Bayesian Probabilities and Odds Ratio Analysis

M. E. Nja[1]; G. M. E. Nja[2]

[1]Department of Statistics,
University of Calabar,
Calabar, Nigeria.
e-mail: mbe_nja@yahoo.com

[2]Department of Public Health,
University of Calabar,
Calabar, Nigeria.
e-mail: glonja2@gmail.com

***Abstract*—The logit is a function of an odd while the posterior distribution is a function of the prior and a reversed conditional probability. These techniques are founded on the probability principle. This paper explores a functional relationship between them. Baye's principle has been used to show the relationship between the beta, uniform and Bernoulli distributions in an effort to establish this link. By designing an odds ratio function, a relationship between Bayesian probabilities and the odds ratio is established. This relationship shows that in any categorical data setting, the posterior probability is approximately equal to the likelihood obtained as a function of the odds ratio.**

Keywords: *Logit, posterior probability, logistic regression, response probability, uniform distribution.*

I. INTRODUCTION

Often when we are faced with several options to select from, limited information poses constraints in the attempt to select the option with the highest probability of success. In Bayesian methods, the problem is solved using posterior probability to update the prior. In logistic regression, the odds ratio is used to provide the solution. It is shown in this paper that both methods update information on the likelihood of the option that should be selected. Additionally, it is shown that the posterior probability is an approximation to the odds ratio. Having the Bernoulli process in the background, we show that the uniform prior on [0,1] is also a Beta (1,1) density and the resulting posterior probability is a Beta (k+1, n+k-1).Bayesian probabilities derive from Bayes' theorem:

$$p(m/x) = \frac{p(x/m)p(m)}{p(x/m)p(m) + p(x/f)p(f)} \qquad (1)$$

where

$p(m/x)$ is the posterior probability, $p(x/m)$ is the reverse conditional probability for males, $p(m)$ is the probability that a male has CA disease, $p(f)$ is the probability that a female has CA disease, and $p(x/f)$ is the conditional probability for female.

II. MATERIALS AND METHODS

A. Data Description

The data of Coronary Artery (CA) disease, presented by [4] is used for illustration. This data was cross tabulated according to sex, ECG status and CA status. A population of 78 people was examined for the incidence of CA disease.

B. The Beta And Uniform Distributions

Several authors including [1], [2], [5],[3] have demonstrated the use of Beta Distribution especially in Bayesian analysis. The Beta distribution is useful in modeling random probabilities and proportions. The Beta distribution with left parameter $\alpha \in (0, \infty)$ and right parameter $\beta \in (0, \infty)$ is the continuous distribution on (0, 1) with probability density function $f(x)$ given as

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1 \qquad (2)$$

where the B(α,β), the beta function is defined as

$$B(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du,\ \alpha > 0,\ \beta > 0 \qquad (3)$$

In terms of the gamma function, the beta function can be written as

$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$; α, β ϵ (0, ∞), B(α, β) is the normalizing constant, ensuring that $f(x)$ is a density function.

C. The Prior And Posterior Probabilities

Joyce (2016) gave a good analysis of Bayesian probabilities by starting with a discrete parameter θ and thereafter treating θ as a continuous parameter.
Consider the Bernoulli process ($x_1, x_2, \ldots, x_n$) where each x_i is a Bernoulli trial having the same parameter θ. If we take a random sample ($x_1, x_2, \ldots, x_n$) and let $f(\theta/x)$ denote the posterior density function, we can write

$$f(\theta/x)=\frac{f(x/\theta)f(\theta)}{f(x)}. \quad (4)$$

where $f(x)$ is a constant, so that $\frac{1}{f(x)}$ is a constant of proportionality. $f(x/\theta)$ is a Bernoulli conditional probability, $f(x/\theta)$ is considered as the reverse of the posterior probability, $f(\theta)$ is the prior probability $f(x/\theta)=\theta^k(1-\theta)^{n-k}$. Therefore, the posterior probability

$$f(\theta/x)=\frac{\theta^k(1-\theta)^{n-k}f(\theta)}{f(x)} \quad (5)$$

If the prior distribution is not known, it can be assumed that all values of θ are equally likely; this implies that the prior $f(\theta)$ is uniform. The posterior probability reduces to $f(\theta/x)=\frac{\theta^k(1-\theta)^{n-k}}{f(x)}$ which is a Beta distribution.

III. Statistical Analysis

Table 1:

Sex	ECG	Disease	No Disease	Total
Female	< 0.1 ST segment depression	4	11	15
Female	≥ 0.1 ST segment depression	8	10	18
Male	< 0.1 ST segment depression	9	9	18
Male	≥ 0.1 ST segment depression	21	6	27

Source: [4]: Categorical Data Analysis Using the SAS system

The SPSS version 21 has been adopted for the analysis using the data of table 1 to obtain the logits and the odds ratios. The SPSS execution is based on the Iterative Weighted Least Squares algorithm. The prior, the posterior and other conditional probabilities were computed in line with relevant Bayesian theorems.

IV. Results

From Table 1, the logit(θ_{ij}) is given from [5] as

$$\log\left(\frac{\theta_{ij}}{1-\theta_{ij}}\right)=-1.1747\ +1.2770\ sex\ +1.0545\ ECG \quad (6)$$

The odds ratio for male versus female is given as

$$\frac{e^{\beta_0+\beta_1}}{e^{\beta_0}}=e^{\beta_1} or \frac{e^{\beta_0+\beta_1+\beta_2}}{e^{\beta_0+\beta_2}}=e^{\beta_1}$$

From equation (6), the odds ratio for male versus female is

$$e^{\beta_1}=e^{1.2770}=3.586$$

From odds ratio analysis, the likelihood (probability) that he will have CA disease is

$$1-\left(1+e^{\beta_1}\right)^{-1}=0.782$$

Going by the Bayesian formula of (2), if a person is randomly selected from the population, the likelihood (posterior probability) that it will be a male is 0.715.

V. Discussion

The odds ratio analysis reveals that men are 3.586 times more likely to CA attack than females; that is for every female who has a CA disease, there will be a minimum of 3men having CA attack. This implies that when a man is randomly selected from this population, the likelihood (probability) that he will have CA disease is

$$1-\left(1+e^{\beta_1}\right)^{-1}=0.782$$

Now using Bayesian method, the likelihood (posterior probability) that a randomly selected male has CA disease is 0.715. These values are obtained using the Baye's formula of (2). For the females the posterior probability is 0.285. It can be seen that the odds ratio probability (0.782) compares favourably with the Bayesian probability (0.715). Thus, Bayesian probabilities can be used to approximate odds ratio probability in a cross tabulated data setting.

V. CONCLUSSIONS

Bayesian probabilities can be obtained from odds ratios. The posterior probability can be obtained using the functional relationship established in this paper. Computational rigours associated with the logistic regression estimation can avoided by using this Bayesian approach.

REFERENCES

[1] Fisher, I. (2012). Bayes' formula. URL: Pages.stern.nyu.edu.

[2] Joyce, D. (2016). A short introduction to Bayesian Statistics, part one. Clark University.

[3] Robinson, D. (2014). Understanding the beta distribution (using baseball statistics). URL: http://varianceexplained.org/statistics/beta_distribution_and_baseball/.

[4] Stokes, M. E., Davis, C. S., Koch, G. G. (2003). Categorical Data Analysis using the SAS system 2nd ed. SAS Inst. Inc. Nc, USA.

[5] Weissten, E. W. (2003). Beta distribution. Wolfram MathWorld. URL: mathworld.wolfram.com/BetaDistribution.hotm

.