Calculation of Theoretical Brand Performance Measures from the Parameters of the Dirichlet Model

Cam Rungie and Gerald Goodhardt

The Dirichlet Model is used in marketing to provide a probability density function for the repeated purchases, by shoppers over a period of time, of the competing brands within a product category. The model is used to analyze a range of measures which report on the performance and competitive status of the brands. This Research Note provides an overview of the algebra and statistical theory of aspects of the Dirichlet Model. Specifically, it provides the algebra for using the parameter of the Dirichlet Model (or estimates) to generate the theoretical brand performance measures. These are marketing metrics such as market share, penetration, 100% loyals, share of category requirements and others. This algebra, which is critical to the further development of the Dirichlet Model, has not been previously documented. Program files to calculate the formulas are provided separately for both Excel and MATLAB.

Keywords: Brand Choice, Buyer Behavior, Choice Models, Data Mining, Marketing Metrics, Dirichlet

Introduction

The Dirichlet Model was developed by Goodhardt, Ehrenberg and Chatfield (Ehrenberg 1959; Goodhardt, Ehrenberg & Chatfield 1984). It is a model for the patterns of repeat purchases of the brands within a product category. It models simultaneously the counts of the number of purchases of each brand, over a period of time. It has been shown to be applicable to many product categories and to have substantial uses, particularly with regards to the analysis of what are known as the 'brand performance measures' (Ehrenberg 1988; Ehrenberg & Uncles 1999; Ehrenberg, Uncles & Goodhardt 2003; Uncles, Ehrenberg & Hammond 1995).

Dirichlet Modelling usually involves applying the model to generate estimates of the brand performance measures including market share, average purchase rate, penetration, purchase frequency, share of category requirements,100% loyals, average portfolio size and repeat rate. These estimates are known as the 'theoreticals' and are compared to the 'observeds' which are calculated directly from the data (Ehrenberg, Uncles & Goodhardt 2003). The procedures for calculating the theoreticals from the parameters of the Dirichlet Model traditionally have been considered to be "fairy complex", which is true (Ehrenberg, Uncles & Goodhardt 2003) and are not well documented. Consequently, the Dirichlet Model is usually applied through packaged software (Hewitt 1990; Kearns 2000; Uncles 1989). However, as the use and development of the model expands, there is a requirement to document the functional forms, and their derivations, for calculating the brand performance measures from the parameters of the Dirichlet Model. This has not been previously done and is presented in this paper.

This Technical Note presents the algebra but not the underlying theoretical structure of the model. In particular, there is no discussion here of the underlying gamma distributions (Ehrenberg, Uncles & Goodhardt 2003; Goodhardt, Ehrenberg & Chatfield 1984; Rungie &

Laurent 2003b; Rungie, Laurent & Stern 2003). Also, this Research Note does not discuss methods for estimating the parameters of the Dirichlet Model (Ehrenberg 1988; Goodhardt, Ehrenberg & Chatfield 1984; Rungie 2004).

The notation used in this Research Note differs slightly from the notation used in the original documentation of the model. This has been to move the notation closer to statistical tradition where parameters are given Greek symbols and to avoid duplication in the use of symbols. The variations in the notation, where they occur, are noted.

Many of the expressions in this research note use the mathematical function gamma. This is given the symbol Γ. It is a function available in Excel as the log of gamma. The Excel function is GAMMALN. Where it is necessary to convert this to the gamma function then take the exponential. Thus the Excel gamma function is Γ=EXP(GAMMALN). A property of the gamma function is that $r\Gamma(r) = \Gamma(r+1)$ and $r! = \Gamma(r+1)$. Do not confuse the gamma function, Γ, with the parameter, γ, nor with the gamma distribution.

The purpose of this Research Note is to document the algebra of the model in a form available to researchers working with computer programs such as Excel. Many researchers fit the model to data by using existing packaged software. This is certainly appropriate. However, should the researcher wish to explore the model further then, with this Research Note, it is possible to access, understand and program the formulas for the theoretical brand performance measures.

Probability Density Function

The Dirichlet Model is a probability density function. It specifies the distribution of purchases by a population of shoppers of each of the brands within a product category over a specified period of time. The category might be, for example, detergents, the period of time a year, the population all households (called shoppers) in France, and the brands the set of about twenty brands of detergents available in supermarkets in that country. The data in this example would record for each household in the sample the purchases of each of the brands over the specified year. The data is multivariate in that there are several brands. It models the count of the purchases where one purchase of one SKU counts as one purchase of the relevant brand, regardless of other characteristics of the SKU which might be, for example, pack size, product formulation, flavor etc. It is discrete, integer and nonnegative because purchases are counts which are whole numbers and, by definition, can't be negative.

The Dirichlet Model is the combination of two probability density functions, the negative binomial distribution (NBD) and the Dirichlet multinomial distribution (DMD). In the Dirichlet Model the category purchase rate is assumed to have a NBD over the population of shoppers. The NBD is derived through specifying that each shopper's purchases for the category follow a Poisson process based on a category propensity. Over the population of shoppers these propensities have a gamma distribution. The purchases of the individual brands are assumed to have a DMD which is conditional on the category purchase rate. The DMD is derived through specifying that each shopper's purchases of the brands (conditional on the category purchase rate) follow a multinomial process based on brand choice probabilities. Over the population of shoppers, these probabilities have a multivariate

Dirichlet distribution. The two distributions the NBD and the DMD are assumed to be independent. In addition their parameters are assumed to have no associations.

Over the population of shoppers let the category purchase rate be a random variable *K*. (This notation differs from the original notation developed for the Dirichlet Model in which the symbol K was used for another purpose.) The Dirichlet Model assumes that the category purchase rate has a NBD; i.e. *K* has a NBD. Over the population of shoppers *K* is a random variable for which an individual shopper has an observation category purchase rate of *k*.

The NBD has two parameters which are both positive:

- the shape[1] parameter γ (In the original notation of the Dirichlet Model this parameter was given the Arabic symbol K. This should not be confused with the random variable referred to above as *K*.)
- the scale[2] parameter (which also influences the shape) β.

The probability density function for the NBD is (Johnson, Kotz & Kemp 1993):

Equation 1 $$f_{\gamma,\beta}(k)=\frac{\Gamma(\gamma+k)}{\Gamma(\gamma)k!}\frac{\beta^k}{(1+\beta)^{(\gamma+k)}} \qquad \text{for } k=0,1,2,...$$

Let there be *h* brands. Over the population of shoppers let the purchase rate of each brand be a set of random variables $R_1,R_2,...,R_h$. Then, the sum of these purchase rates is the category purchase rate; $R_1+R_2+...+R_h = K$. The Dirichlet Model assumes that the purchases of each brand, conditional on the category purchase rate, have a DMD; i.e. $R_1,R_2,...,R_h$, conditional on *K,* has a DMD. Over the population of shoppers $R_1,R_2,...,R_h$ are random variables for which an individual shopper has observed purchase rates for the brands of $r_1,r_2,...,r_h$.

The DMD has *h* parameters, one for each brand. These are $\alpha_1,\alpha_2,...,\alpha_h$ where each is positive.

The probability density function for the DMD is (Johnson, Kotz & Balakrishnan 1997)

Equation 2 $$f_{\alpha_1,\alpha_2,...,\alpha_h}(r_1,r_2,...,r_h \mid r_1+r_2+...+r_h=k)=\frac{\Gamma\left(\sum_{j=1}^{h}\alpha_j\right)k!}{\Gamma\left(\sum_{j=1}^{h}\alpha_j+k\right)}\prod_{j=1}^{h}\frac{\Gamma(\alpha_j+r_j)}{r_j!\Gamma(\alpha_j)}$$

The Dirichlet Model is a probability density function for the purchases of all brands in a product category over a period of time. The model combines these two distributions; the NBD and the DMD. The probability density function for the Dirichlet model is given by:

---

[1] Referred to as the *K* parameter in the original Goodhardt Ehrenberg and Chatfield 1984 paper and often referred to as the α parameter in the statistical literature.

[2] Referred to as the *A* parameter in the original Goodhardt Ehrenberg and Chatfield 1984 paper and often referred to as the β or λ parameter in the statistical literature $\beta=1/\lambda$.

Equation 3 $$f_{\gamma,\beta,\alpha_1,\alpha_2,\ldots,\alpha_h}(r_1, r_2,\ldots, r_h) = f_{\gamma,\beta}(k) f_{\alpha_1,\alpha_2,\ldots,\alpha_h}(r_1, r_2,\ldots, r_h \mid r_1 + r_2 + \ldots + r_h = k)$$

The number of brands in a product category, h, can typically be of the order of 20 to 40 but it can be as low as two or as high as 150 or more. It is rare for the individual brand alphas to be much greater than one and they are often in the range of 0.7 to 0.01 and less. The value of γ is of the order of 0.5 to 1 but it can be higher or lower. It is greater for product categories which are purchased by many shoppers. The value of β is proportional to the length of the time period over which the data is collected and can be of the order of 0.1 to 50.

Estimating the Parameters of the Dirichlet Model

The data should be aggregated such that for each shopper and for each brand there is a count which is the purchase rate. Shoppers who do not purchase (non buyers) should be included. If there are h brands and n shoppers there will be nh counts. It is usual to set this data up in a table with n rows (one per shopper) and h columns (one per brand).

There exists a range of methods for estimating the parameters of the Dirichlet Model. The developers of the model used a combination of the method of zeros and means and the method of moments (Ehrenberg 1988; Goodhardt, Ehrenberg & Chatfield 1984). This is quick and relatively simple. Likelihood theory provides an alternate method which is also now relatively easy to apply (Rungie 2004). The original method has the excellent property that it generates estimates of the market shares for the brands which exactly match the observed market shares. In addition the original method is well matched to the traditional use of the Dirichlet Model in estimating the brand performance measures. The likelihood method is more efficient (less sampling variation) and is well matched to extensions of the Dirichlet Model such as its development into a generalised model (Rungie, Laurent & Stern 2003).

The methods documented and demonstrated in this Research Note do not rely on any one method of estimation for the parameters of the Dirichlet Model. The model has a set of h+2 parameters which are γ, β, α_1, α_2,…, α_h . In order to use the methods presented here these parameters can be estimated using the traditional approach, likelihood theory or any other approach.

Brand Performance Measures

One of the substantive applications of the Dirichlet Model is the estimation of the brand performance measures. These are a series of measures which relate to the characteristics of individual brands. It is customary to calculate the brand performance measures from the raw data, these are known as ‘observed’, and then to also estimate the measures from the Dirichlet Model, these are known as ‘theoretical’ (Ehrenberg, Uncles, & Goodhardt 2003). Each theoretical value is a probability statement.

Definitions

The bulk of the brand performance measures relate to purchases over a fixed period of time such as a month or a year.

Shopper (respondent, panellist, household etc). The unit for which the purchase data is recorded. The shoppers are all the potential buyers regardless of whether or not they buy. A shopper can have a purchase rate of zero.

Buyer. A shopper who makes at least on purchase. Their purchase rate is greater than zero. An individual shopper may be a buyer for one brand but and not another.

Purchase Rate. The count of the quantity purchased by the shopper over the specified time period. Each shopper has a separate purchase rate for the category and for each brand. The category purchase rate is the sum of the brand purchase rates.

Average Purchase Rate. The purchase rate averaged over all shoppers. There is a separate average purchase rate for the category and for each brand. The category average purchase rate is the sum of the brand average purchase rates.

Penetration[3]. The proportion of shoppers who are buyers. There is a separate penetration for the category and for each brand.

Purchase Frequency[4]. The purchase rate averaged over buyers. The average purchase rate for those shoppers who buy the brand. There is a separate purchase frequency for the category and for each brand.

100% Loyals (Sole Buyers)[5]. For each brand, the proportion of buyers who only buy the brand.

Share of Category Requirements (SCR). In Dirichlet modelling SCR is defined as:

$$\frac{\text{Total Purchases of brand } j}{\text{Total Purchases of the category by buyers of brand } j}$$

Average Portfolio Size. Average number of brands purchased by buyers of the category.

---

[3] Referred to as the *B*, for the category, and *b*, for a specified brand, in the original Goodhardt Ehrenberg and Chatfield 1984 paper.

[4] Referred to as the *W*, for the category, and w, for a specified brand, in the original Goodhardt Ehrenberg and Chatfield 1984 paper.

[5] The definitions for measures such as 100% loyals and SCR include the shopper who only buys once and so is 100% loyal.

All of the above definitions refer to purchases over a fixed period of time. Different shoppers make different numbers of purchases. This is not a characteristic used in the specification of the remaining brand performance measures.

Repeat Rate. The proportion of buyers of a brand at the last purchase occasion who then buy the same brand again at the next purchase occasion. A purchase occasion is the event where the shopper makes a purchase from the category. This is an intuitive measure of loyalty. It records how much a brand hangs onto its buyers.

Dirichlet *S* and Category Polarization. Two statistical measure of loyalty which are discussed below.

Summary of the Formulas for the Theoretical Brand Performance Measures

The derivation the formulas for the theoretical brand performance measures are given in the appendix

Average Purchase Rate for brand *j*

$$= \frac{\alpha_j \gamma\beta}{S}$$

Market Share for brand *j*

$$\mu_j = \frac{\alpha_j}{S}$$

Penetration for brand *j*

$$= \sum_{k=1}^{\infty} f_{\gamma,\beta}(k)\left(1 - \frac{\Gamma(S)\Gamma(S-\alpha_j+k)}{\Gamma(S+k)\Gamma(S-\alpha_j)}\right)$$

Purchase Frequency for brand *j*

$$= \frac{\text{Average Purchase Rate for brand } j}{\text{Penetration for brand } j}$$

100% Loyals for brand *j*

$$= \frac{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k)\left(\frac{\Gamma(S)\Gamma(\alpha_j+k)}{\Gamma(S+k)\Gamma(\alpha_j)}\right)}{\text{Penetration for brand } j}$$

Share of Category Requirements for brand *j*

$$= \frac{\text{Average Purchase Rate for brand } j}{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k)k\left(1 - \frac{\Gamma(S)\Gamma(S-\alpha_j+k)}{\Gamma(S+k)\Gamma(S-\alpha_j)}\right)}$$

Average Portfolios Size

$$= \frac{\sum_{j=1}^{h} \text{Penetration for brand } j}{\text{Penetration for the category}}$$

Repeat Rate for brand *j*

$$\rho_j = \frac{\alpha_j + 1}{S+1}$$

Dirichlet *S*

$$S = \sum_{j=1}^{h} \alpha_j$$

Category Polarization, φ

$$\varphi = 1/(S+1)$$

And, where market share = μ, and repeat rate = ρ for any one brand, then

$$\rho = \mu + \varphi - \mu\varphi$$

Some of the formulas above specify infinite summations. These can be numerically approximated with finite summations. This is discussed in the appendix.

Program Files

Two program files accompany this paper. They will be of assistance to researchers wanting to explore the brand performance measures further. These are for the advanced user. Those researchers simply wanting to fit the Dirichlet Model to data would be far better using one of the established software packages.

Excel

An Excel workbook accompanies this research note. It demonstrates each of the steps in using estimates of the parameters of the Model to calculate the theoretical brand performance measures. The name of the workbook is dirichletbpm.xls.

The Excel workbook has three sections:

Section	Work Sheets
Inputs	Inputs and Results
Results	Inputs and Results
Detailed working	All other work sheets

The workbook contains five worksheets. The function of each sheet is described below.

Introduction

Documentation

Inputs and Results

Enter the estimates of the parameter values in this work sheet. This work sheet presents the brand performance measures. An example is given in Table 1. Those measures which have simple closed forms are calculated in this work sheet. Those that have more complex forms requiring numerical approximations draw interim calculations from the three supporting work sheets (1) NA Brand A, (2) NA Brand B and (3) NA Brand C.

Table 1. An example of the results provided by the attached Excel file.

	Category	Brands Brand A	Brand B	Brand C
INPUTS				
Gamma (also known as K)	**2.1000**			
Beta (also known as A)	**0.7000**			
Brand Alphas		**1.1000**	**0.5000**	**0.4000**
RESULTS				
Market Share	100%	55%	25%	20%
Average Purchase Rate	1.47	0.81	0.37	0.29
Penetration	67%	46%	24%	20%
Purchase Frequency	2.19	1.77	1.52	1.48
100% Loyals		61%	44%	42%
Share of Category Requirements		73%	58%	56%
Average Portfolio Size	1.34			
Repeat Rate		70%	50%	47%
Category S (sum of brand alphas)	2.00			
Category Polarization	0.33			

NA Brand A

Numerical Approximations for Brand A. This work sheet generates the interim calculations for the penetration, 100% loyals and the share of category requirements. This involves a sum over many values of the category purchase rate, k. Each row is a value of k starting at $k = 1$. There must be enough rows to ensure that the summations are sufficiently accurate. It is clear that successive rows contribute less and less to the summations and it is, in this case, acceptable to stop the summation after about 100 rows.

NA Brand B   Numerical Approximations for Brand B.

NA Brand C   Numerical Approximations for Brand C.

For extreme values of the parameters it is possible that the numerical approximations in the Excel file will not be sufficiently accurate. This is because an infinite summation is replaced with a summation which has 1000 terms. Under these conditions a warning is automatically given.

MATLAB

The Excel file provided with this research note does not require the user to have access to MATLAB. A researcher who wishes to work extensively with the theoretical brand performance measures will find Excel too limiting. MATLAB is one of many programming platforms which are more suitable to extensive automation of calculations. A MATLAB file also accompanies this research note (known as an m file). Unlike the Excel file the MATLAB file can process any number of brands. Copy the file into the 'work' directory within the MATLAB program directory on your computer. Then, for help on using this file, in MATLAB, type 'help dirichletbpm'.

Figure 1. Help Information for the MATLAB m File

```
 dirichletbpm(parcat,alpha)
    e.g.                    dirichletbpm([1 1],[1 .5 .5])
    Function
        Calculates theoretical brand performance measures from the
parameters of the Dirichlet Model.
    Inputs
        parcat              [gamma, beta] parameters for the NBD; ie [K A]
        alpha               [brandalpha1 brandalpha2 ....]
    Outputs
        Columns             Brands and then category is the final column.
        Rows        1       Market Share
                    2       Average Purchase Rate
                    3       Penetration
                    4       Purchase Frequency
                    5       100% Loyals
                    6       Share of Category Requirements
                    7       Average Portfolio Size (category only)
                    8       Repeat Rate
                    9       Brand alphas (i.e. input) and category S
                    10      Category polarization (category only).
    Author: Cam Rungie 2003. University of South Australia.
cam.rungie@unisa.edu.au
```

The m file as two inputs. Both are vectors. The first, parcat, has two elements which are the parameters of the NBD, γ and β. The second, alpha, contains α_1, α_1,… α_h and has as many elements as there are brands.

In MATLAB type in: dirichletbpm([.21 .7],[1.1 .5 .4])

In this case parcat = [0.21 0.7] and alpha = [1.1 0.5 0.4]. The result is shown in Figure 2. The output has no print statements because the m file has been written to be a subroutine for other program files. The columns and rows of the results are explained in Figure 1.

Figure 2. Output from the MATLAB m file

```
ans =

    0.5500    0.2500    0.2000    1.0000
    0.0809    0.0367    0.0294    0.1470
    0.0635    0.0308    0.0249    0.1054
    1.2733    1.1939    1.1807    1.3941
    0.8250    0.7260    0.7107    1.0000
    0.8460    0.7509    0.7358    1.0000
         0         0         0    1.1302
    0.7000    0.5000    0.4667    1.0000
    1.1000    0.5000    0.4000    2.0000
         0         0         0    0.3333
```

For extreme values of the parameters the numerical approximations in the Excel file automatically adjust. The infinite summations are estimated using finite summations. The number of terms in the finite summation will be increased automatically. For extreme values of the parameters the execution will be slower.

Conclusions

This Research Note has documented the array of formulas used in the Dirichlet Model; the fundamental probability density functions and the expressions for the brand performance measures. The formulas presented here will be of use to researchers who wish to experiment with and get to know the model by programming it themselves in software such as Excel and MATLAB.

References

Bass FM, Jeuland A & Wright GP (1976). Equilibrium stochastic choice and market penetration theories: Derivations and comparisons. Management Science, 22 (10), 1051-63.

Brown B, Rungie CM, Laurent G & Rudrapatna S (2003). A Standard Error Estimator for the Polarization Index: Assessing the Uncertainty in Loyalty. Adelaide, Australia: School of Marketing, University of South Australia.

Chatfield C & Goodhardt GJ (1970). The Beta-Binomial Model for consumer purchasing behaviour. Journal of the Royal Statistical Society: Series C (Applied Statistics), 19 (3), 240-50.

Ehrenberg ASC (1959). The pattern of consumer purchases. Applied Statistics, 8, 26-41.

Ehrenberg ASC (1988). Repeat-buying, Facts, Theory and Applications. New York: New York: Oxford University Press.

Ehrenberg ASC & Uncles MD (1999). Understanding Dirichlet-Type Markets. London: Centre for Research in Marketing, South Bank University.

Ehrenberg ASC, Uncles MD & Goodhardt GG (2003). Understanding brand performance measures: Using Dirichlet benchmarks. Journal of Business Research (forthcoming).

Fader PS & Schmittlein DC (1993). Excess behavioral loyalty for high-share brands: Deviations from the Dirichlet Model for repeat purchasing. Journal of Marketing Research, 30 (4 November).

Goodhardt GJ, Ehrenberg ASC & Chatfield C (1984). The Dirichlet: A comprehensive model of buying behaviour. Journal of the Royal Statistical Society, Series A (General), 147 (Part 5), 621-55.

Hewitt I (1990). DIPRED: A Program for Predicting Buyer Behavior from the Dirichlet Model (Version 2.6). Port Sunlight, UK: Unilever Research Laboratory.

Johnson NL & Kotz S (1972). Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley & Sons, Inc.

Johnson NL, Kotz S & Balakrishnan N (1994a). Continuous Univariate Distributions, Volume 1 (Second ed.). New York: John Wiley & Sons, Inc.

Johnson NL, Kotz S & Balakrishnan N (1994b). Continuous Univariate Distributions, Volume 2 (Second ed.). New York: John Wiley & Sons.

Johnson NL, Kotz S & Balakrishnan N (1997). Discrete Multivariate Distributions. New York: John Wiley & Sons, Inc.

Johnson NL, Kotz S & Kemp AW (1993). Univariate Discrete Distributions (Second ed.). New York: John Wiley & Sons, Inc.

Kalwani MU (1980). Maximum Likelihood Estimation of zero-order models given variable numbers of purchases per household. Journal of Marketing Research, XVII, 547-51.

Kalwani MU & Morrison DG (1977). A parsimonious description of the Hendry System. Management Science, 23 (January), 467-77.

Kalwani MU & Morrison DG (1980). Sample size requirements for zero-order models. Journal of Marketing Research, XVII, 221-27.

Kearns, Zane (2000). Dirichlet No Solver Software. New Zealand: Massey University.

Rungie CM (2003). How to estimate the parameters of the Dirichlet Model using Likelihood Theory in Excel. Marketing Bulletin, 14, Technical Note 1. http://marketing-bulletin.massey.ac.nz.

Rungie CM & Laurent G (2003a). Repeated Binary Logit (RBL): Fitting Explanatory Variables to the Beta Binomial Distribution. Adelaide: School of Marketing, University of South Australia.

Rungie CM & Laurent G (2003b). Repeated Revealed Preference Methods: Dirichlet Multinomial Regression. Adelaide: School of Marketing, University of South Australia.

Rungie CM, Laurent G & Stern P (2003)., Modeling Long-Run Dynamic Markets. Adelaide: School of Marketing, University of South Australia.

Sabavala, Darius J. and Donald G. Morrison (1977), "A Model of TV Show Loyalty," Journal of Advertising Research, 17 (6), 35-43.

Uncles MD (1989). BUYER: Buyer Behavior Software. London: London Business School (refer to web site).

Uncles MD, Ehrenberg ASC & Hammond K (1995). Patterns of buyer behavior: Regularities, models and extensions. Marketing Science, 14 (3, 2 of 2), G71-G78.

Appreciation

The authors wish to thank Helen Monroe and Carl Driesener for their assistance in the development of this paper.

Cam Rungie is in the School of Marketing, University of South Australia, and Gerald Goodhardt is in the Centre for Research in Marketing, South Bank University

Appendix A: Derivation of the Formulas for Theoretical Brand Performance Measures

Statistical Properties of the Dirichlet Model

The properties of the NBD and DMD are useful in developing applications for the Dirichlet Model (Ehrenberg 1959; Ehrenberg 1988; Goodhardt, Ehrenberg & Chatfield 1984; Johnson & Kotz 1972; Johnson, Kotz & Balakrishnan 1994a; Johnson, Kotz & Balakrishnan 1994b; Johnson, Kotz & Balakrishnan 1997; Johnson, Kotz & Kemp 1993).

NBD

$$E[K] = \gamma\beta = \text{average purchase rate for the category}$$

$$\Pr(K=0) = \frac{1}{(1+\beta)^{\gamma}}$$

DMD

$$E[R_j \mid k] = \frac{\alpha_j k}{S} = \text{average purchase rate for brand } j$$

Marginal DMD: If $R'_j = K - R_j$ then R_j and R'_j jointly also have a DMD with parameters α_j and $S - \alpha_j$.

Thus

$$\Pr(R_j = 0 \mid k) = \frac{\Gamma(S)\Gamma(S-\alpha_j+k)}{\Gamma(S+k)\Gamma(S-\alpha_j)}$$

$$\Pr(R_j = k \mid k) = \frac{\Gamma(S)\Gamma(\alpha_j+k)}{\Gamma(S+k)\Gamma(\alpha_j)}$$

Performance Measures

Category Purchase Rate

Average Purchase Rate for the category $= E[K] = \gamma\beta$

Penetration for the category $= \Pr(K>0) = 1 - \Pr(K=0) = 1 - \dfrac{1}{(1+\beta)^{\gamma}}$

Purchase Frequency for the category

$$= \frac{\text{Average Purchase Rate for the category}}{\text{Penetration for the category}}$$

$$= \frac{E[K]}{\Pr(K > 0)}$$

Restatements of the Definitions of the Brand performance Measures

The definitions for the brand performance measures can be rewritten as probability statements:

Average Purchase Rate for brand $j = E[R_j]$

Penetration for brand $j = \Pr(R_j > 0)$

Purchase Frequency for brand j

$$= \frac{\text{Average Purchase Rate for brand } j}{\text{Penetration for brand } j}$$

$$= \frac{E[R_j]}{\Pr(R_j > 0)}$$

100% Loyals for brand j

$$= \frac{\text{Probability of a shopper buying brand } j \text{ only}}{\text{Probability of a shopper buying brand } j}$$

$$= \frac{\Pr(R_j = K > 0)}{\Pr(R_j > 0)}$$

Define a new variable:

$$K' = K \text{ if } R_j > 0 \quad \text{otherwise} \quad K' = 0$$

Share of Category Requirements for brand j

$$= \frac{E[R_j]}{E[K']}$$

Results Conditional on Category Purchase Rate

In order to examine the brand performance measures for a single brand it is useful to first consider the case where the category purchase rate is fixed; $K = k$. These are the theoretical brand performance measures for brand j conditional on the category purchase rate being k.

$$\text{Average Purchase Rate } = E[R_j \mid k] = \frac{\alpha_j k}{S}$$

$$\text{Market Share } = \frac{E[R_j \mid k]}{k} = \frac{\alpha_j}{S}$$

Penetration

$$\begin{aligned} &= \Pr(R_j > 0 \mid k) \\ &= 1 - \Pr(R_j = 0 \mid k) \\ &= 1 - \frac{\Gamma(S)\Gamma(S - \alpha_j + k)}{\Gamma(S + k)\Gamma(S - \alpha_j)} \end{aligned}$$

100% Loyal interim calculation

$$\Pr(R_j = k \mid k) = \frac{\Gamma(S)\Gamma(\alpha_j + k)}{\Gamma(S + k)\Gamma(\alpha_j)}$$

Share of Category Requirements interim calculation

$$E[K \mid k] = k$$

Therefore

$$E[K' \mid k] = k \Pr(R_j > 0 \mid k) = k\left(1 - \frac{\Gamma(S)\Gamma(S - \alpha_j + k)}{\Gamma(S + k)\Gamma(S - \alpha_j)}\right)$$

Results Incorporating the Distribution for the Category Purchase Rate

These conditional results can be converted to results which are not conditional on k by mixing over the probability density function for the category purchase rate; i.e. by mixing over the NBD. These are the unconditional results. These are the theoretical brand performance measures for brand j.

In some cases mixing over the NBD simplifies to a closed form; e.g.

Average Purchase Rate for brand $j = E[R_j] = \sum_{k=1}^{\infty} f_{\gamma,\beta}(k) E[R_j \mid k] = \frac{\alpha_j \gamma\beta}{S}$

Market Share for brand $j = \sum_{k=1}^{\infty} f_{\gamma,\beta}(k) \frac{E[R_j \mid k]}{k} = \frac{\alpha_j}{S}$

Alternatively, Market Share $= \Pr(R_j = 1 \mid k = 1) = \frac{\alpha_j}{S}$

However, for the remaining cases the expressions are infinite sums which can be numerically estimated.

Penetration for brand j

$$
\begin{aligned}
&= \Pr(R_j > 0) \\
&= \sum_{k=1}^{\infty} f_{\gamma,\beta}(k) \Pr(R_j > 0 \mid k) \\
&= \sum_{k=1}^{\infty} f_{\gamma,\beta}(k) \left(1 - \frac{\Gamma(S)\Gamma(S - \alpha_j + k)}{\Gamma(S+k)\Gamma(S-\alpha_j)} \right)
\end{aligned}
$$

Also

Purchase Frequency for brand j

$$
= \frac{E[R_j]}{\Pr(R_j > 0)}
$$

100% Loyals for brand j

$$
\begin{aligned}
&= \frac{\Pr(R_j = K > 0)}{\Pr(R_j > 0)} \\
&= \frac{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k) \Pr(R_j = k \mid k)}{\Pr(R_j > 0)} \\
&= \frac{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k) \left(\frac{\Gamma(S)\Gamma(\alpha_j + k)}{\Gamma(S+k)\Gamma(\alpha_j)} \right)}{\Pr(R_j > 0)}
\end{aligned}
$$

Share of Category Requirements for brand j

$$= \frac{E\lfloor R_j \rfloor}{E[K']}$$

$$= \frac{E\lfloor R_j \rfloor}{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k) E[K' \mid k]}$$

$$= \frac{E\lfloor R_j \rfloor}{\sum_{k=1}^{\infty} f_{\gamma,\beta}(k) k \left(1 - \frac{\Gamma(S)\Gamma(S - \alpha_j + k)}{\Gamma(S + k)\Gamma(S - \alpha_j)}\right)}$$

The above formulae require numerical approximations because it is not possible to sum to infinity. It is necessary to sum over values for the category purchase rates starting from one and ranging through to an upper limit k_s which is large enough to generate accurate estimates. A sensible approach is to select k_s such that $\sum_{k=1}^{k_s} f_{\gamma,\beta}(k) > 0.99999$. This is equivalent to keeping summing from k = 0 until a level of k is reached where the brand performance measures will not change.

Other Results Not Reliant on Distribution of the Category Purchase Rate

Average Portfolio Size

Define a new variable such that:

$R'_j = 1$ if $R_j > 0$ otherwise $R'_j = 0$

Then, as an alternate specification,
Penetration for brand j

$$= E\lfloor R'_j \rfloor$$

Also the average portfolio size amongst all shoppers

$$= E[R'_1 + R'_2 + \ldots + R'_h]$$

$$= E[R'_1] + E[R'_2] + \ldots + E[R'_h]$$

$$= \sum_{j=1}^{h} \text{Penetration for brand } j$$

Finally, the average portfolio size for all buyers

$$= \frac{\text{Average Portfolio Size for all shoppers}}{\text{Category penetration}}$$

Thus, Average Portfolios size over all buyers

$$= \frac{\sum_{j=1}^{h} \text{Penetration for brand } j}{\text{Penetration for the category}}$$

Repeat Rate

Repeat Rate for brand *j* is defined to be

$$= \frac{\Pr(R_j = 2 \mid k = 2)}{\Pr(R_j = 1 \mid k = 1)}$$

Using the probability density function for the DMD:
Repeat Rate for brand *j*

$$= \frac{\alpha_j + 1}{S + 1}$$

Polarization

Repeat rate is an intuitive measure of loyalty. It is a measure of how much a brand hangs onto its customers. However, repeat rate is confounded with market share. Generally, the larger the market share the larger the repeat rate. Polarization is a measure of loyalty which is the repeat rate standardized for the market share.

Define three new symbols:

(1) Let the repeat rate for brand *j* be ρ_j

(2) Let the market share for brand *j* to be μ_j where it is shown above that $\mu_j = \alpha_j/S$

(3) Let polarization, φ be $\varphi = 1/(S+1)$. Then $0 < \varphi < 1$

Then $\rho_j = \mu_j + \varphi - \mu_j \varphi$

Without loss of generalization this can be written as

$$\rho = \mu + \varphi - \mu\varphi$$

Thus, the repeat rate is a function of market share and polarization. The larger the market share the larger the repeat rate.

Also the larger the polarization, φ, the greater the repeat rate and the greater the loyalty. When $\varphi = 1$ there is complete loyalty. When $\varphi = 0$ there is no loyalty(Brown et al. 2003; Rungie and Laurent 2003a).

Also

$$\varphi = (\rho - \mu)/(1 - \mu)$$

Thus, Polarization is the repeat rate standardized for the market share.

Polarization φ is first described as a characteristic of the beta distribution by Sabavala and Morrison (Sabavala & Morrison 1977), and is further investigated by Kalwani and Morrison (Kalwani 1980; Kalwani & Morrison 1980). It relates to other measures of loyalty including Hendry's k (Kalwani & Morrison 1977), Bass's θ (Bass, Jeuland & Wright 1976), Chatfield and Goodhardt's work on the beta binomial distribution (Chatfield & Goodhardt 1970) and Goodhardt and Ehrenberg's Dirichlet S statistic (Goodhardt, Ehrenberg & Chatfield 1984). Under Dirichlet Model conditions $\varphi = \theta = 1/(1+S) = 1-k$. Note that, under traditional Dirichlet Modeling, there is one polarization for the whole category; φ is constant across brands. However, in other traditions of modelling, polarization generally refers to choices between two alternatives. In this manner it is particularly useful as it can describes loyalty in the marginal choice between one specified brand and all others in the category. Thus, there can be one polarization for the category and one for each brand. Both can be estimated by using likelihood maximization based on the DMD applied first to all brands and then separately to each brand (the data is the marginal choice between the brand and all other brands combined). This supplies a method of evaluating one condition of the Dirichlet Model, that φ and loyalty is constant across brands. Also, it is an excellent method for the study of deviations from the model in the form of variations in loyalty and of excess loyalty for big brands (Fader & Schmittlein 1993).