

On the Use of Posterior Probabilistic Clustering

S.S. Abdulkadir
Department of Statistics and Operations Research, Modibbo Adamal University of Technology, Yola Adamawa State.Nigeria saidusauta@yahoo.com

E. T. Jolayemi ,
Department of Statistics, University of Ilorin, Ilorin. Kwara State.Nigeria

ABSTRACT.
Bayesian approach to mixture models makes use of Gibbs sampler, the most common of Markov Chain Monte Carlo (MCMC), for estimation of posterior density and subsequent classification of objects into components of mixture, especially for conjugate priors. In practice conjugacy may not exist and when it does, the time required calculating the posterior density will be far too high for the Bayesian approach to be applied in practice (McLachlan and Peel, 2000). Therefore, we developed a clustering procedure that is a result of using non-conjugate prior distribution of product multinomial to obtain posterior distribution that is hypergeometric, for cross-classifying categorical data. The performance of the scheme was examined through a simulation study of observed tables of counts compared with expected generated by assuming product multinomial to obtain posterior distribution under variety of parameter distributions and loadings. We observed that the approach performed well when the component proportions are properly distinguishable. The approach was illustrated using real life data from social science.

Keywords: Mixture model, Posterior Probability, Dirichlet distribution, Markov Chain Monte Carlo (MCMC).

1. INTROUDCTION.

In some applications of mixture models, questions related to clustering may arise only after the mixture model has been fitted. The reason for fitting the model is to obtain adequate model for the distribution of data. If this were achieved then it may be of interest to consider the problem of identifying the components of the mixture with externally existing groups or subpopulations.

The mixture model can only be used purely as a device for exposing any grouping that may underlie the data. This approach can be used for clustering where an initially specified number of groups are in various proportions (McLachlan and Peel, 2000).

A parametric form is specified for each of component density and a probabilistic clustering of the data is obtained in terms of the fitted posterior probability of component parameters for the data. To estimate the parameters of the mixture models, numerical approach techniques such as Expected Maximization (EM) algorithm developed by Dumpster, Laid and Rubin (1977), the classical ones of scoring for parameters, and Newton-Raphson methods have been discussed in literature (Everitt and Hand, 1981; Woodward *et.al* ,1984; McLachlan and Peel ,2000), to tackle this problem. The estimation is straightforward using EM algorithm (McLachlan and Peel, 2000). In Bayesian approach to mixture models the estimation is feasible using posterior simulation through the development of Markov Chain Monte Carlo (MCMC) methods. The development of MCMC, the Gibbs Sampler, proposed by Tanner and Wong (1987), and Gelfand and Smith (1990) leads to application of Bayesian approach for mixtures in practice. The Gibbs sampler, the most common of MCMC algorithm can be implemented properly for conjugate prior. In many cases conjugacy may not exist in practice and as such the application of Gibbs sampler is not practicable. Damien *et.al* (1999) asserted that practitioner may turn to the Metropolis- Hastings Algorithms. However, the algorithms may be difficult to set up and in particular 'tuning' to achieved satisfactory performance (Bernett *et.al* 1996; Chib and Greenberg 1995). Alternatively, 'black box' random variate generation techniques such as the rejection algorithm (Devroye,1986),adaptive rejection sampling for log-concave densities (Gilks and Wild,1992) or the ratio-of-uniform method(Wakefield *et-al* 1991) may be used. The use of such techniques may be daunting to those who are unfamiliar with their use (Damien *et. al* (1998).

In mixture models, if the component densities belong to the same exponential family and allows conjugate priors for both component parameters and the mixing proportions to derive posterior density, the posterior expectation of these parameters, even though can be written in closed form, the time required to calculate the posterior density will be far too high for the Bayesian approach to be applied in practice, even for moderate sample sizes (McLachlan and Peel, 2000; Cheng and Curie, 2003)

This article focuses on Bayesian approach to mixture models for non-conjugate prior where the central limit theorem is used to sample from posterior distribution for categorical data. This is a form of an adopted EM algorithm for the estimation and classification of objects into components of the mixture.

2. Model Fitting.

Let N be the observations taken from an infinite population and cross-classified using two categorical variables, says X and Y having r and c outcomes respectively. Let $\{n_{ij}\}$ denote the cell counts obtained in the cell (i,j),where i=1,…,r and j=1,…,c. Let θ_{ij} denotes the probability that an observation falls in that particular cell (i,j).Then $\{n_{ij}\}$ has a multinomial distribution which can be displayed in a r by c contingency table. But depending on the method of data collection, the underlying distribution for the table could be independent Poisson, full multinomial, product multinomial, or hypergeometric distribution,Birch(1963),Jolayemi(1982),Agresti(1990),Sanni and Jolayemi(1998) among many other authors. Furthermore, they all asserted that these distributions all have parameters that are fixed but unknown.

Without loss of generality we assume product multinomial of dimension c, where c is unknown. A mixture model (MM) now says that the population having c outcomes actually contain k mixtures. That is, if $\underline{n}$ is an r x c matrix of observations.

$$f(\underline{n}/\theta_i) = \prod_{i=1}^{r} \binom{n_{i.}}{n_{i1}.n_{i2}......n_{ic}} \theta_{i1}^{n_{i1}}\theta_{i2}^{n_{i2}}......\theta_{1c}^{n_{ic}} \quad \text{………………………(2.01)}$$

where $(n_{i1}, n_{i2}, ...n_{ic})$ are the multinomial observations generated as independent multinomial random variables with parameter vector $\alpha_1, \alpha_2, ...\alpha_k$ for each i = (1,….k) and $\sum_j \theta_{ij} = 1$.

Here we assume that θ_{ij} are known and having a distribution rather than fixed. Therefore, since 0< θ_{ij} <1

θ_{ij} ~Dirichlet distribution.i.e.

$$f(\underline{\theta}_i) = \frac{\Gamma(\lambda_{ij})}{\Pi_j^c \Gamma(\lambda_j)} \Pi_j^c \theta_{ij}^{\lambda_j - 1} \quad \text{…………………………………………(2.02)}$$

where $\sum_{j=1}^{c} \alpha_j = 1, \quad \sum_j \theta_{ij} = 1$

The unconditional probability is given by

$$f(\underline{n}) = \int_0^1 p_r(\theta_{ij}) f_i(\underline{n}/\theta_i) d\theta_{ij} \quad \text{……………………………………........(2.03)}$$

where $p_r(\theta_{ij}) and f_i(\underline{n}/\theta_i)$ are respectively the mixing proportion and component density.

Therefore equation (2.03) becomes

$$f(\underline{n}) = \int_0^1 \binom{n_{i.}}{n_{i1}.n_{i2}.........n_{ic}} \theta_{i1}^{n_{i1}}\theta_{i2}^{n_{12}}.........\theta_{ic}^{n_{ic}} * \frac{\Gamma\left(\sum_{i=1}^{r}\lambda_i\right)}{\Pi_{i=1}^{r}\Gamma(\lambda_i)} \Pi_{i=1}^{r} \theta_{ij}^{\lambda_i - 1} d\theta_{ij} \quad \text{……....(2.04)}$$

after some manipulation we have

$$f(\underline{n}) = \frac{n_{i.}!}{\Pi_{i=1}^{r} n_i!}\Pi_{i=1}^{r}\frac{\binom{n_{i.}+\lambda_i-1}{n_{i.}}}{\binom{n_{..}+g-1}{n_{..}}} \ldots\ldots(2.05)$$

which is hypergeometric distribution. It can also be shown that the posterior distribution, hypergeometric, is obtained if any other conditional distribution is assumed. This posterior distribution is not from the same family as prior, therefore we are dealing with Bayesian non-conjugate prior.

The special case of equation (2.01) is where r=c =2, the equation reduces to product binomial for counts $\{n_{ij}\}$ and so $f(n_{11})$ is given by

$$f(n_{11}) = \binom{n_1}{n_{11}}\int_0^1 \theta_{11}^{n_{11}}(1-\theta_{11})^{n_{1.}-n_{11}}\frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\theta_{11}^{\lambda_1-1}(1-\theta_{11})^{\lambda_2-1}d\theta_{11} \ldots\ldots (2.06)$$

$$= \binom{n_{1.}}{n_{11}}\left(\frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\int_0^1 \theta_{11}^{\lambda_1+n_{11}-1}(1-\theta_{11})^{n_{1.}+\lambda_2-n_{11}-1}d\theta_{11}\right)$$

$$= \binom{n_{1.}}{n_{11}}\frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\frac{\Gamma(n_{11}+\lambda_1)\Gamma(n_{1.}+\lambda_2-n_{11})}{\Gamma(n_{1.}+\lambda_1+\lambda_2)}$$

$$= \frac{n_{i.}!}{n_{11}!(n_{1.}-n_{11})!}\frac{(\lambda_1+\lambda_2-1)!}{(\lambda_1-1)!(\lambda_2-1)!}\frac{(n_{11}+\lambda_1-1)!(n_{1.}+\lambda_2-n_{11}-1)!}{(n_{1.}+\lambda_1+\lambda_2-1)!}$$

$$= \binom{n_{1.}}{n_{11}}\frac{\binom{n_{11}+\lambda_1-1}{n_{11}}\binom{n_{1.}+\lambda_2-n_{11}-1}{n_{1.}-n_{11}}}{\binom{n_{1.}+\lambda_1+\lambda_2-1}{n_{1.}}} \ldots\ldots(2.07)$$

The equations (2.05) and (2.07) suggest hypergeometric sampling for classification into components of mixture model for general and special cases respectively.

Therefore expected value of n_{11} from the product binomial denoted by E(n_{11}) is given as

$$\sum n_{11} f(n_{11}) = \sum n_{11}\frac{\binom{n_{1.}}{n_{11}}\binom{n_{.1}}{n_{1.}-n_{11}}}{\binom{n}{n_{11}}}$$

$$= \frac{n_{1.}n_{.1}}{n}\sum_{n_{11}>1}\frac{\binom{n_{1.}-1}{n_{11}-1}\binom{n_{..1}}{n_{.1}-n_{11}}}{\binom{n-1}{n_{.1}-1}}$$

$$= \frac{n_{1.}n_{.1}}{n} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2.08)$$

Also the variance of n_{11} is given as

$\text{Var}(n_{11}) = \text{E}(n_{11}^2) - (\text{E}(n_{11}))^2$ but

$\text{E}(n_{11}^2) = \text{E}(n_{11}(n_{11}\text{-}1)) + \text{E}(n_{11})$

$$= \sum(n_{11}(n_{11}\text{-}1))\frac{\binom{n_{1.}}{n_{11}}\binom{n_{.1.}}{n_{.1}-n_{11}}}{\binom{n}{n_{.1}}}$$

$$= \frac{n_{1.}n_{.1}(n_{1.}-1)(n_{.1}-1)}{n(n-1)}\sum_{n_{11}>2}\frac{\binom{n_{1.}-2}{n_{11}-2}\binom{n_{.1}}{n_{.1}-2}}{\binom{n-2}{n_{.1}-2}}$$

$$= \frac{n_{1.}n_{.1}(n_{1.}-1)(n_{.1}-1)}{n(n-1)}$$

Therefore $\text{Var}(n_{11}) = \text{E}(n_{11}(n_{11}\text{-}1)) + \text{E}(n_{11}) - (\text{E}(n_{11}))^2$

$$= \frac{n_{1.}n_{.1}(n_{1.}-1)(n_{.1}-1)}{n(n-1)} + \frac{n_{1.}n_{.1}}{n} - \left(\frac{n_{1.}n_{.1}}{n}\right)^2$$

$$= \frac{n_{1.}n_{.1}}{n}\left(\frac{(n_{1.}-1)(n_{.1}-1)}{(n-1)} + 1 - \frac{n_{1.}n_{.1}}{n}\right)$$

3. **Design of Simulation Study.**

Illustration of Classification Procedure for objects into components of a mixture.

(i) Specify the number of components in advance

(ii) Give the Sample size or component loadings for each component.

(iii)Estimate the expected and variance values $\{n_{ij}\}$ of product multinomial by

$$E(n_{ij}) = n_{i.}\theta_{ij} \quad \text{...} \quad (3.01)$$

and

$$\text{Var}(n_{ij}) = n_{i.}\theta_{ij}(1-\theta_{ij}) \quad \text{...} \quad (3.02)$$

where $\theta_{ij} = \dfrac{n_{ij}}{n_{i.}}$ (by MLE).

(iv)Compute correlation coefficient between the cell counts. That is,

Let ρ_{ij} be the correlation between n_i and n_j in a product multinomial. Then

$$\rho_{ij} = -\sqrt{\frac{\theta_i\theta_j}{(1-\theta_i)(1-\theta_j)}} \quad \text{...} \quad (3.03)$$

(v) Generate a r x c contingency table to satisfy equations (3.01) to (3.03)

This procedure is an adapted Expectation Maximization (EM) algorithm for assigning observation into components of a mixture and k denotes the number of components in a mixture.

Since the posterior distribution of the product multinomial is hypergeometric, then this assumes fixed marginals. In this regard, it can be shown that

$$E(n_{ij}) = \frac{n_{i.}n_{.j}}{n..} \quad \text{...} \quad (3.04)$$

$$\text{Var}(n_{ij}) = \frac{n_{i.}n_{.j}}{n..}\left(\frac{(n_{i.}-1)(n_{.j}-1)}{(n_{..}-1)} + 1 - \frac{n_{i.}n_{.j}}{n_{..}}\right) \quad \text{.........................} \quad (3.05)$$

A r x c contingency table is generated using the formulas (3.04) and (3.05) as follows

$$n_{ij} = \frac{n_{i.}n_{.j}}{n..} + W\sqrt{\left(\frac{n_{i.}n_{.j}}{n_{..}}\left(\frac{(n_{i.}-1)(n_{.j}-1)}{(n_{..}-1)} + 1 - \frac{n_{i.}n_{.j}}{n_{..}}\right)\right)}$$

where W is sampled from the standard normal variate.

This procedure is an adopted (EM) algorithm for assigning observation into component of the mixture.
To determine whether the conceived r-component mixture obtained through product multinomial and hypergeometric are compatible; a simulation of size 1000 was carried out for the two distributions and the generated two tables were compared using Pearson Chi-squared statistic.
The empirical level of significance α , attained by the statistic was computed as the proportion of the time the value of the test statistic exceeded the critical value $\chi_{\alpha,(r-1)(c-1)}$ for nominal value of α =0.05, where (r-1)(c-1) degrees of freedom.
To determine whether the attained α was reasonably close to the normal value α we adopted Cochran's(1952) suggestion that the attained level should below 60% at 5% level.

4.0 Data Analysis and Result.

We considered the results of the simulation study when product multinomial and hypergeometric distributions were assumed for the observed and expected cell counts, respectively. The component parameters were fixed and arranged in symmetric or asymmetric in some cases. For example Table4.1, the component parameters were fixed at π_{11}=0.9, π_{21}=0.1 for the first component and π_{21}=0.1 , π_{22}=0.9 for the second component. The difference between the adjacent component parameter is 0.8 while the ratio of the loading was in 1:1. For each of the sets the loadings were varied between 1:1 and 1:4. A simulation of 1000 was carried out to validate the scheme.

It is observed that the error of cross-classifying objects into components of a mixture using hypergeometric sampling distribution increased geometrically as a function of the sample size. For instance, in Table 4.1, the error rate of 0.003 was obtained for cross-classifying 20 objects into their components while an error of 0.029 was committed for classifying 60 objects even when the loading remaining as 1:1. As earlier mentioned Cochran (1952) was used to determine unacceptable cross-classification table. The error rate became unpredictable beyond classification of 240 objects. In situation where error rate is not within the Cochran bound it is postulated that the sampling distribution may not provide a good fit.

A close examination of Table4.2 depicts that each of the sets shows that the loading increases from 1:1 and 1:4, the error rate decreases and therefore the performance of the scheme for classification improves. In other words, as the component proportions are well distinguishable the scheme performs creditably.

The performance of the scheme was also tested in higher dimension tables and the results were similar to what obtained under 2x2 contingency table. The example of this is given in Tabl4.3

A sample result of the simulation are Table4.4 and Table 4.5, for this, we assumed that the component (structural) parameters for two–component mixture are π_{11}=0.9and π_{12}=01 for the first component and π_{21}=0.2 and π_{22}=0.8 for the second. The component loading are in the ratio 1:2(20:40). We generated the two tables from product binomial (i.e observed counts) and hypergeometric (i.e. expected counts) as explained in section 3 above.

A comparison of the two tables using X^2 and G^2 statistics gave values 3.27 and 3.69 respectively and the P-value exceed 0.1.Therefore the expected counts (Table4.5) compared favourably well with the observed counts presented in Table4.4.

The proposed model was illustrated using real life data collected from Ilorin and Yola Prison Services,Nigeria on age and offences by prison inmates between the period 2000 and 2004 (Table4.6 and Table 4.7). The assumption here is that the crime/offence pattern in Nigeria is identical. Thus, the loadings would be similar between Ilorin and Yola. However, assuming Ilorin offence pattern is sustained the proposed sampling scheme was used to predict the crime/offence by a distribution for Yola, especially when the age group distribution is assumed to be the only distribution available.

The predicted crime/offence by group is found in Table4.8(X^2=12.34 and P-value =0.250). The result shows that the tables are compatible; hence the new approach is good for classification.

SUMMARY AND CONCUSION

In Bayesian approach to mixture models, Gibbs sampler, the most common MCMC,is used for estimation of posterior density for conjugate priors, and subsequent classification of objects into components of mixture. In practice conjugacy may not exist and when it does the time required calculating posterior density will be too high for Bayesian approach to be applied in practice Therefore, we developed a clustering procedure that is a result of using non-conjugate prior distribution of product multinomial to obtain posterior distribution that is hypergeometric, for classifying categorical data.

We examine the accuracy of the approach through simulation study of observed tables of counts compared with expected generated by assuming product multinomial under a variety of parameters and loadings.

We observed that the approach performed well when the component proportions are properly distinguishable. It was also found that higher number of objects to classify increases the possible errors committed.

The performance of the scheme in higher dimension table is similar to what obtained under 2x2 contingency table. The real life data from social science used, shown that the approach fitted the data, showing that the distribution found in one environment was similar to another.

Table4.1: The distribution of sample size and error rate in a simulation of two-component mixture when the adjacent component parameters differed by 0.8 for loadings in ratio one to one.

Sample Size	Error Rate	Sample Size	Erro Rate	Sample Size	Error Rate
20	0.003	150	0.040	280	0.074
30	0.004	160	0.041	290	0.082
40	0.010	170	0.041	300	0.091
50	0.024	180	0.041	310	0.103
60	0.029	190	0.042	320	0.116
70	0.033	200	0.042	330	0.124
80	0.035	210	0.046	340	0.132
90	0.037	220	0.050	350	0.135
100	0.039	230	0.056	360	0.146
110	0.039	240	0.059	370	0.152
120	0.039	250	0.061	380	0.161
130	0.040	260	0.046	390	0.168
140	0.040	270	0.068		

Component parameters for 2x2: $\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$

Table4.2: The summary showing the distribution of error rate, and ratio of loadings for fixed sample sizes when the component parameters differed by 0.8.

	Ratios of loadings			
Sample Size	1:1	1:2	1:3	1:4
60	0.029	0.021	0.010	0.006
90	0.037	0.032	0.021	0.011
120	0.039	0.034	0.025	0.017
150	0.040	0.035	0.026	0.019
180	0.041	0.037	0.028	0.019
200	0.042	0.039	0.030	0.020
240	0.059	0.040	0.030	0.025
270	0.068	0.043	0.031	0.026
300	0.091	0.048	0.033	0.029
330	0.103	0.049	0.033	0.029
360	0.145	0.051	0.039	0.032
400	0.171	0.073	0.040	0.034
420	0.179	0.129	0.040	0.035
480	0.317	0.206	0.042	0.038
510	0.253	0.241	0.043	0.040
570	0.304	0.293	0.046	0.040
600	0.411	0.364	0.048	0.041

Component parameters for 2x2: $\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.10, 2014

Table4.3: The distribution of sample size and error rate in a simulation of three – component model for loadings in ratio one to three.

Sample Size	Error Rate	Sample Size	Error Rate
60	0.011	600	0.121
90	0.023	660	0.142
120	0.034	720	0.243
180	0.042	780	0.350
240	0.053	840	0.390
300	0.058	900	0.456
360	0.056	960	0.514
420	0.054	1020	0.574
480	0.063	1080	0.690
540	0.084	1140	0.761

Component parameters for 3x3: $\begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$

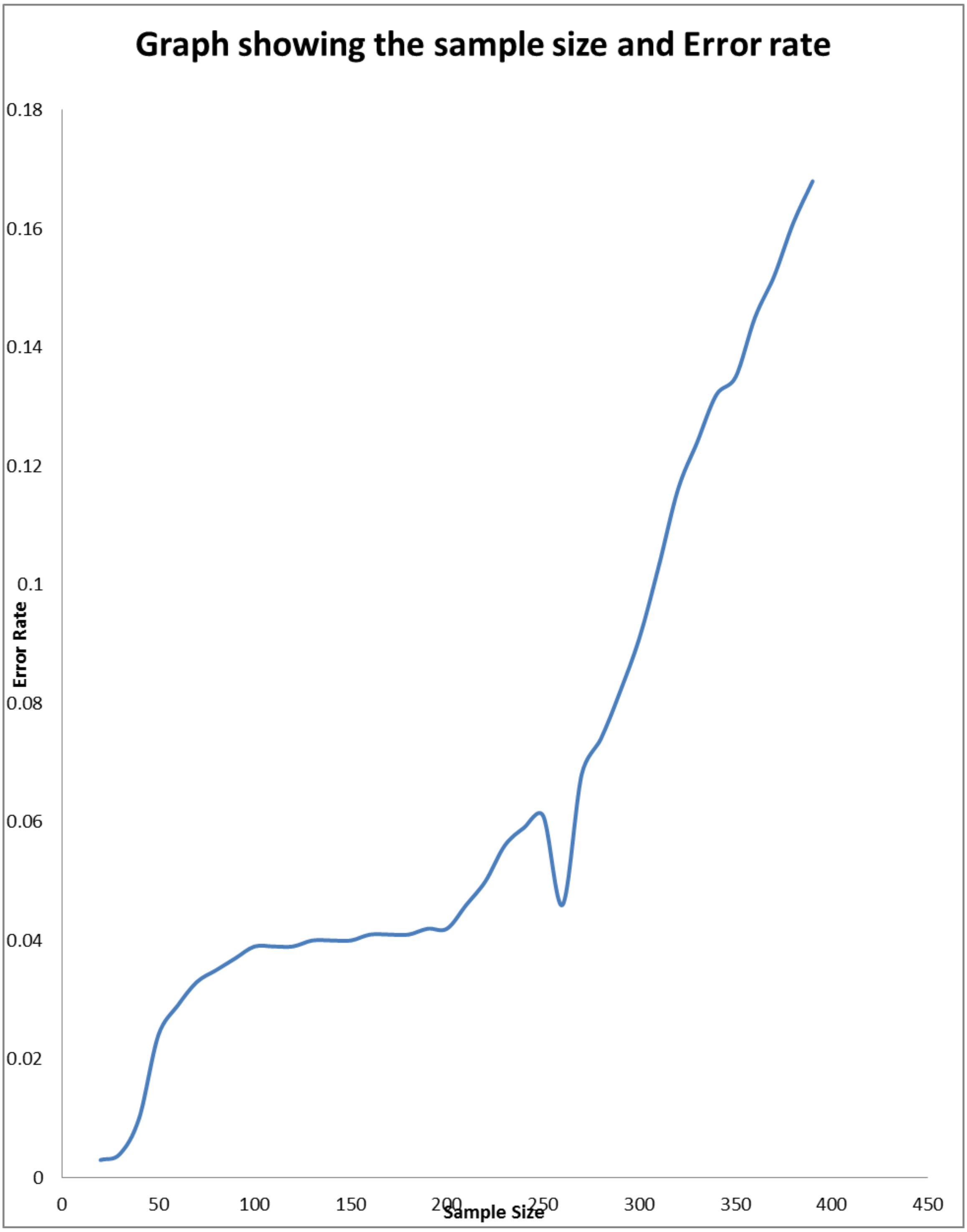

Figure 1: Graph showing the distribution of sample size and error rate in a simulation of two-component mixture when the adjacent component parameters differed by 0.8 for loadings in ratio one to one.

Table 4.4: A sample Configuration from Simulation of Two-component Mixture when Product binomial was assumed for the Count.

Row \ Column	**1**	**2**	**Loadings** α_i	**Total**
1	16	4	$\frac{1}{3}$	20
2	8	32	$\frac{2}{3}$	40
Total	24	36		60

Table4.5; A Sample Configuration from Simulation of Two-Component Mixture when hypergeometric was assumed for the Counts.

Row \ Column	**1**	**2**	**Loadings** α_i	**Total**
1	18	2	$\frac{1}{3}$	20
2	6	34	$\frac{2}{3}$	40
Total	24	36		60

Table 4.6 Ilorin prison inmates Data for a period of 2000 – 2004

Age Group / Offence	≤ 25	26-30	31-35	36-40	40 and Above	Total	Loading α_i
Armed Robbery	76	15	14	10	6	121	0.12
Theft	32	14	2	3	181	232	0.23
Culpable Homicide	31	6	2	4	28	71	0.07
Indian Hemp	21	12	1	3	84	121	0.12
Assult	1	4	1	1	43	50	0.05
Others	171	63	67	35	76	412	0.41
Total	332	114	87	56	418	1007	1

Table 4.7: Yola prison inmates Data for a period of 2000 – 2004

Age Group / Offence	≤ 25	26-30	31-35	36-40	40& Above	Total
ArmedRobbery	56	38	15	7	9	125
Theft	145	53	23	4	26	251
Culpable Homicide	17	17	13	10	18	75
Indian Hemp	68	36	13	8	5	130
Assult	25	11	5	2	5	48
Others	198	95	46	23	83	445
Total	509	250	115	54	146	1074

Table 4.8:Predicted Yola prison inmates Data for a period of 2000 – 2004

Age Group / **Offence**	≤ 25	**26-30**	**31-35**	**36-40**	**40 and Above**	**Total**
Armed Robbery	60	32	16	9	8	125
Theft	139	55	22	5	30	251
Culpable Homicide	23	19	8	9	16	75
Indian Hemp	64	33	15	10	7	130
Assault	22	12	7	1	6	48
Others	201	99	47	20	79	445
Total	509	250	115	54	146	1074

X^2=12.34 P –value>0.250

REFERENCES

Agresti A. (1990) Categorical Data Analysis, New York Wiley.
Bernnet,J.E;Racine-Poon,A,and Wakefield,J.C(1996).MCMC for nonlinear hierarchical models.In Markov Chain Monte Carlo in practice(eds W.R.Gilk,S.Richardson and D.J. Sringelhalter)London Chapman and Hall.
Birch, M.W. (1963) “Maximum likelihood in three- way Contingency tables. Journal Royal Statist Soc.Sci B25 220 – 233
Cheng R.C. H and Currie, C.S. M (2003), Prior and candidate models in the Bayesian Analysis of finite mixture Proceeding of the 2003 inter simulation conference S. Chick, P.S .Sachez, D Ferrin and D.J Morrice.
Chib,S. and Greenberg,E(1995) Undersatndin the Metropolis-Hastings Algorithm.Am. Statistn,49
Cochran,A.C,(1952),The X^2 Test of Goodness - of-fit. Ann.Maths.statist.23,315-345.
Damien,P;Wakefield,J;and Walker,S(1999) Giggs Sampling for Bayesian Non-Conjugate and Hierarchical Model by Auxiliary Variables. J.Royal Statist. Soc. B 61.
Dempster A,P, Laird, N.M and Rubin, D.B (1977),Maximum likelihood Estimation from Incomplete Data via the EM algorithm (with discussion)J Roy Statist, Soci ser B,39,1 – 38.
Devroye,L(1986) Non-uniform Random Variate Generation.New York.Sringer.
Diaconis, P and YIvisaker, D (1985) Conjugate Priors for Exponential Families. American J. of Statist. 7,269-281

Everitt, B.S. and Hand, D.J (1981) .Finite Mixture Distributions. Chapman and Hall London.

Gelfand,A.E. and Smith,A.F.M(1990) Sampling based approaches to calculating Marginal densities.J.American statist. Assoc.

Gilks,W.R; and Wild,P(1992) Adaptive Rejection sampling for Gibbs Samplings. Appl. Statist.44

Jolayemi E.T. (1982) A CP method to select A Log linear model. A Ph.D Thesis submitted to the University of Michigan

McLachlan, G.J; and Peel D; (2000), Finite Mixture Models. Wiley,New York.

Sammi O.O. M and Jolayemi E. T (1998), Robustness of some categorical test Statistics in Small sample Situations. Journal of Nigerian Statistics. Tertiary Publications Limited.,29-38. Statistics in Small sample situations. Journal of Nigeria Statistic. Tertiory Publications limited, 29 – 38.

Tanner,M.A, and Wong,W.H(1987) The Calculation of Posterior Distributions by data Augmentation(with discussion) J.Am.Statist. Assoc.82

Wakefield,J.C;Gelfand,A.E;and Smith,A.F.M(1991)Efficiency Generation of Random Variates via ratio-Uniform Method.Statist. Comput.,1.

Woodward, W, A; Parr, W.C; Schucany, W, R; and Lindsay, H. (1984) A comparison of Minimum Distance and Maximum Likelihood Estimation of a mixture Proportion. Journal of the American statistical Assocation, 79,590- 598.