Islam, *Cogent Mathematics & Statistics* (2018), 5: 1472519
https://doi.org/10.1080/25742558.2018.1472519

cogent mathematics & statistics

Received: 24 July 2017
Accepted: 01 May 2018
First Published: 14 May 2018

*Corresponding author: M. Ataharul Islam, ISRT (Institute of Statistical Research and Training), University of Dhaka, Dhaka-1000, Bangladesh
E-mail: mataharul@yahoo.com

Reviewing editor:
Guohua Zou, Chinese Academy of Sciences, China

Additional information is available at the end of the article

STATISTICS | RESEARCH ARTICLE

A trivariate Bernoulli regression model

M. Ataharul Islam[1*]

Abstract: A trivariate Bernoulli regression model is proposed in this paper. There is extensive need for analysing repeated binary outcomes where correlated binary outcomes are obtained from repeated measures or longitudinal data. The proposed model is based on marginal and conditional probabilities as functions of covariates. The estimation and test procedures are shown. The tests include testing of hypotheses for first- and second-order associations among outcome variables. It is noteworthy that the test procedures for dependence in outcome variables can be demonstrated in terms of vectors of regression parameters for models on outcome variables. The proposed model can be extended for more than three correlated outcome variables conveniently.

Subjects: Science; Mathematics and Statistics; Statistics and Probability; Statistics; Statistical Modeling; Longitudinal Data; Statistics & Probability; Probability; Stochastic Models & Processes; Mathematical Statistics; Multivariate Statistics
Keywords: trivariate Bernoulli; regression model; repeated measures; marginal probability; conditional probability; joint probability; correlated outcomes

1. Introduction

The use of Bernoulli regression model is well established. The logistic regression model has been one of the most extensively used techniques in various applications which is based on univariate Bernoulli distribution. Since the development of generalized linear models, it has been presented more explicitly using the logit link function from exponential family of distributions for binary data. There have been attempts to develop regression models for bivariate and multivariate Bernoulli regression models. Some

ABOUT THE AUTHOR

M. Ataharul Islam is currently the QM Husain Professor, ISRT, University of Dhaka. He was a former professor of statistics at the University Sains Malaysia, King Saud University, University of Dhaka and the East West university. He was a visiting faculty at the University of Hawaii and University of Pennsylvania. He is recipient of the Pauline Stitt Award, the WNAR Biometric Society Award for content and writing, University Grants Commission Award for book and research, Ibrahim Gold Medal for research, etc. He published more than 100 research papers in international journals on various topics, extensively on longitudinal and repeated measures data, including multistate and multistage hazards models, statistical models for repeated measures data, Markov models with covariate dependence, generalized linear models, conditional and joint models for correlated outcomes, etc. He authored several books either published or being published.

PUBLIC INTEREST STATEMENT

This research work provides an important development in modelling correlated outcomes data and shows the test for different types of associations. This model is developed for three binary outcomes. In various fields, the presence of correlation in outcome variables poses formidable difficulty to model the relationship between potential explanatory variables and outcome variables. However, if the dependence in outcome variables is not considered in modelling, the relationships between explanatory and outcome variables may be affected to a large extent resulting in misleading results. This paper provides a trivariate Bernoulli regression model and tests for the potential relationships of first, second and third orders are proposed. The applications of these procedures will make the analysis of correlated binary outcomes possible with appropriate interpretation and underlying mechanism of relationships among outcome variables and between outcome and explanatory variables.

noteworthy works were presented by McCullagh and Nelder (1989), Glonek and McCullagh (1995), Yee and Dirnbock (2009) and Dai, Ding, and Wahba (2013). McCullagh and Nelder (1989) considered proportional odds model as a starting point for constructing proportional odds model for three ordinal categories. Glonek and McCullagh generalized the model for several categorical responses. Alternative conditional models based on Markovian assumptions were proposed by Islam and Chowdhury (2006, 2008, 2017), Islam, Chowdhury, and Huda (2009) and Islam et al. (2012a). On the other hand, Dai et al. (2013) showed a multivariate Bernoulli model to estimate structure of graphs with binary nodes. Islam, Alzaid, Chowdhury, and Sultan (2013) proposed an alternative procedure using marginal-conditional approach to construct a bivariate Bernoulli model and provided tests for dependence in outcome variables. In this paper, a trivariate Bernoulli regression model is shown using marginal-conditional approach. Tests for dependence in trivariate models are also shown. The proposed model can be used very extensively in longitudinal studies with trivariate binary outcomes in various fields.

2. Trivariate Bernoulli distribution

Marshall and Olkin (1985) showed the bivariate Bernoulli form of Y_1 and $Y_2$with Bernoulli marginal. Let us consider three binary variables Y_1, Y_2 and Y_3, which, in longitudinal studies, can be considered as status of outcome variables at time points T_1, T_2 and T_3, respectively. The probability distribution is displayed below:

Y_1	Y_2		Y_3	Total
0	0	0	p_{001}	p_{00}
0	1	p_{000}	p_{011}	p_{01}
1	0	p_{010}	p_{101}	p_{10}
1	1	p_{100}	p_{111}	p_{11}
Total		p_{110}	$p_{..1}$	1

and the trivariate Bernoulli probability can be represented as follows

$$P(Y_1=y_1,Y_2=y_2,Y_3=y_3)=p_{000}^{(1-y_1)(1-y_2)(1-y_3)}p_{001}^{(1-y_1)(1-y_2)y_3}$$
$$p_{010}^{(1-y_1)y_2(1-y_3)}p_{100}^{y_1(1-y_2)(1-y_3)}p_{011}^{(1-y_1)y_2y_3}p_{101}^{y_1(1-y_2)y_3}p_{110}^{y_1y_2(1-y_3)}p_{111}^{y_1y_2y_3}. \tag{1}$$

We can easily find the expression for the above trivariate Bernoulli in exponential family form and after taking log the log-likelihood for n = 1 is

$$l=y_1\ln\left(\frac{p_{100}}{p_{000}}\right)+y_2\ln\left(\frac{p_{010}}{p_{000}}\right)+y_3\ln\left(\frac{p_{001}}{p_{000}}\right)+y_1y_2\ln\left(\frac{p_{110}p_{000}}{p_{100}p_{010}}\right)+$$
$$y_1y_3\ln\left(\frac{p_{101}p_{000}}{p_{100}p_{001}}\right)+y_2y_3\ln\left(\frac{p_{011}p_{000}}{p_{010}p_{001}}\right)+y_1y_2y_3\ln\left(\frac{p_{100}p_{010}p_{001}p_{111}}{p_{110}p_{011}p_{101}p_{000}}\right)+\ln p_{000}. \tag{2}$$

The natural link functions are:

$$\theta_1=\ln\left(\frac{p_{100}}{p_{000}}\right),\theta_2=\ln\left(\frac{p_{010}}{p_{000}}\right),\theta_3=\ln\left(\frac{p_{001}}{p_{000}}\right),$$
$$\theta_{12}=\ln\left(\frac{p_{110}p_{000}}{p_{100}p_{010}}\right),\theta_{13}=\ln\left(\frac{p_{101}p_{000}}{p_{100}p_{001}}\right),\theta_{23}=\ln\left(\frac{p_{011}p_{000}}{p_{010}p_{001}}\right), \tag{3}$$
$$\theta_{123}=\ln\left(\frac{p_{100}p_{010}p_{001}p_{111}}{p_{110}p_{011}p_{101}p_{000}}\right),\theta_0=\ln p_{000}.$$

Islam, Alzaid, Chowdhury, and Sultan (2013) showed for bivariate Bernoulli regression model that underlying relationships can be explored more conveniently if a marginal-conditional approach is employed. We can use the underlying conditional and marginal models for expressing the association parameters in the proposed model. The bivariate Bernoulli model is comprised of 3 (2^2-1) models, two conditional and one marginal. Extension to trivariate Bernoulli shows that there are 7 (2^3–1) models. It appears from the above link functions that there are underlying relationships between link functions of three first, three second and one third orders that emerge from the natural link functions.

The first-order link functions are simply odds between respective cell probability for an outcome variable and baseline probability of non-occurrence of any of the outcomes at three time points.

3. The marginal-conditional models for trivariate Bernoulli

We can express the joint probability of three outcome variables Y_1, Y_2 and Y_3 for given X as follows:

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3|X) = P(Y_1 = y_1|X)P(Y_2 = y_2|y_1, X)P(Y_3 = y_3|y_1, y_2, X) \tag{4}$$

Here is $P(Y_1 = y_1|X) = \pi_{y_1}(X)$ marginal probability for Y_1, $P(Y_2 = y_2|y_1, X) = \pi_{y_1y_2}(X)$ is conditional probability for Y_2 given Y_1, $P(Y_3 = y_3|y_1, y_2, X) = \pi_{y_1y_2y_3}(X)$ is conditional probability for Y_3 given Y_1 and Y_2, and $X = (1, X_1, \ldots, X_p)$. Using the relationship shown in Equation (4), the joint probabilities are

$$\begin{aligned}
p_{000}(X) &= [1-\pi_1(X)][1-\pi_{01}(X)][1-\pi_{001}(X)]\\
p_{100}(X) &= \pi_1(X)[1-\pi_{11}(X)][1-\pi_{101}(X)]\\
p_{010}(X) &= [1-\pi_1(X)]\pi_{01}(X)[1-\pi_{011}(X)]\\
p_{001}(X) &= [1-\pi_1(X)][1-\pi_{01}(X)]\pi_{001}(X)\\
p_{110}(X) &= \pi_1(X)\pi_{11}(X)[1-\pi_{111}(X)]\\
p_{011}(X) &= [1-\pi_1(X)]\pi_{01}(X)\pi_{011}(X)\\
p_{101}(X) &= \pi_1(X)[1-\pi_{11}(X)]\pi_{101}(X)\\
p_{111}(X) &= \pi_1(X)\pi_{11}(X)\pi_{111}(X)
\end{aligned} \tag{5}$$

The marginal model $\pi_1(X)$ is

$$\pi_1(X) = \frac{e^{X\beta_1}}{1 + e^{X\beta_1}} \tag{6}$$

where $\beta_1 = (\beta_{10}, \beta_{11}, \ldots, \beta_{1p})$ and $\pi_0(X) = \frac{1}{1+e^{X\beta_1}}$ as $\pi_0(X) + \pi_1(X) = 1$.

The conditional probabilities can be obtained from the first- and second-order Markov models with covariate dependence (Islam et al., 2012b, 2012a; Islam et al., 2013; Islam & Chowdhury, 2017). The conditional models for first-order Markov chain transition probabilities in the relationships (5) are displayed in Table 1 shown below for outcome variables Y_1 and Y_2:

The transition probabilities are functions of covariates as displayed below

$$\pi_{01}(X) = \frac{e^{X\beta_{01}}}{1 + e^{X\beta_{01}}}, \tag{7}$$

$$\pi_{11}(X) = \frac{e^{X\beta_{11}}}{1 + e^{X\beta_{11}}} \tag{8}$$

where $\beta_{01} = (\beta_{010}, \beta_{011}, \ldots, \beta_{01p}), \beta_{11} = (\beta_{110}, \beta_{111}, \ldots, \beta_{11p})$,

$\pi_{00}(X) = \frac{1}{1+e^{X\beta_{01}}}, \pi_{10}(X) = \frac{1}{1+e^{X\beta_{11}}}$, and $\pi_{00}(X) + \pi_{01}(X) = 1$

and $\pi_{10}(X) + \pi_{11}(X) = 1$

Table 1. Transition probabilities for outcome variables Y_1 and Y_2

Y_1	Y_2	
	0	**1**
0	$\pi_{00}(X)$	$\pi_{01}(X)$
1	$\pi_{10}(X)$	$\pi_{11}(X)$

Four second-order conditional models $\pi_{001}(X), \pi_{011}(X), \pi_{101}(X)$, and $\pi_{111}(X)$ are needed for the trivariate Bernoulli model. The conditional probabilities for second-order Markov models satisfy $\pi_{000}(X) + \pi_{001}(X) = 1$,

$$\pi_{010}(X) + \pi_{011}(X) = 1,\ \pi_{100}(X) + \pi_{101}(X) = 1\ ,\ \pi_{110}(X) + \pi_{111}(X) = 1$$

The second-order transition probabilities are shown in Table 2:

The covariate-dependent second-order models are

$$\pi_{001}(X) = \frac{e^{X\beta_{001}}}{1 + e^{X\beta_{001}}}, \pi_{000}(X) = \frac{1}{1 + e^{X\beta_{001}}}, \tag{9}$$

$$\pi_{011}(X) = \frac{e^{X\beta_{011}}}{1 + e^{X\beta_{011}}}, \pi_{010}(X) = \frac{1}{1 + e^{X\beta_{011}}} \tag{10}$$

$$\pi_{101}(X) = \frac{e^{X\beta_{101}}}{1 + e^{X\beta_{101}}}, \pi_{100}(X) = \frac{1}{1 + e^{X\beta_{101}}} \tag{11}$$

and

$$\pi_{111}(X) = \frac{e^{X\beta_{111}}}{1 + e^{X\beta_{111}}}, \pi_{110}(X) = \frac{1}{1 + e^{X\beta_{111}}}. \tag{12}$$

Here, $\beta_{001} = (\beta_{0010}, \beta_{0011}, \ldots, \beta_{001p}), \beta_{011} = (\beta_{0110}, \beta_{0111}, \ldots, \beta_{011p})$

$\beta_{101} = (\beta_{1010}, \beta_{1011}, \ldots, \beta_{101p}), \beta_{111} = (\beta_{1110}, \beta_{1111}, \ldots, \beta_{111p})$

4. Link functions and estimating equations for trivariate Bernoulli model

The link functions for trivariate Bernoulli regression models are displayed in (3). In Section 3, we have shown the relationship between joint probabilities and marginal-conditional probabilities. The conditional probabilities are obtained from the first- and second-order transition probabilities. We have introduced seven logistic regression models, one marginal model in Equation (6) and six conditional models in Equations 7–12. Using the relationships between joint and marginal-conditional probabilities, we can redefine the link functions which are summarized below:

$$\theta_0 = \ln p_{000} = -\ln(1 + e^{X\beta_1}) - \ln(1 + e^{X\beta_{01}}) - \ln(1 + e^{X\beta_{001}}), \tag{13}$$

$$\theta_1 = \ln\left(\frac{p_{100}}{p_{000}}\right) = X\beta_1 - \left\{\ln(1 + e^{X\beta_{11}}) - \ln(1 + e^{X\beta_{01}})\right\} - \left\{\ln(1 + e^{X\beta_{101}}) - \ln(1 + e^{X\beta_{001}})\right\} \tag{14}$$

$$\theta_2 = \ln\left(\frac{p_{010}}{p_{000}}\right) = X\beta_{01} - \left\{\ln(1 + e^{X\beta_{011}}) - \ln(1 + e^{X\beta_{001}})\right\}, \tag{15}$$

$$\theta_3 = \ln\left(\frac{p_{001}}{p_{000}}\right) = X\beta_{001}, \tag{16}$$

Table 2. Transition probabilities for outcome variables Y_1, Y_2 and Y_3

Y_1	Y_2	Y_3	
		0	**1**
0	0	$\pi_{000}(X)$	$\pi_{001}(X)$
0	1	$\pi_{010}(X)$	$\pi_{011}(X)$
1	0	$\pi_{100}(X)$	$\pi_{101}(X)$
1	1	$\pi_{110}(X)$	$\pi_{111}(X)$

$$\theta_{12} = \ln\left(\frac{p_{110}p_{000}}{p_{100}p_{010}}\right) = X\beta_{11} - X\beta_{01} - \ln(1+e^{X\beta_{111}}) + \ln(1+e^{X\beta_{011}}) + \ln(1+e^{X\beta_{101}}) - \ln(1+e^{X\beta_{001}}) \tag{17}$$

$$\theta_{13} = \ln\left(\frac{p_{101}p_{000}}{p_{100}p_{001}}\right) = X\beta_{101} - X\beta_{001}, \tag{18}$$

$$\theta_{23} = \ln\left(\frac{p_{011}p_{000}}{p_{010}p_{001}}\right) = X\beta_{011} - X\beta_{001}, \tag{19}$$

$$\theta_{123} = \ln\left(\frac{p_{111}p_{100}p_{010}p_{001}}{p_{110}p_{011}p_{101}p_{000}}\right) = X\beta_{111} + X\beta_{001} - X\beta_{101} - X\beta_{011}. \tag{20}$$

The estimating equations are obtained by differentiating the log-likelihood function (2) with respect to regression parameters of seven marginal and conditional models. It may be noted here that for simplicity, the summation signs for i = 1,…,n is ignored here and the equations are shown for n = 1. The estimating equations are:

$$\left[\frac{\partial l}{\partial \beta_{sk}}\right]' = \begin{bmatrix} \frac{\partial l}{\partial \beta_{1k}} \\ \frac{\partial l}{\partial \beta_{01k}} \\ \frac{\partial l}{\partial \beta_{11k}} \\ \frac{\partial l}{\partial \beta_{001k}} \\ \frac{\partial l}{\partial \beta_{011k}} \\ \frac{\partial l}{\partial \beta_{101k}} \\ \frac{\partial l}{\partial \beta_{111k}} \end{bmatrix} = \begin{bmatrix} X_k(y_1 - \pi_1(X)) \\ X_k[(1-y_1)(y_2 - \pi_{01}(X))] \\ y_1X_k(y_2 - \pi_{11}(X)) \\ X_k[(1-y_1-y_2+y_1y_2)(y_3 - \pi_{001}(X))] \\ y_2X_k[(1-y_1)(y_3 - \pi_{011}(X))] \\ y_1X_k[(1-y_2)(y_3 - \pi_{101}(X))] \\ y_1y_2X_k(y_3 - \pi_{111}(X)) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \; k = 0, 1, \ldots, p. \tag{21}$$

where $X_k, k = 1, \ldots, p$ is the kth explanatory variable corresponding to the coefficient β_{sk} , s denotes marginal and conditional models represented by 1, 01,11,001, 011, 101 and 111. The information matrix is comprised of information matrices for seven sets of parameters for marginal and conditional models. The (k, k') th element of the information matrix for model s is $-\frac{\partial^2 l}{\partial \beta_{sk}\partial \beta_{sk'}}$. Let us denote the $7(p+1) \times 7(p+1)$ matrix containing 7 diagonal $(p+1) \times (p+1)$ matrices as shown below:

$$I = \begin{bmatrix} I_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_7 \end{bmatrix} \tag{22}$$

where

$$I_1 = \left[X_kX_k{}'\pi_1(X)(1-\pi_1(X)), k, k' = 0, 1, \ldots, p\right]_{(p+1)\times(p+1)},$$
$$I_2 = \left[X_kX_k{}'(1-y_1)\pi_{01}(X)(1-\pi_{01}(X)), k, k' = 0, 1, \ldots, p\right]_{(p+1)\times(p+1)},$$
$$I_3 = \left[X_kX_k{}'y_1\pi_{11}(X)(1-\pi_{11}(X)), k, k' = 0, 1, \ldots, p\right]_{(p+1)\times(p+1)},$$
$$I_4 = \left[X_kX_k{}'(1-y_1-y_2+y_1y_2)\pi_{001}(X)(1-\pi_{001}(X)), k, k' = 0, 1, \ldots, p\right]_{(p+1)\times(p+1)},$$
$$I_5 = \left[X_kX_k{}'(1-y_1)y_2\pi_{011}(X)(1-\pi_{011}(X)), k, k' = 0, 1, \ldots, p\right]_{(p+1)\times(p+1)},$$
$$I_6 = \left[X_kX_k{}'y_1(1-y_2)\pi_{101}(X)(1-\pi_{101}(X)), k, k' = 0, 1, \ldots, p\right]_{(p+1)\times(p+1)},$$
$$I_7 = \left[X_kX_k{}'y_1y_2\pi_{111}(X)(1-\pi_{111}(X)), k, k' = 0, 1, \ldots, p\right]_{(p+1)\times(p+1)}.$$

5. Tests for models and dependence of outcome variables

For trivariate Bernoulli regression models, we have seven sets of parameters, one set for marginal for Y_1 (Model 1for s = 1), two sets for conditional models from first-order Markov chains for transition from Y_1 to Y_2 (Models 1 and 2 for s = 01, 11, respectively) and four sets from second-order Markov chain for outcomes Y_1, Y_2 and Y_3 (s = 001, 011, 101, 111, respectively) . The likelihood ratio test for overall joint model is

$$-2[\ln L(\beta_0) - \ln L(\beta)] \sim {\chi^2}_{7p}$$

where $\beta_0 = (\beta_{s0}, s = 1, 01, 11, 001, 011, 101, 111)$ and $\beta = (\beta_s, s = 1, 01, 11, 001, 011, 101, 111)$. Let $\beta* = (\beta_s*, s = 1, 01, 11, 001, 011, 101, 111)$ where $\beta_s* = \left(\beta_{s1}, \ldots, \beta_{sp}\right)$. The null hypothesis of the above test is $H_0 : \beta* = 0$.

Tests for the marginal and conditional models can be performed separately as well. In that case, the null hypothesis for each model is $H_0 : \beta_s* = 0$, s = 1, 01, 11, 001, 011, 101, 111. The likelihood ratio test for each model can be shown as

$$-2[\ln L(\beta_{s0}) - \ln L(\beta_s)] \sim {\chi^2}_p$$

The Wald test will be applied to test for significance of a parameter for each model.

The tests for association parameters, $\theta_{12}, \theta_{13}, \theta_{23}, \theta_{123}$, can be performed based on the parameters of regression models (17)—(20). The null hypotheses and test statistics are displayed here:

(i) Test for independence of Y_2 and Y_3

The null hypothesis is

$$H_{01} : \beta_{001} = \beta_{011}$$

The test statistic for equality of parameters

$$A_1 = (\hat{\beta}_{001} - \hat{\beta}_{011})' [V(\hat{\beta}_{001} - \hat{\beta}_{011})]^{-1} (\hat{\beta}_{001} - \hat{\beta}_{011})$$

which is asymptotically chi-square with p degrees of freedom. It may be noted here that $[V(\hat{\beta}_{001} - \hat{\beta}_{011})]^{-1} \simeq I_4 + I_5$.

(ii) Test for independence of Y_1 and Y_3

The null hypothesis is

$$H_{01} : \beta_{001} = \beta_{101}$$

The test statistic for equality of parameters

$$A_2 = (\hat{\beta}_{001} - \hat{\beta}_{101})' [V(\hat{\beta}_{001} - \hat{\beta}_{101})]^{-1} (\hat{\beta}_{001} - \hat{\beta}_{101})$$

which is asymptotically chi-square with p degrees of freedom. It may be noted here that $[V(\hat{\beta}_{001} - \hat{\beta}_{101})]^{-1} \simeq I_4 + I_6$.

(iii) Test for independence of Y_1 and Y_2.

It appears from Equation (17) that independence of Y_1 and Y_2 depends not only on equality of regression parameters from Models 7 and 8 but also on equality of parameters of Models 9 and 11 and also Models 10 and 12.

The null hypotheses are

https://doi.org/10.1080/25742558.2018.1472519

$H_{01}: \beta_{01} = \beta_{11}$
$H_{02}: \beta_{001} = \beta_{101}$
$H_{03}: \beta_{011} = \beta_{111}.$

The test statistic for equality of regression parameters from Models 7 and 8 (H_{01}) can be performed using the following statistic

$$A_3 = (\hat{\beta}_{01} - \hat{\beta}_{11})'[V(\hat{\beta}_{01} - \hat{\beta}_{11})]^{-1}(\hat{\beta}_{01} - \hat{\beta}_{11})$$

The denominator is obtained approximately from

$$[V(\hat{\beta}_{01} - \hat{\beta}_{11})]^{-1} \simeq I_2 + I_3$$

Test for H_{02} is shown in (ii). Similarly, the test for H_{03} is based on Models 10 and 12 and the test statistic is

$$A_4 = (\hat{\beta}_{011} - \hat{\beta}_{111})'[V(\hat{\beta}_{011} - \hat{\beta}_{111})]^{-1}(\hat{\beta}_{011} - \hat{\beta}_{111})$$

where $[V(\hat{\beta}_{011} - \hat{\beta}_{111})]^{-1} \simeq I_5 + I_7$.

(iv) Test for independence of Y_1, Y_2 and Y_3

It appears from Equation (20) that independence of Y_1, Y_2 and Y_3 depends on equality of regression parameters from Models 9 and 11 but also on equality of parameters of Models 10 and 12.

The null hypotheses are

$H_{01}: \beta_{001} = \beta_{101}$
$H_{02}: \beta_{011} = \beta_{111}$

The first null hypothesis is shown in (ii) which is the test for independence of Y_1 and $Y_3(A_2)$ and the second null hypothesis is discussed in (iii) for partial test for independence of Y_1 and Y_2 (A_4). In other words, independence of Y_1, Y_2 and Y_3 depends on both (a) conditional independence of Y_1 and Y_2 (A_4) and (b) independence of Y_1 and Y_3 (A_2). Independence of three outcome variables depends on these pairwise conditional or unconditional independence of outcome variables that are shown in previous tests.

An alternative null hypothesis is quite straightforward from the model shown in Equation (20). Let us define $\beta_{...} = \beta_{111} + \beta_{001} - \beta_{101} - \beta_{011}$ then the independence of Y_1, Y_2 and Y_3 can be tested alternatively for null hypothesis $H_{01}: \beta_{...} = 0$. We can test the null hypothesis using

$$A_5 = \hat{\beta}_{...}{}'[V(\hat{\beta}_{...})]^{-1}\hat{\beta}_{...}$$

where $[V(\hat{\beta}_{...})]^{-1} \simeq I_4 + I_5 + I_6 + I_7$.

6. Concluding remarks

We need to analyse binary repeated measures data in many instances where the outcome variables are correlated. The modelling of correlated outcome variables have been of interest in many fields due to recent emergence of need for analysing repeated measures data in the presence of correlation among outcome variables in addition to models for identifying explanatory variables associated with outcome variables. Several attempts have been made in the past to model such data but due to inbuilt complexity in modelling multivariate data with correlated outcomes it remained a challenge for a long time. This study shows an alternative approach based on marginal-conditional formulation to describe a joint model and provides a set of marginal and conditional models that can provide joint probabilities for a trivariate binary case. This procedure can be extended for more than three correlated outcomes easily using the same approach which is not shown in this paper to keep the exposition simple. Several tests are displayed in this paper for testing the overall model for trivariate Bernoulli as well as for examining dependence in outcome variables.

Acknowledgements
This work was supported by the HEQEP sub-project CP 3293 sponsored by UGC, Bangladesh and World Bank. I would like to express my gratitude to the anonymous reviewers for their helpful suggestions.

Funding
This work was supported by the HEQEP sub-project [Grant Number CP 3293] sponsored by UGC, Bangladesh and World Bank.

Author details
M. Ataharul Islam[1]
E-mail: mataharul@yahoo.com
[1] ISRT (Institute of Statistical Research and Training), University of Dhaka, Dhaka, Bangladesh.

Citation information
Cite this article as: A trivariate Bernoulli regression model, M. Ataharul Islam, *Cogent Mathematics & Statistics* (2018), 5: 1472519.

References

Dai, B., Ding, S., & Wahba, G. (2013). Multivariate Bernoulli distribution. *Bernoulli, 19*(4), 1465–1483. doi:10.3150/12-BEJSP10

Glonek, G. F. V., & McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*, 533–546.

Islam, M. A., Alzaid, A. A., Chowdhury, R. I., & Sultan, K. S. (2013). A generalized bivariate Bernoulli model with covariate dependence. *Journal of Applied Statistics*, *40* (5), 1064–1075. doi:10.1080/02664763.2013.780156

Islam, M. A., & Chowdhury, R. I. (2006). A higher order Markov model for analyzing covariate dependence. *Applied Mathematical Modeling*, *30*, 477–488. doi:10.1016/j.apm.2005.05.006

Islam, M. A., & Chowdhury, R. I. (2008). Chapter 4: First and higher order transition models with covariate dependence. In F. Columbus (Ed.). *Progress in applied mathematical modeling* (pp. 153–196). Hauppage, NY: Nova Science Publishers.

Islam, M. A., & Chowdhury, R. I. (2017). *Analysis of repeated measures data*. Singapore: Springer.

Islam, M. A., Chowdhury, R. I., & Alzaid, A. A. (2012b). Tests for dependence in binary repeated measures data. *Journal of Statistical Research*, *46*, 203–217.

Islam, M. A., Chowdhury, R. I., & Briollais, L. (2012a). A bivariate binary model for testing depedence in outcomes. *Bulletin of Malaysian Mathematical Sciences Society*, *35*, 845–858.

Islam, M. A., Chowdhury, R. I., & Huda, S. (2009). *Markov models with covariate dependence for repeated measures*. New York: Nova Science Publishers.

Marshall, A. W., & Olkin, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. *Journal of the American Statistical Association*, *80*, 332–338. doi:10.1080/01621459.1985.10478116

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

Yee, T. W., & Dirnbock, T. (2009). Models for analyzing species' presence/absence data at two time points. *Journal of Theoretical Biology*, *259*, 684–694. doi:10.1016/j.jtbi.2009.05.004