International Journal of Statistics and Applications 2018, 8(2): 59-64
DOI: 10.5923/j.statistics.20180802.04

Bayesian Estimation of Proportion of Women Engaged in Sudan Agricultural Development: An R-WinBUGS Application

Siraj Osman Omer*, Mai Ibrahim Eljack, Zahia Sayed

Experimental Design and Analysis Unit, Agricultural Research Corporation, Wad Medani, Sudan

Abstract The objective of this paper is to estimate the proportion of women engaged in Sudan agricultural development. The study was used classical and Bayesian approaches where the best of the three priors were chosen for posterior estimation proportion. The agricultural studies graduate's data from governmental universities in Sudan for 2010-2011 were used. R2WinBUGS software for Bayesian inference on Binomial parameters can be used. The results showed that there was a significant difference (P<0.01) between classical estimation of female (67%) points compared that with male students (33%). The Bayesian estimate of the proportion of women was found as 64% when the prior distribution was beta (0.5, 0.5) and 65% for beta (0, 10). Bayesian approach for statistical inference in this study is useful and acceptable in gender studies. Bayesian estimation has been outlined for an estimation of the proportion of women engaged in Sudan agricultural development.

Keywords Bayesian estimation, Binomial distribution, R-language, WINBUGS

1. Introduction

Women play an indispensable role in farming and in improving the quality of life in rural areas, therefore, women are increasingly more likely to engage in agricultural fields compared to males [1]. The successful application of Bayesian data analysis has appeared in many different fields, including biological, environmental and social sciences, health sciences, business, computer science [2]. In Bayesian statistics, all uncertainties and all information are incorporated through the use of probability distributions, and all conclusions obey the laws of probability theory [3]. Many applications of the Bayesian method involve estimation of an unobservable parameter [4]. Bayesian approach demonstrates estimation by providing a natural and coherent approach for representing and manipulating all forms of uncertainty in modeling [5]. The probability mass function for a binomial model tells us how the probabilities of the different possible outcomes of an experiment vary as functions of fixed values of the parameters [6]. In the case of gender issue we might have data from prior experiments with similar populations which suggest that some values of the estimation probability are probably than others, because information can be incorporated into the analysis reason in a Bayesian inferences and decisions, because the Bayesian paradigm is the most acceptable approach, so far invented for quantifying all relevant sources of uncertainty in complicated problems and making choices in the face of such uncertainty [7]. Use of Markov Chain Monte Carlo (MCMC) methods, including WinBUGS for Bayesian inference requires knowledge and skill beyond statistical modeling [8]. In Binomial model, the choice of initial values may affect convergence substantially [9]. Literature containing recommendations on how to deal with these issues in MCMC use includes Brooks (1998), and Gilks, Richardson, and Spiegelhalter [10] sampling from the posterior Bayesian inference is usually presented as a method for determining how scientific belief should be modified by data [12]. R2WinBUGS package in R will be implemented, because it offers a versatile approach for making MCMC computations within WinBUGS and returning them to R [13]. The feature of R2WinBUGS software uses a text file containing the data model with parameters, and information on the parameters' a priori distribution [14]. In Binomial probability, Bayesian inference can be used to estimate the proportion (p) of female and male. Bayesian approach estimate is necessary to candidate the best prior for posterior probability distribution [11]. estimate of proportion (p) [15]. In this study a case of the simple binomial model can be computed the posterior summaries such as mean, mode, standard deviation and percentile intervals using WinBUGS software model development. The objective of this paper is to estimate the proportion of women engaged in Sudan agricultural development. In this study a Bayesian approach considering

* Corresponding author:
sirajstat@yahoo.com (Siraj Osman Omer)
Published online at http://journal.sapub.org/statistics

the using uniform prior.

2. Methodologies

2.1. Dataset

Data were obtained from the website http://www.mohe.gov.sd/ of the Ministry of Higher Education and Scientific Research on data agricultural studies graduates from governmental universities in Sudan for 2010-2011. The agricultural students graduate take as random sample from different agricultural colleges with different disciplines, despite starting with priors having quite different shapes in binomial model of the population contains a fixed number of departments/ colleges (n) in assumed that every student has the same probability of engaged in Sudan agricultural development, so the in binomial model, Bayesian posterior density is a true probability density which can be used to make direct probability statements about a parameter.

2.2. Statistical Methods

2.2.1. Binomial Distribution

In this study the conditional distributions of observed data say Y and let parameters θ to be the probability of success in a total of n independent trials of *binomial distribution* (n, θ). The conditional probability function for Y given θ is given by

$$P[Y=k|\theta]=\binom{n}{k}\ \theta^{k}\ (1-\theta)^{n-k} \qquad (1)$$

Here θ is held as fixed at the probability distribution of Y over its possible values k=1…n.

2.2.2. Binomial Distribution in WinBUGS

WinBUGS is a software package that uses Markov chain Monte Carlo (MCMC) methods to fit Bayesian statistical models, has facilitated Bayesian analysis in a wide variety of application areas [11]. R2WinBUGS is an R package for executing WinBUGS from R [16] Walters (1985) used the uniform prior and its implied posterior distribution in constructing a confidence interval for a binomial parameter (in Bayesian terminology, a credible region"). Also Diaconis and Freedman (1990) investigated the degree to which posterior distributions put relatively greater mass close to the sample proportion p as n increases. Bayesian Inference for a prior distribution for a proportion is given as fallow

$$\theta \sim beta[a,b] \qquad (2)$$

Where Beta[a, b] represents a beta distribution with properties:

$$p(\theta|a,b)=\frac{A(a,b)}{A(a).A(b)}\theta^{a-1}(1-\theta)^{b-1}\ , where\ \theta\epsilon(0,1) \qquad (3)$$

$$E(\theta|a,b)=\frac{A}{a+b}$$

$$V(\theta|a,b)=\frac{ab}{(a+b)^{2}(a+b+1)} \qquad (4)$$

Where E(.) and V(.) stand for the expected value and variance of the random variable in the parenthesis under the condition if any. The binomial distribution in WinBUGS parameterized in terms of its proportion, the model being contains (alpha and beta) [17]. In Bayesian estimation beta-binomial distribution can be used very shortly mentioned as the predictive distribution for the binomial distribution, given the conjugate prior distribution, the beta distribution. WinBUGS notation: theta~dbeta(a, b). In R2WinBUGS, the good-ness-of fit statistics can be obtained using deviance information criterion (DIC) and effective number of parameters (pD), and expected values (posterior estimates) of the parameters of interest, MC error, and quantiles of their posterior distribution. DIC, smaller the better, was used to choose the prior [18].

2.2.3. Prior Distribution

Prior information in Bayesian analyses is represented by probability distributions for the parameters and uniform prior can be used of previous data on the impacts of gender issues in human development[19]. The uniform prior of Beta (2*,* *2*) =U (0, 1) corresponds to having 2 prior experiments, one of which was a woman and the other man. Moreover, with Beta (0, 0) prior the posterior is not defending if the observed data happen to be either 0 or *N* under a Binomial such as (n;k) model. For single parameters, the Jffreys' prior is sometimes used, but for Multiparameter problems the results are more controversial, and a hierarchical modeling approach is more common. Uncertainty about parameter can be updated repeatedly when new data are available for taking a current posterior distribution as prior and compute new posterior distribution conditional on new data [20].

2.3. Bayesian Estimation of Binomial Proportion

Bolstad (2004) has derived the Bayesian estimator of a binomial proportion for Bayesian estimation of binomial proportion where, a proportion of the population has some attribute will came from large population. In study a random sample from the population will be taken and let *Y* be the observed number of the population. The conditional distribution of the observation *Y,* the total number of successes in *n* trials given the parameter, is binomial (n, k). The conditional probability function or likelihood function for y given θ is given by

$$p(k)=\binom{n}{k}\ \theta^{k}\ (1-\theta)^{n-k}\ \text{for k=1,…, k} \qquad (5)$$

Here θ is assuming as a fixed, and looking at the probability distribution of *y* over its possible values. Therefore, relationship between θ and *k*, can be holding k fixed parameters at the number of successes we observed. To use Bayes' theorem, assuming g(θ) is prior distribution that gives our belief about the possible values of the parameter θ before taking the data. It is important to realize that the prior must not be constructed from the data. Bayes' theorem is summarized by posterior is proportional to the prior times the likelihood.

$$p(\theta|y) \propto p(\theta) \times f(\mathrm{y}|\theta) \quad (6)$$

To get the actual posterior, we need to divide equation (5) this by some constant k to make sure it is a probability distribution, meaning that the area under the posterior integrates to 1. We find k by integrating p. So $p(\theta) \times f(\mathrm{y}|\theta)$ over the whole range. So, in general,

$$p(\theta|y) = \frac{p(\theta) \times f(\mathrm{y}|\theta)}{\int_0^1 p(\theta) \times f(\mathrm{y}|\theta)dx}$$

This requires integration. Depending on the prior $p(\theta)$ chosen, there may not necessarily be a closed form for the integral, so it may be necessary to the integration numerically. Suppose a beta(a,b) prior density is used for θ:

$$p(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\ for\ 0 \le \theta \le 1 \quad (7)$$

The posterior is proportional to prior times likelihood. The constants in the prior and likelihood, can be ignored, when constants don't depend on the parameter, because, the multiplying the prior or the likelihood by a constant won't affect the results of Bayes' theorem. This gives

$$p(\theta|y) \propto \theta^{a+y-1}(1-\theta)^{b+n+y-1}\ for\ 0 \le \theta \le 1 \quad (8)$$

Where $p(\theta|y)$ is the shape of the posterior as a function of θ. beta distribution can recognized with parameters $a' = a + y$ and $b' = b + n - y$. That is, we add the number of successes to a and number of failures to b:

$$p(\theta|y) = \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)}\theta^{y+a-1}(1-\theta)^{n-y+b-1}$$
$$for\ 0 \le \theta \le 1 \quad (9)$$

We note that the uniform prior is a special case of the *beta (a,b)* prior where *a=1* and *b=1*.

The posterior mean is a very frequently used measure of location. It is the expected value,

or mean of the posterior distribution.

$$m' = \int_0^1 \theta p(\theta|\mathrm{y})dx \quad (10)$$

When the posterior $p(\theta|y)$ is *beta*(a',b') the posterior mean equals

$$m' = \frac{a'}{a'+b'} \quad (11)$$

The posterior mean as the Bayesian estimate for θ, can be estimated use the *beta(1,1)* prior (uniform prior).

3. Results and Discussion

3.1. Case Study: Gender approaches in Agricultural Studies Graduates

Students graduated in agricultural studies, both male and female, need a job doing. In agricultural practices men are more willing to share decision-making with their wives, though working hardly in any situation or bad situation they generally still consider themselves household. The example view that, the number of females is more than male into all disciplines and gardening, it is important to note that these changes have occurred over a period as short as two or three years. Labor market needs an effective graduates depend upon qualification, experiences, availability and should be considered in the broader sense to include markets for inputs also required a good strategy, concepts of 'participation' and 'gender' has been a part of emancipator discourse and practices with increasingly visible as an issue in development. Gender has become increasingly visible as an issue in development and widely used in agricultural development, referring primarily to participation in projects or in the community [21]. Women participation needs be processed in inclusion and critical reflection encouraged by women participatory [22].

Table 1. Total number of graduates of male and female from agricultural colleges during 2010-2011

Discipline	nm	nf	Total	p
Agricultural studies	506	1193	1699	0.70
Veterinary	132	243	375	0.65
Animal production	158	276	434	0.64
Natural recourses	152	299	451	0.66
Environmental studies	107	151	258	0.59
Forestry	21	30	51	0.59
Total	1076	2192	3268	0.67
Means	179.33	365.33	544.67	0.64
SD	167.60	417.32	584.50	0.043
Variance	28091.07	174153.07	341635.47	0.002

Where um= number of males, nf= number of females, p= proportion

Table 1 show the analysis of binary graduate student of gender data published by 2012 in which the numbers of female more than male at N=6 different disciplines are recorded, agricultural graduates are more likely to be female with means 365 with standard division 417.3 while Male 159 with SD 167. The results highlight that the comparison between male and female of number total and their mean estimates for each specialization. The observed number of female nf at each disciplines y_i is assumed as binomial distribution with sample size n_i and true rate p_i. Based on the Bayesian approach the estimation can be conducted the main focus, which has been given to women of agricultural graduates, the proportion parameters indicated contribution role that women take in the farming live, in the family decision and also in the market selling process.

3.2. Selections of Priors

The analysis components computed are presented along with the statistics using Bayesian approach in the Table 2, Table 3 and Table 4. The choices of priors for Bayesian analysis were made from the statistics given in Table 1. The values of DIC and pD are reasonably close to each other for the three priors sets P_1, P_2 and P_3 of beta distribution respectively. However, the prior setb P_2 seems to have numerically lowest value of DIC (48.068). We took P_2 for further estimation of the estimation of proportion of women.

Table 2. DIC values for selection of the priors of beta priors

Prior model for θ	Dbar ($\overline{D}$)	Dhat ($\hat{D}$)	p_D	DIC
P_1 =Beta (0.5, 5)	43.083	37.258	5.825	48.908
P_2= Beta (0, 10)	43.142	38.217	4.926	48.068
P_3= Beta (0.5, 10)	43.106	37.264	5.842	48.947

$\overline{D}$ =posterior mean of (- 2 × log-likelihood). $\hat{D}$ = - 2 × log-likelihood at posterior means of parameters. p_D = effective number of parameters *DIC* = Deviance information criterion.

Table 3. Description of the marginal of posterior distributions for odds ratio (od) and proportion of females (p) from MCMC simulations: means, standard deviation and 95% credible intervals (2.5%, 50% and 97.5 percentiles) on beta (0, 10)

Node	Mean	SE	MC error	Percentile		
				2.50%	median	97.50%
deviance	43.14	3.32	0.07	38.41	4257.00%	51.5
od_1	1185.00	19.02	0.42	1146	1185.00	1223.00
od_2	243.20	8.11	0.22	227.5	243.30	258.70
od3	277.30	9.17	0.18	259.2	277.40	294.70
od4	297.70	9.13	0.20	279.8	297.90	316.00
od5	156.20	7.80	0.17	140.2	156.60	170.70
od6	31.87	2.41	0.06	26.44	32.07	36.09
p1	0.70	0.011	0.0003	0.67	0.70	0.72
p2	0.65	0.022	0.0006	0.61	0.65	0.69
p3	0.64	0.021	0.0004	0.60	0.64	0.68
p4	0.66	0.020	0.0004	0.62	0.66	0.70
p5	0.61	0.030	0.0008	0.54	0.61	0.66
p6	0.62	0.047	0.0012	0.52	0.63	0.71

Where od_i=odds ratio of females for i-th discipline class (i=1,2,…6), p_i= proportion of females in i-th discipline class

Table 2 gives a summary of the marginal posterior distribution for all the parameters in the model. The true values are always within their interval estimations, deviance in the posterior distributions. The results show that The Bayesian posterior observed larger means than the maximum likelihood or REML estimates, which correspond to the *modes* of the likelihood. As expected, the empirical Bayes point estimates of the proportions very sequence proportion as follow Agricultural studies (70%), Veterinary (65%), Animal production (64%) Natural recourses (66%), Environmental studies (61%) and Forestry (62%) respectively. The results highlight shat using Bayesian approach depend on posterior means of point estimates of proportion of women engaged in Sudan agricultural development are very high, This is evidence that our MCMC sampler is indeed drawing from a reasonable approximation to the true posterior distribution. (Additional appropriate comparisons, such as posterior standard deviations with frequentist standard errors are shown. The posterior probability that the proportion percentage of agriculture from the total agricultural graduated is between 0.80- 0.95 this interval is wide, which the proportion that the experiment provides little information on female proportion. However, this is more valuable than the simple p= (67%) obtained in table 1.

Table 4. Description of the marginal of Beta posterior distributions for proportions from MCMC simulations of female (P)

Discipline	Beta distribution		
	(0.5, 5)	(0, 10)	(0.5, 10)
Agricultural studies	0.70	0.70	0.70
Veterinary	0.65	0.65	0.65
Animal production	0.64	0.64	0.64
Natural recourses	0.66	0.66	0.66
Environmental studies	0.59	0.61	0.59
Forestry	0.59	0.62	0.59
Average proportion of Female	0.64	0.65	0.64
Average proportion of Male	0.36	0.35	0.36

Table 4 represented statistical model and posterior parameter estimation of Beta distributions (0.5, 5), (0, 10) and (0.5, 10) respectively. Bayesian methods are sometimes criticized because of the possibility of subjectivity in defining the prior distributions. The results highlight the important concept of conjugacy in Bayesian statistics. When the prior and likelihood are of such a form that the posterior distribution follows the same form as the prior, the prior and likelihood are said to be conjugate. To illustrate some of these ideas, table 4 plots the beta distribution for (0.5, 0.5), (0, 10) and (0.5, 10) as the number female proportion for each discipline, the variance of the beta distribution decrease as the number of data items increases. Also the posterior distribution of female and male in a scenario in which Bayesian Estimation of proportion of Women engaged in agricultural development shows the probability assigned to that value of the beta posterior distribution. However, all the beta probabilities have been shown different scaled. The prior distribution is a beta distribution, for example (0, 10) the mean of the posterior distribution is a compromise between the mean of the prior distribution and the mean of the data. The posterior distribution can be very complicated f models that include large numbers of parameters and this may make it impossible to compute summary statistics analytically. Statistical inference can be made about the parameter values. Our results comparing the classical and Bayesian estimation of the proportion of women engaged in Sudan agricultural development using R2WinBUGSApplication. The number of male students will go down. An increasing number of female are moving towards using general statements about the qualities of their graduates as a part of their policy and quality assurance frameworks. According to the results of study, highly encouraged, the role of women to be engaged in agricultural development. It is also realized that growing cereal crop and increase the production has relative advantage of a woman; therefore agricultural female graduates will make essential contributions to agricultural and rural economies in Sudan.

4. Conclusions

The paper view, a framework developed for Bayesian binomial model for gender issue. In this study the data were analyzed using three candidate beta models. Bayesian estimation has been outlined of the proportion of women engaged in Sudan agricultural development. R2WinBUGS software package and R-environmental have been used with Markov chain Monte Carlo (MCMC) methods to fit Bayesian statistical models, has facilitated Bayesian analysis in a wide variety of applications, Bayesian estimation of proportion to the Binomial model based on percentage value is very close to the frequentist approach. Bayesian approach provides information on precision (in term of standard deviation and credible interval) of proportion of parameters.

Appendixes

WinBugs code for binomial distribution

```
#load packs
library(lattice)
library(coda)
library(R2WinBUGS)
#data from bdata3.........................................
bdata3<- read.table("bdata3.txt", header=TRUE)
bdata3

nm<- bdata3$nm
nf<- bdata3$nf
K<- length(nm)
K
t<- array(0, dim=K)
t<- (nm+nf)
t
a<- 5
b<- 10
print(cbind(nm,nf,t))
data<- list("nf","t", "K","a","b")
data

inits1<-    list(p=c(rep(.12,K)))
inits2<-    list(p=c(rep(.10,K)))
inits3<-    list(p=c(rep(.11,K)))
inits <- list(inits1, inits2, inits3)
inits

parameters <- list("p", "od")
parameters

gender3.sim<- bugs(data, inits, parameters,
"Gender3.bug", n.chains=3, n.iter=100000,
n.sims=5000, debug=TRUE)

#    rcb.sim<- bugs(data, inits, parameters,
"Gender3.bug", n.chains=3, n.iter=100000,
n.sims=3000, debug=TRUE)
```

: The r-codes were used to carry out calculations of standard error of differences and the posterior average and quantiles of ranks distribution

WinBugs code for Bayesian binomial data analysis

```
######## bug model data analysis
#This file is dbino.bug

model {

  # 1. data model

    for (i in 1 : K){
                        nf[i] ~ dbin(p[i] , t[i])

                          p[i]    ~ dbeta( 2, 10)
                          }

# parameters of interest

     for(i in 1: K){    nfhat[i] <- t[i] * p[i]
           od[i]<- p[i]/(1-p[i])}

          }
# end of BUGS codes
```

REFERENCES

[1] Raney, T., Anríquez, G., Croppenstedt, A., Gerosa, S., Lowder, S., Matuscke, I., Skoet, J. and Doss, C. FAO, ESA Working Paper No. 11-02, 2011. available at http://www.fao.org/publications/sofa/en/.

[2] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. Bayesian Data Analysis. Second Edition.Chapman & Hall, Boca Raton, FL, 2004.

[3] Koop. Bayesian Analysis, Computation and Communication Software. Journal of Applied Econometrics, vol.14, pp. 677-689.

[4] Sevin., V. and Ergun, G., “Usage of different prior distribution Bayesian vector Autoregressive Models,” VoL.38, pp.85 – 93, 2009.

[5] William, M. B., “Introduction to Bayesian Statistics,” ISBN, John Wiley & Sons, Inc, 2004.

[6] DeCoursey W. J. Statistics and Probability for Engineering Applications With Microsoft® Excel, Library of Congress Cataloging-in-Publication Data, USA. ISBN: 978-0-7506-7618-2, 2003.

[7] Draper, D., “Bayesian Statistical Analysis in Medical Research,” ROLE Steering Committee Meeting, New York NY, 2007.

[8] Rahardja., D., Zhao, Y.D. and Zhang, H. Bayesian Credible Sets for a Binomial Proportion Based on One-Sample Binary Data Subject to One Type of Misclassi. Journal of Data Science, vol.10, pp. 51-59, 2012.

[9] Stefano., C. and Spezzaferri, F., “A Bayesian Approach to Subset Analysis with Application to Binomial Count Data,” Translated: 2008-05-28 / SLB, 2005.

[10] Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. BUGS 0.5: Bayesian inference Using Gibbs Sampling - Manual (version ii), Medical Research Council Biostatistics Unit, Cambridge, 1996.

[11] Cowles., M.K., “Review of WinBUGS 1.4,” the American Statistician, Vol. 58, pp.4, 2014.

[12] Goodman, S. N., “Toward Evidence-Based Medical Statistics: The Bayes Factor,” Ann Intern Med, vol.130, pp.1005-1013, 1999.

[13] Ntzoufras, I. Bayesian Modeling Using WinBugs, First Edition, ISBN John Wiley & Sons, Inc, 2007.

[14] Agresti, A. and Hitchcock. D., “Bayesian inference for categorical data analysis, Statistical Methods and Application (Journal of the Italian Statistical Society),” vol.14, pp.297-330, 2005.

[15] Singh., M, Al-Yassin, A. and Omer, S. Bayesian Estimation of Genotypes Means, Precision and Genetic Gain due to Selection from Routinely Used Barley Trials. Crop Science, vol. 55, pp. 501–513, 2015.

[16] Sturtz, S., Liggesy, U. and Gelman, A. R2OpenBUGS: A Package for Running WinBUGS from R. Journal of Statistical Software, vol. 12, pp. 1-16, 2005.

[17] Lunn, D., Thomas, A., Best, N., Spiegelhalter, D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, 325–337, 2000.

[18] Cowles, MK and Carlin, BP (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association, 91, 883–904.

[19] Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515–533, 2009.

[20] Jacobs., R., “Bayesian Statistics: Beta-Binomial Model,” Department of Brain & Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA, 2008.

[21] World Bank, FAO & IFAD. Gender in Agriculture Sourcebook. The World Bank, Washington D.C, 2009.

[22] Supriya., A., “Gender and participation overview report,” Institute of Development Studies, ISBN 1 85864 385, 2001.