs(cdf1(minMax) - cdf2(minMax)) == max(abs(cdf1(minMax) -
            cdf2(minMax))))]

  y0 <- cdf1(x0)
  y1 <- cdf2(x0)
  KS <- y1 - y0
  message(paste('KS = ', KS))

  # Install pROC package if not.
  # install.packages("pROC")
  library(pROC)
  auc(roc(df[[target]], df[[score]]))
}
```

References

Breslow, N.E. & Day, N.E. (1980). *Statistical Methods in Cancer Research. Volume 1-The Analysis of Case-Control Studies*. Lyon, France: IARC. 249–251.

Bewick, V., Cheek, L., & Ball, J (2003). Statistics review 7: Correlation and regression: Presenting and summarizing data. *Critical Care,* 7:451-459.

Le Cessie, S. & van Houwelingen, J. C. (1992). Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 41(1): 191-201.

Chatelain, J. B. & Ralf, K. (2014). Spurious Regressions and Near- Multicollinearity, with an Application to Aid, Policies and Growth. *Journal of Macroeconomics*, 39(PA): 85-96.

Cincotta, K. (2011). Multicollinearity in Zero Intercept Regression: They Are *Not* Who We Thought They Were. *Presented at the Society of Cost Estimating and Analysis (SCEA) Conference.* June, 2011, Albuquerque, NM.

Dohoo, I. R., Ducrot C., Fourichon C., Donald A., & Hurnik D. (1997). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive Veterinary Medicine,* 29(3): 221–239.

Duffy. D. E. & Santner, T. J. (1989). On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models. *Communications in Statistics - Theory and Methods*,18(3): 959-980, DOI: 10.1080/03610928908829944

Eisenhauer, J. G. (2003). Regression through the Origin. *Teaching Statistics*, 25(3): 76-80.

Farrar, D. & Glauber, R. (1967). Multicollmearity in Regression Analysts the Problem Revisited, *Review of Economics and Statistics* (February 1967), 92-107.

Kozak, A. & Kozak, R. A. (1995). Notes on regression through the origin. *THE FORESTRY CHRONICLE*, 71(3): 326-330.

Lee, A. H. & Silvapulle, M. J. (1988). Ridge estimation in logistic regression, *Communications in Statistics - Simulation and Computation*, 17(4): 1231-1257, DOI: 10.1080/03610918808812723

Othman, S. A. (2014). Comparison between Models With and Without Intercept. *Gen. Math. Notes,* 21(1): 118-127.

SAS Institute Inc. 2013. *SAS/STAT 13.1 User's Guide*. Cary, NC: SAS Institute Inc.

Shen, J. & Gao, S. (2008). A Solution to Separation and Multicollinearity in Multiple Logistic Regression. *Journal of Data Science,* 6: 515-531.

Tian, G.L, Fang, H.B, Liu, Z.Q, & Ming T. Tan (2009). Regularized (bridge) logistic regression for variable selection based on ROC criterion. *Statistics and Its Inference, 2:* 493-502.

Vatcheva, K. P., Lee, M., McCormick, J. B, & Rahbar˒ M. H. (2016). Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology* (Sunnyvale), 6(2): 227.

Voss, D. S. (n.d.) Multicollinearity. In The Encyclopedia of Social Measurement, ed. Kempf-Leonard, Kimberly. San Diego, CA: Academic Press.

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. (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 - Theory and Methods,* 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. & Zeng, E. (2019a). On the Three-way Equivalence of AUC in Credit Scoring with Tied Scores. *Communications in Statistics - Theory and Methods,* 48(7): 1635-1650. doi: 10.1080/03610926.2018.1435814.

Zeng, G. & 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.

Zeng, G. (2020). On the Confusion Matrix in Credit Scoring and Its Analytical Properties. *Communications in Statistics - Theory and Methods,* 49(9): 2080-2093. doi: 10.1080/03610926.2019.1568485.

Zeng, G. (2021). On the Existence of an Analytical Solution in Multiple Logistic Regression. *International journal of applied mathematics & statistics*, 60(1): 53-67.

Zeng, G. (2021). On the analytical properties of category encodings in logistic regression. *Communications in Statistics - Theory and Methods.* doi: 10.1080/03610926.2021.1939382