Global Scientific JOURNALS

GSJ: Volume 10, Issue 1, January 2022, Online: ISSN 2320-9186
www.globalscientificjournal.com

Bayesian Methods for Diagnostic Test Using Negative Binomial and Multinomial Models

Abdussalam Muhammad Lawan [1], Yusuf A. Mohammed [2], Imam Akeyede [3], Abdulmalik Mohammed [4], Nuruddeen Shehu [5]

[1] *Department of Health Information Management, Abubakar Tafawa Balewa University Teaching Hospital Bauchi, Nigeria.*
[2] *Department of Mathematical Sciences, University of Maiduguri, Nigeria.*
[3] *Department of Mathematical Sciences, Federal University Lafia, Nigeria.*
[4] *Department of Mathematical Sciences, Abubakar Tafawa Balewa University Bauchi, Nigeria.*
5 Department of Mathematics Education, Federal College of Education Gombe, Nigeria.

Corresponding Author: *Abdussalam Muhammad Lawan E-mail: abdussalami2012@gmail.com and abdussalami@atbuth.ng +2348039372518*

ABSTRACT

Background:

This paper considers two Bayesian models (Negative binomial and Multinomial) for diagnostics test evaluation. Prior data on HCV, HBsAg and HIV screened using RAPID test and ELIZA test obtained were fitted into the models and posterior data obtained using Monte Carlo Markov Chain (MCMC). The Bayesian approach is taken because of its efficient use of prior information, and the analysis is executed with a Bayesian software package OpenBUGS.

Objective:

In this paper, we will discuss how the two bayesian models will be fitted in order to come up with the most fitted model for diagnostic test evaluation among them.

Result:

The Negative binomial DIC value of 6.868 indicates that Negative binomial Bayesian model is more fitted to Bayesian diagnostics test evaluation over multinomial Bayesian model with DIC value of 8.08.

Keywords: Bayesian model, Negative binomial, Multinomial, MCMC, OpenBUGS, Prior, Posterior, DIC.

1. Introduction

The use of Bayesian inference in diagnostic test accuracy has recently increased, diagnostic-test evaluation is particularly suited to the Bayesian framework because prior scientific information about the sensitivities and specificities of the tests and prior information about the prevalence of the sampled populations can be incorporated [1]. A review article by [2] described many frequentist and Bayesian diagnostic-test modeling

strategies. Different statistical distribution models have been used at different times by researchers in evaluating diagnostic test. [3] Described multinomial and beta models by maximum likelihood estimate. While, [4] present the computational aspects involved in the used of models described by [3].

The Negative binomial distribution is suitable to diagnostic test because it described count data, in recent years, several researchers have used bayesian inference for the negative binomial distribution [5]. However, none of these have obtained diagnostic test sensitivity and specificity.

Prior information about model parameters is essential for obtaining solutions to accuracy-estimation problems when there are more parameters than degrees of freedom provided by the data [4]. This occurs in the simplest case (one test in one population) and when there are two correlated tests used in one or more populations as in our case. In contrast, models for data based on two or more independent tests and two or more sampled populations are identifiable (i.e. have sufficient information to estimate all parameters of interest). The parameters estimated are Se, Sp, and the prevalence, p. Denote by y the number of reactors (test positive) out of the n sampled patients. The probability of a sample testing positive is given by $= P(T^+) = \pi S_e + (1-\pi)\big(1-S_p\big)$.

Under bayesian perspective, we let *y* denote the observations, π and $\pi(y)$ the prior and posterior probabilities. Then the prior and posterior distribution can be define by $(1-\pi)$ and $[1-\pi(y)]$. The prior and posterior odds is define by $p = \pi(1-\pi)^{-1}$ and $p(y) = \pi[1-\pi(y)]^{-1}$. Many problems in medicine involve well known standard problem of the test for independence in a 2 x 2 contingency table. We used OpenBUGS Bayesian statistical package for the analysis. Sampling from posterior distribution can, in many cases, be accomplished by using OpenBUGS. In OpenBUGS, the likelihood and prior distributions are specified by the user. While, the posterior distributions are sampled automatically.

2. Negative Binomial Distribution Test Evaluation

The probability density function of the (discrete) negative binomial distribution (NBD) is given by

$$p(y|r,p) = \left\{\frac{\Gamma(r+y)p^r(1-p)^y}{\Gamma(r)\Gamma(y+1)}\right\}_{y\geq 0}^{y<0} \qquad 2.1$$

Where the notation $y \mid r, p$ means "y given r and p" with r and p being parameters of the density function and y being the outcome variable. r is the number of events until the experiments is stopped and p is the probability of success in each event. Since NBD is a discrete distribution we have,

$$\sum_{y=0}^{\infty} p(y|r,p) = \sum_{y=0}^{\infty} \frac{\Gamma(r+y)p^r(1-p)}{\Gamma(r)\Gamma(y+1)} \quad 2.2$$

The log-likelihood function of the considered density function is given by

$$l(r,p|x) = \ln[p(r,p|y)]$$

$$= \ln[p(y|r,p)]$$

$$= \ln[\Gamma(\mathrm{r}+\mathrm{y})p^r(1-p)^y] - \ln[\Gamma(r)\Gamma(y+1)] \quad 2.3$$

Considering the value of the parameters r, p given the vector of observed data $y_1, y_{2\ldots}y_N$

$$l(r,p|x_i,\ldots,N) = \sum_{i=1}^{N} l(r,p|x)$$

$$= \sum_{i=0}^{N} l(r,p|x_i) \sum_{i=1}^{N} \ln[\Gamma(r+x_i) + x_i \ln(1-p) - \ln(\Gamma(x_i+1)] \quad 2.4$$

Then, the vectors (y_{11}, y_{12}, y_{21}, *and* y_{22}) can be given as,

$$y_{11} = \pi(1-\pi)Se_1 + \left(1-Sp_1\right)\left(1-Sp_2\right) \quad 2.5$$

$$y_{12} = \pi\left(1-Sp_1\right) + (1-Se_1)Se_1 \quad 2.6$$

$$y_{21} = \pi(1-\pi)Se_2 + \left(1-Sp_1\right)\left(1-Sp_2\right) \quad 2.7$$

$$y_{22} = \pi(1-Sp_2) + (1-Se_2)Se_2 \quad 2.8$$

$$Se_1 \sim NBD(\pi_1,\pi_2) \ and \ Se_2 \sim NBD(\pi_1,\pi_2) \quad 2.9$$

$$Sp_1 \sim NBD(\pi_1,\pi_2) \ and \ Sp_2 \sim NBD(\pi_1,\pi_2) \quad 2.10$$

3. Multinomial Distribution Test Evaluation

The multinomial model described by [3] and computational aspect presented by [4] provides a solution diagnostic test evaluation.

Let $y_1 = (y_{112}, y_{122}, y_{212}, y_{222})$ denote the analogous data sampled from population. The data y_1 and y_2 are assumed to have independent multinomial sampling distributions:

$y_k \sim multinomial\left(n_k, (p_{11k}, p_{12k}, p_{21k}, p_{22k})\right), k = 1,2$ Where the multinomial cell probabilities for population k are given by

$$p_{11k} = \pi_k S_{e_1} S_{e_2} + (1-\pi_k)\left(1-S_{p_1}\right)\left(1-S_{p_2}\right)\left(1-S_{p2}\right) \quad 3.1$$

$$p_{12k} = \pi_k S_{e_1}\left(1-S_{e_2}\right) + (1-\pi_k)\left(1-S_{p_1}\right)S_{p2} \quad 3.2$$

$$p_{21k} = \pi_k (1-S_{e_1}) S_{e_2} + (1-\pi_k) S_{p_1}\left(1-S_{p_2}\right) \quad 3.3$$

$$p_{22k} = \pi_k (1-S_{e_1})\left(1-S_{e_2}\right) + (1-\pi_k) S_{p_1} S_{p2} \quad 3.4$$

To allow for zero-infection prevalence, π_1 and π_2 are mixture of point mass at zero and a continuous beta distribution so that $\pi_k = \pi_k^*$ with probability $\tau_k$and $\pi_k = 0$ with probability$1-\tau_k$. Uncertainty about the values of π_1^* and π_2^* and the two sensitivities and two specificities is modeled using independent beta prior distributions.

4. Model Discrimination

There are several information criteria available to determine the best model Bayesian framework. All of them are likelihood based; the deviance information criterion (DIC) is a hierarchical modeling generalization of the Akaike information criterion (AIC) [6]. It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. DIC is

an asymptotic approximation as the sample size becomes large, like AIC. It is only valid when the posterior distribution is approximately multivariate normal.

Deviance information criterion $D(\theta) = -2\log\big(p(y/\theta)\big) + C$ 4.1

Where y are the data, θ are the unknown parameters of the model and p(y/θ) is the likelihood function. C is a constant that cancels out in all calculations that compare different models and which therefore does not need to be known.

DIC is described by [7] as;

$$PD = \overline{D\,(\theta)} - D(\bar{\theta}) \quad 4.2$$

Where $\bar{\theta}$ is the expectation of θ.

5. Result and Discussion

The models (Negative binomial and multinomial) were set as a posterior distribution. While, the beta distribution as a prior distribution. The models were fitted using the 200 blood samples data screened for HCV, HBsAg and HIV obtained from Abubakar Tafawa Balewa University Teaching Hospital Bauchi, Nigeria blood bank as prior data. The sample were screened using RAPID test and ELIZA test. The data was stimulated 10,000 to obtained posterior data. Sensitivity, specificity, mean, standard deviation, median and 95% credible interval were obtained for both the negative binomial model and multinomial model.

Table 1. Negative binomial distribution

Parameter	μ	σ	Median	95% Credible Interval	
Sensitivity	0.3134	0.2523	0.254	0.005357	0.8881
Specificity	0.07184	0.011	0.05922	2.649E-4	0.09067
Theta[1]	0.5003	0.08768	0.4984	0.3299	0.668
Theta[2]	0.497	0.8934	0.496	0.3225	0.6668
Theta[3]	0.5104	0.08277	0.506	0.3537	0.6715
Theta[4]	0.7486	0.1122	0.7512	0.5444	0.9545

The table 1 above reports the OpenBUGS output for negative binomial distribution after 10,000 simulations. The specificity 0.07184 shows how accurate a negative test can be dictate which is similar to [8] findings, with 0.05922 median, also 0.3134 sensitivity mean and 0.2523 standard deviation shows how positive test are diagnose accurately. The 95% credible intervals of theta [1] - [4] range 03225 – 0.9545.

Table 2. Multinomial distribution

Parameter	μ	σ	Median	95% Credible Interval	
Sensitivity	0.5061	0.3397	0.5233	0.003991	0.997
Specificity	0.744	0.2105	0.790	0.3024	0.9979
Theta[1]	17.25	6.685	16.59	6.821	32.22
Theta[2]	27.29	8.676	27.08	10.95	44.52
Theta[3]	42.04	8.850	42.02	24.85	59.55
Theta[4]	57.60	9.023	57.73	40.27	74.78

A casual look at table 2 above reports the analysis which was executed with 10,000 observations generated from the posterior distribution of the multinomial parameters, the sensitivity mean 0.5061 is as close as median 0.5233. Also, the specificity have almost the same median 0.790 as mean 0.744. Theta [1] – [4] mean range (17.25, 57.60), the median also, show the same pattern as the mean (6.821, 57.73) implying that the posterior distribution is symmetric about the means.

5.1 Best Model selection

When the goal is to pick a model with the best out of sample predictive power then selection can be made on the basis of the deviance information criterion (DIC) [6]. DIC is a combined measure of goodness fit and model complexity [7]. The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset of the same structure as that currently observed.

Table 3. Deviance information criterion

Model	Dbar	Dhat	pD	DIC
Negative binomial	5.884	4.901	0.9831	6.867
Multinomial	8.08	36.02	0.2374	8.08

The Table 3 above summarizes the complexity of the two fitted models. Negative binomial, DIC = 6.867. While, multinomial model have DIC = 8.08.

6. Conclusion

This paper considered two Bayesian diagnostic test model. The study concluded that, all two bayesian model approaches can be used to determine the accuracy of diagnostic test. The negative binomial model is considered the most preferred model between the two models for diagnostic test evaluation, because the values of the DIC corresponding to the two models represented in table 3 above are 6.867 and 8.08 respectively. Therefore, of the models under consideration, negative binomial exhibits the best predictive accuracy for diagnostic accuracy test with DIC = 6.867.

7. Reference

[1] Gardner, (2002). ''The utility of Bayes' theorem and Bayesian inference in veterinary clinical practice and research,'' Aust. Vet. J. 80, 758–761.

[2] Bouma, A., Stegeman, J.A., Engel, B., de Kluijver, E.P., Elbers, A.R.W., De Jong, M.C.M., (2001). ''Evaluation of diagnostic tests for the detection of classical swine fever in the field without a gold standard,'' J. Vet. Diagn. Invest. 13, 383–388.

[3] Hui, S. L., & Walter, S. D. (1980). ''Estimating error rates of diagnostic tests. Biometrics,'' Vol. 36 (1) 167- 171.

[4] Branscum, A.J., Gardner, I.A., Johnson, W.O. (2005). ''Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling,'' Preventive Veterinary Medicine, 145-163

[5] Bradlow, E.T., Hardie, B.G.S., Fader, S.F. (2002). ''Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions,'' Journal of Computational and Graphical Statistics; 11(1), 189-201

[6] Scollnik, M., & Peter, D. (2017). ''Bayesian Inference for Three Bivariate Beta Binomial Models,'' The Open Statistics & Probability Journal, 2017, 08, 27-38.

[7] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van der Linde, A., (2002). ''Bayesian Measures of Model Complexity and Fit,'' Journal of The Royal Statistical Society B (2002) 64, part 4 pp. 583 – 639

[8] Broemeling, L.D., (2011). ''Bayesian Methods for Medical Test Accuracy,'' Diagnostics Journal; Diagnostics 2011, (1) 1-35.

[9] Basilio, D.P., Carlos, D.P., (2005). ''A Likelihood Approach to diagnostic test in clinical medicine,'' Statistical Journal, 3(1) 77-98.

[10] Lio, Y. L. (2009). ''A Note on Bayesian Estimation for the Negative Binomial Model,'' Pliska Studia Mathematica Bulgaria. 19, 207-216

[11] Lawan, A.M., (2022). '' A Bayesian Approach for The Evaluation of Diagnostic Test –Sensitivity And Specificity Of Blood Screen at Abubakar Tafawa Balewa University Teaching Hospital Bauchi,'' MSc dissertation, Dept. of Mathematical Sciences., University of Maiduguri, Nigeria., 2022. (Thesis or dissertation)