Correction of Verification Bias using Log-Linear Models for a Single Binary Scale Diagnostic Tests

Haresh Rochani  
*Georgia Southern University, hrochani@georgiasouthern.edu*

Hani M. Samawi  
*Georgia Southern University, hsamawi@georgiasouthern.edu*

Robert L. Vogel  
*Georgia Southern University, rvogel@georgiasouthern.edu*

Jingjing Yin  
*Georgia Southern University, yin@georgiasouthern.edu*

Follow this and additional works at: [https://digitalcommons.georgiasouthern.edu/biostat-facpubs](https://digitalcommons.georgiasouthern.edu/biostat-facpubs)  
Part of the [Biostatistics Commons](https://digitalcommons.georgiasouthern.edu/biostat-facpubs) and the [Public Health Commons](https://digitalcommons.georgiasouthern.edu/biostat-facpubs)

**Recommended Citation**  
[https://digitalcommons.georgiasouthern.edu/biostat-facpubs/116](https://digitalcommons.georgiasouthern.edu/biostat-facpubs/116)

This article is brought to you for free and open access by the Biostatistics, Department of at Digital Commons@Georgia Southern. It has been accepted for inclusion in Biostatistics Faculty Publications by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact digitalcommons@georgiasouthern.edu.
Correction of Verification Bias using Log-linear Models for a Single Binary-scale Diagnostic Tests

Haresh Rochani, Hani Samawi, Robert Vogel and Jingjing Yin
Department of Biostatistics, Jiann-Ping Hsu College of Public Health, USA

Abstract

In diagnostic medicine, the test that determines the true disease status without an error is referred to as the gold standard. Even when a gold standard exists, it is extremely difficult to verify each patient due to the issues of cost-effectiveness and invasive nature of the procedures. In practice some of the patients with test results are not selected for verification of the disease status which results in verification bias for diagnostic tests. The ability of the diagnostic test to correctly identify the patients with and without the disease can be evaluated by measures such as sensitivity, specificity and predictive values. However, these measures can give biased estimates if we only consider the patients with test results who also underwent the gold standard procedure. The emphasis of this paper is to apply the log-linear model approach to compute the maximum likelihood estimates for sensitivity, specificity and predictive values. We also compare the estimates with Zhou’s results and apply this approach to analyze Hepatic Scintigraph data under the assumption of ignorable as well as non-ignorable missing data mechanisms. We demonstrated the efficiency of the estimators by using simulation studies.

Keywords: Verification bias; Diagnostic tests; Log-linear models; Missing data

Introduction

Diagnostic testing in medicine is the process of identifying the patients with and without a particular disease. Accuracy of the diagnostic test is the ability of the test to correctly identify the true disease status of the patient [1]. The test or procedure that determines the true disease status without an error is called the gold standard test. However, even with the existence of a gold standard test, verification for the disease status of each patient may not be obtained due to various reasons such as the invasive nature or too costly gold standard test. For example, the Prostate Specific Antigen (PSA) blood test is used as a screening test for diagnosis of prostate Cancer with a ranging cost from $60 to $80. However, the true diagnosis is generally confirmed by invasive procedures such as prostate biopsy with a ranging cost from $1600 to $1800. Thus patients with a high risk or a prostate screening test positive are more likely to be offered the gold standard test than those with low risk or a negative screening test. Furthermore, the inference about the measures of diagnostic accuracy such as sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) may be biased because individuals who are selected for the gold standard test based on the diagnostic test results are not a random sample [2]. Many authors have referred to this bias as, work-up bias [3] or verification bias [4]. Therefore, in the presence of verification bias, the efficiency of a diagnostic test heavily depends only on those patients whose disease status has been verified. In addition if there is a positive association between patient selection for verification and the test results will produce more bias results [5].

When the true disease status is missing among patients who were not selected for verification, the framework of the missing data mechanism proposed by Rubin can be applied for verification bias correction. If the probability of selecting patients for verification of the disease status which results in verification bias for diagnostic tests. The ability of the diagnostic test to correctly identify the patients with and without the disease can be evaluated by measures such as sensitivity, specificity and predictive values. However, these measures can give biased estimates if we only consider the patients with test results who also underwent the gold standard procedure. The emphasis of this paper is to apply the log-linear model approach to compute the maximum likelihood estimates for sensitivity, specificity and predictive values. We also compare the estimates with Zhou’s results and apply this approach to analyze Hepatic Scintigraph data under the assumption of ignorable as well as non-ignorable missing data mechanisms. We demonstrated the efficiency of the estimators by using simulation studies.

*Corresponding author: Haresh Rochani, Assistant Professor, Department of Biostatistics, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA, 30460, USA, Tel: 912.478.1101; E-mail: hrochani@georgiasouthern.edu

Received October 30, 2015; Accepted December 11, 2015; Published December 18, 2015


Copyright: © 2015 Rochani H, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
diagnostic accuracy especially under MNAR mechanism. However, by using the log-linear model approach, the explicit solutions for the measures of diagnostic accuracy estimates can be computed by the use of a simple algebraic formula (computationally convenient) under all missing data mechanisms.

In this paper, we provide the log-linear model approach to correct for verification bias in order to compute the estimates of diagnostic measures (i.e. Sensitivity, Specificity, PPV and NPV) under various assumptions of missing data mechanisms. We also confirm by simulations that the estimators are asymptotically unbiased and more efficient than complete cases estimates. In next section, we will show the impact of verification bias using by hypothetical example. The following sections will focus on the method of log-linear models to correct the verification bias for diagnostic measures and comparison with Zhou [7] results. Furthermore, we illustrate the log-linear model approach by application to the Hepatic Scintigraph data.

Impact of Verification Bias

In diagnostic medicine, it is a common practice to refer only patients who are test positives for a gold standard test for disease verification. To illustrate the impact of bias, we considered a hypothetical study of 1000 patients to measure the sensitivity, specificity, PPV and NPV (Tables 1 and 2).

The hypothetical data are summarized in Table 1. If every patient is being referred for verification for the gold standard then the true sensitivity from Table 1 is 0.71. Furthermore, the true specificity, PPV and NPV are 0.87, 0.37 and 0.96 respectively. However, if we consider that only 10% of those who have a test negative result receive the gold standard for verification, then the data are summarized in Table 2. From Table 2, the observed sensitivity, specificity, PPV and NPV are 0.96, 0.4, 0.37 and 0.96 respectively. Overall if positive test results are more likely to receive disease verification, (i.e. MAR) then by using only verified cases, the estimates for sensitivity and specificity are biased, in fact, sensitivity is being overestimated and specificity is being underestimated. However, under the MAR assumption, the estimates of PPV = P(D+ |T+) and NPV = P(D− |T−) from complete cases are unbiased [8,15,16].

Methodology

Log-linear parameterization

Baker et al. [17] used the homogeneous log-linear model approach to adjust the expected cell counts for a two way table with missing data. We can use the similar approach (BRD model approach) to compute the estimates for measures of diagnostic accuracy correcting for verification bias. To demonstrate this idea, define two random variables I and J, each with two levels i = 1,2 and j = 1,2. Let $R_i$ be the missing data indicator for variable J, such that $R_i = 1$ if J is being observed and $R_i = 2$, if J is missing. Similarly, let $R_i = 1$ if we observe in variable I. Consider $I \times J$ cross-classified Table 3 with total sample size of $n_{ij}$, each with two levels with one supplemental margin (in the terminology of Dempster and Baker 1992) with cell count $n_{ij}$. The cell counts with $R_i = R_j = 1$ is denoted as $n_{i1}$. The supplemental margin for variable J when $R_j = 2$ is denoted by $n_{i.1}$. In these cases, the symbol ‘+’ denotes the summation over the corresponding index. Furthermore, denote the expected cell counts by $\mu_{ij}$, cell probabilities for I, J, $R_i$ and $R_j$ by $\pi_{ijk}$ and the marginal probabilities by $\pi_i$ for Table 3.

Log-linear models are widely used to analyze contingency tables. Missing data indicators for contingency tables with supplemental margins can be incorporated in log-linear models in the same way as they incorporate the other variables. In terms of $\lambda$, the homogenous log-linear model (presence of only two-way interactions) for partially observed one variable is $\log \mu_{ij} = \lambda + \lambda_i + \lambda_j + \lambda_{ij}$ (1)

Because this is an over parameterized model, we add the side conditions that the sum of each $\lambda$ term over each of the indicated subscripts is zero. For computational purposes, it is convenient to use an alternative parameterization of the model. Let $m_0 \geq 0$, $b_i \geq 0$ and $\sum_{i=1}^{2} m_{i} (1+b_i) = n_{i+}$, such that $\mu_{i+} = m_{i}$ and $\mu_{+i} = m_{+.i}$. In terms of log-linear models

$$m_i = \exp(\lambda_i + \lambda_j + \lambda_{ij} + \lambda_{.i} + \lambda_{.j} + \lambda_{..} + \lambda_{i..} + \lambda_{.j..} + \lambda_{i..} + \lambda_{.j..} + \lambda_{i.j..})$$

Taking advantage of the side conditions, we can derive the following expression for $b_i$:

$$b_i = -2 \exp(\lambda_i) + \lambda_j + \lambda_{ij} + \lambda_{.i} + \lambda_{.j} + \lambda_{..} + \lambda_{i..} + \lambda_{.j..} + \lambda_{i..}$$

The detailed derivation can be found in the Appendix. Denote,

$$p_{i} = Pr(R_i = 1, R_j = 1 | I = i, J = j) = \frac{\mu_{ij}}{n_{i+}}$$

$$p_{+i} = Pr(R_i = 1, R_j = 2 | I = i, J = j) = \frac{\mu_{i+2}}{n_{i+}}$$

In probabilistic terms, $m_i$ and $b_i$ can be expressed as follows,

$$m_i = n_{i+} p_i$$

$$b_i = \frac{\mu_{i+2}}{\mu_{i+1}} = \frac{\mu_{i+2}}{\mu_{i+1}} = \frac{p_{+i}}{p_i}$$

For Table 3, the total expected count is

$$\mu_{ij} = \sum_{i=1}^{2} \sum_{j=1}^{2} \mu_{i+j} = \sum_{i=1}^{2} \sum_{j=1}^{2} \mu_{i+j} = \sum_{i=1}^{2} \sum_{j=1}^{2} m_{i} (1+b_i)$$

Table 3: Two way table with one supplementary margin.
and in general notation the marginal and conditional probabilities can be expressed as:

\[ \pi_{ij} = \Pr(I=i, J=j) = \frac{m_{ij}(1+b_i)}{n_{c+}} \]

\[ \pi_{i.} = \Pr(I=i) = \sum_{j} \frac{m_{ij}(1+b_i)}{n_{c+}} \]

\[ \pi_{.j} = \Pr(J=j) = \sum_{i} \frac{m_{ij}(1+b_i)}{n_{c+}} \]

The likelihood ratio statistic with respect to a model fitting the data perfectly is:

\[ G^2 = 2 \left[ \sum_{ij} n_{ij} \log \left( \frac{n_{ij}}{\hat{m}_{ij}} \right) + \sum_{i} n_{i.} \log \left( \frac{n_{i.}}{\hat{m}_{i}} \hat{m}_{.j} \right) + \sum_{j} n_{.j} \log \left( \frac{n_{.j}}{\hat{m}_{j}} \hat{m}_{i.} \right) \right] \]

Models parameter estimates

Baker et al. [17] identified nine different models for the two-way table with three supplementary margins based on the dependence of missingness on one or the other or its own realization. For Table 3, when \( b_i \) equals to \( \beta \), \( \beta_i \) or \( \beta_j \), there are three different identifiable models. First and second subscripts to the parameter \( \beta \) correspond to variable I and J respectively. The subscript " \( . \) " indicates that parameter is constant over corresponding index. In addition, each model has a unique interpretation. For example, Model (\( \beta \)) can be interpreted as the missingness in J is constant (MCAR). Similarly, Model (\( \beta_i \)) can be interpreted as the missingness in J depends on the realization of I (MAR) and when the missingness in J depends on its own realization, then we consider Model \( \beta_j \) to be MNAR.

Assuming Poisson distribution for cell counts, the maximum likelihood estimates for \( m_i \) and \( \beta \) can be derived from log-likelihood functions for each model. Detailed derivation of the estimates are provided in Appendix. The likelihood ratio statistics for each model can also be found in the Appendix.

**MCAR estimates**

\[ \hat{\beta} = \frac{n_{c+}}{n_{++}}, \hat{m}_i = \frac{n_{i.} \cdot n_{.+} - n_{c+}}{n_{i.} \cdot n_{.+} - n_{c+}} \]

\[ \hat{\pi}_{ij} = \frac{\hat{m}_i (1 + \hat{\beta})}{\sum_i \hat{m}_i (1 + \hat{\beta})} \]

\[ \hat{\pi}_{i.} = \frac{\hat{m}_i (1 + \hat{\beta})}{\sum_i \hat{m}_i (1 + \hat{\beta})} \]

\[ \hat{\pi}_{.j} = \frac{\hat{m}_j (1 + \hat{\beta})}{\sum_j \hat{m}_j (1 + \hat{\beta})} \]

From equation 2, the estimates for Sensitivity and Specificity are

\[ \text{Sensitivity(MCAR)} = \hat{\pi}_{i.} = \frac{\hat{m}_i (1 + \hat{\beta})}{\hat{m}_i + \hat{m}_j} \]

\[ \text{Specificity(MCAR)} = \hat{\pi}_{.j} = \frac{\hat{m}_j (1 + \hat{\beta})}{\hat{m}_i + \hat{m}_j} \]

Similarly, the estimates for PPV (\( \hat{\pi}_{+j} \)) and NPV (\( \hat{\pi}_{-j} \)) can be computed from equation 3.

**MAR estimates**

\[ \hat{\beta}_i = \frac{n_{i+}}{n_{i+} - n_{c+}}, \hat{m}_i = \frac{n_{i.} \cdot n_{.+} - n_{c+}}{n_{i.} \cdot n_{.+} - n_{c+}} \]

\[ \hat{\pi}_{ij} = \frac{\hat{m}_i (1 + \hat{\beta}_i)}{\sum_i \hat{m}_i (1 + \hat{\beta}_i)} \]

\[ \hat{\pi}_{i.} = \frac{\hat{m}_i (1 + \hat{\beta}_i)}{\sum_i \hat{m}_i (1 + \hat{\beta}_i)} \]

\[ \hat{\pi}_{.j} = \frac{\hat{m}_j (1 + \hat{\beta}_j)}{\sum_j \hat{m}_j (1 + \hat{\beta}_j)} \]

From equation 4 the estimates of the Sensitivity and Specificity can be obtained as follows

\[ \text{Sensitivity(MAR)} = \hat{\pi}_{i.} = \frac{\hat{m}_i (1 + \hat{\beta}_i)}{\hat{m}_i (1 + \hat{\beta}_i) + \hat{m}_j (1 + \hat{\beta}_j)} \]

\[ \text{Specificity(MAR)} = \hat{\pi}_{.j} = \frac{\hat{m}_j (1 + \hat{\beta}_j)}{\hat{m}_i (1 + \hat{\beta}_i) + \hat{m}_j (1 + \hat{\beta}_j)} \]

**MNAR estimates**

\[ \hat{\pi}_{ij} = \frac{\hat{m}_i (1 + \hat{\beta}_j)}{\sum_i \hat{m}_i (1 + \hat{\beta}_j)} \]

\[ \hat{\pi}_{i.} = \frac{\hat{m}_i (1 + \hat{\beta}_j)}{\sum_i \hat{m}_i (1 + \hat{\beta}_j)} \]

\[ \hat{\pi}_{.j} = \frac{\hat{m}_j (1 + \hat{\beta}_j)}{\sum_j \hat{m}_j (1 + \hat{\beta}_j)} \]

For Model (\( \hat{\beta}_j \)), equation 8 provides the estimates of Sensitivity and Specificity as follows

\[ \text{Sensitivity(MNAR)} = \hat{\pi}_{i.} = \frac{\hat{m}_i (1 + \hat{\beta}_j)}{\hat{m}_i (1 + \hat{\beta}_j) + \hat{m}_j (1 + \hat{\beta}_j)} \]

\[ \text{Specificity(MNAR)} = \hat{\pi}_{.j} = \frac{\hat{m}_j (1 + \hat{\beta}_j)}{\hat{m}_i (1 + \hat{\beta}_j) + \hat{m}_j (1 + \hat{\beta}_j)} \]

while equation 9 provides the estimates of PPV (\( \hat{\pi}_{+j} \)) and NPV (\( \hat{\pi}_{-j} \))

It is possible to obtain negative maximum likelihood estimates for \( \hat{\beta}_j \) in MNAR model. If any solution is negative (\( \hat{\beta}_j < 0 \)), the estimates still can be computed by maximizing the likelihood function by using the limited memory algorithm for constrained optimization Byrd [18]. Optim function can be utilized in statistical software R 2.15.1 or higher versions.

**Inference about diagnostic measures**

In this section, the multivariate delta method is used to draw the
inferences about the diagnostic measures for Model (β_i) (MAR). A similar method can be used for Model (β_i) (MCAR) and Model (β_i) (MNAR). To derive the asymptotic variance of the estimates, we have

\[ \frac{\hat{\sigma}_{ij} - \pi_{ij}}{\sqrt{\text{var}(\hat{\sigma}_{ij})}} \sim N(0,1). \]

However, to ensure better normal approximation, we derived the asymptotic distribution of

\[ \log(1 - \hat{\sigma}_{ij}) - \log(1 - \pi_{ij}). \]

By using multivariate delta method for Model (β_i), we have

\[ \text{logit}(\hat{\pi}_{ij}) = \text{logit}(\pi_{ij}) \cdot \mathbf{C}^\mathbf{I}_{1i} \mathbf{C}^\mathbf{t}. \]

(10)

where \( \mathbf{C} = \left( \begin{array}{cccc} \log \frac{\pi_{11}}{\pi_{00}} & \log \frac{\pi_{10}}{\pi_{01}} & \log \frac{\pi_{12}}{\pi_{02}} & \log \frac{\pi_{22}}{\pi_{21}} \end{array} \right) \) and \( \mathbf{I}_{1i} \) is the Fisher information. The detailed derivation of \( \mathbf{I}_{1i} \) is provided in the Appendix. Therefore, the 100(1−\( \delta \))% confidence interval for \( \pi_{ij} \) is

\[ \left[ \frac{\exp \left( \log \hat{\pi}_{ij} + \frac{1}{2} \text{var} \log \hat{\pi}_{ij} \right)}{1 + \exp \left( \log \hat{\pi}_{ij} + \frac{1}{2} \text{var} \log \hat{\pi}_{ij} \right)} \right] \]

In case of negative estimates for MNAR model, the variances for the measures of the diagnostic accuracy can be estimated by bootstrap method.

Comparison with Zhou's result

Zhou [7] derived the estimates for Sensitivity and Specificity under MAR and MNAR assumptions. The estimates obtained in previous section for Sensitivity and Specificity under MAR assumption are equivalent to Zhou's estimates (Appendix). However, under MNAR assumption Zhou [7] assumes that the two ratios

\[ \epsilon_0 = \frac{P(R_i = 1 | J = 1, J = 1)}{P(R_i = 1 | J = 2, J = 1)} \quad \text{and} \quad \epsilon_1 = \frac{P(R_i = 1 | J = 1, J = 1)}{P(R_i = 1 | J = 1, J = 2)} \]

are known to estimate the Sensitivity and Specificity. However, the log-linear model approach does not require such assumptions to estimate MNAR parameters.

Simulation Study

A simulation study was conducted to illustrate the reduction in bias and root mean squared error (RMSE) of the estimators for sensitivity and specificity for MAR model compared to complete case analysis (Table 4).

Two variables "Test Results" and "Disease Status" with different sample sizes (N) were generated from the multinomial distribution (Table 4). In this section, we will demonstrate the log-linear modeling approach to compute the measures of diagnostic accuracy under various missing mechanisms by using the Hepatic Squaligraphy data published by [19]. In the Hepatic Squaligraphy data, 650 patients underwent hepatic squaligraphy of which 429 patients had positive scans while 221 patients had negative hepatic scans. Only 263 patients out of 429 and 81 patients out of 221 were referred for status verification using procedures such as biopsy, exploratory laparotomy or autopsy within 6 weeks of their scans. Table 5 represents the verified and unverified cases by hepatic scan results. Table 6 shows the Maximum likelihood estimates for model parameters (Tables 5-7).

Table 7 represents the estimates of sensitivity, specificity, PPV and NPV and 95% confidence intervals under three different models by using the log-linear model approach to Table 5. By using Zhou [7] method the estimates of Sensitivity and Specificity can range from (0.68,0.95) and (0.37,0.86) respectively depending upon the values of \( \epsilon_0 \) and \( \epsilon_1 \). From Table 7, it can be confirmed that under MAR and MCAR assumptions the estimates of PPV and NPV and under MNAR assumption the estimates of sensitivity and specificity can be obtained from complete cases [8].

Final Remarks

Verification bias is an extremely common problem in diagnostic medicine. This paper shows that how the log-linear model approach in single binary-scale diagnostic tests correct the verification bias for estimating the diagnostic measures. Log-linear models also reduces...
the bias and improves the performance of the estimators compared to complete case analysis. In addition to this, explicit estimates for the measures of diagnostic accuracy under different missing mechanisms can be computed from simple algebraic formulas, while other available methods require iterative methods to estimate the measures of diagnostic accuracy. Furthermore, this approach will allow us to test for MCAR assumption for particular data. However, the only way to confirm for MAR or MNAR missing mechanisms is to recollect the missing data. Although in diagnostic medicine due to the issue of cost-effectiveness, we do not have the luxury of getting hold of the missing data. As a result, careful modeling of missing mechanisms to reduce the bias is more an important issue. Therefore, if the data are not MCAR then by using the scientific knowledge about the data we can assume a MAR or MNAR missing mechanism to compute estimates for the measures of diagnostic accuracy with log-linear models.

Conflict of Interest

The authors have declared no conflict of interest.

References

Appendix

Log-linear parameterization

Expected counts for all verified cases can be modeled by using equation 1 in the following way:

\[ \mu_{ij} = \exp\{ \alpha + \lambda_i(j) + \lambda_i(j) + \lambda_{ij}(1) + \lambda_{ij}(1) + \lambda_{ij}(1) + \lambda_{ij}(1) \} \]

Expected counts for all un-verified cases can be modeled as the following way:

\[ \mu_{ij} = \exp\{ \alpha + \lambda_i(j) + \lambda_i(j) + \lambda_{ij}(1) + \lambda_{ij}(1) + \lambda_{ij}(1) + \lambda_{ij}(1) \} \]

Some of the side conditions (the sum of each \( \lambda \)-term over each of the indicated subscript is zero) for the log-linear models are

\[ \sum_{k} \lambda_{ij}(k) = 0 \quad \sum_{j} \lambda_{ij}(j) = 0 \quad \sum_{k} \lambda_{ij}(k) = 0 \quad \sum_{k} \lambda_{ij}(k) = 0 \]

By using these conditions, we can get

\[ b_{ij} = \frac{\mu_{ij}}{\mu_{ij}} = -2 \exp\{ \lambda_{ij}(1) + \lambda_{ij}(1) + \lambda_{ij}(1) + \lambda_{ij}(1) \} \]

Maximum likelihood (ML) Estimates for Model (\( \beta \)) (MCAR)

\[ L = \sum_{ij} n_{ij} \log(m_{ij}) + \sum_{ij} n_{ij} \log(m_{ij} - m_{ij}) - \sum_{ij} m_{ij}(1 + \beta) \]

\[ \frac{\partial L}{\partial \beta} = \sum_{ij} \left( \frac{n_{ij}}{m_{ij}} - \frac{m_{ij}}{m_{ij} - m_{ij}} \right) - \sum_{ij} m_{ij} \]

\[ 0 = \frac{n_{ij}}{m_{ij}} - \hat{m}_{ij} \]

Maximum likelihood Estimates for Model (\( \beta \)) (MAR)

\[ L = \sum_{ij} n_{ij} \log(m_{ij}) + \sum_{ij} n_{ij} \log(m_{ij} - m_{ij}) - \sum_{ij} m_{ij}(1 + \beta) \]

\[ \frac{\partial L}{\partial \beta} = \sum_{ij} \left( \frac{n_{ij}}{m_{ij} - m_{ij}} - m_{ij} \right) \]

First solving for \( \hat{m}_{ij} \)

\[ \hat{m}_{ij} = \left( \frac{n_{ij}}{m_{ij}} + \frac{n_{ij}}{m_{ij}} \right) \frac{n_{ij}}{n_{ij}} \]

Substituting \( \hat{m}_{ij} \) from equation 11:

\[ \hat{m}_{ij} = \left( \frac{n_{ij}}{n_{ij}} + \frac{n_{ij}}{n_{ij}} \right) \frac{n_{ij}}{n_{ij}} \]

Maximum likelihood Estimates for Model (\( \beta \)) (MAR)

\[ L = \sum_{ij} n_{ij} \log(m_{ij}) + \sum_{ij} n_{ij} \log(m_{ij} - m_{ij}) - \sum_{ij} m_{ij}(1 + \beta) \]

\[ \frac{\partial L}{\partial \beta} = \sum_{ij} \left( \frac{n_{ij}}{m_{ij} - m_{ij}} - m_{ij} \right) \]
Therefore, the ML estimates for $\beta_i$ are:

$$\hat{\beta}_i = \frac{n_{i1} + n_{i2}}{\hat{m}_i} \left(1 + \hat{\beta}_j\right)$$

ML estimates for Model ($\beta$) can be derived by taking the derivative of $L$ with respect to $\beta_j$:

$$\frac{\partial L}{\partial \beta_j} = \sum_j \left(\frac{n_{i1} + n_{i2}}{m_j} \hat{m}_j - \sum_i m_i \left(1 + \hat{\beta}_j\right)\right)$$

$$0 = \sum_i \left(\frac{n_{i1} + n_{i2}}{m_j} \hat{m}_j - \sum_i \hat{m}_j \hat{\beta}_j\right)$$

$$0 = \frac{n_{i1} + n_{i2}}{\hat{m}_0} \sum_i \hat{m}_i \hat{\beta}_j - \sum_i \hat{m}_j \left(1 + \hat{\beta}_j\right)$$

$$\hat{m}_0 \left(1 + \hat{\beta}_j\right) = n_{i1} + \frac{n_{i2}}{\hat{m}_j} \hat{\beta}_j$$

$$\hat{m}_0 = n_{i1}$$

From equation 6, we can guess ML estimates for $\hat{\beta}_j$ as:

$$\hat{\beta}_j = \hat{\beta}_j$$

Comparison of MAR Estimates with Zhou’s Result

The estimates for Sensitivity is:

$$Sensitivity(Zhou) = \frac{(n_{i11}, n_{i21})}{n_{i11} \cdot n_{i21} + n_{i11} \cdot n_{i22} + n_{i21} \cdot n_{i12}}$$

$$= \frac{n_{i11} \cdot n_{i21}}{n_{i11} \cdot (n_{i12} + n_{i21}) + n_{i21} \cdot (n_{i11} + n_{i22})}$$

From equation 6 the Sensitivity estimates from log-linear model are:

$$Sensitivity(MAR) = \frac{\hat{m}_0 \left(1 + \hat{\beta}_j\right)}{\hat{m}_1 \left(1 + \hat{\beta}_j\right) + \hat{m}_2 \left(1 + \hat{\beta}_j\right)}$$
Similarly, we can show that the estimates for Specificity from log-linear models are equivalent to sensitivity estimates.

Information Matrices:

\( \frac{\partial^2 L}{\partial \beta^2} = \frac{n_{1111}}{\beta^2} \)

\[-E \left( \frac{\partial^3 L}{\partial \beta^3} \right) = E \left( \frac{n_{1122}}{\beta^2} \right) = \frac{m_y \beta}{\beta^2} - \frac{m_x}{\beta} \]

\[-E \left( \frac{\partial^3 L}{\partial m_1 \partial \beta} \right) = \frac{1}{m_y} + \frac{\beta}{m_x} \]

\[-E \left( \frac{\partial^3 L}{\partial m_\beta} \right) = 1 \]

\[ I_1 = \left( \begin{array}{cccc} -E \left( \frac{\partial^2 L}{\partial m_1 \partial m_1} \right) & -E \left( \frac{\partial^2 L}{\partial m_1 \partial m_{12}} \right) & -E \left( \frac{\partial^2 L}{\partial m_1 \partial m_{21}} \right) & -E \left( \frac{\partial^2 L}{\partial m_1 \partial m_{22}} \right) \\
-E \left( \frac{\partial^2 L}{\partial m_{12} \partial m_1} \right) & -E \left( \frac{\partial^2 L}{\partial m_{12} \partial m_{12}} \right) & -E \left( \frac{\partial^2 L}{\partial m_{12} \partial m_{21}} \right) & -E \left( \frac{\partial^2 L}{\partial m_{12} \partial m_{22}} \right) \\
-E \left( \frac{\partial^2 L}{\partial m_{21} \partial m_1} \right) & -E \left( \frac{\partial^2 L}{\partial m_{21} \partial m_{12}} \right) & -E \left( \frac{\partial^2 L}{\partial m_{21} \partial m_{21}} \right) & -E \left( \frac{\partial^2 L}{\partial m_{21} \partial m_{22}} \right) \\
-E \left( \frac{\partial^2 L}{\partial m_{22} \partial m_1} \right) & -E \left( \frac{\partial^2 L}{\partial m_{22} \partial m_{12}} \right) & -E \left( \frac{\partial^2 L}{\partial m_{22} \partial m_{21}} \right) & -E \left( \frac{\partial^2 L}{\partial m_{22} \partial m_{22}} \right) \end{array} \right) \]

Model (\( \beta \))

\[-E \left( \frac{\partial^2 L}{\partial \beta^2} \right) = \frac{m_y}{\beta} \]

\[-E \left( \frac{\partial^2 L}{\partial m_1 \partial \beta} \right) = \frac{1}{m_y} + \frac{\beta}{m_x} \]

\[-E \left( \frac{\partial^2 L}{\partial m_\beta} \right) = 1 \]
\[
\frac{\partial^2 L}{\partial \beta_i^2} = \frac{1}{m_i} \sum \left( m_i - \sum m_{ij} \beta_j \right)
\]

\[
\frac{\partial^2 L}{\partial \beta_j \partial \beta_i} = \frac{1}{m_i} \sum \left( m_i - \sum m_{ij} \beta_j \right)
\]

\[
L_i = \left\{ -E \left( \frac{\partial^2 L}{\partial \beta_i^2} \right), \ -E \left( \frac{\partial^2 L}{\partial \beta_j \partial \beta_i} \right), \ -E \left( \frac{\partial^2 L}{\partial m_i \partial \beta_i} \right), \ -E \left( \frac{\partial^2 L}{\partial m_j \partial \beta_i} \right), \ -E \left( \frac{\partial^2 L}{\partial m_{ij} \partial \beta_i} \right), \ -E \left( \frac{\partial^2 L}{\partial \beta_j \partial \beta_i} \right) \right\}
\]

Likelihood Ratio Statistics

Model (\(\beta_i\))

\[
G^2 = 2 \left\{ \sum_{i} n_{i1} \log \left( \frac{n_{i1}}{\hat{m}_i / \beta_i} \right) + \sum_{j} n_{ij} \log \left( \frac{n_{ij}}{\hat{m}_j} \right) + \sum_{k} n_{ik} \log \left( \frac{n_{ik}}{\hat{m}_k} \right) \right\}
\]

Model(\(\beta_i\)) \(G^2\) = \[
2 \left\{ \sum_{i} n_{i1} \log \left( \frac{n_{i1}}{\hat{m}_i / \beta_i} \right) + \sum_{j} n_{ij} \log \left( \frac{n_{ij}}{\hat{m}_j} \right) + \sum_{k} n_{ik} \log \left( \frac{n_{ik}}{\hat{m}_k} \right) \right\}
\]

Model(\(\beta_i\)) \(G^2\) = \[
2 \left\{ \sum_{i} n_{i1} \log \left( \frac{n_{i1}}{\hat{m}_i / \beta_i} \right) + \sum_{j} n_{ij} \log \left( \frac{n_{ij}}{\hat{m}_j} \right) + \sum_{k} n_{ik} \log \left( \frac{n_{ik}}{\hat{m}_k} \right) \right\}
\]