Dirichlet - Multinomial Model: Its Mixture And Application Using Bayesian Approach

[1]akomolafe Abayomi. A., And [2]yussuf Tajudeen. O

[1] Department Of Statistics, Federal University Of Technology, Akure, Nigeria
[akomolafe01@yahoo.com],
[2]National institute for educational planning and administration, Ondo
[seunyusuf2015@gmail.com]

Abstract: The distribution found by compounding dirichlet distribution with Multinomial distribution give a better distribution than each of them performing individually in terms of the estimates of their characteristics. We provide a comprehensive description of the mathematical precision of the distribution along with its efficient, consistence and flexibility behavior. The usefulness of Dirichlet-Multinomial model was illustrated using real data.
[akomolafe Abayomi. A., And yussuf Tajudeen. O. **Dirichlet - Multinomial Model: Its Mixture And Application Using Bayesian Approach.** *Rep Opinion* 2018;10(2):1-15]. ISSN 1553-9873 (print); ISSN 2375-7205 (online). http://www.sciencepub.net/report. 1. doi:10.7537/marsroj100218.01.

Keyword: Dirichlet Distribution, Standard Error, Generalized Variance, Mean Vector and Dirichlet-Multinomial Model

1. Introduction

Marketing scientists aim to develop reliable models, ones that have been tested and shown to yield accurate estimations in a wide range of circumstances. If a model always fits, then its estimations tell us little that we did not already know from our observed data. Consumers are seen as mostly choosing from personal split-loyalty repertoires, typically buying one brand more often than another. The Dirichlet model describes how frequently-bought branded consumer products are purchased when the market is stationary and unsegmented. This is the common situation where, over the time-periods analysed, the sales of each brand show little variation, and different brands show no special groupings.

Sales of a brand are determined by measures such as how many customers buy the brand, how often, and how much they also buy other brands. To study the consumer buying behavior, it therefore becomes imperative to the producers and the economy planner to track their behavior as regards their product, for future planning, and optimum profitability. This can be modeled using a *Dirichlet multinomial distribution* since buyers have steady but divided loyalties. The Bayesian approach to nonparametric inference, however, faces challenging issues since construction of prior distribution involves specifying appropriate probability measures on function spaces where the parameters lie. Typically, subjective knowledge about the minute details of the distribution on these infinite-dimensional spaces is not available for nonparametric problems. A prior distribution is generally chosen based on tractability, computational convenience and desirable frequentist behavior, except that some key parameters of the prior may be chosen subjectively. To study frequentist properties, it is assumed that there is a true value of the unknown parameter which governs the distribution of the generated data. We are interested in knowing whether the posterior distribution eventually concentrates in the neighborhood of the true value of the parameter. Lack of consistency is extremely undesirable, and one should not use a prior if the corresponding posterior is inconsistent. Consistency of Bayesian nonparametric procedures has been the focus of a considerable amount of research in recent years. Most contributions in the literature exploit the "frequentist" approach to Bayesian consistency, also termed the "what if" method according to Diaconis and Freedman (1986). It is on this background, a Dirichlet multinomial Model is being employed in this study because of its capability of capturing buying behavior pattern of consumers. The computed predictions are also compared with those obtained on the basis of Dirichlet model so as to determine the consistency of Dirichlet multinomial Model using Bayesian Method.

Packaging attributes can persuade consumers to purchase the product. According to Dantas *et al.*, (2004), packages and labels have only a few seconds to make an impact on the consumer's mind; during that time, it must catch the consumer's eye, and convince the shopper that it is the optimum option on the shelf (Rowan,2000). Any piece of information associated with a brand can become part of consumer memory and a key component of brand knowledge (Keller, 2003). These links in memory are also a key source of brand equity, because they result in differential consumer responses to marketing initiatives (Keller, 2003). Therefore, successful brands should aim to be linked to multiple and accessible attributes (Krishnan, 1996). The more attributes linked to the brand name, the greater the number of potential pathways to retrieve the brand or evaluate it as having the necessary qualities for purchase (Romaniuk, 2003).

Under a classic Empirical then Theoretical (E-T) approach as described by Bass and Wind (1995), the aim is to determine which distribution, if any, fits the data better and from that draw conclusions about the nature of attribute responses. Until now, much of the application of mathematical models to marketing concerns consumer purchase behavior data. The most famous of these applications is the Dirichlet model (Goodhardt et al., 1984). The use of probability distributions and functions allows us to predict recurring behavioural patterns. Similar to buying behaviour modeling techniques, this study examines the predictability of patterns in brand attribute responses. Nigerian government has been conducting studies and experiment concerning buying behavior data. In many of these studies, significant relationship between buying behavior of durable goods and subsequent purchases were found to be significant in various economic and statistical models used on the data (Akomolafe & Amahia,2009). Based on the studies carried out by these researchers, it was noted that intenders' purchase rates are higher than those that drop-out of the interaction. In fact, for most of the durable goods, majority of households report that consumers do not actually show their intention carry out most of the actual purchases. Several other studies have examined the theory and practical application of their developed models towards the tracking and forecasting of future buying- behavior of consumers. This may be responsible for incorporating the instability in the consumer buying-behavior, drop-out effect as well as measuring the effect of heterogeneity on the purchase behavior (Akomolafe et al. 2010; Peter,2008; Schmittlem & Petterson 1994; Metheringgham,1988; Davis,1989; Ishii & Hayakawa, 1990; Colombo & Weima,1990).

2. Methodology

This research review the statistical technique used to analyzed the data collected from the retail outlet on buying habit of the consumer with respect to their Zones in which a careful and accurate analysis of the data was carried out using the **Dirichlet and Dirichlet multinomial model** and necessary prediction will be followed to make reasonable inference on the whole populace regarding the buying habit and future behaviors with respect to different Zones across the Country.

The Dirichlet distribution is a multivariate generalization of the beta distribution. Recall that the beta distribution arose as follows; suppose that v_1 and v_2 are independent Gamma random variables with $v_1 \sim_{Ga\,(\propto_{1,}\,\beta)}, v_2 \sim_{Ga\,(\propto_{2,}\,\beta)}$. Then if X is defined by $X = \frac{V_1}{(V_1+V_2)}$, we have that $X \sim Be\,(\propto_1, \propto_2)$. Now consider a generalization; suppose that $v_1, v_2, v_3 \dots v_{k+1}$ are independent Gamma random variables with $v_i \sim_{Ga\,(\propto_{i,}\,\beta)}, for\ i = 1, \dots, k+1$.

Define $X_i = \frac{V_i}{V_1+V_2+\cdots+V_{k+1}} for\ i = 1, \dots, k + 1$. Then the joint distribution of vector $X = (X_1, \dots, X_k)^T$ is given by density

$$fX_{1\dots,}X_k(x_{1,\dots,}x_k) = \frac{\gamma(\propto)}{\gamma(\propto_1)\dots\gamma(\propto_K)\gamma(\alpha_{K+1})} x_1^{\propto 1-1} \dots x_k^{\propto k-1} x_{k+1}^{\propto k+1-1},$$

for $0 \le x_i \le 1$ for all i such that $x_i + \cdots + x_k + x_{k+1} = 1$, where $\propto = \propto_1 + \cdots + \propto_{k+1}$ and where x_{k+1} is defined by $x_{k+1} = 1 - (x_1 + \cdots x_k)$. This is the density function which reduces to the beta distribution if k = 1.

Dirichlet distribution: Let U = ($U_1, U_2, U_3, \dots, U_k$) be a random pmf, that is $U_i \ge 0$ for i = 1, 2,…, k and $\sum_{i=1}^{K} U_i = 1$ In addition, suppose that $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ with $\alpha_i > 0$ 0 for each i, and let $\alpha_0 = \sum_{i=1}^{k} \alpha_i$. Then, **U** is said to have a Dirichlet distribution with parameter α, which is denoted by U ~Dir (α), if it has f (u; α) = 0 if u is not a pmf, and if **u** is a pmf then

$$f[u;\ \alpha] = \frac{\Gamma(\Sigma\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \prod_{i=1}^{k} \mu_i^{\propto_i - 1} \quad (1)$$

The parameters can be estimated from a training set of proportions: $D = (U_1, \dots, U_N)$. The maximum-likelihood estimate of α maximizes p (D/ $\alpha) = \prod_i p(U_i / \alpha).$ The log-likelihood can be written as:

$$\text{Log p} (D/\alpha) = N \log \Gamma\left(\sum_i \alpha_i - N \sum_i \log \Gamma(\alpha_i) + N \sum_i (\alpha_i - 1)\right) \log \overline{U}_i \quad (2)$$

Where $\log \overline{U}_i = \frac{1}{N} \sum_k \log U_{ki}$

However, the mean and variance of dirichlet distribution can be therefore be obtained by the following expectation procedures

$$\sum_{i=1}^{k} \mu_{i=1}\ \epsilon[\mu_i]$$

$$\mu = [\mu_1, \dots\dots, \mu_k]^T$$

$$\epsilon[\mu_i] = \int \mu_i p_k(\mu_i) d\mu_i$$

$$= \frac{\Gamma(\Sigma\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \int .. \int \mu_i \prod_{i=1}^{k} \mu_i^{\propto_i - 1}\, d\mu_i \ldots d\mu_k$$

$$= \frac{\Gamma(\Sigma\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \int .. \int \mu_i \prod_{i=1, i\neq 1}^{k} \mu_i^{\propto_i - 1}\, \mu_i^{\alpha_{i+1-1}} d\mu_i \ldots . d\mu_k$$

$$= \frac{\Gamma(\Sigma\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \frac{\prod_{i=1, i\neq 1}^{k}\Gamma\alpha_i\Gamma(\alpha_{i+1})}{\Gamma\sum\alpha_i + 1}$$

Recall the relation

$\Gamma x + 1 = x\Gamma(x)$

$$\epsilon(\mu_i) = \frac{\Gamma(\Sigma\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} \frac{\prod_{i=1, i\neq 1}^{k}\Gamma\alpha_i.\alpha_i\Gamma\alpha_i}{\sum_{i=1}^{k}\alpha_i\,\Gamma(\sum_{i=1}^{k}\alpha_i)}$$

$$\epsilon(\mu_i) = \frac{\alpha_i}{\sum_{i=1}^{k}\alpha_i} \quad (3)$$

Dirichlet Distribution is a special case of beta distribution when k = 2

Variance

$\mu_i \sim Dir(\mu/\alpha)\mu_i \in \mu$

$Var[\mu_i] = \epsilon[\mu_i^2] - \left(\epsilon(\mu_i)\right)^2$

$$\epsilon[\mu_i^2] = \int_0^1 \mu_1^2 p_k[\mu_i] d\mu_i \quad [4]$$

$$\epsilon[\mu_i^2] = \int \mu_i^2 p_k[\mu_i \ldots . \mu_k] d\mu_i \ldots .. d\mu_k$$

$$Dir(\mu/\alpha) = C_m \prod_{i=1}^{k} \mu_i^{\alpha_{i-1}}$$

$$= \int \mu_i^2\, C_m \prod_{i=1}^{k} \mu_i^{\alpha_{i-1}}\, d\mu_i \ldots .. d\mu_k$$

$$= C_m \int \prod_{i=1}^{k} \mu_i^{\alpha_{i-1}}\, \mu_i^{\alpha_{m+2-1}} d\mu_i \ldots .. d\mu_k$$

$$= C_m . \frac{1}{C_m^1} \quad [5]$$

$$C_m^1 = \frac{\Gamma\Sigma_{i=1}^{k}(\alpha_i + 2)}{\prod_{i=1, i\neq 1}^{k}\Gamma\alpha_i\Gamma(\alpha_{i+2})}$$

$$= \frac{\Gamma\sum_{i=1}^{k}\alpha_i}{\prod_{i=1}^{k}(\Gamma\alpha_i)} \frac{\prod_{i=1, i\neq 1}^{k}\Gamma\alpha_i\Gamma(\alpha_{i+2})}{\Gamma\Sigma_{i=1}^{k}(\alpha_k + 2)} \quad [6]$$

Using the relation

$\Gamma x + 1 = x\Gamma(x)$

$$= \frac{\Gamma(\Sigma\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)} . \frac{(\alpha_i + 1)\alpha_i\Gamma\alpha_i\prod_{i=1, i\neq 1}^{k}\Gamma(\alpha_i)}{(\sum_{i=1}^{k}\alpha_{i+1})(\sum_{i=1}^{k}\alpha_i)\Gamma(\sum_{i=1}^{k}\alpha_i)} \quad [7]$$

$$\epsilon[\mu_i^2] = \frac{\alpha_i\left(1 + \alpha_j\right)}{\sum_{i=1}^{k}\alpha_i\left(1 + \sum_{i=1}^{k}\alpha_i\right)}$$

$$\text{var}\ (u_i) = \frac{\alpha\ (1 + \alpha_i)}{\sum_{i=1}^{k}\alpha_i\left(1 + \sum_{i=1}^{k}\alpha_i\right)} - \left(\frac{\alpha_i}{\sum_{i=1}^{k}\alpha_i}\right)^2$$

$$= \frac{\sum_{i=1}^{k}(\alpha_i - \alpha_j)\alpha_i}{\left(\sum_{i=1}^{k}\alpha_i\right)^2\left(1 + \sum_{i=1}^{k}\alpha_i\right)} \quad (8)$$

Generating the Dirichlet from Gamma RVs

Generating samples from the Dirichlet distribution using Gamma random variables is more computationally efficient than both the urn-drawing method and the stick-breaking method (Bela *et al*, 2010). This method has two steps which we explain in more detail and prove below:

Step 1**:** Generate gamma realizations: for i = 1, 2,…, k, draw a number z_i from $\Gamma(\alpha_i, 1)$

Step 2**:** Normalize them to form a pmf: for i = 1, 2,…, k, set $u_i = \frac{z_i}{\sum_{j=1}^{k} z_j}$. Then u is a realization of Dir (α). The Gamma distribution $\Gamma(k, \theta)$ is defined by the following probability density:

$$f(x; k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k\Gamma(k)} \quad (9)$$

k > 0 is called the shape parameter, and θ>0 is called the scale parameter. Suppose $X_i \sim \Gamma(k_i, \theta)$ are independent for i = 1, 2,…, n; which implies that they are on the same scale but can have different shapes. Then, S=$\sum_{i=1}^{n} X_i \sim \Gamma(\sum_{i=1}^{n} k_i, \theta)$

Prove of the procedure generating Dirichlet samples from Gamma r.v. draws works, using the change-of-variables formula to show that the density of U is the density corresponding to the Dir (α) distribution

Proof**:** Recall that the original variables are $(Z_i)_1^k$, and the new variables are Z, $U_1, \ldots, U_{k-1}$. Using the transformation T:

$$(Z_1, \ldots, Z_k) = T(Z, U_1, \ldots, U_{k-1}) = (ZU_1, \ldots, ZU_{k-1}, Z(1 - \sum_{i=1}^{k-1} U_i)).$$

The Jacobian matix (matrix of first derivatives) of this transformation is:

$$J(T) = \begin{pmatrix} U_1 & Z & 0 & 0 & . & . & . & 0 \\ U_2 & 0 & Z & 0 & . & . & . & 0 \\ U_3 & 0 & 0 & Z & . & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ U_{K-1} & 0 & 0 & . & . & . & 0 & Z \\ 1-\sum_1^{K-1} U_i & -Z & -Z & -Z & . & . & . & -Z \end{pmatrix} \quad (10)$$

Which has determinant Z^{k-1}

The standard change-of-variables formula tells us that the density of $(Z, U_1, \ldots, U_{K-1})$ is $f = goTx|\det(T)|$, where

$$g(z_1, z_2, \ldots, z_k; \alpha_1, \ldots, \alpha_k) = \prod_{i=1}^{k} z_i^{\alpha_i - 1} \frac{e^{-z_i}}{\Gamma(\alpha_i)} \quad (11)$$

Is the joint density of the original (independent) random variables? Substituting (11) into the change of variables formula, we find the joint density of the new random variables.

$$f(z, u_1, \ldots, u_{k-1}) = \left(\prod_{i=1}^{k} (zu_i)^{\alpha_i - 1} \frac{e^{-zu_i}}{\Gamma(\alpha_i)}\right)\left[\left(z\left(1 - \sum_{i=1}^{k-1} u_i\right)\right)^{\alpha_k - 1} \frac{e^{-z\left(1-\sum_{i=1}^{k=1} u_i\right)}}{\Gamma(\alpha_i)}\right] z^{k-1} \quad [12]$$

$$= \frac{\left(\prod_{i=1}^{k-1} u_i^{\alpha_i - 1}\right)\left(1 - \sum_{i=1}^{k-1} u_i\right)^{\alpha_k - 1}}{\prod_{i=1}^{k} \Gamma(\alpha_i)} z\left(\sum_{i=1}^{k} \alpha_i\right) - 1e^{-z} \quad [13]$$

Integrating over z, the marginal distribution of $\{U_i\}_{i=1}^{k-1}$ is

$$f(u) = f(u_1, \ldots, u_{k-1}) = \int_0^\infty f(z, u_1, \ldots, u_{k-1}) dz$$

$$= \frac{\left(\prod_{i=1}^{k-1} u_i^{\alpha_i - 1}\right)\left(1 - \sum_{i=1}^{k-1} u_i\right)^{\alpha_k - 1}}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \int_0^\infty z\left(\sum_{i=1}^{k} \alpha_i\right) - 1e^{-z} dz \quad [14]$$

$$= \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)}\left(\prod_{i=1}^{k-1} u_i^{\alpha_i - 1}\right)\left(1 - \sum_{i=1}^{k=1} u_i\right)^{\alpha_k - 1} \quad [15]$$

Which is the same as the Dirichlet density in (1)

The Multinomial Distribution

The multinomial distribution is a multivariate generalization of the binomial distribution. Recall that the binomial distribution arose from an infinite Urn model with two types of objects being sampled with replacement. Suppose that the proportion of "Type 1" objects in the urn is U (so $0 \leq U \leq 1$) and hence the proportion of "Type 2" objects in the urn is $1-U$. Suppose that n objects are sampled, and X is the random variable corresponding to the number of "Type 1" objects in the sample. Then, $X \sim Bin(n, U)$ and the PDF is given as:

$$fX(x) = \binom{n}{x} U(1-U)^{n-x} \quad , \quad x \in \{0, 1, \ldots, n\} \quad . \quad [16]$$

Now consider a generalization; suppose that the Urn contains k + 1 types of objects (k = 1, 2,...), with U_i being the proportion of Type i objects, for i = 1,..., k + 1. Let X_i be the random variable corresponding to the number of type i objects in a sample of size n, for i = 1,..., k. Then the joint distribution of vector $X = (X_1, \ldots, X_k)^T$ is given by

$$fX_{1,\dots,}X_k(x_{1,\dots,}x_k) = \frac{n!}{x_1!\dots x_k!x_{k+1}!}U_1^{x_1}\dots U_k^{x_k}U_{k+1}^{x_{k+1}} = \frac{n!}{x_1!\dots x_k!x_{k+1}!}\prod_{i=1}^{k+1}U_i^{x_i} \quad (17)$$

$0 \le U_i \le 1$ for all i, and $U_i + \cdots + U_k + xU_{k+1} = 1$ and x_{k+1} is defined by $x_{k+1} = n - (x_1 + \cdots x_k)$. Hence, this is the mass function for the multinomial distribution which reduces to the binomial if k = 1.

Moments

Recall for random variable x with mass density function fx, the moment generating function of x, k_x, is defined by

$$M_x(t) = E[\exp\{tn^T\}] = \sum_n \exp\{tn^T\}f(N;n,U) \quad [18]$$

$$= \sum_n \exp\{tn^T\}\left(\frac{N!}{\prod_{i=1}^k x_i!}\right)\prod_{i=1}^k U_i^{x_i} \quad [19]$$

$$= \left(\sum_{i=1}^k U_i e^{t_i}\right)^n$$

$$M_x(t) = \left(U_1e^{t_1} + U_2e^{t_2} + \dots + U_k\right)^N$$

$$M_x(t) = \left(U_1e^{t_1} + U_2e^{t_2} + \dots + 1 - \sum_{i=1}^{k-1}U_i\right)^N$$

Where $U_i = 1 - \sum_{i=1}^{k-1}U_i$

$$\frac{\partial}{\partial t}M_x(t) = NU_ie^t\left(U_1e^{t_1} + U_2e^{t_2} + \dots + 1 - \sum_{i=1}^{k-1}U_i\right)_{N-1}$$ (Erik, 2014, page 26)

$$\frac{\partial}{\partial t}M_x(t=0) = NU_i$$

$$\epsilon[U_i] = NU_i \quad (20)$$

$$\text{Variance }[U_i] = \varepsilon(U_i)^2 - (\varepsilon(U_i))^2 = NU_i(1 - U_i) \quad (21)$$

Dirichlet-Multinomial Distribution

When nominal data shows a lot of variance, the multinomial distribution will not be a adequate model category counts because of its limited variance-covariance structure. In search of a more adequate distribution that can model nominal data, one can let ***U*** be a random variable which follows the Dirichlet distribution. The Dirichlet-multinomial is achieved through integrating the product of the multinomial and dirichlet distribution over the simplex ω. The distribution is also called Polya distribution (Minka, 2012). The compound probability mass function for the Dirichlet-multinomial is given as

$$f(x \mid \alpha) = \int_\Omega f(x \mid u)f(x \mid \alpha)du =$$

Combine the integral of (1) and (2)

$$= \int_\Omega \frac{N!}{\prod_{i=1}^k x_i!}\prod_{i=1}^k U_i^{x_j}\frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)}\prod_{i=1}^k U_i^{\alpha_i - 1}\partial u =$$

$$= \frac{N!}{\prod_{i=1}^k x_i!}\frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)}\int_\Omega \prod_{i=1}^k U_i^{x_i+\alpha_i-1}\partial u$$

Focusing on the integral alone,

$$\int_\Omega \prod_{i=1}^k U_i^{x_i+\alpha_i-1}\partial u = \frac{\prod_{i=1}^k \Gamma(\alpha_i + x_i)}{\Gamma(\sum_{i=1}^k \alpha_i + x_i)}\left(\int_\Omega \frac{(\Gamma(\sum_{i=1}^k \alpha_i + x_i)}{\prod_{i=1}^k \Gamma(\alpha_i + x_i)}\prod_{i=1}^k U_i^{x_i+\alpha_i-1}\partial u\right) \quad [22]$$

Where the term in brackets on the RHS is a PDF of the Dirichlet ($(x_i + \alpha - 1)$ whose the integral evaluated to be 1. Hence,

$$f(x \mid \alpha) = \frac{N!}{\prod_{i=1}^{k} x_i!} \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\Gamma(\sum_{i=1} \alpha_i + x_i)} \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + x_i)}{\Gamma(\alpha_i)} \tag{23}$$

α can be estimated from a training set of count vectors: D = $(Y_1, \dots, Y_N)$. The likelihood is

$$x_i = \sum_i ki \tag{24}$$

$$p(D/\alpha) = \prod_k p(y_i/\alpha)$$

$$= \prod_k \left(\frac{N!}{\prod_{i=1}^{k} x_i!} \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\Gamma(\sum_{i=1} \alpha_i + x_i)} \prod_{i=1}^{k} \frac{\Gamma(\alpha_i + x_i)}{\Gamma(\alpha_i)} \right) \quad [25]$$

The gradient of the log-likelihood is

$$g_i = \frac{\partial \log p(D/\alpha)}{\partial \alpha_I} =$$

$$\sum_k \Psi\left(\sum_i \alpha_i\right) - \Psi\left(x_i + \sum_i \alpha_i)\right) + \Psi(x_i + \alpha_i) - \Psi(\alpha_i) \tag{26}$$

The maximum can be computed via the fixed-point iteration

$$\alpha_i^{new} = \alpha_i \frac{\sum_i \Psi(x_i + \alpha_k) - \Psi(\alpha_i)}{\sum_i \Psi(x_i + \sum_i \alpha_i) - \Psi(\sum_i \alpha_i)} \tag{27}$$

where Ψ is the digamma function

α can also be estimated from a training set of count vectors: D = $(Y_1, \dots, Y_N)$ using simplified Newton iteration method and EM algorithm method see Minka (2000).

Moments Of Dirichlet-Multinomial Distribution

Notations of methods of moments were taken from Wang Ng. *et al.* (2012).

Recall that:

$$\varepsilon(X) = \varepsilon[\varepsilon(X/Y)]$$

Therefore,

$$\varepsilon(X_i) = \varepsilon[\varepsilon(X_i/U_i)] = N\varepsilon(U_i)$$

From equation (3.1.3), $\epsilon(\mu_i) = \frac{\alpha_i}{\sum_{i=1}^{k} \alpha_i}$

Hence, $\varepsilon(X_i) = N\varepsilon(U_i) = \frac{N\alpha_i}{\sum_{i=1}^{k} \alpha_i}$ (28)

Variance

$$Var(X) = Var(\varepsilon[X/Y]) + \varepsilon[Var(X/Y)]$$

$$Var(X_i) = Var(\varepsilon[X_i/U_i]) + \varepsilon[Var(X_i/U_i)]$$

$$= \varepsilon[NU_i(1-U_i)] + Var(NU_i)$$

$$= N\varepsilon(U_i) - N\left[Var(U_i) + \varepsilon(U_i)^2\right] - N^2 Var(U_i)$$

$$= N\varepsilon(U_i)[1 - \varepsilon(U_i)] + N(N-1)Var(U_i)$$

From equation (3) and (4)

$$= \frac{N\alpha_i}{\sum_{i=1}^{k}\alpha_i}\left[1-\frac{\alpha_i}{\sum_{i=1}^{k}\alpha_i}\right]+N(N-1)\frac{\alpha_i\left(\sum_{i=1}^{k}\alpha_i-\alpha_j\right)}{\left(\sum_{i=1}^{k}\alpha_i\right)^2\left(1+\sum_{i=1}^{k}\alpha_i\right)}$$

$$= \left(\frac{N\left(N+\sum_{i=1}^{k}\alpha_i\right)}{1+\sum_{i=1}^{k}\alpha_i}\right)\left[\frac{\alpha_i}{\sum_{i=1}^{k}\alpha_i}\right]\left(1-\frac{\alpha_i}{\sum_{i=1}^{k}\alpha_i}\right) \tag{29}$$

Conjugate Priors

When doing Bayesian inference, we are often interested in a few key distributions. Below, D refers to the dataset. We have at hand, and we typically assume each $x_i \in D$ is drawn independently and identically distributed iid from the same distribution $f(x;\theta)$, where parameterizes the likelihood model. That is, treating θ as an unknown, but postulating that θ follows some prior distribution $f(\theta;\alpha)$, where α parameterizes the prior distribution (and is often called the hyper-parameter). Conjugate priors, however, are the class of models where we can analytically compute distributions of interest. These distributions are:

Posterior predictive: This is the distribution denoted notationally by $f(x \mid D)$ where n is a new data point of interest. By the assumption, we have

$$f(x \mid D) = \int f(x,\theta \mid D)d\theta = \int f(x \mid \theta)f(\theta \mid D)d\theta$$

The distribution $f(\theta \mid D)$, often called the posterior distribution, which can be broken down further

$$f(\theta \mid D) = \frac{f(\theta,D)}{f(D)} \alpha f(\theta,D) = f(\theta \mid \alpha)\prod_{x_i \in D} f(x_i \mid \theta) \quad [30]$$

Marginal distribution of the data

This is denoted $f(D)$, and is derived by integrating out the model parameter.

$$f(D) = \int f(D,\theta)d\theta = \int f(\theta \mid \alpha)\prod_{n_i \in D} f(x_i \mid \theta)d\theta$$

Posterior

$$f(\theta \mid D) \alpha f(\theta,D)$$

$$= f(U_1,\ldots,U_k \mid \alpha_1,\ldots,\alpha_k)\prod_{x_i \in D} f(x_i \mid U_1,\ldots,U_k)\alpha\prod_{i=1}^{k} U_i^{\alpha_i-1}\prod_{n_i \in D}\prod_{i=1}^{k} U_i^{x_i(k)} \tag{31}$$

$$= \prod_{i=1}^{k} u_i^{\alpha_j-1} + \sum_{n_i \in D} x_i^{(k)}$$

This density is exactly that of a Dirichlet distribution, except we have

$$\alpha_i = \alpha_i + \sum_{x_i \in D} x_i^{(k)}$$

That is, $f(\theta \mid D) = Dir(\alpha_1^{'},\ldots,\alpha_k^{'})$.

Posterior Predictive

$$f(x \mid D) = \int f(x \mid \theta)f(\theta \mid D)d\theta$$

$$= \int f(x_i \mid U_1,\ldots,U_k)f(U_1,\ldots,U_k \mid D))dS_k$$

$$= \int \frac{\Gamma(x+1)}{\prod_{i=1}^{K}\Gamma\left(x^{(i)}+1\right)} \prod_{i=1}^{K} U^{x^{(k)}} \frac{\Gamma\left(\sum_{i=1}^{K}\alpha_i'\right)}{\prod_{i=1}^{K}\Gamma\left(\alpha_i\right)} \prod_{i=1}^{K} U_i^{\alpha_j'-1} dS_K$$

$$= \frac{\Gamma(x+1)}{\prod_{i=1}^{K}\Gamma\left(x^{(i)}+1\right)} \frac{\Gamma\left(\sum_{i=1}^{K}\alpha_i'\right)}{\prod_{i=1}^{K}\Gamma(\alpha i)} \int \prod_{i=1}^{K} U_i^{x^{(i)}+\alpha_i'-1} dS_K \quad [32]$$

$$= \frac{\Gamma(x+1)}{\prod_{i=1}^{K}\Gamma\left(x^{(i)}+1\right)} \frac{\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)}{\prod_{i=1}^{K}\Gamma\left(\alpha_i'\right)} \frac{\prod_{i=1}^{K}\Gamma\left(x^{(i)}+\alpha i\right)}{\Gamma\left(x+\sum_{i=1}^{K}\alpha_i'\right)} \quad (33)$$

where dS_K denotes integrating $\left(U_1,\ldots,U_k\right)$ with respect to the $(K-1)$ simplex.

Marginal

$$f(D) = \int f(\theta \mid \alpha) \prod_{n_i \in D} f(n_i \mid \theta) d\theta$$

$$= \int f\left(U_1,\ldots,U_k \mid \alpha_1,\ldots,\alpha_k\right) \prod_{x_i \in D} f\left(x_i \mid U_1,\ldots,U_k\right) dS_K \quad (34)$$

$$= \int \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{K}\Gamma\left(\alpha_i\right)} \prod_{i=1}^{K} U_i^{\alpha_i-1} \prod_{x_i \in D} \frac{\Gamma(n+1)}{\prod_{i=1}^{K}\Gamma\left(x_i^{(k)}+1\right)} \prod_{i=1}^{K} U_i^{x_i^{(k)}} dS_K \quad [35]$$

$$= \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{K}\Gamma\left(\alpha_i\right)} \left[\prod_{x_i \in D} \frac{\Gamma(n+1)}{\prod_{i=1}^{K}\Gamma\left(x_i^{(k)}+1\right)}\right] \int \prod_{i=1}^{K} U_k^{\sum_{x_i \in D} x_i^{(k)}+\alpha_i-1} dS_K$$

$$= \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{K}\Gamma\left(\alpha_i\right)} \left[\prod_{x_i \in D} \frac{\Gamma(n+1)}{\prod_{i=1}^{K}\Gamma\left(x_i^{(k)}+1\right)}\right] \frac{\prod_{i=1}^{K}\Gamma\left(\sum_{x_i \in D} x_i^{(k)}+\alpha_i\right)}{\Gamma\left(|D|x+\sum_{i=1}^{K}\alpha_i\right)} \quad [36]$$

Goodness-of-fit tests

Before we estimate the parameters of the Dirichlet or Dirchlet – Multinomial Distribution, we need to test whether the data came from a population that has a, Dirichlet, or Dirchlet – Multinomial Distribution. The null hypothesis of interest for a goodness-of-fit test is that a given random variable or vector *X* follows a specified Dirichlet or Dirchlet – Multinomial Distribution (*x*). The parameters of the hypothesized distribution may be known or unknown.

If $X_1,\ldots,X_n$ is a random sample with common Dirichlet or Dirchlet – Multinomial Distribution distribution function *F*, then we would like to test *F* = F_0 against $F \neq F_0$, where F_0 is a specified distribution with known or unknown parameters.

3. Data Analysis and Results

The implementation of consistency of dirichlet multinomial model in tracking purchasing pattern on the secondary data collected for 10 products (Trophy Lager Beer, Castle Milk Stout, Castle lager, Hero Lager Beer, MGD RB 600mlx12, Eagle Lager, Betamalt, Grand Malt, AFB and Voltic Water from International Breweries Ilesha) for ten years.

Fitting Dirichlet Distribution to the Buying Behavior Data

Table 1: Parameter Estimate of Dirichlet Distribution

coefficients	Estimate	St. Error	Z-value	Pr (>|Z|)
α_1	1.06177	0.03222	-31.562	2e-16
α_2	1.00756	0.03739	26.946	2e-16
α_3	0.03470	0.03467	1.001	0.317

From table above we obtain the parameter vector to be α = (1.06177 1.00756 0.03470), the statistical significance test shows that the first and second parameters are more statistically significant: P-value (2e-16) < α (0.05), and these determines significantly the shape of the dirichlet distribution.

Parameters and Mean Vectors for Dirichlet Distribution

The parameters of the dirichlet distribution is represented by the vector α_i = (1.06177 1.00756 0.03470), the corresponding mean vector is represented by m_i = (0.50463643 0.47887150 0.01649216). From the parameter vector, we can infer that the purchase rate for Trophy and CastleLager beer is significantly higher than that of Voltic water. The corresponding mean purchase for Trophy and Lager Beer are 0.50463643 and 0.47887150 respectively which are significantly higher than that of Voltic water (0.01649216). This shows that there will be an increment of 50.46% in purchase of Trophy and over 47.89% increment will be expected in Lager Beer while Voltic Water shows little public acceptance with the increment in purchase of about 1.65% over the period. The normalized parameter vector described the rate of purchase of each of the flavor under investigation, and it shows that Trophy has the highest purchase rate.

Variance and Bayes Factor

The vector d = (0.080533534 0.080396641 0.005225519) represents the variance of dirichlet distribution over the flavor under investigation.

Test for Goodness of Fit Dirichlet Distribution [DD]

H_0: the fit is not good for DD (α =0.05) vs H_1: the fit is good for DD

Decision Rule**:** If the Bayes factor falls in the critical region (10<BF<30), we do not accept H_0, Otherwise accept H_0.

Decision**:** Since BF =25.98807 lies in the region (10<BF<30), then we have statistical reason no accept H_0 and conclude that Dirichlet model is not a better fit for the buying behavior data.

Precision

If the precision parameter S <1, then the model is adequate and consistent for the buying behavior data, it is not consistent if otherwise. Since the precision parameter S (1.46129) >1, then the model is not consistent and it has inflated the parameter vector α = (α_i) by over 46.129% and this constitute to the low consistency of the model as regards the flavors under investigation which in turn vehemently affect the efficiency of the model.

Fitting Dirichlet-Multinomial Distribution on Buying Behaviour Data

Parameter Estimate of Dirichlet-Multinomial Model

Table 2: Parameter Estimate of Dirichlet-Multinomial Model

Coefficients	Estimate	St. Error	Z-value	Pr (>|Z|)
α_1	2.9747	0.2628	-11.320	<2e-16
α_2	0.5015	0.1492	3.362	0.000773
α_3	0.1169	0.1661	0.704	0.481590
α_4	0.1845	0.1626	1.135	0.256395

From table 4. 2 above, we can obtain the parameter vector to be α = (2.9747 0.5015 0.1169 0.1845), the statistical significance test shows that the Trophy and castle-lager parameters are more statistically significant: P-value (2e-16, 0.00773) < α (0.05), and these determines significantly the shape of the dirichlet distribution. Trophy and castle-lager have more purchase rate and purchase percentage of about 2.9747 and 50.15% respectively as compared to voltic water and betamalt. Betamalt contributes about 18.45% to sales volume while voltic water contribute the least to total sale of International brewery Ilesha.

Parameters and Mean Vectors for Dirichlet-Multinomial Distribution [DMD]

The parameters of the dirichlet multinomial distribution is represented by the vector α_i = (2.9747 0.5015 0.1845 0.1169), the corresponding mean vector is represented by m_i = (7.78745765 1.3275625 0.3094557 0.4884053) which depict the generalized mean purchase or sales volume at the first ten period of production. We can deduced that throughout the period, the sale average purchase for trophy is significantly higher than the rest of the product while other product maintain relatively small uniform sales throughout the period, with a little bit increase in the sale of castle lager within the period of investigation.

Bayes Factor

Test for Goodness of fit

H_0: the fit is not good for DMD (α =0.05) Versus H_1: the fit is good for DMD

Decision Rule**:** If the Bayes factor falls in the critical region (10<BF<30), we do not accept H_0, this is because there exist ample and moderate evidence that support not to accept H_0.

Decision**:** Since BF =16.3421 lies in the region (10<BF<30), then we have statistical reason no accept H_0 and conclude that Dirichlet-Multinomial Distribution is a better fit for the buying behavior data. The model is very efficient and consistent for making prospective decision as regards the flavors under investigation.

Precision

If the precision parameter S <1, then the model is adequate and consistent for the buying behavior data, it is not consistent if otherwise.

Since the precision parameter S (0.004487474) <1, then the model is consistent and efficient as compared to Dirichlet. Due to the mixture of Dirichlet and Multinomial, the adequacy of the model has improved significantly, because there is no exaggeration in the parameter vector of the model. Since the parameter of the model is model, this constitute to over 93.2% to the model adequacy and its predictive estimate can be totally relied upon in decision making and prospective planning.

Comparison Criteria for Dirichlet Model [DM] and Dirichlet-Multinomial Model [DMM]

Table 3: Shows comparison criteria for Dirichlet and Dirichlet-Multinomial Model

	Bayes Factor [10<BF<30	**Precision Parameter [S<1]**
DMM	**16.3421[it has a better fit]**	**0.00449[highly consistent and efficient]**
DM	**25.9881**	**1.46129**

Models	**Dirichlet [DM]**	**Dirichlet-multinomial [DMM]**
AIC	**1138.99**	**390.04**
BIC	**1473.23**	**131.453**

From the table above, it can be deduced that the Dirichlet-Multinomial fits the purchase data reasonably well. The Predictive Dirichlet-Multinomial Model stands the test of Akaike information criteria [AIC] and Bayesian information criteria [BIC]. This test states that model with the lowest AIC and BIC will produce best result when fitted to the data sets. Dirichlet-Multinomial is efficient for predicting the expected sales volume and the general behavior of the sales pattern in the prospective marketing period, thus it is also consistent and efficient as compared to Dirichlet Model.

The generalized mean at different proposed sales volume

Table 4: Shows generalized mean for different flavors

Proposed Sales (“000”)	Trophy	Castlelager	Betamalt	Volticwater
10	7.8745765	1.3275625	0.3094557	0.48844053
20	15.74915	2.265515129	2.6189452	1.4652160
30	23.62372	3.9826874	0.9283672	0.9768107
40	31.4983	5.30205	1.9536213	1.237823
50	39.3722882	6.637812	2.5442027	1.547727

From table 4 above, we can deduce that the average sale for all the flavors increases with period, Trophy will have the most sales and the castle lager follows.

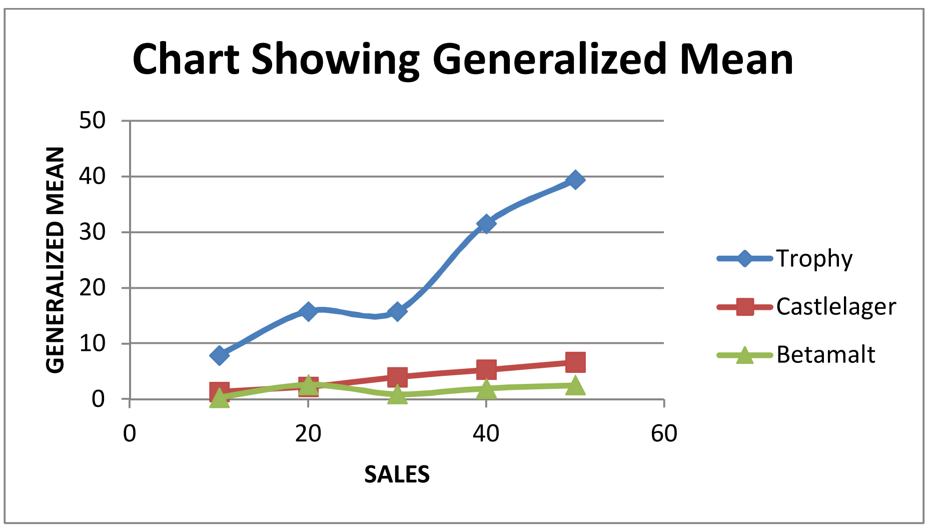

Fig 1: The plot showing the chart for generalized mean

From the fig 1 above, we can deduce that there is increase in the mean sales of Trophy as compare to other flavor under investigation.

Generalized Variance

Table 5: Showing generalized variance

Proposed Sales (“000”)	Trophy	Castlelager	Betamalt	Volticwater
10	11.8706950	9.5052054	0.2862015	0.3040550
20	7.9732630	6.7437828	1.7200977	2.2362936
30	4.3077148	1.715732	1.5343782	1.0494548
40	2.874572	0.859748	2.65151518	3.8644753
50	1.6722381	0.421054	4.07147436	5.9340347

From table 5 above, we can deduce that the variance reduces with increasing in sales for Trophy and Castle lager but shows increase with increasing in sales for other two flavors.

Standard error

Table 6: shows the standard error

Proposed Sales (“000”)	Trophy	Castlelager	Betamalt	Volticwater
10	3.4453875	3.08254019	0.53497804	0.55141182
20	2.82369667	2.59687944	1.31152495	1.49542422
30	2.07550351	1.30985953	1.23870021	1.02442901
40	1.69545628	0.92722597	1.62834738	1.96582687
50	1.29315046	0.64888674	2.01778947	2.43598742

From the table 5, we can deduce the pattern of the generalized variance of Dirichlet-multinomial distribution and their corresponding standard error, the standard error of Trophy and Castle lager reduces with increasing in future sales while the standard errors for voltic water and betamalt maintain almost uniform pattern with little increase at sales of about fifty thousand crates.

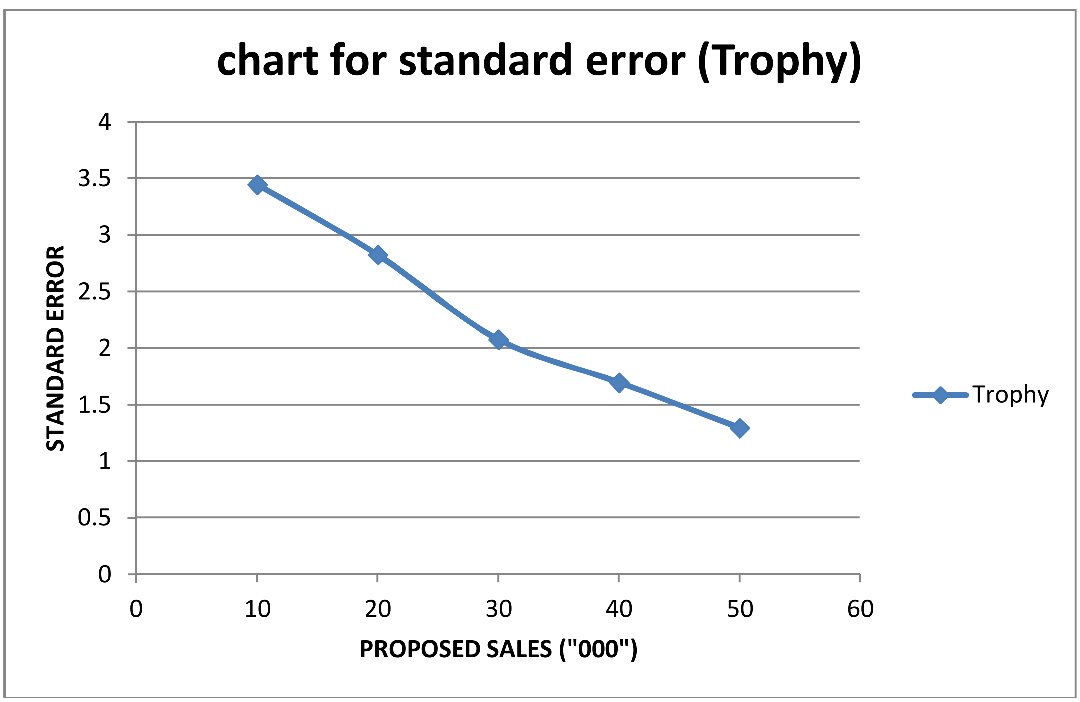

Fig 2: shows the standard error at different proposed sales of Trophy

The fig 2 above shows the downward slope for the standard error. This implies the standard error will be reducing with advancing sales volume of Trophy. There is steepness in the slope of the pattern of movement, the performance of Dirichlet-multinomial is significant in prospective planning and evidence based decision making as regards Trophy.

4. Conclusion

Based on the result of the analysis, Dirichlet-Multinomial model provides a good basis for relying on the predicted values regarding the four products under investigation. The Dirichlet-Multinomial Model can therefore be relied on for accurate, adequate and consistent decision making and planning.

References

1. Agresti, Alan (2013). *Categorical data analysis*. Third edition. Wiley series in Probability and statistics. Wiley.
2. Akomolafe A. A. (2016). Consistency of a Mixture Model of two Different Distributions Approach to the Analysis of Buying Behaviour Data. International Journal of Mathematical Vol. 5(2): 179-187.
3. Akomolafe A. A, Awogbemi C. A and Alagbe S. A (2012). Counting Dropout Rate of Consumers: A case study of durable goods. Mathematical Theory and Modeling. Vol.2(11).
4. Banelis, M Riebe, E and Rangie, C. M (2013). Empirical evidence of repertoire size. Australasian Marketing Journal, 21:59-65.
5. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. The Journal of machine Learning research, 3, 993-1022.
6. Colombo, R. and Morrison, D. A Brand Switching Model with Implications for Marketing.
7. Strategies. Marketing Science. 8 (Winter).
8. Ehrenberg ASC, Goodhardt GJ, Barwise P. T. (1990). Double jeopardy revisited. J Mark; 54:82 – 91.
9. Ehrenberg Andrew S. C, Uncles M. D, Carrie D. (1994). Armed to the teeth: An exercise in brand management. Cranfield (UK): European Case Clearing House.
10. Ehrenberg ASC; Uncles M. D & Goodhardt G. J. (2004). Understanding brand performance measures: using Dirichlet benchmarks. *Journal of Business Research, 57,* 1307-1325.
11. Erik Thorsen (2014). Multinomial and Dirichlet Multinomial Modeling of categorical time series. Mathematical statistics Stockholm University (Bachelor Thesis).
12. Fader, P. and Hardie, B. (1996). Modelling Consumer Choice among SKU's. Journal of Marketing Research.
13. Hariharan, H. S., & Velu, R. P. (1993). On estimating dirichletparametersa comparison of initial values. Journal of statistical computation and simulation, 48(1-2), 47-58.
14. Narayanan, A. (1991). Algorithm AS 266: maximum likelihood estimation of the parameters of the Dirichlet distribution. Journal of the Royal Statistical Society. Series C (Applied Statistics), 40(2),365-374.
15. Ng, K. W., Tian, G. L., & Tang, M. L. (2011). Dirichlet and Related Distributions: Theory, Methods and Applications (Vol. 888). John Wiley & Sons.
16. Paul, S. R., Balasooriya, U., & Banerjee, T. (2005). Fisher information matrix of the Dirichlet-multinomial distribution. *Biometrical Journal, 47*, 230−236.
17. Raj, S. P. (1985). Striking a balance between brand 'popularity' and brand loyalty. *Journal of Marketing, 49*(Winter), 53–59.
18. Rothmane,. D, D. F. Sing and A. R. Templeton (19 74). A model for analysis of population structure. Genetics 79: 943-960.
19. Ronning, G. (1989). Maximum likelihood estimation of Dirichlet distributions. Journal of statistical computation and simulation, 32(4), 215-221.
20. Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 1 9 395-420.
21. Wright, Malcolm and Anne Sharp (1999). "New Brand Effects in a Dirichlet Market," in 28th European Marketing Academy Conference Vol. CD proceedings. Berlin, Germany: Institute of Marketing II, Humboldt-University.
22. Yamamoto, M., Sadamitsu, K., & Mishina, T. (2003). Context modeling using Dirichlet mixtures and its applications to language models. *Information Processing Society of Japan-SIGSLP*, *104*, 29–34.

Appendix

```
DIRICHLET-MULTINOMIAL
#a=trophy
#b=castlelager
#c=betamalt
#d=volticwater
```

```
> a<-c ()
> b<-c ()
> c<-c ()
> d<-c ()
> z<-data.frame (a,b,c,d)
> m<-vglm (cbind (a,b,c,d)~1,dirmultinomial,data=z,trace=TRUE)
VGLM    linear loop 1: loglikelihood = -5690.066
VGLM    linear loop 2: loglikelihood = -5689.545
VGLM    linear loop 3: loglikelihood = -5689.516
VGLM    linear loop 4: loglikelihood = -5689.5066
VGLM    linear loop 5: loglikelihood = -5689.5018
VGLM    linear loop 6: loglikelihood = -5689.4993
VGLM    linear loop 7: loglikelihood = -5689.498
VGLM    linear loop 8: loglikelihood = -5689.4973
VGLM    linear loop 9: loglikelihood = -5689.4969
VGLM    linear loop 10: loglikelihood = -5689.4967
VGLM    linear loop 11: loglikelihood = -5689.4966
> summary ()

Call:
vglm (formula = cbind (a, b, c, d) ~ 1, family = dirmultinomial, data = z, trace = TRUE)

Pearson residuals:
          Min      1Q       Median   3Q       Max
log (prob [,1]/prob [,4]) -0.8588 -0.6949 -0.37001 0.08266 2.365
log (prob [,2]/prob [,4]) -2.0672 -0.6752 0.03811 0.83565 1.540
log (prob [,3]/prob [,4]) -1.5047 -0.3818 -0.12386 0.68682 1.338
logit (phi)  -1.0643 -0.9676 -0.11644 0.20749 2.514

Coefficients:
          Estimate Std. Error z value Pr (>|z|)
(Intercept):1     2.9747   0.2628   -11.320 < 2e-16 ***
(Intercept):2     0.5015   0.1492   3.362 0.000773 ***
(Intercept):3     0.1169   0.1661   0.704 0.481590
(Intercept):4     0.1845   0.1626   1.135 0. 256395
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of linear predictors: 4

Names of linear predictors:
log (prob [,1]/prob [,4]), log (prob [,2]/prob [,4]), log (prob [,3]/prob [,4]), logit (phi)

Dispersion Parameter for dirmultinomial family:       1

Log-likelihood: -5689.497 on 36 degrees of freedom

Number of iterations: 11
> coef ()
(Intercept):1 (Intercept):2 (Intercept):3 (Intercept):4
2.9746778 0.5014967 0.1168609 0.1845160
> coef (.,matrix=TRUE)
          log (prob [,1]/prob [,4]) log (prob [,2]/prob [,4])
(Intercept) 0.184516   0.5014967
          log (prob [,3]/prob [,4]) logit (phi)
(Intercept) 0.1168609 2.974678
```

```
> head (fitted (.))
          a         b         c         d
1 0.2416004 0.3317116 0.2257956 0.2008924
2 0.2416004 0.3317116 0.2257956 0.2008924
3 0.2416004 0.3317116 0.2257956 0.2008924
4 0.2416004 0.3317116 0.2257956 0.2008924
5 0.2416004 0.3317116 0.2257956 0.2008924
6 0.2416004 0.3317116 0.2257956 0.2008924
#dirichlet
> AIC (.,correlate=TRUE)
[1] 11386.99
#dirmultinomial
> AIC (m)
[1]390.04
> BIC (.)
[1] 131.2
> vcov (.)
               (Intercept):1 (Intercept):2 (Intercept):3 (Intercept):4
(Intercept):1 0.0264309161 -0.009787742 -0.0081494998 0.0003234312
(Intercept):2 -0.00978774230.022247135 -0.0095433840 -0.0024824285
(Intercept):3 -0.0081494998 -0.009543384 0.0275743493 0.0009368328
(Intercept):4 0.0003234312 -0.002482429 0.0009368328 0.0690527667
> depvar (.)
          a         b          c          d
1 0.1829268 0.3211382 0.22357724 0.27235772
2 0.2328358 0.2686567 0.23283582 0.26567164
3 0.3726708 0.2080745 0.27950311 0.13975155
4 0.1723077 0.1384615 0.42153846 0.26769231
5 0.1901709 0.4059829 0.19017094 0.21367521
6 0.5051699 0.3412112 0.06646972 0.08714919
7 0.1867089 0.4936709 0.22468354 0.09493671
8 0.2686567 0.2925373 0.19402985 0.24477612
9 0.1559953 0.5017462 0.10826542 0.23399302
10 0.1382979 0.3546099 0.37234043 0.13475177
>
    > sqrt (diag (vcov (.)))
    (Intercept):1 (Intercept):2 (Intercept):3 (Intercept):4
         0.1625759 0.1491547 0.1660553 0.2627789

alpha=c (2.9747,0.5015,0.1169,0.1845)
> #precision
> s<-sum (alpha)
> mean_diri<-((n*alpha)/s)
> mean_diri
[1] 39.372882 6.637812 1.547279 2.442027
> #standard error dirichlet
> a<-((s*s)*(s+1))
> var<-((alpha*(1-alpha)*(n*(n+s))/a))
> var
[1] 31.672238       9.859748   4.071497   5.934034
> #bayes factor
> b<-((gamma (2.9747)*gamma (0.5015)*gamma (0.1169)*gamma (0.1845))/gamma (s))
> b
[1] 16.342
> #PRECISION
> prob=c (0.03317116,0.02417116, 0.02257956,0.02008924)
```

```
> k=4
> s_cap<-(((k-1)/2)/(-sum (mean_diri*(log (prob/mean_diri)))))
> s_cap
[1] 0.004487474
> consumption_habbit<-rdirichlet (49,c (1.06177,1.00756,0.03470))
> plot3d (consumption_habbit,col="red")
>
> g
```

	(Intercept):1	(Intercept):2	(Intercept):3	(Intercept):4
(Intercept):1	1.000000000	-0.40363552	-0.30187157	0.007570687
(Intercept):2	-0.403635515	1.00000000	-0.38531216	-0.063335787
(Intercept):3	-0.301871566	-0.38531216	1.00000000	0.021469350
(Intercept):4	0.007570687	-0.06333579	0.02146935	1.000000000

1/25/2018