



RESEARCH ARTICLE

Correlation Statistic for Assessing a Diagnostic Value Based on a Bivariate Bernoulli Distribution with Application to a Retrospective Single-Center Clinical Study

Xiaoyong Wu, Jayesh P Rai², Guohua Yan³, Srivastava DK⁴ and Shesh N Rai^{1,5*}

¹Biostatistics Shared Facility, University of Louisville, Louisville, KY, USA

²Division of Cardiology, University of Louisville, Louisville, KY, USA

³Department of Mathematics & Statistics, University of New Brunswick, Fredericton, NB, Canada

⁴Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA

⁵Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA

*Corresponding author: Shesh N Rai, Biostatistics Shared Facility, Department of Bioinformatics and Biostatistics, JG Brown Cancer Center, University of Louisville, Clinical Translational Research Building, 505 Hancock St., Louisville, KY, USA, 40202, Tel: 502-852-4030, Fax: 502-852-7979



Abstract

Assessing the performance of diagnostic tests for repeated binary measurements is very important in clinical trials and diagnostic medicine. The gastric-emptying studies [1,2] involve 4-hourly measurements of emptying. Empirical results suggest that the gastric emptying results over time are correlated and therefore failure to consider the correlation in statistical measures may not produce satisfactory assessments. If these measurements are highly correlated/associated, early gastric emptying can be used to predict late gastric emptying which offers an opportunity to reduce the number of repeats. Although the likelihood ratio has been the most widely used statistical measures for assessing the test performance, the choice of the compromise between the likelihood ratio positive and the likelihood ratio negative needs investigators' clinical experience. In this article, we propose a correlation statistic to assess the performance of the test for repeated binary measurements. The statistical measure aims to identify an acceptable balance between sensitivity and specificity and produce a unique assessment for the test performance by using the estimator of the correlation coefficient based on a correlated bivariate Bernoulli distribution. Simulation studies are conducted to compare our proposed statistical measure to other measures. The simulated results indicate that our proposed statistical measure performs well. Using our theoretical approach, we find that the greatest correlation statistic always corresponds to the smallest average absolute likelihood ratio and are approximately close to the

greatest Youden's index. We further analyze a real gastric emptying data set and identify a threshold using our proposed statistical measure.

Keywords

Repeated binary measurements, Bivariate bernoulli distribution, Correlation statistic, Gastric emptying

Introduction

A diagnostic test is an approach used to gather clinical information for making a clinical diagnosis. Assessment of diagnostic tests, especially the likelihood ratios, is very important in clinical trials and diagnostic medicine [3-6]. There exist two versions of likelihood ratios known as positive likelihood ratio and negative likelihood ratio. A positive likelihood ratio is the ratio of the probability that an individual with the disease tests positive to the probability that an individual without the disease tests positive; a negative likelihood ratio is the ratio of the probability that an individual with the disease tests negative to the probability that an individual without the disease tests negative. An asymptotic hypothesis test is constructed to compare the positive and negative likelihood ratios [7]. The likelihood ratios have

been paid much attention to measures diagnostic tests for repeated binary outcomes [1,8-10].

Although the likelihood ratios have their many advantages in clinical findings, laboratory tests, imaging studies, etc, they have some limitations [11-16]. Furthermore, identification of an appropriate compromise between the likelihood ratio positive and the likelihood ratio negative may depend on the investigators' knowledge and experience in clinical trials. Youden's index is widely used as a single statistic to evaluate different test algorithms [17], but it may not perform well when a subject has a low prevalence.

Considering the limitations, we propose a correlation statistic based on a bivariate Bernoulli distribution to assess the diagnostic tests for repeated binary outcomes. A bivariate Bernoulli distribution for binary variables was studied [18], but the authors did not give an explicit expression of bivariate distribution. In contrast with the likelihood ratio positive and the likelihood ratio negative, our correlation statistic strikes a sort of balance between sensitivity and specificity. It is based on the difference between the chance of two independent subjects with both true positive and true negative and chance of two independent subjects with both false positive and false negative.

This paper is organized as follows. In Section 2, we present a new form of bivariate Bernoulli distribution with an explicit expression and give the maximum likelihood estimators of the parameters. The correlation statistics proposed as a measure for the evaluation of the test performance in Section 3.

Through the simulated studies in Section 4, we demonstrate that the estimators perform well and that our proposed statistical measure is advantageous when the prevalence is low. The advantage of our statistical measure is further demonstrated by analyzing a data set on gastric emptying in Section 5. Some discussion is given in Section 6. Lemmas and theorems and their proofs are provided in the [Appendix](#).

Bivariate Bernoulli Distribution and Parameter Estimation

First, we introduce a motivating example for our paper. Delayed gastric emptying results from a variety of chemical and functional etiologies. Gastric emptying studies are the standard of measuring gastric emptying, the protocol has been standardized [19]. Our aim is to establish the predictors to assess the possibility of decreasing the study time. This study is the retrospective chart review study design from Feb 2009 to May 2011. There are 600 subjects in our study. Their gastric emptying is the percentage of remaining radioactivity in the stomach, measured at 0, 1, 2, 3 and 4 hours after ingestion.

According to the standard protocol, a gastric

emptying result for a subject is abnormal (or normal) at 2 and 4 hours if the percentage of remaining radioactivity is greater than 60% and 10% (or less), respectively. Here the gastric emptying results measured at 4 hours are used as subjects' actual results while the gastric emptying results measured at 2 hours are used as predicted results. The gastric emptying results at 2 hours and at 4 hours on the same subject are repeated measurements over time and then are often correlated to each other.

This motivates us to develop a unique and feasible statistical measure that accounts for this correlation in the assessment of test performance. Let Y_1 and Y_2 be the actual result and predicted result for a subject, respectively, where each of the actual and predicted results is either positive or negative. For simplicity, let 1 and 0 denote a positive result and a negative result, respectively. As we've known that if one of $p_1 = P(Y_1 = 1)$ and $p_2 = P(Y_2 = 1)$ is 0 or 1, then Y_1 and Y_2 are independent. Here to avoid the degenerate cases, we assume from now on that $0 < p_1 < 1$ and $0 < p_2 < 1$. The correlation coefficient of Y_1 and Y_2 is given by

$$\rho = \frac{p_{11} - p_1 p_2}{\sqrt{p_1(1-p_1)p_2(1-p_2)}} \quad (1)$$

Or equivalently,

$$\rho = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{p_1(1-p_1)p_2(1-p_2)}} \quad (2)$$

Where $p_{y_1 y_2} = P(Y_1 = y_1, Y_2 = y_2)$, $y_1, y_2 = 0, 1$ and $p_{11} - p_1 p_2$ is the probability that two independent subjects have both true positive and true negative and probability that two independent subjects have both false positive and false negative. The correlation coefficient may then be viewed as a standardized difference on a certain range. The following lemma shows the range for the correlation coefficient.

LEMMA 1.

$$(i) \quad -\min \left\{ \sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}, \sqrt{\frac{(1-p_1)(1-p_2)}{p_1 p_2}} \right\} \leq \rho \leq \min \left\{ \sqrt{\frac{p_1(1-p_2)}{(1-p_1)p_2}}, \sqrt{\frac{(1-p_1)p_2}{p_1(1-p_2)}} \right\}$$

$$(ii) \quad \rho = \min \left\{ \sqrt{\frac{p_1(1-p_2)}{(1-p_1)p_2}}, \sqrt{\frac{(1-p_1)p_2}{p_1(1-p_2)}} \right\} \text{ if and only if } p_{11} = \min \{p_1, p_2\}.$$

In particular, $\rho = 1$ if and only if $p_{11} = p_1 = p_2$.

$$(iii) \quad \rho = -\min \left\{ \sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}, \sqrt{\frac{(1-p_1)(1-p_2)}{p_1 p_2}} \right\} \text{ if and only if } p_{11} = \max \{0, p_1 + p_2 - 1\}.$$

In particular, $\rho = -1$ if and only if $p_{10} = p_1 = 1 - p_2$.

$$(iv) \quad \rho = 0 \text{ if and only if } Y_1 \text{ and } Y_2 \text{ are independent.}$$

Proof. See the [Appendix](#). The results from Lemma 1 inspire us to consider the following bivariate function

$$p(y_1, y_2, p_1, p_2, \rho) = p_1^{y_1} (1-p_1)^{1-y_1} p_2^{y_2} (1-p_2)^{1-y_2} + (-1)^{y_1+y_2} \rho \sqrt{p_1(1-p_1)p_2(1-p_2)}, \quad (3)$$

Where $y_1, y_2 = 0, 1$, p_1, p_2 and ρ are defined as above. To make this bivariate function to become a probability distribution and use it statistical inference, we need to show the following theorem holds true.

Theorem 1

This bivariate function given in equation (3) is a well-defined and identifiable probability distribution.

The proof is given in the [Appendix](#). Any two correlated categorical variables may be assumed to have the bivariate probability distribution with the parameters p_1, p_2 and ρ , which are the marginal probabilities of the variables and their correlation coefficient. Such distribution may be viewed as a generalization of the well-known univariate Bernoulli distribution to 2-dimension and we call it as a bivariate Bernoulli distribution denoted by $(Y_1, Y_2) \sim BB(p_1, p_2, \rho)$.

Now we present the maximum likelihood estimators of the parameters p_1, p_2 and ρ . Let $(y_{i1}, y_{i2}), i = 1, \dots, n$, be the sample from the bivariate Bernoulli distribution $BB(p_1, p_2, \rho)$. The likelihood function is given by

$$L = \prod_{i=1}^n p(y_{i1}, y_{i2}, p_1, p_2, \rho) = (p(1,1,p_1,p_2,\rho))^{N_{11}} (p(1,0,p_1,p_2,\rho))^{N_{10}} (p(0,1,p_1,p_2,\rho))^{N_{01}} (p(0,0,p_1,p_2,\rho))^{N_{00}}, \quad (4)$$

Where N_{11}, N_{10}, N_{01} and N_{00} are the numbers of subjects with true positive, false negative, false positive and true negative, respectively. The standard maximum likelihood methods often involve optimization of the log-likelihood function. The optimization is reached by taking the derivative of the log-likelihood function with respect to parameters. The Newton-Raphson iteration is frequently used to find the maximum likelihood estimators. However, this iteration finds an approximation to the maximum likelihood estimators. More importantly, it is difficult to show this iteration converges and lead to time-consuming computation. All these motivate us to find an explicit and closed-form expression for the maximum likelihood estimators. Now we use the following lemma to derive the maximum likelihood estimators.

LEMMA 2

Suppose that k is a positive integer and for $t = 1, \dots, k, j_t$ is a nonnegative integer such that $\sum_{t=1}^k j_t < n$. Then

$$(\theta_1, \dots, \theta_k) \in \Theta_k \left(\prod_{t=1}^k \theta_t^{j_t} \right) \left(1 - \sum_{t=1}^k \theta_t \right)^{n - \sum_{t=1}^k j_t} = \left(\prod_{t=1}^k \left(\frac{j_t}{n} \right)^{j_t} \right) \left(1 - \frac{\sum_{t=1}^k j_t}{n} \right)^{n - \sum_{t=1}^k j_t}, \quad (5)$$

Where

$$\Theta_k = \left\{ (\theta_1, \dots, \theta_k) : \sum_{t=1}^k \theta_t < 1, 0 < \theta_t < 1, t = 1, \dots, k \right\}.$$

The proof is given in the [Appendix](#). From this lemma, we can derive the following theorem.

Theorem 2

The maximum likelihood estimators of the

parameters are $\hat{p}_1 = \frac{N_{11} + N_{10}}{n}$, $\hat{p}_2 = \frac{N_{11} + N_{01}}{n}$ and

$$\hat{\rho} = \frac{N_{11}N_{00} - N_{10}N_{01}}{\sqrt{(N_{11} + N_{10})(N_{01} + N_{00})(N_{11} + N_{01})(N_{10} + N_{00})}}, \quad \text{where } N_{11},$$

N_{10}, N_{01} and N_{00} are the numbers of subjects with true positive, false negative, false positive and true negative, respectively.

The proof is given in the [Appendix](#).

Correlation Statistic

The repeated binary outcomes are often correlated and failure to consider their correlation in statistical measures may lead to poor assessments for association analysis. Now we see that correlation coefficient determines the strength and direction of association between two categorical variables.

Let Y_1 and Y_2 be the actual and predicted results for a subject, respectively. Assume that (Y_1, Y_2) has bivariate Bernoulli distribution $BB(p_1, p_2, \rho)$, where $p_1 = P(Y_1 = 1)$, $p_2 = P(Y_2 = 1)$ and ρ is the correlation coefficient. We discuss correlation coefficient in the following three cases. (i). A zero-correlation coefficient implies that $P(Y_2 = 1 | Y_1 = 1) = P(Y_2 = 1 | Y_1 = 0) = P(Y_2 = 1)$, which suggests any actual result has no effect on a positive predicted result. Also, a zero-correlation coefficient implies that $P(Y_2 = 0 | Y_1 = 1) = P(Y_2 = 0 | Y_1 = 0) = P(Y_2 = 0)$, which suggests any actual result has no effect on a negative predicted result. Thus, a zero-correlation coefficient implies that the two results are independent. (ii). A positive correlation coefficient implies that $P(Y_2 = 1 | Y_1 = 1) = P(Y_2 = 1 | Y_1 = 1) > P(Y_2 = 1)$ and $P(Y_2 = 0 | Y_1 = 0) > P(Y_2 = 0)$, which suggests the actual result has an increased effect on the predicted result when the two results is the same. Also, a positive correlation coefficient implies that $P(Y_2 = 1 | Y_1 = 0) < P(Y_2 = 1)$ and $P(Y_2 = 0 | Y_1 = 1) < P(Y_2 = 0)$, which suggests the actual result has a decreased effect on the predicted result when the two results is different. Thus, a positive correlation coefficient implies that the two results are positively associated. (iii). A negative correlation coefficient implies that $P(Y_2 = 1 | Y_1 = 1) < P(Y_2 = 1)$ and $P(Y_2 = 0 | Y_1 = 0) < P(Y_2 = 0)$, which suggests the actual result has an decreased effect on the predicted result when the two results is the same. Also, a negative correlation coefficient implies that $P(Y_2 = 1 | Y_1 = 0) > P(Y_2 = 1)$ and $P(Y_2 = 0 | Y_1 = 1) > P(Y_2 = 0)$, which suggests the actual result has an increased effect on the predicted result when the two results is different. Thus, a negative correlation coefficient implies that the two results are negatively associated. Therefore, the correlation coefficient provides the direction of association between the actual and predicted results.

However, for a nonzero correlation coefficient, whether the correlation coefficient is positive or negative, its absolute value measures the strength of the association between the actual and predicted

results. The greater the absolute value of the correlation coefficient is, the stronger the association between the actual and predicted results becomes. Thus, the correlation coefficient provides the strength and direction of association between categorical variables.

Since the correlation coefficient is unknown, we use its maximum likelihood estimator

$$CS = \frac{N_{11}N_{00} - N_{10}N_{01}}{\sqrt{(N_{11} + N_{10})(N_{01} + N_{00})(N_{11} + N_{01})(N_{10} + N_{00})}}, \quad (6)$$

as a statistical measure for the performance of the diagnostic test and call it as correlation statistic. Such measure may have a potential to evaluate the association between correlated binary variables.

Compared with other measures such as likelihood ratio positive, likelihood ratio negative, average absolute likelihood ratio and Youden's index, our CS shows advantages for assessing the test performance. As we know, the CS is a statistical measure for estimating correlation between the actual result and predicted result from a general bivariate Bernoulli model. Therefore, the CS has potential to provide a direct evaluation of association between true and predicted results. However, other measures depend on sensitivity and specificity, which is an indirect evaluation of association between true and predicted results. These may lead to a conclusion that the CS would achieve better prediction of true result than other measures,

Table 1: Simulated means, standard errors (SE) and mean squared errors (MSE).

Scenario	n	Parameter	MLE	SE	MSE
(a)	60	$p_1 = 0.2$	0.207	0.052	0.003
		$p_2 = 0.4$	0.403	0.064	0.004
		$\rho = 0.3$	0.291	0.124	0.016
	120	$p_1 = 0.2$	0.202	0.037	0.001
		$p_2 = 0.4$	0.399	0.044	0.002
		$\rho = 0.3$	0.301	0.084	0.007
	180	$p_1 = 0.2$	0.201	0.031	0.001
		$p_2 = 0.4$	0.401	0.036	0.001
		$\rho = 0.3$	0.298	0.071	0.005
(b)	60	$p_1 = 0.6$	0.603	0.062	0.004
		$p_2 = 0.4$	0.399	0.063	0.004
		$\rho = 0.6$	0.599	0.093	0.009
	120	$p_1 = 0.6$	0.601	0.044	0.002
		$p_2 = 0.4$	0.402	0.047	0.002
		$\rho = 0.6$	0.593	0.064	0.004
	180	$p_1 = 0.6$	0.6	0.038	0.001
		$p_2 = 0.4$	0.402	0.036	0.001
		$\rho = 0.6$	0.593	0.049	0.002
(c)	60	$p_1 = 0.2$	0.202	0.049	0.002
		$p_2 = 0.8$	0.802	0.048	0.002
		$\rho = -0.8$	0.762	0.113	0.014

which is shown from simulation studies in Section 4.

Simulation Studies

To evaluate the performance of the proposed \hat{P}_1 , \hat{P}_2 and \hat{P} , we conduct a moderate scale simulation experiment. We generate $n = 60, 120, 150$ observations from the bivariate Bernoulli distribution:

$$P_{y_1, y_2} = P_1^{y_1} (1 - P_1)^{1 - y_1} P_2^{y_2} (1 - P_2)^{1 - y_2} + (-1)^{y_1 + y_2} \rho \sqrt{P_1(1 - P_1)P_2(1 - P_2)}, \quad y_1, y_2 = 0, 1.$$

We consider the following four scenarios: (a) $p_1 = 0.2, p_2 = 0.4, \rho = 0.3$. (b) $p_1 = 0.6, p_2 = 0.8, \rho = 0.5$. (c) $p_1 = 0.2, p_2 = 0.4, \rho = -0.3$. (d) $p_1 = 0.6, p_2 = 0.8, \rho = -0.2$. We generate 1000 data sets in each of the configurations. The simulation results on the simulated means, standard errors and mean squared errors are summarized in Table 1.

To compare the performance between our proposed statistical measure and the likelihood ratio, we conduct a moderate scale simulation experiment. We generate $n = 300$ in (a) and 600 in (b) observations from the bivariate normal distribution:

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{(y_1-\mu_1)^2}{\sigma_1^2} - \frac{2r(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right)\right\}.$$

We consider the following two scenarios: (a) $\mu_1 = 8, \mu_2 = 8, \sigma_1 = 3, \sigma_2 = 3, r = 0.8$. (b) $\mu_1 = 30, \mu_2 = 30, \sigma_1 = 6, \sigma_2 = 6, r = 0.9$. For each of (a) and (b), we classify Y_1 and Y_2 into abnormal or normal results based on the threshold, that is, Y_1 (or Y_2) is abnormal if it is less than or equal to the threshold and otherwise, it is normal. Also, we treat Y_1 as the predictor and treat Y_2 as actual status. We generate 1000 data sets in each of the configurations. For each simulation run, the thresholds for Y_1 are the quantiles which correspond to the probabilities 10%, 20%, ..., 90%, respectively while the thresholds for Y_2 are the quantiles which correspond to the probabilities 30% and 60%, respectively. We use our proposed method to identify the appropriate threshold which leads a good predictor. The simulation results on the sensitivity and specificity (SENS and SPEC), likelihood

	120	$p_1 = 0.2$	0.199	0.037	0.001
		$p_2 = 0.8$	0.801	0.036	0.001
		$\rho = -0.8$	-0.787	0.069	0.005
	180	$p_1 = 0.2$	0.201	0.031	0.001
		$p_2 = 0.8$	0.799	0.029	0.001
		$\rho = -0.8$	-0.798	0.053	0.003
(d)	60	$p_1 = 0.6$	0.603	0.06	0.004
		$p_2 = 0.2$	0.207	0.055	0.003
		$\rho = -0.5$	-0.489	0.112	0.012
	120	$p_1 = 0.6$	0.6	0.045	0.002
		$p_2 = 0.2$	0.204	0.035	0.001
		$\rho = -0.5$	-0.493	0.072	0.005
	180	$p_1 = 0.6$	0.6	0.039	0.001
		$p_2 = 0.2$	0.201	0.029	0.001
		$\rho = -0.5$	-0.495	0.057	0.003

Table 2: Various measures in a simulated study to identify the appropriate threshold where the best value for each measure is boldfaced.

Scenario	n	THRES 1	THRES 2	SENS	SPEC	LRP	LRN	AALR	YOU DEN	CS
(a)	300	6.43	4.16	0.767	0.905	Inf	0.258	Inf	0.673	0.441
			5.48	0.825	0.799	4.408	0.219	4.46	0.624	0.540
			6.43	0.873	0.703	3.030	0.181	3.842	0.577	0.576
			7.24	0.913	0.616	2.411	0.142	4.47	0.529	0.566
			8.00	0.942	0.544	2.083	0.106	Inf	0.486	0.531
			8.76	0.965	0.478	1.857	0.073	Inf	0.443	0.473
			9.57	0.982	0.424	1.711	0.042	Inf	0.406	0.406
			10.52	0.993	0.374	1.591	0.018	Inf	0.367	0.321
			11.84	0.998	0.332	1.498	0.005	Inf	0.331	0.216
		8.76	4.16	0.443	0.995	Inf	0.560	Inf	0.438	0.267
			5.48	0.494	0.977	Inf	0.518	Inf	0.471	0.385
			6.43	0.554	0.954	Inf	0.468	Inf	0.507	0.474
			7.24	0.616	0.921	8.622	0.417	6.145	0.537	0.536
			8.00	0.682	0.882	6.169	0.36	4.499	0.565	0.576
			8.76	0.753	0.837	4.751	0.295	3.988	0.590	0.590
			9.57	0.823	0.782	3.838	0.226	4.426	0.605	0.566
			10.52	0.895	0.723	3.267	0.146	Inf	0.618	0.504
			11.84	0.959	0.663	2.865	0.063	Inf	0.621	0.38
	600	6.43	4.16	0.767	0.906	Inf	0.258	Inf	0.673	0.440
			5.48	0.825	0.801	4.286	0.219	4.354	0.626	0.547
			6.43	0.873	0.705	3.000	0.180	3.796	0.578	0.578
			7.24	0.911	0.619	2.409	0.144	4.321	0.530	0.566
			8.00	0.942	0.544	2.074	0.106	6.052	0.486	0.530
			8.76	0.966	0.478	1.856	0.071	Inf	0.444	0.474
			9.57	0.982	0.420	1.697	0.043	Inf	0.402	0.402
			10.52	0.993	0.373	1.585	0.019	Inf	0.366	0.32
			11.84	0.998	0.334	1.500	0.005	Inf	0.332	0.217
		8.76	4.16	0.444	0.994	Inf	0.559	Inf	0.439	0.268
			5.48	0.495	0.978	Inf	0.516	Inf	0.474	0.387
			6.43	0.552	0.954	13.649	0.470	10.197	0.506	0.473
			7.24	0.614	0.923	8.389	0.419	5.993	0.537	0.537
			8.00	0.680	0.882	5.911	0.363	4.345	0.562	0.574
			8.76	0.754	0.834	4.602	0.295	3.903	0.588	0.587
			9.57	0.824	0.782	3.812	0.225	4.332	0.606	0.567
			10.52	0.895	0.723	3.255	0.145	6.675	0.619	0.505
			11.84	0.959	0.662	2.848	0.061	Inf	0.621	0.380
(b)	300	26.85	22.31	0.776	0.979	Inf	0.229	Inf	0.755	0.495
			24.95	0.85	0.900	Inf	0.167	Inf	0.750	0.654
			26.85	0.911	0.792	4.58	0.113	5.991	0.702	0.702
			28.48	0.954	0.677	3.014	0.069	8.579	0.631	0.675
			30.00	0.979	0.579	2.346	0.037	Inf	0.558	0.609
			31.52	0.992	0.495	1.976	0.015	Inf	0.488	0.521
			33.15	0.998	0.429	1.754	0.005	Inf	0.427	0.426
			35.05	1.000	0.374	1.602	0.000	Inf	0.374	0.327
			37.69	1.000	0.331	1.498	0.000	Inf	0.331	0.217
		31.52	22.31	0.444	1.000	Inf	0.556	Inf	0.444	0.271
			24.95	0.498	0.998	Inf	0.503	Inf	0.496	0.404

			26.85	0.568	0.990	Inf	0.436	Inf	0.558	0.521
			28.48	0.648	0.971	Inf	0.362	Inf	0.619	0.618
			30.00	0.736	0.936	13.001	0.282	8.323	0.672	0.686
			31.52	0.827	0.885	7.557	0.196	6.220	0.712	0.712
			33.15	0.906	0.817	5.067	0.115	8.423	0.723	0.676
			35.05	0.966	0.741	3.777	0.046	Inf	0.707	0.577
			37.69	0.995	0.667	3.009	0.007	Inf	0.662	0.404
	600	26.85	22.31	0.776	0.978	Inf	0.230	Inf	0.754	0.493
			24.95	0.851	0.900	9.283	0.166	8.643	0.751	0.656
			26.85	0.911	0.792	4.484	0.112	5.877	0.703	0.703
			28.48	0.953	0.680	3.01	0.069	7.993	0.633	0.677
			30.00	0.979	0.579	2.334	0.036	18.319	0.558	0.608
			31.52	0.993	0.495	1.971	0.015	Inf	0.487	0.521
			33.15	0.998	0.426	1.742	0.004	Inf	0.424	0.424
			35.05	1.000	0.374	1.599	0.001	Inf	0.374	0.327
			37.69	1.000	0.333	1.501	0.000	Inf	0.333	0.217
		31.52	22.31	0.445	1.000	Inf	0.556	Inf	0.444	0.271
			24.95	0.500	0.998	Inf	0.501	Inf	0.498	0.406
			26.85	0.568	0.990	Inf	0.437	Inf	0.558	0.522
			28.48	0.647	0.971	Inf	0.364	Inf	0.618	0.618
			30.00	0.734	0.936	12.069	0.284	7.818	0.670	0.684
			31.52	0.828	0.884	7.285	0.195	6.061	0.712	0.711
			33.15	0.907	0.817	5.011	0.114	8.058	0.724	0.677
			35.05	0.966	0.741	3.750	0.046	Inf	0.707	0.577
			37.69	0.995	0.666	2.992	0.007	Inf	0.661	0.405

ratio positive and likelihood ratio negative (LRP and LRN), average absolute likelihood ratio (AALR), Youden's index (YOUDEN), correlation statistic (CS) and assigned threshold values (THRES 1 and THRES 2) are summarized in Table 2. In the table, an entry is the average over 1000 simulations for the corresponding measure. For the CS, YOUDEN and LRP, the greater the measure becomes, the better the test is. But for the measures LRN and AALR, the smaller the measure becomes, the better the test is. From Table 2, we can see that in all the simulated cases, the CS is able to identify the correct threshold. Also, the CS may draw an acceptable balance between sensitivity and specificity. The greatest CS approximately corresponds to the greatest sensitivity (or specificity) given that specificity (or sensitivity) is greater than or equal to around 0.8. The well-known YOUDEN which is defined as SENS + SPEC-1, is often used to identify the threshold. However, it does not perform well to make prediction especially for the cases (Scenario (a), n (300), THRES 1 (8.76)) and (Scenario (a), n (600), THRES 1 (8.76)). It should be noted that the greatest YOUDEN does not achieve a good balance between sensitivity and specificity. For example, for the case (Scenario (a), n (300), THRES 1 (8.76)), the greatest YOUDEN approximately corresponds to a relatively low specificity (0.663). If we use likelihood ratio positive or likelihood ratio negative to identify the threshold individually, it misses the correct value in

each case. Also, the greatest LRP or smallest LRN does not achieve an acceptable balance between sensitivity and specificity in many cases. It still remains challenge to compromise between the likelihood ratio positive and the likelihood ratio negative for balance between sensitivity and specificity. One disadvantage of the likelihood ratio positive is that its value may be infinity in some extreme cases. Although the smallest AALR corresponds to the greatest CS, surprisingly, the average absolute likelihood ratio, which is conceptually an average of likelihood ratios over a population/sample, is also able to identify the correct threshold, if we are willing to use it in a counterintuitive way: The smaller the measure, the better the test.

Application to Gastric Emptying Data

Gastric emptying study is the standard of measuring gastric emptying and the aim of this study is to establish the predictors and to assess the possibility of shortening the study time [2,19,20].

In our data analysis, we consider the gastric emptying at 2 hours as predictor and the gastric emptying at 4 hours as actual status. Although the gastric emptying results at 2 hours and at 4 hours are count variables, we classify them based on the thresholds. The gastric emptying results at 4 hours is defined as abnormal if the percentage of remaining radioactivity is greater than 10% and otherwise it is

Table 3: Comparison between the correlation statistic and likelihood ratio measure in the gastric emptying study.

THRES	SENS	SPEC	LRP	LRN	AALR	YOUNDEN	CS
30%	0.952	0.543	2.084	0.088	5.997	0.495	0.432
40%	0.925	0.687	2.951	0.109	6.293	0.612	0.528
50%	0.823	0.819	4.547	0.216	4.602	0.642	0.584
60%	0.633	0.922	8.188	0.398	3.727	0.555	0.583
70%	0.374	0.982	21.186	0.637	3.634	0.356	0.5
80%	0.211	0.996	47.765	0.792	3.807	0.207	0.389

defined as normal. For the gastric emptying results at 2 hours, the possible thresholds are 30%, 40%, ..., 80%. The gastric emptying results at 2 hours are defined as abnormal if it is greater than a threshold and otherwise it is defined as normal. We use our proposed method to identify the threshold of 50% to classify the gastric emptying results at 2 hours. The results about sensitivity and specificity shows the threshold of 50% is feasible. Some simulation results are summarized in Table 3 where THRES is the threshold value. The area under the ROC curve (AUC) is the average sensitivity of the biomarker over the range of specificities, which is often used as an evaluation of the overall performance of the biomarker. Our data show the estimated AUC is 0.816, which provides evidence that our correlation statistic is useful for correctly classifying subjects with predicted positive and subjects with predicted negative. From the table and resulting AUC value above, we conclude the threshold of 50% is best choice. As we have known, the sensitivity and specificity are true positive rate and are true negative rate, respectively. The positive likelihood ratio considers only predicted positive and ignore predicted negative while the negative likelihood ratio considers only predicted negative and ignore predicted positive. However, our correlation statistic provides comprehensive and powerful measures and has a potential to correctly and reliably assess a diagnostic test. It can be used as alternative to average absolute likelihood ratio and Youden's index when they are not applicable. Also, our correlation statistic could provide the direction of association between two categorical variables.

Discussion

Due to the importance of assessment of test in medicine and clinical trials, many researchers have paid attention to statistical measures. The likelihood ratio positive and the likelihood ratio negative are the most commonly used measures of diagnostic test. However, the likelihood ratio positive and the likelihood ratio negative often present the reverse tendencies and to determine the compromise between the likelihood ratio positive and the likelihood ratio negative relies on the investigators' sound knowledge and experience in medicine and clinical trials. In this article, we proposed a statistical measure. Compared with the likelihood ratio positive and the likelihood ratio negative, our measure could have a unique solution and produce an acceptable

balance between sensitivity and specificity.

We should also point out that the bivariate Bernoulli distribution for correlated binary variables focuses on construction of statistical measure for test performance, but it is important to explore the applicability of our proposed bivariate Bernoulli distribution in bivariate interobserver agreement since the correlation coefficient can be used to indicate the direction and strength of the difference between the chance of two uncorrelated individuals with both true positive and negative rates and the chance of the two individuals with false positive and negative rates. It can be used as alternative to average absolute likelihood ratio (AALR) and Youden's index when they are not applicable. In addition, we can explore the applicability of our proposed bivariate Bernoulli distribution in the other literature such as logistic regression for two binary outcomes, repeated measurements, etc., and extend the bivariate distribution to higher-dimension.

Acknowledgements

Dr. Rai is grateful to generous support from Dr. DM Miller, Director James Graham

Brown Cancer Center and Wendell Cherry Chair in Clinical Trial Research.

Conflict of Interest

The authors have declared no conflict of interest.

References

1. Bonta D, Lee H, Ziessman H (2011) Shortening the 4-hour gastric-emptying protocol. *Clinical Nuclear Medicine* 36: 283-285.
2. Bhatt G, Li XF, Wu XY, Rai SN, Civelek AC (2014) Factors affecting 4-Hour Solid Gastric Emptying Study - Evaluating the potential to reduce the study time. Report.
3. Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, et al. (2006) Evaluation of diagnostic tests for infectious diseases: General principles. *Nature Reviews Microbiology* 4: S21-S31.
4. Rudwaleit M, Sieper J (2007) A case of axial undifferentiated spondyloarthritis diagnosis and management. *Nature Clinical Practice Rheumatology* 3: 298-303.
5. deFranchis R (2009) Cirrhosis: Screening for esophageal varices. *Nature Reviews Gastroenterology & Hepatology* 6: 449-450.
6. Peeling RW, Artsob H, Pelegrino JL, Buchy P, Cardosa MJ,

- et al. (2010) Evaluation of diagnostic tests: dengue. *Nature Reviews Microbiology* 8: S30-S38.
7. Luts J, Nofuentes JAR, Luna JDL, Van S (2011) Asymptotic hypothesis test to compare likelihood ratios of multiple diagnostic tests in unpaired designs. *Journal of Statistical.*
 8. Fink HA, Lederle FA, Roth CS, Bowles CA, Nelson DB, et al. (2000) The accuracy of physical examination to detect abdominal aortic aneurysm. *Archives of Internal Medicine* 160: 833-836.
 9. Hahne AJ, Keating JL, Wilson SC (2004) Do within-session changes in pain intensity and range of motion predict between-session changes in patients with low back pain? *Australian Journal of Physiotherapy* 50: 17-23.
 10. Green DM, Cox CL, Zhu L, Krull KR, Srivastava DK, et al. (2012) Risk factors for obesity in adult survivors of childhood cancer: A report from the Childhood Cancer Survivor Study. *Journal of Clinical Oncology* 30: 246-255.
 11. Langlotz CP (2003) Fundamental measures of diagnostic examination performance: usefulness for clinical decision making and research. *Radiology* 228: 3-9.
 12. Grimes DA, Schulz KF (2005) Refining clinical diagnosis with likelihood ratios. *Lancet* 365: 1500-1505.
 13. Liang H (2006) Checking linearity of non-parametric component in partially linear models with an application in systemic inflammatory response syndrome study. *Statistical Methods in Medical Research* 15: 273-284.
 14. Koul HL, Song W (2010) Model checking in partial linear regression models with Berkson measurement errors. *Statistica Sinica* 20: 1551-1579.
 15. McGee SR (2002) Simplifying likelihood ratios. *Journal of General Internal Medicine* 17: 647-650.
 16. Marill KA (2012) Diagnostic testing and the average absolute likelihood ratio: application to diagnosing wide QRS complex tachycardia and other ED diseases. *American Journal of Emergency Medicine* 30: 1895-1906.
 17. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF (2008) Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal* 50: 419-430.
 18. Marshall AW, Olkin I (1985) A family of bivariate distributions generated by Bernoulli distribution. *Journal of the American Statistical Association* 80: 332-338.
 19. Tougas G, Eaker EY, Abell TL, Abrahamsson H, Boivin M, et al. (2000) Assessment of gastric emptying using a low fat meal: Establishment of international control values. *The American Journal of Gastroenterology* 95: 1456-1462.
 20. Abell TL, Camilleri M, Donohoe K, Hasler WL, Lin HC, et al. (2008) Consensus recommendations for gastric emptying scintigraphy: A joint report of the American Neurogastroenterology and Motility Society and the Society of Nuclear Medicine. *Journal of Nuclear Medicine Technology* 103: 753-763.

Appendix

Proof of Lemma 1.

(i). Since the correlation coefficient is given by

$$\rho = \frac{p_{11} - p_1 p_2}{\sqrt{p_1(1-p_1)p_2(1-p_2)}},$$

It follows that

$$\rho \leq \frac{\min\{p_1 p_2\} - p_1 p_2}{\sqrt{p_1(1-p_1)p_2(1-p_2)}} = \min \left\{ \sqrt{\frac{p_1(1-p_2)}{(1-p_1)p_2}}, \sqrt{\frac{(1-p_1)p_2}{p_1(1-p_2)}} \right\}.$$

It can be seen that

$$\rho \geq \min \left\{ -\frac{p_1 p_2}{\sqrt{p_1(1-p_1)p_2(1-p_2)}}, \frac{p_1 + p_2 - 1 - p_1 p_2}{\sqrt{p_1(1-p_1)p_2(1-p_2)}} \right\} = -\min \left\{ \sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}, \sqrt{\frac{(1-p_1)(1-p_2)}{p_1 p_2}} \right\}.$$

(ii). It is evident that $\rho = \min \left\{ \sqrt{\frac{p_1(1-p_2)}{(1-p_1)p_2}}, \sqrt{\frac{(1-p_1)p_2}{p_1(1-p_2)}} \right\}$ holds if and only if

$$p_{11} = p_1 p_2 + \sqrt{p_1(1-p_1)p_2(1-p_2)} \times \min \left\{ \sqrt{\frac{p_1(1-p_2)}{(1-p_1)p_2}}, \sqrt{\frac{(1-p_1)p_2}{p_1(1-p_2)}} \right\} = \min\{p_1, p_2\}.$$

In particular, it is evident that $\rho = 1$ holds if $p_{11} = p_1 = p_2$. Conversely, if $\rho = 1$ holds, then

$$p_{11} = p_1 p_2 + \sqrt{p_1(1-p_1)p_2(1-p_2)},$$

Which implies that $p_1 p_2 + \sqrt{p_1(1-p_1)p_2(1-p_2)} \leq \min\{p_1, p_2\}$, i.e.,

$$\sqrt{p_1(1-p_1)p_2(1-p_2)} \in \{p_1(1-p_2), (1-p_1)p_2\}.$$

This is equivalent to $p_2 \leq p_1$ and $p_1 \leq p_2$, which imply that $p_{11} = p_1 = p_2$.

(iii). It is evident that $\rho = -\min \left\{ \sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}, \sqrt{\frac{(1-p_1)(1-p_2)}{p_1 p_2}} \right\}$ holds if and only if

$$p_{11} = p_1 p_2 - \sqrt{p_1(1-p_1)p_2(1-p_2)} \times \min \left\{ \sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}, \sqrt{\frac{(1-p_1)(1-p_2)}{p_1 p_2}} \right\} = \max\{0, p_1 + p_2 - 1\}.$$

In particular, it is evident that $\rho = -1$ holds if $p_{10} = p_1 = 1 - p_2$.

Conversely, if $\rho = -1$ holds, then

$$p_{11} = p_1 p_2 - \sqrt{p_1(1-p_1)p_2(1-p_2)}$$

Note that $p_{11} \geq p_1 + p_2 - 1$, it follows that

$$\sqrt{p_1(1-p_1)p_2(1-p_2)} \leq \min\{p_1 p_2 - p_1 - p_2 + 1, p_1 p_2\},$$

or equivalently, $p_1 + p_2 = 1$.

This in turn implies that $p_{11} = 0$ holds and therefore $p_{10} = p_1 - p_{11} = p_1 = 1 - p_2$ holds.

(iv) If Y_1 and Y_2 are independent, then it is easy to see $\rho = 0$. Otherwise, if $\rho = 0$, then $p_{11} = p_1 p_2$.

Therefore, $p_{10} = p_1 - p_{11} = p_1(1-p_2)$, $p_{01} = p_2 - p_{11} = (1-p_1)p_2$ and

$$P_{00} = P_1 - P_{11} = P_1 (1 - P_2), P_{01} = P_2 - P_{11} = (1 - P_1) P_2$$

And $P_{00} = 1 - P_{11} - P_{10} - P_{01} = (1 - P_1) (1 - P_2)$, which imply that Y_1 and Y_2 are independent.

Proof of Lemma 2

We show this lemma by mathematical induction.

(i). Let $g(\alpha) = \alpha^u (1 - \alpha)^v$, where $0 < \alpha < 1, u \geq 0, v \geq 0, u + v \neq 0$. It can easily be shown that

$g(\alpha)$ reaches the maximum at $\alpha = \frac{u}{u + v}$, which means that

$$\theta_1 \in \Theta_1 \theta_1^{j_1} (1 - \theta_1)^{n - j_1} = \left(\frac{j_1}{n}\right)^{j_1} \left(1 - \frac{j_1}{n}\right)^{n - j_1}.$$

Therefore, the formula is valid when $k = 1$.

(ii). Assuming the formula is true for $k = m - 1$, i.e.,

$$(\theta_1, \dots, \theta_{m-1}) \in \Theta_{m-1} \left(\prod_{t=1}^{m-1} \theta_t^{j_t}\right) \left(1 - \sum_{t=1}^{m-1} \theta_t\right)^{n - \sum_{t=1}^{m-1} j_t} = \left(\prod_{t=1}^{m-1} \left(\frac{j_t}{n}\right)^{j_t}\right) \left(1 - \frac{\sum_{t=1}^{m-1} j_t}{n}\right)^{n - \sum_{t=1}^{m-1} j_t}$$

We will show that it is still true for $k = m$.

Obviously,

$$\left(\prod_{t=1}^m \theta_t^{j_t}\right) \left(1 - \sum_{t=1}^m \theta_t\right)^{n - \sum_{t=1}^m j_t} = \left(\prod_{t=1}^{m-1} \theta_t^{j_t}\right) \left(1 - \sum_{t=1}^{m-1} \theta_t\right)^{n - \sum_{t=1}^{m-1} j_t} \times \left(\frac{\theta_m}{1 - \sum_{t=1}^{m-1} \theta_t}\right)^{j_m} \left(1 - \frac{\theta_m}{1 - \sum_{t=1}^{m-1} \theta_t}\right)^{n - \sum_{t=1}^{m-1} j_t}$$

Since $(\theta_1, \dots, \theta_m) \in \Theta_m$, it is easy to see $0 < \left(\frac{\theta_m}{1 - \sum_{t=1}^{m-1} \theta_t}\right) < 1$ and $n - \sum_{t=1}^{m-1} j_t > 0$. According to the Result in (i), we have

$$\left(\frac{\theta_m}{1 - \sum_{t=1}^{m-1} \theta_t}\right)^{j_m} \left(1 - \frac{\theta_m}{1 - \sum_{t=1}^{m-1} \theta_t}\right)^{n - \sum_{t=1}^{m-1} j_t} \leq \left(\frac{j_m}{n - \sum_{t=1}^{m-1} j_t}\right)^{j_m} \left(1 - \frac{j_m}{n - \sum_{t=1}^{m-1} j_t}\right)^{n - \sum_{t=1}^{m-1} j_t}$$

Combining the two results above and the assumption in (ii), we have

$$\left(\prod_{t=1}^m \theta_t^{j_t}\right) \left(1 - \sum_{t=1}^m \theta_t\right)^{n - \sum_{t=1}^m j_t} \leq \left(\prod_{t=1}^m \left(\frac{j_t}{n}\right)^{j_t}\right) \left(1 - \frac{\sum_{t=1}^m j_t}{n}\right)^{n - \sum_{t=1}^m j_t},$$

Or equivalently,

$$(\theta_1, \dots, \theta_m) \in \Theta_m \left(\prod_{t=1}^m \theta_t^{j_t}\right) \left(1 - \sum_{t=1}^m \theta_t\right)^{n - \sum_{t=1}^m j_t} = \left(\prod_{t=1}^m \left(\frac{j_t}{n}\right)^{j_t}\right) \left(1 - \frac{\sum_{t=1}^m j_t}{n}\right)^{n - \sum_{t=1}^m j_t}$$

Which means the formula is valid when $k = m$. Combining the results of parts (i) and (ii), we conclude by mathematical induction that the formula is valid for all positive integers.

Proof of Theorem 1

The function $P(y_1, y_2, p_1, p_2, \rho)$ can specifically be written as

$$P(1, 1, P_1, P_2, \rho) = P_1 P_2 + \rho \sqrt{P_1 (1 - p_1) P_2 (1 - p_2)}$$

$$P(1, 0, P_1, P_2, \rho) = P_1 (1 - P_2) - \rho \sqrt{P_1 (1 - p_1) P_2 (1 - p_2)}$$

$$P(0, 1, P_1, P_2, \rho) = (1 - P_1) P_2 - \rho \sqrt{P_1 (1 - p_1) P_2 (1 - p_2)}$$

$$P(0, 0, P_1, P_2, \rho) = (1 - P_1) (1 - P_2) + \rho \sqrt{P_1 (1 - p_1) P_2 (1 - p_2)}$$

According to Lemma 1, it is easily seen that $P(y_1, y_2, p_1, p_2, \rho) \geq 0$ for all y_1 and y_2 and

$P(1,1,P_1,P_2, \rho) + P(1,0,P_1,P_2, \rho) + P(0,1,P_1,P_2, \rho) + P(0,0,P_1,P_2, \rho) = 1$, which implies that the distribution $P(y_1, y_2, p_1, p_2, \rho)$ is a probability distribution.

Now we show the probability distribution $P(y_1, y_2, p_1, p_2, \rho)$ is also identifiable. Suppose that

$$p(y_1, y_2, p_1, p_2, \rho) = p(y_1, y_2, \tilde{p}_1, \tilde{p}_2, \tilde{\rho}), y_1 = 0, 1, y_2 = 0, 1$$

Which implies that

$$p_1 p_2 + \rho \sqrt{p_1(1-p_1)p_2(1-p_2)} = \tilde{p}_1 \tilde{p}_2 + \tilde{\rho} \sqrt{\tilde{p}_1(1-\tilde{p}_1)\tilde{p}_2(1-\tilde{p}_2)},$$

$$p_1(1-p_2) - \rho \sqrt{p_1(1-p_1)p_2(1-p_2)} = \tilde{p}_1(1-\tilde{p}_2) - \tilde{\rho} \sqrt{\tilde{p}_1(1-\tilde{p}_1)\tilde{p}_2(1-\tilde{p}_2)},$$

$$(1-p_1)p_2 - \rho \sqrt{p_1(1-p_1)p_2(1-p_2)} = (1-\tilde{p}_1)\tilde{p}_2 - \tilde{\rho} \sqrt{\tilde{p}_1(1-\tilde{p}_1)\tilde{p}_2(1-\tilde{p}_2)},$$

$$(1-p_1)(1-p_2) + \rho \sqrt{p_1(1-p_1)p_2(1-p_2)} = (1-\tilde{p}_1)(1-\tilde{p}_2) + \tilde{\rho} \sqrt{\tilde{p}_1(1-\tilde{p}_1)\tilde{p}_2(1-\tilde{p}_2)}.$$

The sum of the first two equations leads to $p_1 = \tilde{p}_1$ and the sum of the first and third equations leads to $p_2 = \tilde{p}_2$, which in turn implies that $\rho = \tilde{\rho}$. Thus, the probability distribution is identifiable.

Proof of Theorem 2

It is easily seen that

$$L = (P(1,1,P_1,P_2, \rho))^{N_{11}} (P(1,0,P_1,P_2, \rho))^{N_{10}} (P(0,1,P_1,P_2, \rho))^{N_{01}} (P(0,0,P_1,P_2, \rho))^{N_{00}},$$

Where N_{11} , N_{10} , N_{01} and N_{00} are the numbers of subjects with true positive, false negative, false positive and true negative, respectively.

Lemma 2 shows the function $(\prod_{t=1}^k \theta_t^{j_t}) (1 - \sum_{t=1}^k \theta_t)^{n - \sum_{t=1}^k j_t}$ reaches the maximum at $\theta_t = \frac{j_t}{n}, t = 1, \dots, k..$

Thus, the maximum likelihood estimators of p , q and ρ should satisfy the following equations

$$\hat{p}_1 \hat{p}_2 + \hat{\rho} \sqrt{\hat{p}_1(1-\hat{p}_1)\hat{p}_2(1-\hat{p}_2)} = \frac{n_{11}}{n},$$

$$\hat{p}_1(1-\hat{p}_2) - \hat{\rho} \sqrt{\hat{p}_1(1-\hat{p}_1)\hat{p}_2(1-\hat{p}_2)} = \frac{n_{10}}{n},$$

$$(1-\hat{p}_1)\hat{p}_2 - \hat{\rho} \sqrt{\hat{p}_1(1-\hat{p}_1)\hat{p}_2(1-\hat{p}_2)} = \frac{n_{01}}{n}.$$

Direct calculus can give the proposed maximum likelihood estimators.