Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 8, Issue 2 (2015): 13-20

BAYESIAN ANALYSIS OF INFANT MORTALITY BY STOCHASTIC MULTINOMIAL MODEL

G.P. Singh[1], A. Tripathi[1*], S.K.Singh[2] and Umesh Singh[2]
1. Department of Community medicine and DST-CIMS, BHU, Varanasi, India
2. Department of Statistics and DST-CIMS, BHU, Varanasi, India
E Mail: singhgpbhu@gmail.com , ashoka07bhu@yahoo.in*,
singhsk64@gmail.com , umeshsingh52@gmail.com

*Corresponding Author

Received January 11, 2015
Modified August 09, 2015
Accepted September 20, 2015

Abstract

In this study an effort was made to fit a multinomial model on the data, which came out from case control study design on the infant death in India. The Bayesian approach was used for data analysis. Some independent (Exposure) variables were taken to check whether they affect infant mortality. The considered independent variables were mother’s schooling, mother’s age, number of ANC, delivery conducted (Place), delivery conducted (Person), check up after delivery and breastfeeding initiation. It was found that mother’s age, delivery place and breastfeeding initiations significantly affect the infant mortality.

Key Words: Stochastic Model, Case-Control Study Design, Bayesian Approach

1. Introduction

The most common problem in infant mortality analysis is associated with the data of deaths during infanthood. In such situations, development of stochastic model is the most appropriate way to minimize the effect of these errors [1]. Stochastic model is an idealized mathematical description of random phenomena. Such a model has three essential components: the set of all possible outcomes, event of interest and a measure of uncertainty i.e. probability. The most interesting and difficult part of stochastic modelling is the assignment of probabilities to events [2]. If we take the case of infant mortality that a woman experienced then one may find the probability that the woman belongs to illiterate or non-illiterate group. For calculating such probability, we take a sample of women from population and by use of statistical inference procedure, we estimate this probability. In statistical inference, there are two broad categories of interpretation of probability: Bayesian inference and frequentist inference. These views often differ with each other on the fundamental nature of probability. Frequentist inference loosely defines probability as the limit of an event's relative frequency in a large number of trials. Bayesian inference, on the other hand, is able to assign probabilities to any statement. The basis for Bayesian inference is derived from Bayes' theorem. The Bayesian approach to statistical inference regarding a parameter θ collate all pre-existing information in form of prior distribution $p(\theta)$, reflecting both evidence based on past studies and current beliefs. The new evidence from the data collected during the current study is summarized by the likelihood $L(x|\theta)$ and the last step in the

Bayesian process is to combine the prior distribution with the likelihood using Bayes' theorem to get the conditional probability function p(θ |x) of θ, known as posterior probability function of parameter θ. Thus,

$$P(\theta|x) = \frac{L(x|\theta)P(\theta)}{\int L(x|\theta)P(\theta)\, d\theta}$$

From a mathematical point of view, weighting the prior distribution by the likelihood function forms the posterior probability distribution that is used to draw inference and thus forms conclusion about the relevant quantity of interest.

The use of stochastic model in field of infant mortality was probably initiated by [3]. After this many authors contributed towards modeling of infant and child mortality, most of them used classical approach to draw conclusion about parameters of model. But nowadays Bayesian approach of modeling is being used by various authors. Recently Bayesian inference procedure in case of stochastic modeling was used by [4]. They developed a frailty model successfully for child survival. [1] used binomial model in case of child mortality experienced by women during her total reproductive life span. As concern to infant mortality, infant mortality of a nation is widely accepted and long standing indicator of well-being of infants. Probability of dying before the age one year is infant mortality rate (express per 1000 live births). A high infant mortality rate is an indicator of risk of death during the first year of life and is indicative of unmet health needs and unfavorable environmental factors [5].The health needs and its utilization to care infant and mother is widely measured on the basis of number of ANC visits, Delivery Place, breastfeeding practices etc. So to find out the effect of such health facility on infant mortality, this study has been performed. In this study, a multinomial model is adopted for the women who experienced infant death in their last delivery and Bayesian analysis was done to find out effect of health facility on the infant mortality.

1.1 Objective

To study the affect of maternal and infant health practices on infant among rural families having poorest wealth index with the outcome of event (infant mortality) as binary and response (maternal and infant health practise) as polytomous.

1.2 Data and Covariates

The Data for study was taken from District Level Household and Facility Survey (DLHS-3) [6] which were collected by "International Institute for Population Science", Mumbai during 2007-2008. The aim of survey was to provide estimates on maternal and infant health, family planning and other reproductive health indicators. The data according to the objective are taken from DLHS-3. The descriptive statistics regarding selected data are given in Table-1. The data taken here represent women having poorest wealth index, giving their recent birth between 2004 -2005 and belonging to rural area of state of Uttar Pradesh in India. Out of a total of 2726 women found, 74 experienced infant deaths while 2652 did not experience so. After this, 74 women who experienced infant death were taken as case and remaining group of women who did not experience infant death were considered as control. As the number of women considered as control (2652) was large as compared to case, hence a random sample of 300 women was selected out of 2652 considering four controls for each case [7]. The reason of taking such type of data is that the lowest wealth indexed women found to experience more infant deaths in rural area of Uttar Pradesh [6]. Thus an inspection about health services used during and after the pregnancy and its relationship with infant death is important. Seven variables (covariates or factors) were taken namely “mother’s schooling”, “mother’s age”, “Number of ANC”, “delivery conducted (Place)”, “delivery conducted (Person)”, “check up after delivery” and “breastfeeding initiation” in study.

Variables	**Case**	**Control**	**Total**
Mother’s schooling			
0 years of schooling	62(83.3)	260(86.7)	322(86.1)
At least one year schooling	12(16.2)	40(13.3)	52(13.9)
Mother’s age(year)			
<=20	15(20.3)	34(11.3)	49(13.1)
More than 20	59(79.7)	266(88.7)	325(86.9)
Number of ANC			
0	47(63.5)	173(57.7)	220(58.8)
1	24(32.4)	89(29.7)	113(30.2)
More than 1	3(4.1)	38(12.7)	41(11.0)
Delivery Conducted(Place)			
Home	59(79.7)	266(88.7)	325(86.9)
Hospital	15(20.3)	34(11.3)	49(13.1)
Delivery Conducted(Person)			
Others(Not known)	16(21.6)	40(13.3)	56(15.0)
Friend/relative	38(51.4)	149(49.7)	187(50.0)
Health care practitioners	3(4.1)	9(3.0)	12(3.2)
Dai	17(23.0)	102(34.0)	119(31.8)
Checkup after Delivery			
No checkup	20(27.0)	66(22.0)	86(23.0)
checkup	54(73.0)	234(78.0)	288(77.0)
Breastfeeding initiation			
>3 day	44(59.5)	89(29.7)	133(35.6)
2-3 days	26(35.1)	142(47.3)	168(44.9)
Within one day	2(2.7)	37(12.3)	39(10.4)
Immediate	2(2.7)	32(10.7)	34(9.1)
Total	**74(100)**	**300(100)**	**374(100)**

Table-1: Distribution of case and control with various exposure factors.

2. Methodology

The women who experienced infant death in their recent birth are taken as cases and women who did not experience infant death in their recent birth are taken as control. The Seven variables have been taken as exposure factor separately to develop a case–control study design. The aim of case-control study design is to check the association between exposure factor and outcome of interest (i.e. infant mortality). In case-control study design, the outcome of interest plays a role of prediction for any subject to find out to which level of exposure subject belongs. The standard layout of design is given in design-1 and modeling process defined below.

Design-1

Exposure factor	Case	Control
Level-1	$N_1(\theta_1)$	$M_1(\pi_1)$
Level-i	$N_i(\theta_i)$	$M_i(\pi_i)$
Level-k	$N_k(\theta_k)$	$M_k(\pi_k)$
Total	N(1)	M(1)

Let there are k (say) groups or the category or level of exposure variables, and the cases and the controls are divided among these k categories of exposure variables. N_i and M_i are the numbers of cases and control corresponding to ith (i=1,2,...,k) level of a particular exposure variable such that $\sum N_i = N$ and $\sum M_i = M$. Let θ_i be the probability of ith level of an exposure variable under cases and π_i is similar probability under controls. Now we can say that out of N independent women having experience of infant death (cases) N_i belongs to ith (i=1,2,...,k) level of a particular exposure with probability θ_i (i=1,2,...,k) such that $\sum_{i=0}^{k} \theta_i=1$. In this way N_i can be assumed to have multinomial distribution. On the similar explanation, M_i can also be assumed to have multinomial distribution. The joint likelihood function for case and control can be written as.

$L(\underline{\theta}, \underline{\pi}; N_i, M_i) = L(\underline{\theta}; N_i) * L(\underline{\pi}; M_i)$
$= \prod_{i=1}^{k} \theta_i^{N_i} * \prod_{i=1}^{k} \pi_i^{M_i}$

where, $\theta_k = 1-\sum_{i=1}^{k-1} \theta_i$, $\pi_k = 1-\sum_{i=1}^{k-1} \pi_i$, $N_k = N - \sum_{i=1}^{k-1} N_i$ and $M_k = M-\sum_{i=1}^{k-1} M_i$ [8]

As in any Bayesian analysis, we formalize prior belief over unknown parameter. Thus the joint conjugate prior for $\underline{\theta}$ and $\underline{\pi}$.

$P(\underline{\theta}, \underline{\pi}) = P(\underline{\theta}) * P(\underline{\pi})$; ($\underline{\theta}$ and $\underline{\pi}$ are independent)
If $\underline{\theta} \sim$ Drich($\underline{\alpha}$) , $\underline{\pi} \sim$ Drich($\underline{\beta}$)
$P(\underline{\theta}) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}$, $P(\underline{\pi}) \propto \prod_{i=1}^{k} \pi_i^{\beta_i - 1}$
Then $P(\underline{\theta}, \underline{\pi}) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i - 1} * \prod_{i=1}^{k} \pi_i^{\beta_i - 1}$

Dirichlet distribution is the multivariate generalization of beta distribution and it works as natural conjugate prior for multinomial distribution. When we set all hyperparameters equal to unity then the Dirichlet distribution becomes uniform distribution. In other words we can say that all parameters vary all over their range and we have no information to restrict their boundaries. The main reason behind choosing Dirichlet prior is to make the posterior distribution and prior distribution of same family and ease the posterior analysis.

Combining likelihood and prior distribution by Bayes theorem, the karnel density of joint posterior distribution of $\underline{\theta}$ and $\underline{\pi}$ can be written as

$$P(\underline{\theta}, \underline{\pi} | \alpha_i, \beta_i, N_i, M_i) \propto P(\underline{\theta}, \underline{\pi}) * L(\underline{\theta}, \underline{\pi}; N_i, M_i)$$

$$P(\underline{\theta}, \underline{\pi} | \alpha_i, \beta_i, N_i, M_i) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i+N_i-1} * \prod_{i=1}^{k} \pi_i^{\beta_i+M_i-1}$$

$$\propto P(\theta_i | \alpha_i, N_i) * P(\pi_i | \beta_i, M_i)$$

Above equation is Dirichlet distribution with updated parameters (α_i+N_i) and (β_i+M_i), i = 1,2,…,k. An important fact with the Dirichlet distribution is that all the marginal posteriors are available in nice closed forms and follow beta distributions [9]. Thus the marginal posterior distributions of $\underline{\theta}$ and $\underline{\pi}$ are

$$P(\theta_i | \alpha_i, N_i) \quad \propto \quad \left[\boldsymbol{\theta_i}^{(\boldsymbol{\alpha_i+N_i-1})}(\mathbf{1}-\boldsymbol{\theta_i})^{(\boldsymbol{\alpha_0-(\alpha_i+N_i)-1})}\right]$$

$$P(\pi_i | \beta_i, M_i) \quad \propto \quad \left[\pi_i^{(\boldsymbol{\beta_i+M_i-1})}(\mathbf{1}-\pi_i)^{(\boldsymbol{\beta_0-(\beta_i+M_i)-1})}\right]$$

Where $\alpha_0 = \sum_{i=1}^{k}(\alpha_i + N_i)$ and, $\beta_0 = \sum_{i=1}^{k}(\beta_i + M_i)$ If we assume squared error loss function, the expression for the Bayes estimator of θ_i and π_i can be written as follows.

$$\theta_i = \frac{\alpha_i+N_i}{\alpha_0} \quad \text{and} \quad \pi_i = \frac{\beta_i+M_i}{\beta_0}$$

Loss functions are the mathematical formulae that define risk in estimation of parameters. In square error loss function the risk of over estimation and under estimation of parameters is equally likely.

Further for checking association between outcomes of interest with various level of exposures of factor the odds ratio is used as a tool. The odds ratio is defined as:

$$OR = \left(\frac{(\frac{\theta_i}{1-\theta_i})}{(\frac{\pi_i}{1-\pi_i})}\right) = \frac{\theta_i(1-\pi_i)}{\pi_i(1-\theta_i)}$$

Where $\frac{\theta_i}{1-\theta_i}$ is known as odds of exposure in favour of case and $\frac{\pi_i}{1-\pi_i}$ is odds of exposure factor to control so odds ratio is the ratio of these two odds.

In Bayesian analysis the posterior distribution of odd ratio does not come in nice closed form so in place of calculating odds the logarithm of this is used as a measure of association [10],[11]. Thus we define,

$$LOR_i = \text{In}\left[\frac{\theta_i(1-\pi_i)}{\pi_i(1-\theta_i)}\right]$$

The LOR denote log of OR which is differential of log of two odds. The posterior densities of log odds are well approximated by a normal density and using the fact that difference of two normal variables has normal distribution, the posterior distribution of log of OR follows normal distribution[10], with mean (D) and variance (T^2) as given below [11].

$$D_i = \ln\left(\frac{(\alpha_i+N_i+0.5)(\beta_0-(\beta_i+M_i)+0.5)}{(\alpha_0-(\alpha_i+N_i)+0.5)(\beta_i+M_i+0.5)}\right)$$

$$T_i^{\,2} = \left[\frac{1}{\alpha_i+N_i+1}+\frac{1}{\alpha_0+(\alpha_i+N_i)+1}+\frac{1}{\beta_i+M_i+1}+\frac{1}{\beta_0+(\beta_i+M_i)+1}\right]$$

The unit OR is treated as neutral effect of exposure on event, in other words one can say that log value of OR is equal to zero treated as neutral effect of exposure. Similarly positive value of log of OR shows that exposure is a risk factor for event (infant death), however the negative value shows a protective factor for event. If 95% highest density interval (HDI) of log of OR contains the zero, then log of OR is not able to give any authentic conclusion about exposure variable. In results and discussion, the lower and upper limit of 95% HDI are shown by LL and UL.

3. Results and Discussion

To analyze the data of Table 1 by the proposed methodology, first we take the values of hyperparameters. The choice of hyperparameters is the crucial part of Bayesian analysis and until and unless, a strong prior evidence is available about parameters under study, the non informative or vague prior is used to show the prior belief. Beta (1,1) the uniform distribution probability density was proposed by Thomas Bayes to represent ignorance of prior probabilities in Bayesian inference. It describes not only a state of complete ignorance, but also the state of knowledge in which we have observed at least one success and one failure, and therefore we have prior knowledge that both states are physically possible. In present study we have considered α_i and β_i 's equal to unity in every case. It shows that our prior information is not very strong about parameter of interest. So the result derived from analysis of data is not much affected by prior information. Further analysis is given in the Table 2.

Mother's schooling	$\boldsymbol{\theta_i}$	π_i	**D**	**T^2**	**LL**	**UL**
0 years of schooling	0.829	0.864	-0.292	0.115	-0.956	0.371
At least one year of schooling	0.171	0.136	0.292	0.115	-0.371	0.956
Mother's age						
$\leq$ 20	0.211	0.116	0.720	0.107	0.080	1.361
More than 20	0.789	0.884	-0.720	0.107	-1.361	-0.080
Number of ANC						
0	0.623	0.574	0.199	0.067	-0.309	0.707
1	0.325	0.297	0.136	0.073	-0.393	0.666
More than 1	0.052	0.129	-0.892	0.242	-1.856	0.073
Delivery conducted (place)						
Home	0.789	0.884	-0.720	0.107	-1.360	-0.079
Hospital	0.211	0.116	0.720	0.107	0.080	1.360
Delivery conducted (Person)						
Others(Not known)	0.218	0.135	0.589	0.099	-0.028	1.207

Friend/relative	0.500	0.493	0.025	0.063	-0.467	0.518
Health care practitioners	0.051	0.033	0.517	0.308	-0.569	1.605
Dai	0.231	0.339	-0.519	0.084	-1.086	0.047
Checkup after Delivery						
No checkup	0.276	0.222	0.301	0.082	-0.261	0.863
Checkup	0.724	0.778	-0.301	0.082	-0.863	0.261
Breastfeeding initiation						
>3 day	0.577	0.296	1.169	0.067	0.661	1.675
2-3 days	0.346	0.470	-0.502	0.068	-1.021	0.001
Within one day	0.038	0.125	-1.139	0.293	-2.199	-0.079
Immediate	0.038	0.109	-0.982	0.296	-2.049	0.084
Total	**1**	**1**				

Table-2 Outcome of Bayesian analysis

Table-2 shows that where mother's schooling is taken as exposure variable, probability of cases going in 0 years of mother's schooling category is 0.829 and that of controls is 0.864. These probabilities for at least one year of mother's schooling category are 0.171 and 0.136 respectively. The log of OR did not show any significant effect of mother's schooling because 95% HDI contains the value zero. In Table-2, the exposure variable mother's age has two levels of exposure or category (mother's age ≤ 20 and more than 20). The probabilities of cases going in both categories are 0.211 and 0.789 respectively and for controls, these probabilities are 0.116 and 0.884 respectively. The value of log of OR showed that the mother's age ≤ 20 years is risk factor for infant mortality. The number of ANC has been classified into 3 categories viz. 0, 1 and more than 1, the probabilities for cases being 0.623, 0.325 and 0.052 and for controls 0.574, 0.297 and 0.129 respectively. The log of OR did not show any significant effect for any category. The variable Delivery Place has been classified into two categories viz. home and hospital. The probabilities of cases and controls going in first category (home) are 0.789 and 0.884 respectively and for second category, these probabilities are 0.211 and 0.116 respectively. The value of log of OR shows that the home is safer place of delivery. It is a spurious association [12] came out from study due to the fact that women who have lower wealth status go to hospital delivery only in complicated delivery situation. The exposure variable "Delivery conducted by person" was classified into four categories viz. others (not known), friend/relative, Health care practitioners and Dai. The probabilities of cases going in these categories are 0.218, 0.500, 0.051 and 0.231 whereas for controls these probabilities are 0.135, 0.493, 0.033 and 0.339. The log of OR did not show significant value for any category of exposure variable. The exposure variable check-up after delivery was classified into two classes viz. No ckeckup and Checkup. The probabilities of cases for these classes are 0.276 and 0.724, whereas the probabilities for two classes of controls are 0.222 and 0.778. The log of OR showed no role of check-up after delivery in infant mortality. The exposure variable Breastfeeding initiation was classified into four categories viz. initiation after 3 days, initiation between 2-3 days, initiation within one day and the initiation of breastfeeding immediate after delivery. The probabilities of cases going in such category are 0.577, 0.346, 0.038 and 0.038 respectively, whereas for control group these probabilities are 0.296, 0.470, 0.125 and 0.109 respectively. The Log of OR showed that early starting of breast feeding is a protective factor from infant mortality.

Conclusion

In this paper, we have assessed the effect of maternal and infant health related practises during delivery and after delivery on infant mortality in lower wealth status women of rural areas of Uttar Pradesh by use of multinomial model and analysis has been carried out Bayesian setup. It is found that mother's age, Delivery place and breastfeeding initiation have significant effect on the infant mortality. Mothers having age up to 20 years were more prone to experience infant death which is due to early marriage of girls and avoidance of family planning methods. Early breastfeeding showed a protective role in infant mortality. as the early breast milk after delivery has many types of antibodies and raise immunity in newly born babies. The delivery conducted in home is found to be safer which shows spurious association. The variable education of mother does not play a significant role in infant mortality, it is probably due to the fact that the percentage of literate mothers is very low in lower wealth status of rural population of Uttar Pradesh. The exposure variables Number of ANC visit of mothers, check-up after delivery and person who conducted the delivery were not found to have significant effect on infant mortality. It is probably due to the poor level of utilization of such practises and also due to the fact that majority of birth deliveries are carried out at home and are conducted by untrained birth attendant or traditional birth attendant.

References

1. Singh, G. P., Singh, S. K. and Rahul (2011). Modelling of child deaths experienced by women in their reproductive life span. Book Chapter: Population and Reproductive Health: Prospective and Issues, p.206-215, New Delhi, India, Hindustan Publishing Corporation.
2. Rohatgi, V. K. (1984). Statistical Inference, New York, John Wiley & Sons.
3. Keyfitz, N. (1977). Introduction to the Mathematics of the Population with Revisions, Massachusetts, Addition-Wesley Publishing company,
4. Koissi, M. C. and Hogans, G. (2005). Using WinBUGS to study family frailty in infant mortality, African population studies, (20), p. 1-27.
5. Health Report (2012).Infant and Child Mortality in India: Levels, Trends and Determinants, Published by NIMS, ICMR and UNICEF India.
6. International Institute for Population Sciences (IIPS), (2010). District Level Household and Facility Survey (DLHS-3), 2007-08: India. Mumbai: IIPS.
7. Gordis L. (2009). Epidemiology, Philadelphia, Elsevier Publisher.
8. Lipschutz, S. (1982). Theory and Problems of Probability, Singapore, McGraw-Hill International.
9. Bansal, A. K. (2007). Bayesian Parametric Inference, New Delhi, India, Narosa Publishing House.
10. Ashby, D., Hutton, J. L. and McGee, M. A. (1993).Simple Bayesian Analyses for Case-Control Studies in Cancer Epidemiology. The Statistician, l (42), p. 385-397.
11. Jackman, S. (2009). Bayesian Analysis for the Social Sciences, United Kingdom, John Wiley & Sons.
12. Park K. (2010). Park's Textbook of Preventive and Social Medicine, Jabalpur, India, Banarasidas Bhanot Publisher.