Extension of Comparative Analysis of Estimation Methods for Dirichlet Distribution Parameters

M.A. Halid and A.A. Akomolafe*

Department of Statistics, Federal University of Technology, Akure, Nigeria.

E-mail: akomolafeayotade@gmail.com*

ABSTRACT

The Dirichlet distribution is a generalization of the Beta distribution. This research deals with the estimation of scale parameter for Dirichlet distribution with known shapes. We examined three methods to estimate the parameters of Dirichlet distribution which are Maximum Likelihood Estimator, Method of Moment Estimator, and Quasi-Likelihood Estimator. The performance of these methods were compared at different sample sizes using Bias, Mean Square Error, Mean Absolute Error and Variance criteria, an extensive simulation study was carried out on the basis of the selected criterion using statistical software packages as well as the application of the criterion to real life data, all these were done to obtain the most efficient method.

The simulation study and analysis revealed that Quasi-Likelihood Estimator performed better in terms of bias while Method of Moment Estimator is better than the other two methods in terms of variance; the Maximum Likelihood Estimation was the best estimation method in terms of the Mean square Error and Mean Absolute Error; while Quasi- Likelihood Estimation method was the best estimation method with real life data.

(Keywords: Dirichlet distribution, parameter estimation, maximum likelihood estimator, method of moment estimator and quasi-likelihood estimator).

INTRODUCTION

In Bayesian statistics, the Dirichlet distribution is a popular conjugate prior for multinomial distribution. The Dirichlet distribution has a number of applications in various fields. Samuel S. Wilk (1962), gave an example, where he applied the Dirichlet distribution in deriving the distribution of order statistics.

Kenneth Lange (1995), also used the Dirichlet distribution in biology to demonstrate and to compute forensic match probabilities from several distinct populations. In addition, Brad N (2009), used the Dirichlet distribution to model a player`s abilities in Major League Baseball. It is shown that the Dirichlet distribution can be used to model consumer behavior Gerald et al (1984).

Dirichlet distribution can be extended to various fields of study such as biology, astronomy, text mining and so on. The Dirichlet distribution (DD) is usually employed as a conjugate prior for the multinomial modeling and Bayesian analysis of complete contingency tables (Agresti, 2002). Gupta and Richards (1987, 1991, and 1992) extended the DD to the Liouville distribution. Fang, Kotz, and Ng (1990) gave an extensive exposition of the Liouville family and its ramifications.

The problem of estimating the parameters which determine a mixture has been the subject of diverse studies (Redner and Walker, 1984). During the last two decades, the method of maximum likelihood (ML) (Bishop, 1995) and (Rao, 1987) has become the most common approach to this problem. Of the variety of the iterative methods which has been subjected as an alternative to optimize the parameters of a mixture, the one most likely used is the expectation maximization (EM). EM was originally proposed by Dempster, et al. (1977) for estimating the maximum likelihood estimator (MLE) of stochastic models. This algorithm gives an iterative procedure, and the practical form is usually simple and easy to implement .The EM algorithm can be viewed as an approximation of the Fisher scoring method (Ikeda, 2000).

In this research we showed that the Dirichlet distribution can be a very good choice for modelling data, MLE was used to estimate the parameters of the Dirichlet Mixture Model

alongside with EM algorithm. This mixture decomposition algorithm incorporates a penalty term in the objective function to find the number of components required to model the data. This algorithm suffers some set back: the need to specify the number of components each time, which will be determine by selected criterion functions such as AIC, BIC, MDL which has been in existence to validate the model and justify the more efficient one.

This research centered on studying how the different estimators of the unknown parameters of a Dirichlet distribution can behave for different sample sizes. Here, we are mainly comparing the Maximum Likelihood Estimator, Method of Moment Estimator and Quasi-Likelihood Estimator with respect to efficiency, bias, mean absolute error and variance using extensive simulation techniques as well as application of the estimation methods to real life data set.

LITERATURE REVIEW

The Dirichlet model describes patterns of repeat purchases of brands within a product category. It models simultaneously the counts of the number of purchases of each brand over a period of time, so that it describes purchase frequency and brand choice at the same time. It assumes that consumers have an experience of the product category, so that they are not influenced by previous purchase and marketing strategies; for this reason, consumer characteristics and marketing-mix instruments are not included in the model. As the market is assumed to be stationary, these effects are already incorporated in each brand market share which influences other brand performance indexes calculated by the model.

The market is also assumed to be unsegmented. The theory and development of the model is fully described in Ehrenberg (1972). Goodhardt, Ehrenberg and Chatfield (1984), summarize the situation by stating that the Dirichlet model makes explicit that there are simple, general and rather precise regularities in a substantial area of human behavior where this has not always been expected. In setting the context for this particular approach to the modeling of consumer behavior viz. the largely explanatory models of consumer behavior, Ehrenberg (1988) claims that it describes how consumers behave, rather than why, and takes into account only those factors necessary for an adequate description.

Many aspects of buyer behaviors can be predicted simply from the penetration and the average purchase frequency of the item, and even these two variables are interrelated (Ehrenberg, 1988, pg. ii). The Dirichlet model integrates the reported regularities and predicts many aggregate brand performance measures. These measures are the distribution of purchases for a brand, the proportion of a brand's buyers buying that brand only, and the proportion of people purchasing a brand, given that they have previously purchased that brand. When these predictions are compared with observed figures, Ehrenberg claims that it is not unreasonable to expect to obtain correlations in the order of 0.9 and sometimes much higher, (Ehrenberg, 1975, Ehrenberg and Bound, 1993).

Applications and theory can be used to provide norms for examining brand performance, or diagnostic information for the "health" of a brand. In addition, the Dirichlet model can provide interpretative norms for evaluating situations where some trend in sales has occurred, say after a promotion or advertising scheme. Ehrenberg also claims that the Dirichlet model provides valuable insights into the nature and implications of brand-loyalty (e.g., Ehrenberg and Uncles, 1995; Ehrenberg and Uncles, 1999).

The use of likelihood theory to estimate the parameters of the Dirichlet model, providing an alternative to the standard procedure based on the method of zeros and ones and on marginal moments (Rungie, 2003b). In order to write the likelihood function, the data should be in the form of joint frequencies, like those contained in a contingency table with n-rows, representing the number of consumers, and g columns, for the number of brands.

Alternatively, the iterative procedures based on the approach that computations are easy to use, and require only aggregated data as input, as access to original panel data is not necessary as proposed by Goodhardt, Ehrenberg, and Chatfield (1984). Raw panel data cannot always be used since panel operators who measure sales and household consumption provide information only in some aggregate format such as market share, penetration, and average purchase rate with reference to the various brands (Wright, et al., 2002). In these situations, the only way to estimate the Dirichlet model is to use the traditional method.

Dirichlet modeling continues to be a successful and influential approach and is increasingly being used to provide norms against which brand performance can be interpreted (Uncles, et al., 1995; Bhattacharya, 1997; Ehrenberg, et al., 2000). Dirichlet model is useful for the provision of norms for stationary markets, to supply baselines for interpreting change (i.e., non-stationary situations) without having to match the results against a control sample, to help strategic decision-making, and to understand the nature of markets.

There are diverse ways of applying the distribution, where the Dirichlet has proved to be particularly useful is in modeling the distribution of words in text documents [9]. If we have a dictionary containing k possible words, then a particular document can be represented by a probability mass function [pmf] of length k-produced by normalizing the empirical frequency of its words. A group of documents produces a collection of pmfs, and we can fit a Dirichlet distribution to capture the variability of these pmfs.

METHODOLOGY

Deriving the Dirichlet Distribution

Let X_i be a random variable from the Gamma distribution $G(\alpha_i, 1), i = 1, \dots, k$, and let $X_1, \dots, X_k$ be independent. The joint pdf of $X_1, \dots, X_k$ is:

$$f(x_1, \dots, x_k) = \begin{cases} \prod_{i=1}^{k} \frac{1}{\Gamma(\alpha_i)} x_i^{\alpha_i - 1} e^{-x_i}, & \text{if } 0 < x_i < \infty \\ \\ 0, & \text{otherwise} \end{cases}$$

Let,

$$Y_i = \frac{X_i}{X_1 + X_2 + \dots + X_k}, \quad i = 1, 2, \dots, k-1$$

and

$$Z_k = X_1 + X_2 + \dots + X_k .$$

By using the change of variables technique, this transformation maps $M = \{(x_i, \dots, x_k): 0 < x_i < \infty, i = 1, \dots, k\}$ onto $N = \{(y_i, \dots, y_{k-1}, z_k): y_i > 0, i = 1, \dots, k-1, 0 < z_k < \infty, y_1 + \dots + y_{k-1} < 1\}$. The inverse functions are $x_1 = y_1 z_k, x_2 = y_2 z_k, \dots, x_{k-1} = y_{k-1} z_k, x_k = z_k(1 - y_1 - \dots - y_{k-1})$. Hence, the Jacobian is:

$$J = \begin{vmatrix} z_k & 0 & \dots & & 0 & y_1 \\ 0 & z_k & \dots & & 0 & y_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & Z_k & & y_{k-1} \\ -z_k & -z_k & \dots & -z_k & (1 - y_1) - \dots - y_{k-1} \end{vmatrix} = z_k^{k-1}$$

Then, the joint pdf of $Y_1, \dots, Y_{k-1}, Z_k$ is:

$$f(y_1, \ldots, y_{k-1}, z_k) = \frac{y_1^{\alpha_1-1} \cdots y_{k-1}^{\alpha_{k-1}-1} (1-y_1-\cdots-y_{k-1})^{\alpha_k-1}}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_k)} e^{-z_k} z_k^{\alpha_1+\cdots+\alpha_{k-1}}$$

By integrating out z_k, the joint pdf of $Y_1, \ldots, Y_{k-1}$ is:

$$f(y_1, \ldots, y_{k-1}) = \frac{\alpha_1 + \cdots + \alpha_k}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_k)} y_1^{\alpha_1-1} \cdots y_{k-1}^{\alpha_{k-1}-1} (1-y_1-\cdots-y_{k-1})^{\alpha_k-1} ,$$

where $y_i > 0, y_1 + \cdots + y_{k-1} < 1, i = 1, \ldots, k-1$. The joint pdf of the random variables $Y_1, \ldots, Y_{k-1}$ is known as the pdf of the Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_k$. Furthermore, it is clear that Z_k has a Gamma distribution $G(\sum_{i=1}^{k} \alpha_i , 1)$ and Z_k is independent of $Y_1, \ldots, Y_{k-1}$. (Hogg and Craig, 1970).

Moment Generating Function

The moment generating function of $Y^k = [Y_1, \ldots, Y_k]$. Let $t = (t_1, \ldots, t_k)^T \in \mathfrak{R}^k$.

The moment generating function of Y^k at t is $\quad E\left(e^{t^T Y^k}\right) = \int \ldots \int e^{t^T y} f(y^k) dy_1 \ldots dy_k$

$$= \int \ldots \int \sum_{m=0}^{\infty} \frac{(t^T y^k)^m}{m!} f(y^k) dy_1 \ldots dy_k \tag{1}$$

$$= \sum_{m=0}^{\infty} \frac{1}{m!} \int \ldots \int (t^T y^k)^m f(y^k) dy_1 \ldots dy_k \tag{2}$$

     Step (a)

$$= \sum_{m=0}^{\infty} \frac{1}{m!} \left[\int \ldots \int \sum_{n_1+n_2+\cdots+n_k=m} \frac{m!}{n_1! n_2! \ldots n_k!} \times \prod_{i=1}^{k} (t_i y_i)^{n_i} f(y^k) dy_1 \ldots dy_k \right]$$

$$= \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{m=0}^{\infty} \frac{1}{m!} \left[\sum_{n_1+n_2+\cdots+n_k=m} \frac{m!}{n_1! n_2! \ldots n_k!} \prod_{i=1}^{k} (t_i)^{n_i} E\left(\prod_{i=1}^{k} Y_i^{n_i} \right) \right]$$

$$= \sum_{m=0}^{\infty} \frac{1}{m!} \left[\sum_{n_1+n_2+\cdots+n_k=m} \frac{m!}{n_1! n_2! \ldots n_k!} \prod_{i=1}^{k} (t_i)^{n_i} \times \left[\frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1+n_1+\cdots\alpha_k+n_k)} \prod_{i=1}^{k} \frac{\Gamma(\alpha_i+n_i)}{\Gamma(\alpha_i)} \right] \right] \tag{3}$$

In step (a), we apply the multinomial theorem:

$$(x_1 + x_2 + \cdots + x_k)^m = \sum_{n_1+n_2+\cdots+n_k=m} \frac{m!}{n_1! n_2! \ldots n_k!} \prod_{i=1}^{k} x_i^{n_i} \tag{4}$$

for any positive integer k and any non-negative integer m.

Maximum Likelihood Estimation

The ML estimation method concerns choosing parameters to maximize the joint density function of the sample (likelihood function). Therefore, we consider:

$$\max_{\Theta} p(x^k|\Theta) \tag{5}$$

with constraints $\sum_{j=1}^m p(j) = 1$ and $p(j) > 0$ for $j = 1,2,\dots,m$. We can consider $p(j)$ as prior probabilities under these constraints. Now suppose we have a sample that contains n random vectors X_i^k, which are i.i.d., $i = 1,\dots,n$. We maximize the following function with respect to Θ and Λ:

$$\phi(x^k,\Theta,\Lambda) = \sum_{i=1}^n \ln\left(\sum_{j=1}^m p(x^k|\Theta_j)p(j)\right) + \Lambda\left(1 - \sum_{j=1}^m p(j)\right) + \mu\left(\sum_{j=1}^m p(j)\ln(p(j))\right) \tag{6}$$

The first term of equation 8 is the log-likelihood function. Λ is the Lagrange multiplier in the second term. In the last term of eq. 8, we use an entropy-based criterion. Also, μ is the ratio of the first term to the last term in of each iteration t by Nizar Bouguila, Djemel Ziou, and Jean Vaillancourt (2004).

$$\mu(t) = \frac{\sum_{j=1}^m \ln\left(\sum_{j=1}^m p^{t-1}\left(x_i^k|\Theta_j\right)p^{t-1}(j)\right)}{\left(\sum_{j=1}^m p^{t-1}(j)\ln\left(p^{t-1}(j)\right)\right)}, \tag{7}$$

In order to optimize (8), we need to solve the following equations:

$$\frac{\partial}{\partial\Theta}\phi(x^k,\Theta,\Lambda) = 0$$

$$\frac{\partial}{\partial\Lambda}\phi(x^k,\Theta,\Lambda) = 0$$

It is shown that the estimator of the prior probability $p(j)$ is:

$$p(j)^{new} = \frac{\sum_{i=1}^n p^{old}\left(j|x^k,\Theta_j\right) + \mu\left[p(j)^{old}\left(1 + \ln p(j)^{old}\right)\right]}{n + \mu\sum_{j=1}^m p(j)^{old}\left(1 + \ln p(j)^{old}\right)}, \quad j = 1,2,\dots,m. \tag{8}$$

Note that μ is defined by (4.3) and $p(j|x_i^k,\Theta_j)$ is the posterior probability where:

$$p(j|x_i^k,\Theta_j) = \frac{p(x_i^k,\Theta_j)p(j)}{p(x_i^k,\Theta)}, i = 1,\dots,n, \; j = 1,2,\dots,m. \tag{9}$$

Now we want to estimate the parameters α_j^k, $j = 1,2,\dots,m$. The Fisher scoring method is used to find these estimates. Denote α_{jl} as one element of the parameter vector α_j^k for each component $j = 1,2,\dots,m$. The derivative of $\phi(x^k,\Theta,\Lambda)$ with respect to α_{jl} is:

$$\frac{\partial}{\partial\alpha_{jl}}\phi(x^k,\Theta,\Lambda) = \sum_{i=1}^n p(j|x_i^k,\alpha_j^k)(\ln x_{il}) + \left[\psi\left(\alpha_{0j}\right) - \psi(\alpha_{jl})0\right]\sum_{i=1}^n p(j|x_i^k,\alpha_j^k), \tag{10}$$
$$l = 1,\dots,k, \; j = 1,\dots,m,$$

where $\psi(.)$ is the Digamma function. However, α_{jl} can become negative during iterations. In order to keep α_{jl} strictly positive, set $\alpha_{jl} = e^{\beta jl}$. βjl is any real number. Then, the derivative of $\phi(x^k, \Theta, \Lambda)$ with respect to βjl is:

$$\frac{\partial}{\partial \beta jl}\phi(x^k, \Theta, \Lambda) = \alpha_{jl}[\sum_{i=1}^{n} p(j|x_i^k, \alpha_j^k)(\ln x_{il}) + [\psi(\alpha_{0j}) - \psi(\alpha_{jl})]\sum_{i=1}^{n} p(j|x_i^k, \alpha_j^k)], \quad \textbf{(11)}$$
$$l = 1, \dots, k, \quad j = 1, \dots, m.$$

By using the iterative scheme of the Fisher scoring method, we obtain:

$$\begin{pmatrix} \hat{\beta}jl \\ \vdots \\ \hat{\beta}jk \end{pmatrix}^{new} = \begin{pmatrix} \hat{\beta}jl \\ \vdots \\ \hat{\beta}jk \end{pmatrix}^{old} + \begin{pmatrix} VAR(\hat{\beta}jl) & \dots & COV(\hat{\beta}jl,\hat{\beta}jk) \\ \vdots & \ddots & \vdots \\ COV(\hat{\beta}jk,\hat{\beta}jl) & \dots & VAR(\hat{\beta}jk) \end{pmatrix}^{old} \times \begin{pmatrix} \frac{\partial}{\partial \hat{\beta}_{jk}}\phi(x^k,\Theta,\Lambda) \\ \vdots \\ \frac{\partial}{\partial \hat{\beta}_{jl}}\phi(x^k,\Theta,\Lambda) \end{pmatrix}^{old}$$
$$\textbf{(12)}$$

$$j = 1, \dots, m.$$

Note that the variance-covariance matrix is obtained by the inverse of the Fisher information matrix I and:

$$\text{I} = -E\left[\frac{\partial^2}{\partial \beta_{jl_1}\partial \beta_{jl_2}}\phi(x^k, \Theta, \Lambda)\right]. \qquad \textbf{(13)}$$

ANALYSIS AND RESULTS

In this chapter, the results of the simulation study on the basis of the entire criterion at different sample sizes are presented and Performance of parameter estimation method in terms of Bias as the sample size and parameter dimension varies was discussed.

Table 1: Results of the Bias at Different Alpha Level as Sample Size Varies.

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.15$	10	0.2328	0.3125	0.0994
	20	0.1013	0.1848	0.1
	30	0.067	0.151	0.1
$\alpha\_2 = 0.30$	40	0.0524	0.1371	0.0999
	50	0.0401	0.1247	0.1
	75	0.0286	0.1136	0.1
$\alpha\_3 = 0.45$	100	0.0219	0.1074	0.1
	250	0.01	0.0964	0.1001

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.45$	10	0.4683	0.4771	1.1
	20	0.2223	0.2324	1.1
	30	0.1521	0.1638	1.1
$\alpha\_2 = 0.75$	40	0.1072	0.1195	1.1
	50	0.0853	0.0977	1.1
	75	0.0615	0.0743	1.1
$\alpha\_3 = 0.90$	100	0.0442	0.0567	1.1
	250	0.0235	0.0362	1.0999

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.90$	10	0.6564	0.6576	1.889
	20	0.3141	0.3152	1.889
	30	0.185	0.1862	1.888
$\alpha\_2 = 0.99$	40	0.1462	0.1476	1.8891
	50	0.1264	0.1277	1.8891
	75	0.0862	0.0876	1.8891
$\alpha\_3 = 0.999$	100	0.0638	0.0652	1.889
	250	0.0324	0.0339	1.889

Based on the above table, it was observed that the level of alpha the Quasi-likelihood estimator (QLE) has the least value of Bias as the sample size increases and performs better compared with the Method of Moment and Maximum likelihood estimator.

Table 2: Results of the Variance at Different Alpha Level as Sample Size Varies.

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.15$	10	0.108	0.1167	0.0325
	20	0.0355	0.0379	0.0158
	30	0.0187	0.02	0.0103
$\alpha\_2 = 0.30$	40	0.0139	0.0148	0.0082
	50	0.0108	0.0116	0.0065
	75	0.0063	0.0067	0.0041
$\alpha\_3 = 0.45$	100	0.0048	0.0051	0.0033
	250	0.0018	0.0019	0.0013

Alpha	n	QLE	MLE	MOM
$\alpha\_1 = 0.45$	10	0.4132	0.4128	0.0207
	20	0.144	0.1444	0.0101
	30	0.0863	0.0858	0.0071
$\alpha\_2 = 0.75$	40	0.0565	0.0563	0.0051
	50	0.043	0.0426	0.004
	75	0.0279	0.0277	0.0028
$\alpha\_3 = 0.90$	100	0.0201	0.0199	0.0021
	250	0.0076	0.0075	0.0008

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.90$	10	0.6887	0.6885	0.0173
	20	0.2398	0.2396	0.0086
	30	0.1286	0.1285	0.0056
$\alpha\_2 = 0.99$	40	0.0926	0.0925	0.0043
	50	0.0772	0.0772	0.0034
	75	0.05	0.0498	0.0023
$\alpha\_3 = 0.999$	100	0.0353	0.0353	0.0018
	250	0.0134	0.0134	0.0006

Based on the above table and Appendix 2, it was observed that the performance of parameter estimation method in terms of Variance, as the sample size increases Method of Moment performs better compared with the Quasi-likelihood estimator and Maximum likelihood estimator.

Table 3: Results of the Mean Absolute Error (MAE) at Different Alpha Level as Sample Size Varies.

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.15$	10	0.3652	0.3982	0.2668
	20	0.2171	0.2484	0.1915
	30	0.1671	0.2003	0.1633
$\alpha\_2 = 0.30$	40	0.1438	0.1792	0.1512
	50	0.1262	0.1626	0.1396
	75	0.0995	0.1396	0.1233
$\alpha\_3 = 0.45$	100	0.0873	0.1292	0.1181
	250	0.052	0.104	0.1043

Alpha	n	QLE	MLE	MOM
$\alpha\_1 = 0.90$	10	1.0665	1.0665	1.8917
	20	0.6602	0.6601	1.889
	30	0.4866	0.4865	1.889
$\alpha\_2 = 0.99$	40	0.4212	0.4211	1.8891
	50	0.3827	0.3825	1.8891
	75	0.3059	0.3057	1.8891
$\alpha\_3 = 0.999$	100	0.2622	0.2622	1.889
	250	0.1588	0.1588	1.889

Alpha	n	QLE	MLE	MOM
$\alpha\_1 = 0.45$	10	0.7774	0.7778	1.1001
	20	0.486	0.4859	1.1
	30	0.3818	0.3815	1.1
$\alpha\_2 = 0.75$	40	0.3138	0.3147	1.1
	50	0.2748	0.2748	1.1
	75	0.2216	0.2225	1.1
$\alpha\_3 = 0.90$	100	0.186	0.1863	1.1
	250	0.1157	0.1168	1.0999

Performance of parameter estimation method in terms of Mean Absolute Error (MAE) as the sample size increases, Maximum likelihood estimator performs better than Quasi-likelihood estimator and Method of moment as shown in Table 3.

Table 4: Results of the Mean Square Error (MSE) at Different Alpha Level as Sample Size Varies.

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.15$	10	0.1335	0.1569	0.0375
	20	0.0401	0.0503	0.0199
	30	0.0207	0.0281	0.0144
$\alpha\_2 = 0.30$	40	0.0152	0.0214	0.0122
	50	0.0115	0.017	0.0105
	75	0.0066	0.0112	0.0081
$\alpha\_3 = 0.45$	100	0.0051	0.0091	0.0071
	250	0.0018	0.0051	0.0052

Alpha	N	QLE	MLE	MOM
$\alpha\_1 = 0.45$	10	0.4945	0.4969	0.4532
	20	0.1624	0.1633	0.4421
	30	0.095	0.0957	0.439
$\alpha\_2 = 0.75$	40	0.0608	0.0615	0.4371
	50	0.0457	0.0461	0.4363
	75	0.0293	0.0298	0.435
$\alpha\_3 = 0.90$	100	0.0207	0.021	0.445
	250	0.0078	0.008	0.4328

| Alpha	n	QLE	MLE	MOM
$\alpha\_1 = 0.90$	10	0.8341	0.8342	1.2088
	20	0.273	0.273	1.2005
	30	0.1401	0.1401	1.1976
$\alpha\_2 = 0.99$	40	0.0999	0.0999	1.1961
	50	0.0827	0.0827	1.1953
	75	0.0525	0.0525	1.1944
$\alpha\_3 = 0.999$	100	0.0367	0.0367	1.1938
	250	0.0137	0.0137	1.1927

Performance of parameter estimation method in terms of Mean Square Error (MSE) as the sample size increases the Quasi- likelihood estimator and Maximum likelihood estimator performs better as compare to Method of Moment. as shown in Table 4.

Conclusively, the best method for each criterion was based on the modal class for the entire criterion as summarized in Table 5.

Table 5: Shows the Count of Quasi-Likelihood Estimator, Maximum Likelihood Estimator and Method of Moment using Bias, Variance, Mean Absolute Error and Mean Square Error.

	Count			Best
Criterion	QLE	MLE	MOM	Method
Bias	77	0	2	QLE
MSE	57	35	11	QLE
MAE	34	48	6	MLE
VAR	4	11	66	MOM

GRAPHICAL RERESENTATION OF THE CRITERION
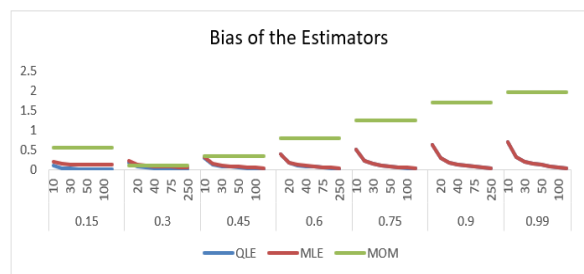
Figure 1: Bias of the Estimator of QLE, MLE, MOM at Different Sample Sizes.
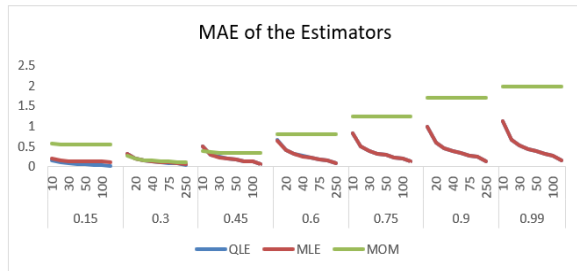
The graph shows the Bias of the estimator of QLE, MLE, MOM at different sample sizes. Judging by the bias criterion, the Quasi-Likelihood method (QLE) was the best for the lower and medium level of alpha, but for the higher level of alpha, Method of Moment performs better.
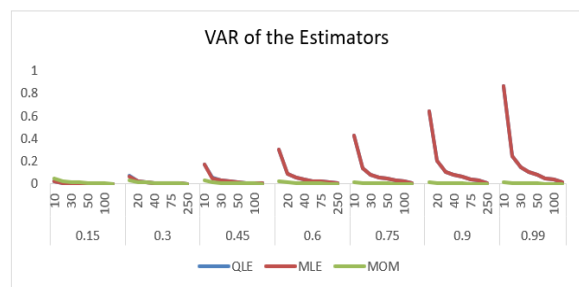
Figure 2: Variance of the estimator of QLE, MLE, MOM at Different Sample Sizes.
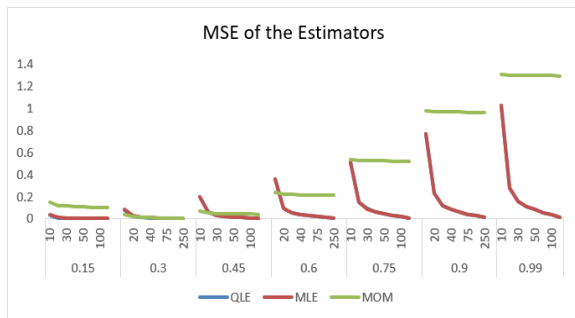
The graph shows the Variance of the estimator of QLE, MLE, MOM at different sample sizes. From the graph, Method of moment consistently performed better across the alpha (parameter) level which implies that the method of Moment is the best method.

Figure 3: Mean Absolute Error (MAE) of the Estimator of QLE, MLE, MOM at Different Sample Sizes.

The graph shows the Mean Absolute Error (MAE) of the estimator of QLE, MLE, MOM at different sample sizes. The Quasi-Likelihood method (QLE) outperformed the other methods for lower level of alpha but as the alpha level increases (medium level and above) the Maximum Likelihood method (MLE) and the QLE has just a slight difference in their estimates. Out of the three methods considered, the method of Moment consistently gives the higher estimate of Mean Absolute Error.

Figure 4: Mean Square Error (MSE) of the Estimator of QLE, MLE, MOM at Different Sample Sizes.

The graph shows the Mean Square Error (MSE) of the estimator of QLE, MLE, MOM at different sample sizes. QLE method was the best for lower level of alpha, but as for the medium and higher level of alpha the QLE and the MLE does not give a significant different estimate. While on the other

hand, the MOM gives a consistently higher estimate of Mean square error.

REAL LIFE DATA RESULTS

Fitting Dirichlet Model to Data Containing Selected Agricultural Products in Nigeria (2008-2017).

Table 6: Parameter Estimation (QLE).

Coefficients	Estimate	Std. Error
α_1	0.2260442	0.1438496
α_2	0.1729296	0.1390305
α_3	0.2410169	0.1453349

From the table above, we obtained the estimates of the mean and standard error of millet, millet rice, and sorghum. The parameter estimate of millet rice is more efficient due to its lowest standard error as compared to others.

Table 7: Parameter Estimation (MLE)

Coefficients	Estimate	Std. Error
α_1	0.03325620	0.455928
α_2	0.02179287	0.401919
α_3	0.03987597	0.471691

From the table above, we obtained the estimates of the mean and standard error of millet, millet rice, and sorghum. The parameter estimate of millet rice is more efficient due to its lowest standard error as compared to others.

Table 8: Parameter Estimation (MOM).

Coefficients	Estimate	Std. Error
α_1	0.04000534	0.4591688
α_2	0.01302276	0.309861
α_3	0.05461017	0.4750334

From the table above, we obtained the estimates of the mean and standard error of millet, millet rice, and sorghum. The parameter estimate of

millet rice is more efficient due to its lowest standard error as compared to others.

Conclusively, the quasi-likelihood estimator performs the best as compared to others.

CONCLUSION

The Dirichlet distribution is a multivariate generalization of the Beta distribution. In this research, we introduced three methods of estimation for Dirichlet distribution which are maximum likelihood estimator (MLE), Method of Moment (MOM) and Quasi-likelihood estimator. This was done in other to obtain the most efficient method. An extensive simulation study was carried out on the basis of selected criterion (Bias, Variance, Mean absolute error and Mean square error) considering various sample sizes, also the methods were subjected to real life data.

The performance of these methods was compared at different sample sizes it shows that the Quasi-likelihood estimator performs better in terms of Bias, than the other methods, while Method of Moment performs better in terms of Variance, than the other methods. Maximum likelihood estimator performs better in terms of Mean Absolute Error (MAE) and (MSE) than the other methods.

The real-life results show that Quasi-likelihood estimator performs better as compared to Method of moment and Maximum likelihood estimator, also the Bayes factor of Dirichlet distribution gives 57.95215, which implies a very strong evidence of the Goodness of Fits. Hence, the Dirichlet distribution is efficient based on what we have done with higher precision and more adequacies in the estimate of the model, also the estimate of the model should be used in taking any prospective decision and can be reliable if large samples is involved.

REFERENCES

1. Aitkin, M. and D. Rubin. 1985. "Estimation and Hypothesis Testing in Finite Mixture Models". *Journal of the Royal Statistical Society B*. 47: 67–75.

2. Akaike, H. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle". In: *Second International Symposium on Information Theory*. 267–281. MR0483125.

3. Andrieu, C. and A. Doucet. 2001. "Recursive Monte Carlo Algorithms for Parameter Estimation in General State-Space Models". *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*. pp. 14–17.

4. Antoniak, C.E. 1974. "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *Annals of Statistics.* 2(6): 1152-1174.

5. Banfield, J.D. and A.E. Raftery. 1993. "Model-Based Gaussian and Non-Gaussian Clustering". *Biometrics.* 49: 803–821. MR1243494.

6. Biernacki, C., G. Celeux, and G. Govaert. 1999. "An Improvement of the NEC Criterion for Assessing the Number of Clusters in a Mixture Model". *Pattern Recognition Letters.* 20: 267–272.

7. Biernacki, C., G. Celeux, and E.M. Gold. 2000. "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood". *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 22: 719– 725.

8. Choi, S.C. and R. Wette. 1969. "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias". *Technometrics.* 11(4):683–690.

9. Escobar, M.D. and M. West. 1995. "Bayesian Density Estimation and Inference using Mixtures". *Journal of the American Statistical Association.* 90:577– 588, 1995.

10. Kai, Wang Ng, T. Guo-Liang, and T. Man-Lai. 2011. *Dirichlet and Related Distributions: Theory, Methods and Applications.* 888, Chapter 1, 2. John Wiley & Sons: New York, NY.

11. Kass, R.E. and A.E. Raftery. 1995. "Bayes Factors". *Journal of the American Statistical Association.* 90: 773–795.

12. Ken, A., D. DeWayne, and T. Peterson. 2014. "Model Selection for Ecologists: The Worldviews of AIC and BIC". *Ecology.* 95(3):631–636, 2014.

13. Koyejo, S.O., A.A. Akomolafe, C.A. Awogbemi, and O.O. Oladimeji. 2020. "Extension of Comparative Analysis of Estimation Methods for Frechet Distribution Parameters". *International Journal of Research and Innovation in Applied Science (IJRIAS).* 5(4). |ISSN 2454-6194.

14. Liu Z, J. Almhana, V. Choulakian, an R. McGorman. 2005. "Online EM Algorithm for Mixture with Application to Internet Traffic Modeling". *Comput Stat Data Anal.* 50:1052–1071.

15. Longford, N.T. 1987. "A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Random Effects". *Biometrika*. 74(4):817–827.

16. MacEachern, S. 1999. "Dependent Nonparametric Processes". *Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association.

17. Mario, A., T. Figueiredo and A.K. Jain. 2009. "Unsupervised Learning of Finite Mixture Models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24(3):381–396.

18. Maitra, R. 2009. "Initializing Partition-Optimization Algorithms". *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 6: 144–157.

19. Marlin, B. 2004. "Collaborative Filtering: A Machine Learning Perspective". Master's thesis, Univ. Toronto: Toronto, Canada.

20. Maitra, R. and V. Melnykov. 2010a. "Assessing Significance in Finite Mixture Models". Tech. Rep. 10-01, Department of Statistics, Iowa State University.

21. Maitra, R. and V. Melnykov. 2010b. "Simulating Data to Study Performance of Finite Mixture Modeling and Clustering Algorithms". *Journal of Computational and Graphical Statistics*.

22. McLachlan, G. and T. Krishnan. 1997. *The EM Algorithm and Extensions*. John Wiley & Sons: New York, NY. MR1417721.

23. Narayanaswamy, B. and V.B. Nevzorov. 2004. *A Primer on Statistical Distributions*. Chapter 27. John Wiley & Sons: New York, NY.

24. Neal, R M. 1991. "Bayesian Mixture Modeling". *Proc. the 11th International Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis*. Seattle, WA. June, 1991,pp.197-211.

25. Nizar, B. and D. Ziou. 2006. "A Hybrid SEM Algorithm for High Dimensional Unsupervised Learning using a Finite Generalized Dirichlet Mixture". *IEEE Transactions on Image Processing*. 15(9):2657–2668.

26. Nizar, B., D. Ziou, and J. Vaillancourt. 2004. "Unsupervised Learning of a Finite Mixture Model Based on the Dirichlet Distribution and its Application". *IEEE Transactions on Image Processing*. 13(11):1533–1543.

27. Pearson, K. 1894. "Contribution to the Mathematical Theory of Evolution". *Philosophical Transactions of the Royal Society*. 185: 71–110.

28. Roeder, K. and L. Wasserman. 1997. "Practical Bayesian Density Estimation using Mixtures of Normals". *Journal of the American Statistical Association*. 92: 894–902. MR1482121.

29. Ray, S. and B. Lindsay. 2008. "Model Selection in High Dimensions: A Quadratic-Risk-Based Approach". *Journal of Royal Statistical Society (B)*. 70: 95–118. MR2412633.

30. Sato, M. and S. Ishii. 2000. "On-Line EM Algorithm for the Normalized Gaussian Network". *Neural Comput*. 12:407–432.

31. Schwarz, G. 1978. "Estimating the Dimensions of a Model". *Annals of Statistics*. 6: 461–464. MR0468014.

32. Sethuraman, J. and R.C. Tiwari. 1982. "Convergence of Dirichlet Measures and the Interpretation of Their Parameter". In: *Statistical Decision Theory and Related Topics*. S. Gupta and J.O. Berger (eds.). Academic Press: London, UK. 2: 305-315.

33. Yeung, K.Y., C. Fraley, A. Murua, A. Raftery, and W.L. Ruzzo. 2001. "Model-Based Clustering and Data Transformations for Gene Expression Data". *Bioinformatics*. 17: 977–987.

34. Zivkovic, Z. and F. van der Heijden. 2004. "Recursive Unsupervised Learning of Finite Mixture Models". *IEEE Trans Pattern Anal Mach Intell*. 26: 651–656.

35. Zoubin G. 2004. "Unsupervised Learning". *Advanced Lectures on Machine Learning*. 72–112. Springer: Berlin, Germany.

SUGGESTED CITATION

Halid M.A. and A.A. Akomolafe. 2020. "Extension of Comparative Analysis of Estimation Methods for Dirichlet Distribution Parameters". *Pacific Journal of Science and Technology*. 21(2):159-169.

Pacific Journal of Science and Technology