*International Journal of Statistical Sciences* *ISSN 1683–5603*
*Vol. 6 (Special Issue), 2007, pp 151-161*


# Testing for Dependency of Bernoulli Variables

**Daniel Bruce and Hans Nyquist**
*Dept. of Statistics*
Stockholm University
106 91 Stockholm
Sweden

[Received December 14, 2005; Revised June 28, 2007; Accepted August19, 2007]

**Abstract**

The aim of this paper is to derive test procedures for studies where data consist of pairs of Bernoulli variables. Applications exist in, for example, ophthalmology and studies on matched pairs. Score tests and likelihood ratio tests are derived for testing the dependency between the Bernoulli variables. Multinomial logit models are used to incorporate explanatory variables. Test statistics for two particular models are thoroughly outlined. Numerical illustrations of these test statistics are presented in three examples, including one with visual impairment data.

**Keywords and Phrases:** Dependent binary data; Generalized linear models.

**AMS Classification:** 62J12.

## 1 Introduction

There are many applications in which pairs of Bernoulli variables are observed and where the problem is to test whether the variables are independent or not. An example is when testing if the occurrence of a particular disease in one eye is independent of the occurrence of the disease in the other eye, for the same individual. For modelling this situation we suppose that $S_1$ and $S_2$ are two equally distributed and possibly dependent Bernoulli variables such that $P(S_1 = 0, S_2 = 0) = \pi_0$, $P(S_1 = 1, S_2 = 0) = P(S_1 = 0, S_2 = 1) = \pi_1/2$ and $P(S_1 = 1, S_2 = 1) = \pi_2$. A possible and often used model is to consider the joint probability distribution of the bivariate vector $(S_1, S_2)$ as a log linear model. Other possibilities include the use a multinomial model (see e.g.

(2)) and a bivariate logistic model ((8)). Reviews on models for dependent Bernoulli variables are also in (4) and (9). Here we define the sum $S = S_1 + S_2$ and model the random variable $S$ by using a multinomial logit model as in (5). Independence will then imply a particular pattern on the probabilities $\pi_0$, $\pi_1$, and $\pi_2$. We therefore assume we have observations $s_i$, $i = 1, 2, \ldots, N$, on $S$ and construct a test to see whether data support a probability structure that is consistent with independence. In this paper we consider two tests for independence, viz. the score test and the likelihood ratio test, both tests being asymptotically equivalent. The performance of the score test and the likelihood ratio test in small samples can be improved by adjusting the critical value of the test, see (6) for a recent application. Furthermore, we consider the case when the probabilities depend on an explanatory variable $x$. In the example with an eye disease, the explanatory variable might be age.

The next section describes the model used and the test statistics are derived, while some numerical examples are provided in the third section. The final section contains some concluding remarks.

## 2 Model and test

Let $Y = (Y_1, Y_2)^T$ be the trinomial response vector, such that $Y_1$ takes the value 1 and $Y_2$ takes the value 0 if $S = 1$; and $Y_1$ takes the value 0 and $Y_2$ takes the value 1 if $S = 2$. In the case $S = 0$ both $Y_1$ and $Y_2$ are 0. The vector of expectations of $Y$ is then $E[Y] = (P(S=1), P(S=2))^T = (\pi_1, \pi_2)^T = \pi$, and we use the outcome $S = 0$, with $P(S=0) = \pi_0 = 1 - \pi_1 - \pi_2$, as a reference category when modelling the distribution of $Y$.

In the simplest case, $N$ observations, say $y_i = (y_{1i}, y_{2i})^T$, $i = 1, 2, \ldots, N$, are drawn independently from the same distribution and the log likelihood becomes

$$\ell(\pi; \mathbf{y}) = \sum_{i=1}^{N} y_{0i} \ln \pi_0 + y_{1i} \ln \pi_1 + y_{2i} \ln \pi_2.$$

The scores, $U_j(\pi) = \partial \ell / \partial \pi_j$, $j = 1, 2$, are found as

$$U_j(\pi) = \sum_{i=1}^{N} \left( \frac{y_{ji}}{\pi_j} - \frac{y_{0i}}{\pi_0} \right), \ j = 1, 2,$$

and the information matrix, $I_{jk}(\pi) = E\left[-\partial^2 \ell / \partial \pi_j \partial \pi_k\right]$, $j, k = 1, 2$, as

$$I(\pi) = N \begin{pmatrix} \frac{1}{\pi_1} + \frac{1}{\pi_0} & \frac{1}{\pi_0} \\ \frac{1}{\pi_0} & \frac{1}{\pi_2} + \frac{1}{\pi_0} \end{pmatrix}.$$

We will now derive the test statistics for the score test and the likelihood ratio test for testing the null hypothesis that the variables $S_1$ and $S_2$ are independent. The test

statistic for the score test is defined as

$$T_S = U^T(\widetilde{\pi})\, I^{-1}(\widetilde{\pi})\, U(\widetilde{\pi}),$$

where $U(\pi) = (U_1(\pi), U_2(\pi))^T$ and $\widetilde{\pi} = (\widetilde{\pi}_1, \widetilde{\pi}_2)^T$ is the estimated vector of probabilities under the null hypothesis. The independence hypothesis implies the restrictions $\pi_0 = (1-\theta)^2$, $\pi_1 = 2\theta(1-\theta)$, and $\pi_2 = \theta^2$, where $\theta = P(S_1 = 1) = P(S_2 = 1)$ is the marginal probability to observe a "success". Under this hypothesis, the maximum likelihood estimator of $\theta$ is evidently

$$\widetilde{\theta} = (2N)^{-1} \sum_{i=1}^{N} (y_{1i} + 2y_{2i}) = \frac{r_1 + 2r_2}{2N}, \tag{1}$$

where $r_j$ is the number of observed pairs that results in $y_j = 1$. Hence, the estimator $\widetilde{\theta}$ equals the total number of "successes" divided by the number of observed variables. Maximum likelihood estimators of $\pi_0$, $\pi_1$, and $\pi_2$ are accordingly

$$\widetilde{\pi}_0 = \left(1 - \widetilde{\theta}\right)^2, \ \widetilde{\pi}_1 = 2\widetilde{\theta}\left(1 - \widetilde{\theta}\right), \text{ and } \widetilde{\pi}_2 = \widetilde{\theta}^2, \tag{2}$$

respectively. By inserting the expressions for the scores and the information, we easily find that

$$T_S = \sum_{j=0}^{2} \frac{(r_j - N\widetilde{\pi}_j)^2}{N\widetilde{\pi}_j}, \tag{3}$$

which coincides with the $\chi^2$-test statistic for testing the goodness of fit of a trinomial distribution with probabilities restricted as described above. Asymptotically, $T_S$ has a $\chi^2$ distribution with 1 degree of freedom, the approximation being good provided the expected frequencies, $N\widetilde{\pi}_j$, $j = 0, 1, 2$, are sufficiently large.

The test statistic for the likelihood ratio test is

$$\begin{aligned} T_{LR} &= 2\left(\ell(\widehat{\pi}; \mathbf{y}) - \ell(\widetilde{\pi}; \mathbf{y})\right) \\ &= 2\sum_{j=0}^{2} r_j \ln \frac{\widehat{\pi}_j}{\widetilde{\pi}_j}, \end{aligned}$$

where $\widehat{\pi}$ is the unrestricted maximum likelihood estimator of $\pi$,$\widehat{\pi}_j = \frac{r_j}{N}$, $j = 0, 1, 2$.

This simple case generalizes straightforwardly to the case with several, say $K$, groups with $N_k$ observations in each group. The distribution of the trinomial response vector in each group is here defined by the vector $(\pi_{0k}, \pi_{1k}, \pi_{2k})^T$, $k = 1, 2, \ldots, K$, of probabilities. The test statistic for the score test now becomes

$$T_S = \sum_{k=1}^{K} \sum_{j=0}^{2} \frac{(r_{jk} - N_k\widetilde{\pi}_{jk})^2}{N_k\widetilde{\pi}_{jk}}, \tag{4}$$

where $r_{jk}$ is the observed frequency of category $j$, $j = 0, 1, 2$, in group $k$, $k = 1, 2, \ldots, K$,

$$\widetilde{\pi}_{0k} = \left(1 - \widetilde{\theta}_k\right)^2, \widetilde{\pi}_{1k} = 2\widetilde{\theta}_k \left(1 - \widetilde{\theta}_k\right), \widetilde{\pi}_{2k} = \widetilde{\theta}_k^2, \tag{5}$$

and

$$\widetilde{\theta}_k = \left(r_{1k} + 2r_{2k}\right) / \left(2N_k\right). \tag{6}$$

Similarly, the test statistic for the likelihood ratio test becomes

$$T_{LR} = 2 \sum_{k=1}^{K} \sum_{j=0}^{2} r_{jk} \ln \frac{\widehat{\pi}_{jk}}{\widetilde{\pi}_{jk}}. \tag{7}$$

where the unrestricted estimator $\widehat{\pi}$ is

$$\widehat{\pi}_{jk} = \frac{r_{jk}}{N_k}.$$

The test statistics $T_S$ and $T_{LR}$ are asymptotically equivalent and has a $\chi^2$ distribution with $K$ degrees of freedom, asymptotically. Here it is important for the approximation to be good that each $N_k \widetilde{\pi}_{jk}$ is sufficiently large.

A more structured model is obtained if the vector of probabilities $\pi$ is governed by a vector of explanatory variables $x$. For example, in a multinomial logit model the probabilities are defined by the vector valued logit link function

$$g \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} \ln \frac{\pi_1}{\pi_0} \\ \ln \frac{\pi_2}{\pi_0} \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \eta,$$

where $\eta = (\eta_1, \eta_2)^T$ is the vector valued linear predictor with

$$\eta_j = x_j^T \beta_j, \ j = 1, 2,$$

$x_j$ and $\beta_j$ being vectors of explanatory variables and associated parameters used for determining the probability $\pi_j$. By defining the block diagonal matrix $\mathbf{x} = diag\,(x_1, x_2)$ and the parameter vector $\beta = (\beta_1, \beta_2)^T$ the linear predictor can be written more compactly as

$$\eta = \mathbf{x}\beta.$$

The probabilities for the response categories are

$$\begin{aligned} \pi_0(x) &= \frac{1}{1 + e^{\eta_1} + e^{\eta_2}} \\ \pi_1(x) &= \frac{e^{\eta_1}}{1 + e^{\eta_1} + e^{\eta_2}} \\ \pi_2(x) &= \frac{e^{\eta_2}}{1 + e^{\eta_1} + e^{\eta_2}} \end{aligned}$$

With $N$ independent observations on $Y$, say $y_i = (y_{1i}, y_{2i})^T$, $i = 1, 2, \ldots, N$, the log likelihood function becomes

$$l(\beta; \mathbf{y}) = \sum_{i=1}^{N} \left( y_{1i}\eta_{1i} + y_{2i}\eta_{2i} - \ln\left(1 + e^{\eta_{1i}} + e^{\eta_{2i}}\right)\right).$$

The score vector, $U(\beta) = \partial l(\beta)/\partial\beta$, is readily found to be

$$U(\beta) = \sum_{i=1}^{N} \mathbf{x}_i^T \left(y_i - \pi(x_i)\right)$$

and the Fisher information matrix, $I(\beta) = E\left[U(\beta)\, U^T(\beta)\right]$ is

$$I(\beta) = \sum_{i=1}^{N} \mathbf{x}_i^T D(x_i)\, \mathbf{x}_i$$

where

$$D(x) = \begin{pmatrix} \frac{\partial \pi_1}{\partial \eta_1} & \frac{\partial \pi_1}{\partial \eta_2} \\ \frac{\partial \pi_2}{\partial \eta_1} & \frac{\partial \pi_2}{\partial \eta_2} \end{pmatrix} = \begin{pmatrix} \pi_1(x)\left(1 - \pi_1(x)\right) & -\pi_1(x)\,\pi_2(x) \\ -\pi_1(x)\,\pi_2(x) & \pi_2(x)\left(1 - \pi_2(x)\right) \end{pmatrix},$$

for details, see (5). Maximum likelihood estimation of this model can be done by using ordinary algorithms for estimating multivariate logit models.

We now consider two particular models within this framework. The first being the model where the two vectors of explanatory variables are identical and consist of dummy variables $x_1 = x_2 = (d_1, d_2, \ldots, d_K)^T$, where each $d_k$ is either 1 or 0, indicating if an observation comes from response group $k$ or not, respectively, $j = 1, 2, \ldots, K$. In this case the model reduces to the case with $K$ response groups discussed above and the test statistics for independence are (4) and (7), respectively.

The second particular case we consider appears when the linear predictors consist of an intercept and a single explanatory variable, $z$, the same variable in both linear predictors, so that $\eta_j = x^T\beta_j$, $x = (1, z)^T$ and $\beta_j = (\beta_{j0}, \beta_{j1})^T$, $j = 1, 2$. In this model, the explanatory variable $z$ may influence the success probabilities $\pi_1$ and $\pi_2$ differently. However, the odds ratio of $S_1$ and $S_2$ does not depend on $z$ if $\beta_{21} = 2\beta_{11}$, i.e. the dependence of $S_1$ and $S_2$ is constant over different values of $z$. If, in addition, $\beta_{20} = 2\beta_{10} - \ln 4$, then $S_1$ and $S_2$ are independent as well, making their sum $S$ to a binomially distributed random variable with parameters $(2, \theta)$. For details, see (5). It is therefore of interest to test whether data give enough support to reject the hypothesis

$$H_0 : \beta_{20} = 2\beta_{10} - \ln 4 \text{ and } \beta_{21} = 2\beta_{11}.$$

Suppose now that the variable $z$ takes $K$ different values, and that $N_k$ observations are made at $z = z_k$, so that $\sum_{k=1}^{K} N_k = N$. Furthermore, let $r_{1k} = \sum y_{1i}$ and

$r_{2k} = \sum y_{2i}$, the sums being taken over all the observations with $z = z_k$, be the observed frequencies of the two possible responses $Y_1 = 1$ and $Y_2 = 1$, respectively. Then, $r_{1k} \sim bin\,(N_k, \pi_{1k})$ and $r_{2k} \sim bin\,(N_k, \pi_{2k})$.

Under $H_0$, the model reduces to an ordinary logit model. Denoting the maximum likelihood estimator of the parameter vector $\beta$ under $H_0$ by $\widetilde{\beta}$, the score test statistic becomes

$$T_S = U^T\left(\widetilde{\beta}\right) I^{-1}\left(\widetilde{\beta}\right) U\left(\widetilde{\beta}\right). \tag{8}$$

which is asymptotically $\chi^2$-distributed with 2 degrees of freedom, provided that the number of observations at each $z_k$ tends to infinity. This follows since $U\left(\widetilde{\beta}\right)$ is approximately normal with zero mean and variance $I\,(\beta)$ if $H_0$ is true. Furthermore, the hypothesis $H_0$ is rejected on the $\alpha \cdot 100\%$ level if the observed value of $T_S$ exceeds the critical value $c = \chi^2_{2,1-\alpha}$, the $(1-\alpha)\cdot 100$th percentile of a $\chi^2$ distribution with 2 degrees of freedom.

When computing the test statistic for the likelihood test, the model needs to be estimated both under the restrictions imposed by $H_0$, yielding the univariate logit estimator $\widetilde{\beta}$, and without these restrictions, yielding the bivariate logit estimator $\widehat{\beta}$. The test statistic is then obtained by evaluating the log likelihood function at these two estimates

$$T_{LR} = 2\left(l\left(\widehat{\beta};\mathbf{y}\right) - l\left(\widetilde{\beta};\mathbf{y}\right)\right). \tag{9}$$

Asymptotically, both $T_S$ and $T_{LR}$ are equivalent. In particular, also $T_{LR}$ is asymptotically $\chi^2$-distributed with 2 degrees of freedom and the null hypothesis is rejected at the level $\alpha \cdot 100\%$ for observed values exceeding the critical value $c$.

## 3 Examples

This section presents three examples to illustrate the test procedures described above. Example 1 is a visual impairment data, while the other two examples are artificially created data materials. Dependency is tested using the score test and the likelihood ratio test for all these data materials. Significance level is chosen to be 5%.

### Example 1

The data material on visual impairment data is taken from (7). 5199 people are subject to a visual examination, measuring if the left eye and/or the right eye has a visual impairment or not. The outcome for each eye is binary, where " + " indicates visual impairment and " − " no visual impairment. Age is used as explanatory variable in this example, see Table 1. In the table there are, for example, 3627 out of 3958 people in age 40-70 that have no visual impairment.

Before a model can be fitted to the data an assumption has to be made. The probability that the left eye is visually impaired is assumed to be equal to the probability

| Left | Right | Age: $40-70$ | Age: $71+$ | Total |
|---|---|---|---|---|
| $-$ | $-$ | 3627 | 913 | 4540 |
| $+$ | $-$ | 122 | 89 | 211 |
| $-$ | $+$ | 133 | 104 | 237 |
| $+$ | $+$ | 76 | 135 | 211 |
| | Total | 3958 | 1241 | 5199 |

Table 1: Joint distribution of visual impairment for both eyes, for the two age groups $40-70$ and over 70, respectively. Data are taken from Liang et al.(1992).

that the right eye is visually impaired. This assumption is reasonable since the risk of visual impairment (in percent) is similar for the left and the right eye in both groups.

Let $S_1$ and $S_2$ be Bernoulli variables for visual impairment of the left eye and the right eye, respectively. The elements of the response vector $y_i = (y_{1i}, y_{2i})^T$, $i = 1, 2, \ldots, 5199$, are indicator variables. For a given person $Y_1 = 1$ if only one eye is visually impaired and $Y_2 = 1$ if both eyes have a visual impairment. The vector of explanatory variables consists of dummy variables $d_1$ and $d_2$ since there are two independent groups. The link function is therefore

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \beta_{11} d_1 + \beta_{12} d_2 \\ \beta_{21} d_1 + \beta_{22} d_2 \end{pmatrix}.$$

Suppose now that primary interest is in the possible dependency between $S_1$ and $S_2$. In this model $S_1$ and $S_2$ are independent if the parameter restrictions

$$\beta_{2j} = 2\beta_{1j},\ j = 1, 2$$

are satisfied. As stated previously the score test statistic for independence is given by (4). The test statistic has a $\chi^2$ distribution with 2 degrees of freedom, asymptotically. The observed test statistic for the data material in Table 1 becomes, using (4),

$$T_S \approx 751.22$$

Hence the hypothesis about independence is rejected since the critical value on 5% level is 5.991. The observed likelihood ratio test statistic can be derived by estimating the model both under the restrictions imposed by $H_0$ and without these restrictions. Alternatively, the test statistic in (7) give an equivalent result. The value on the test statistic is

$$T_{LR} \approx 465.35,$$

so the hypothesis about independence is rejected when using the likelihood ratio test as well.

**Example 2**

Data consist of 100 pairs of Bernoulli variables. Each pair is associated with a single covariate, $z$, ranging between zero and ten, see Figure 9. This example refers to the second particular case above where the restrictions $\beta_{20} = 2\beta_{10} - \ln 4$ and $\beta_{21} = 2\beta_{11}$ are tested. Ignoring the covariate $z$, the observed frequencies for $S = 0, 1, 2$ are $49, 17$, and 34, respectively. $S_1$ and $S_2$ seem dependent by only looking at these observed frequencies. The goodness of fit test given in (3) confirms this. The observed test statistic is

$$\chi^2_{obs} \approx 42.54.$$

Clearly, the conclusion based only on this test would be that $S_1$ and $S_2$ are dependent. It is not sufficient to look at observed marginal frequencies only. When testing for independency one has to study how the probabilities $\pi_0(z), \pi_1(z)$, and $\pi_2(z)$ change when taking account of the covariate, $z$. The relative low frequency of pairs where $S = 1$ is explained by the fact that many observations are taken at $z-$values where $\pi_1(z)$ is small.

The score test statistic and the likelihood ratio test statistic given in (8) and (9), take covariates into account in the test procedures. The observed test statistics for the two tests become

$$T_S \approx 0.0340$$

and

$$T_{LR} \approx 0.0338,$$

respectively. Because the critical value is 5.991 the hypothesis of independence can not be rejected in either of the tests.

Another good indicator of the possible dependency between $S_1$ and $S_2$ is the estimated probability distribution of $S$, given in Figure 9

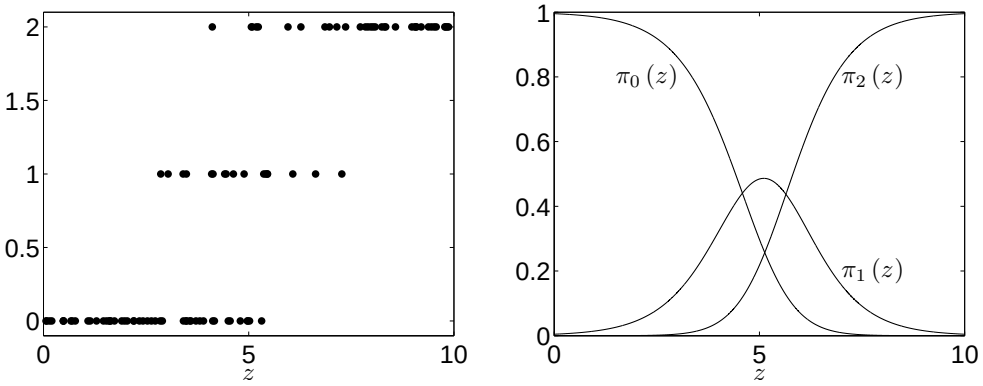


Figure 9: a) The 100 observations on $S$ for Example 2. b) Probabilities $\pi_0, \pi_1$, and $\pi_2$ as functions of $z$. The parameter values used are the bivariate logit estimates $\widehat{\beta}$.

The probability distribution closely resembles the appearance of a distribution for independent Bernoulli variables. A model for independent data has several symmetry properties, see (5) for a more comprehensive discussion. Two of these properties are clearly shown in Figure 9. First the maximum value of $\pi_1(z)$ is close to 0.5, and secondly $\pi_1(z)$ is a symmetric function around $\arg\max_z \pi_1(z)$. This example emphasizes the importance of including existing covariates in the analysis.

**Example 3**

The data in the third example have a similar structure as the data in Example 2. Data are given by 100 pairs of Bernoulli variables, where each pair is associated with a single covariate. Thus, the same model can be fitted to this data material as to the previous data material. Figure 10 shows that the data in Example 3 resemble the data in Example 2. Nevertheless, the score test statistic and the likelihood ratio test statistic are given by

$$T_S \approx 6.3066$$

and

$$T_{LR} \approx 7.7791,$$

respectively. The hypothesis of independence is rejected in both tests because the observed test statistics exceed the critical value. Figure 10 presents the probability distribution of $S$ based on the bivariate logit estimator $\widehat{\beta}$.

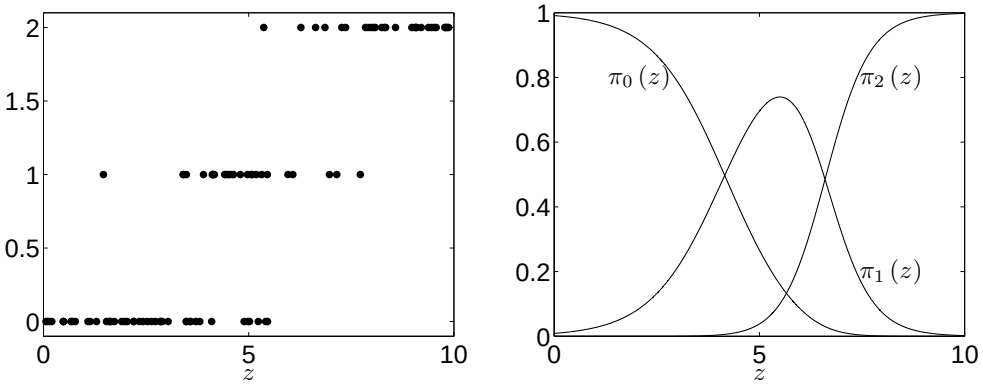


Figure 10: a) The 100 observations on $S$ for Example 3. b) Probabilities $\pi_0, \pi_1$, and $\pi_2$ as functions of $z$. The parameter values used are the bivariate logit estimates $\widehat{\beta}$.

The probability distribution does not share the symmetry properties that independent Bernoulli variables would have generated. Maximum value of $\pi_1(z)$ is relatively far from 0.5 and $\pi_1(z)$ is not a symmetric function around $\arg\max_z \pi_1(z)$.

## 4 Concluding remarks

Models for equally distributed and possibly dependent binary variables are examined in this paper. In particular, test procedures for testing the possible dependence between the binary variables, $S_1$ and $S_2$, are derived. The test procedures are important since the complexity of the model is greatly reduced if the variables are independent. In these situations where the bivariate logit model is difficult to estimate, a score test is preferable since the model only need to be estimated under the restrictions of independence.

A limitation with the models is that they are only applicable to data where equally distributed binary variables exist. The models are inappropriate in situations where $P(S_1 = 1)$ differ from $P(S_2 = 1)$. Nevertheless, in examples with eye data like the one above, the assumption of equally distributed variables (eyes) is often fulfilled. It remains to investigate how poorly the models fit when the assumption is not valid.

In some situations, where it is possible to conduct an experiment, the values of the covariate can be controlled. The distributions of the test statistics depend on the values of the covariate. Hence, different set of values on the covariate generates different powers of the tests. In this framework, a favourable power function can be generated if the values of the covariate, i.e. the design, are chosen in an optimal way. The design consists of the choice of values for the covariate (design points) and the corresponding proportion of observations (design weights), see (3). Finding the design that maximizes the power is not trivial. The optimal design depends on the design criterion, the unknown parameters, and the alternative hypothesis.

It should be noted that the test statistics are sensitive against small expected frequencies. The performance of the test statistics are like other asymptotically $\chi^2$ distributed test statistics for contingency tables, affected by too small expected frequencies.

The models do not account for variations among individuals with the same covariates. A further development would be to include, for example random effect parameters in the linear predictor. These parameters would then account for the individual effects. Parameters for modelling the heterogeneity among individuals are included in the general model for dependent binary response given in (1).

## References

[1] Agresti, A. (1997), A Model for Repeated Measurements of a Multivariate Binary Response. *Journal of the American Statistical Association*, **92**, 315-321.

[2] Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons.

[3] Atkinson, A. and Donev, A. (1992). *Optimum Experimental Design*. Oxford science publications.

[4] Bonney, G.E. (1987). Logistic Regression for Dependent Binary Observations. *Biometrics*, **43**, 951-973.

[5] Bruce, D. (2005). D-optimal Designs for Dependent Binary Variables. Department of Statistics, Stockholm University.

[6] Hoque, Z., Khan, S., and Billah, B. (2006). Edgeworth Size Corrected W, LR and LM Tests in the Formation of the Preliminary Test Estimator. *Journal of Statistical Research*, **40,** 2, 55-64.

[7] Liang, K.Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate Regression Analyses for Categorical Data. *Journal of Royal Statistical Society*, Series B, **54**, 1, 3-40.

[8] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. 2nd ed. Chapman & Hall.

[9] Murtaugh, P.A. and Fisher, L.D. (1990). Bivariate binary models of efficacy and toxicity in dose-renging trials. *Communications in Statistics - Theory and Methods*, **19**, 2003-2020.