How to Estimate the Parameters of the Dirichlet Model using Likelihood Theory in Excel

Cam Rungie

This technical note provides a description of the algebra and an overview of the statistical theory of the Dirichlet Model. It gives a description of the use of likelihood theory to estimate the parameters of the model. The model traditionally has been estimated using the method of zeros and ones and marginal moments. Likelihood theory provides an alternate which is now, in general, the standard for modeling. It has different statistical properties, the strength and weaknesses of which are discussed in the paper. A description and example is given of how to use Excel to generate the likelihood estimates. This will allow greater understanding of the application of the likelihood method to the Dirichlet Model and greater diversity for the model.

Keywords: Dirichlet Model, Negative Binomial Distribution, Likelihood, Discrete Choice, Revealed Preference, Repeated Choice, Count Data.

Introduction

The Dirichlet Model was developed by Goodhardt, Ehrenberg and Chatfield (Ehrenberg 1959; Goodhardt, Ehrenberg et al. 1984). It is a model for the patterns of repeat purchases of the brands within a product category. 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; Uncles, Ehrenberg et al. 1995; Ehrenberg & Uncles 1999; Ehrenberg, Uncles et al. 2003; Rungie, Goodhardt et al. 2003).

The Dirichlet Model is usually applied using packaged software (Uncles 1989; Hewitt 1990; Kearns 2000). The purpose of this document is to layout the algebra and procedures for estimating the parameters of the model. This technical note has been written to demonstrate, to those researchers who are inclined, how to estimate the parameters of the Dirichlet Model in Excel, or a similar package, using likelihood theory.

The notation used in this technical note differs slightly from the notation used in the original documentation of the model and in the statistical literature. 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. Specifically:

Construct	Original Notation	Notation Used Here
Random variable for the Category Purchase Rate		K
Shape parameter for the distribution of the category purchase rate (negative binomial distribution)	K	γ
Scale parameter for the distribution of the category purchase rate (negative binomial distribution)	A	β
Random variables for the purchases of individual brands		$R_1, R_2, R_3,\ldots$
Parameters for the distribution of purchases of the individual brands conditional on the category purchase rate (Dirichlet multinomial distribution)	$\alpha_1, \alpha_2, \alpha_3,\ldots$	$\alpha_1, \alpha_2, \alpha_3,\ldots$

This technical note presents the algebra but not the underlying structure of the model. In particular there is no discussion here of the interpretation of the parameters of the model and the underlying latent gamma distributions (Goodhardt, Ehrenberg et al. 1984; Ehrenberg, Uncles et al. 2003; Rungie, Laurent et al. 2003; Rungie & Laurent 2003).

Dirichlet Modelling usually involves applying the model to generate estimates of the brand performance measures such as purchase rate, market share, penetration, purchase frequency, share of category requirements and 100% loyals. These estimates are known as the 'theoreticals' and are compared to the 'observeds' which are calculated directly from the data (Ehrenberg, Uncles et al. 2003). A separate research note documents the procedures for converting the parameters into the 'theoretical' estimates of the brand performance measures (Rungie, Goodhardt et al. 2003).

Many of the expressions in this technical 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 paper starts with a discussion of why a newer alternate method of estimation for the Dirichlet Model is useful and of the strengths and weaknesses of the traditional and likelihood methods. It then describes the Dirichlet Model using algebra and statistical theory. It sets up the algebra for likelihood estimation and gives a worked example. The algebra and likelihood estimation method are then demonstrated in an Excel workbook which is attached. The work book is applicable to a specific data set. It is not a general program for fitting the Dirichlet Model. It demonstrates the methods discussed here. The name of the workbook is Dirichlet likelihood.xls**.**

The Features of the Estimation Methods

The traditional approach to estimating the parameters of the Dirichlet Model, as developed and recommended by the model's authors, is the method of zero and ones and marginal moments. This is an appropriate method. However, given the capabilities which now exist in statistical theory and computing, it is sensible to consider alternate approaches to estimation. Likelihood theory is now a standard for estimation in modelling generally and it is a relatively uncomplicated procedure to apply to the Dirichlet Model and data sets. The two methods, the traditional approach and likelihood estimation, have different statistical properties which are discussed below.

The traditional method of zeroes and ones and marginal moments, as developed by the model's authors, has several relevant properties (Goodhardt, Ehrenberg et al. 1984; Ehrenberg 1988). (1) It is computationally easy to use and quick. (2) It uses only the aggregate results from the raw data, such as the observed penetrations and purchase frequencies. Access to the original unit record raw data is not required. While unit record data is now more and more available there are some panel market research companies who still do not fit the Dirichlet Model to their data and who also only publish aggregate results. Thus there is still a need for methods, like the traditional method of zeros and ones, which estimate the parameters of the model from aggregated results.

The alternate estimation method based on likelihood theory has several relevant properties. (1) It is known that estimates from likelihood theory have a unique statistical property in that they are 'efficient'. Over several samples the estimates of the parameters from likelihood theory will vary less. They are the estimates with the least sampling variation. In this manner they are more accurate (Edwards 1976; Kalwani 1980; Kalwani & Morrison 1980; Eliason 1993; Chickamenahalli 2000; Brown, Rungie et al. 2003). (2) Likelihood theory uses the original unit record raw data. Intuitively (and in reality) this can be expected to generate more accurate estimates but it is computationally more complex. With the power of modern computers the added calculations are not problematic. (3) Again intuitively, the accuracy of likelihood theory can be seen by considering a data set where there are very large purchase rates. As the length of time over which the data is collected is increased the penetrations of all the brands will move closer to each other and towards the category purchase rate. Despite the fact that the volume of data available for estimation is increasing, the decline in the differences between the penetrations will limit the accuracy of the traditional method of zeros and ones. As a comparison, likelihood theory just goes on getting more accurate as the volume of data increases.

The Dirichlet Model is often used to estimate the brand performance measures. The traditional method for estimating the parameters of the model was designed to ensure that, for one specific measure, market shares, the estimate from the parameters would exactly match the observed. This was seen as desirable as it would reduce potential objections from users, such as brand managers, to a model which did not accurately estimate the market shares. Furthermore, in Dirichlet Modelling, the market shares are sometimes seen as inputs to the model, as if they are explanatory variables or covariates, and so there has been a specific focus on market shares. As the use of the model expands this emphasis on the one brand performance measure, market shares, may not be sustained. Likelihood theory focuses on the estimation of all the parameters of the model equally regardless of the potential use of the estimates. In this sense likelihood theory is more robust. As the applications of the Dirichlet Model develop and the use of the model diversifies this robustness of likelihood estimation may well become increasingly important.

One development of the Dirichlet Model which is particularly important is its generalization; the inclusion of covariates such as marketing mix and the characteristics of the shoppers. Under these conditions the traditional method of zeros and ones can not be used whereas likelihood theory can be (Rungie, Laurent et al. 2003). When comparing the original Dirichlet Model and its generalized form it is advisable to use the same estimation methods for both. Thus, in this field, the use of likelihood theory is likely to become more prevalent, even for the original Dirichlet model.

Finally, likelihood theory is now a standard for estimating for many models in marketing. It certainly should be at least available for the Dirichlet Model. Also, over time, it may evolve that likelihood theory becomes the standard estimation procedure for traditional Dirichlet Model.

Probability Density Functions

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 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 is discrete, integer and nonnegative because purchases are counts which are whole numbers and 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 purchases of the individual brands are assumed to have a DMD which is conditional on the category purchase rate. Within the Dirichlet Model the two distributions, the NBD and the DMD, are assumed to be independent. Also, their parameters are assumed to have no associations.

Over the population of shoppers let the category purchase rate be a random variable *K*[1]. The Dirichlet Model assumes that the category purchase rate has a NBD; i.e. *K* has a NBD.

The NBD has two parameters which are both positive:
the shape[2] parameter γ and
the scale[3] parameter (which also influences the shape) β.

The probability density function for the NBD is (Johnson, Kotz et al. 1993):

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

Let there be h brands. Over the population of shoppers let the purchases 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.

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

---

[1] In this paper this random variable which is the category purchase rate is given the symbol *K*. In the original documentation of the Dirichlet Model the symbol K was used for one of the parameters of the distribution of this category purchase rate. The change has been made because (1) parameters should have Greek symbols not Arabic and (2) the category purchase rate is used as the total purchases in the DMD which, in the statistical literature, is refereed to as *k*.

[2] 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.

[3] 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 $\beta=1/\lambda$.

The probability density function for the DMD is (Johnson, Kotz et al. 1997)

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

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:

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

Estimating the Parameters from Data using Likelihood Theory

Likelihood theory can be briefly summarised. If the probability density function for a distribution over a random variable K is $f(k)$ and if an independent sample of n observations is drawn, $k_1,k_2,\ldots,k_n$, then the joint probability density function, L, for the sample is the product of all the individual density functions $L=f(k_1)\,f(k_2)\ldots f(k_n)$. This is known as the Likelihood Function. The estimates for the parameters of $f(k)$ are then generated by finding the values of these parameters which maximize L. This is equivalent to finding the joint probability density function such that the most likely observed value is the observed data set. Often, L is a very small number and it is easier and more accurate to maximize its natural log, $LL=\log(L)$. It is often easier to calculate the natural logs of the individual probability density functions.

Equation 4 $$LL=\sum_{i=1}^{n}\log\left(f\left(k_i\right)\right)$$

This is particularly the case for the Dirichlet Model. Its probability density functions includes the gamma function and it is easier and more accurate to calculate the log of the gamma function. The task, in Excel, is to calculate LL and then find the parameter values which maximize its value.

Likelihood estimates for the Dirichlet Model can be generated from raw panel data by fitting the NBD to the distribution of category purchase rates and the DMD to the distributions of purchases of individual brands. Thus there are two procedures and two probability density functions (1) the NBD and (2) the DMD.

The algebraic expressions for LL can be a little complex but calculating them in Excel (as is demonstrated below) is not overly difficult.

For the NBD, LL can be summarized as:

Equation 5 $$LL=\sum_{i=1}^{n}\ln\left(f_{\gamma,\beta}\left(k_i\right)\right)$$

Where:

Equation 6
$$\ln\left(f_{\gamma,\beta}\left(k_i\right)\right)=\ln\left(\Gamma\left(\gamma+k_i\right)\right)+k_i\ln(\beta)-\ln\left(\Gamma(\gamma)\right)-\ln\left(\Gamma\left(k_i+1\right)\right)-\left(\gamma+k_i\right)\ln(1+\beta)$$

The likelihood estimates of parameters γ and β for the NBD are the values for these parameters which maximize *LL* as defined in Equation 5 and Equation 6.

For the DMD, *LL* can be summarized as

Equation 7 $$LL = \sum_{i=1}^{n} \ln\left(f_{\alpha_1,\alpha_2,\ldots,\alpha_h}\left(r_{1,i}, r_{2,i},\ldots, r_{h,i}\right)\right)$$

Where

Equation 8 $$\ln\left(f_{\alpha_1,\alpha_2,\ldots,\alpha_h}\left(r_{1,i}, r_{2,i},\ldots, r_{h,i}\right)\right)$$

$$= \ln\left(\Gamma\left(\sum_{j=1}^{h}\alpha_j\right)\right) + \ln(\Gamma(k_i + 1)) - \ln\left(\Gamma\left(k_i + \sum_{j=1}^{h}\alpha_j\right)\right) + \sum_{j=1}^{h}\left\{\ln\left(\Gamma\left(\alpha_j + r_{j,i}\right)\right) - \ln\left(\Gamma\left(r_{j,i} + 1\right)\right) - \ln\left(\Gamma\left(\alpha_j\right)\right)\right\}$$

The likelihood estimates of parameters α_1, α_2,… α_n for the DMD are the values for these parameters which maximize *LL* as defined in Equation 7 and Equation 8.

The Data

The data should be such that for each shopper and for each brand there is a count, which is the purchase rate over a specified time period. Shoppers who do not purchase (non buyers) should be included. If there are *h* brands and *n* shoppers there will be *nh* counts. Set this data up in a table with *n* rows (one per shopper) and *h* columns (one per brand).

It is assumed that the data is independent. The purchases of any one shopper are independent of the purchases of all other shoppers. This is a condition assumed in many statistical procedures and is usually achieved through proper random sampling and data collection procedures.

Solver in Excel

The task is to (1) calculate *LL* in Excel for a particular set of values for the parameters and then (2) vary the parameters to find the maximum for *LL*. The parameter values will each be in a cell in Excel and the value of *LL* will also be in a cell.

Solver is a function in Excel which will identify the parameter values which maximize *LL*. Solver is usually found under the 'Tools' menu. If not, then it can be installed. Go to 'Tools' and then to 'add-ins' (Lilien & Rangaswamy 2003).

In the example Excel workbooks discussed below Solver has been set up to automatically find the parameter values which maximize *LL*. Note that, in each case, constraints have been specified within Solver to ensure that the parameter values are all greater than 0.000001. This is because for both the NBD and the DMD the parameters must be positive.

A Worked Example

An Excel workbook accompanies this Technical Note.
It demonstrates each of the steps in fitting the model to data using likelihood maximization. The name of the workbook is **Dirichlet Likelihood.xls**

The Excel workbook cannot be considered to be packaged software for the Dirichlet Model. The workbook applies the model to a data set in order to demonstrate the methods as a one-off example. The workbook has not been setup to adjust automatically to other data sets. The data set here has 2500 shoppers and 3 brands. Considerable effort would be required to adjust the workbook to data sets with a different number of shoppers or brands.

The workbook contains four work sheets. The function of each sheet is described below. The work sheets are presented in the order in which the calculations are undertaken. Thus the main results are in the last work sheet.

Documentation

Data A data set which records the purchases of 2500 shoppers from a product category with three brands. The data was simulated with parameter values $\gamma = 1$, $\beta = 5$, $\alpha_1 = 1$, $\alpha_2 = 0.67$, $\alpha_3 = 0.33$ and n=2500. These are the true parameter values. Normally the true values of the parameters would not be known, and they are known in this case only because the data was simulated.

NBD Likelihood The category purchase rates for the shoppers is extracted from the data set. The likelihood function (*LL*) for the NBD is calculated. If the user runs SOLVER, which is an Excel optimising command, then the likelihood function will be maximised by varying the parameter values γ and β. The result will be estimates of the parameters for fitting the NBD to the category purchase rate. These are the 'likelihood' estimates of the parameters. The estimates for γ and β are 0.97 and 5.2.

DMD Likelihood The purchase rates for each of the brands for the shoppers are extracted from the data set. The likelihood function (*LL*) for the DMD is calculated. If the user runs SOLVER then the likelihood function will be maximised by varying the parameter values $\alpha_1,\alpha_2,...,\alpha_h$. The result will be estimates of the parameters for the DMD for the brand purchase rates. These are 'likelihood' estimates of the parameters. The estimates for α_1, α_2 and α_3 are 0.94, 0.67 and 0.31.

Parameters This work sheet extracts the parameter estimates from the two previous work sheets. The estimates of the parameters γ and β for the NBD are extracted from NBD likelihood. The estimates of the parameters $\alpha_1,\alpha_2,...,\alpha_h$ for the DMD are extracted from DMD likelihood. Because the data was simulated the true parameter values are known (and are specified above). As is shown in the work sheet, the differences between the true and estimated parameter values are small. These differences are due to sampling variation.

Conclusion

The paper has discussed the estimation of the parameters of the Dirichlet Model. In the traditional use of the model these parameters are used to generate estimates, known as the theoreticals, for the brand performance measures such as purchase rate, market share, penetration, purchase frequency, share of category requirements and 100% loyals (Rungie, Goodhardt et al. 2003).

Likelihood theory is a method for estimating the parameters of the Dirichlet Model which is an alternate from the method originally developed and recommended by the authors of the model. Likelihood estimation, unlike the traditional method, requires access to the original raw unit record data and is computationally more intense which, with modern computing systems, is not problematic. Likelihood estimation is statistically more efficient; it has less sampling variation. This property alone has led to likelihood estimation becoming a standard in modelling. However, the real motivation for considering it as an alternate to the traditional method of zeros and ones is that increasingly the Dirichlet Model is being used in a wider range of applications. New developments for the model are appearing, such as its generalized form in which covariates are introduced. The model is taking on greater diversity. These developments are increasingly using likelihood estimation.

This Technical Note has demonstrated the use of likelihood theory to generate estimates of the parameters for the Dirichlet Model. The algebraic formulas have been specified. A worked example in Excel has been given with 2500 shoppers and three brands. The example uses simulated data. The estimates from likelihood estimation have been compared with the original parameter values used in the simulation. The researcher, using the instructions in the technical note can now use likelihood theory to apply the Dirichlet Model to data in Excel.

References

Brown B; Rungie CM et al. (2003). *A Standard Error Estimator for the Polarization Index: Assessing the Uncertainty in Loyalty*. Adelaide, Australia, School of Marketing, University of South Australia.

Chickamenahalli S (2000). *Estimation of Heterogeneity in Brand Choice Models.* Masters Dissertation. School of Mathematics. Adelaide, University of South Australia.

Edwards AWF (1976). *Likelihood - An account of the statistical concepts of likelihood and its application to scientific inference.* Bath, UK, Cambridge University Press.

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 et al. (2003). Understanding Brand Performance Measures: Using Dirichlet Benchmarks. *Journal of Business Research* (forthcoming).

Eliason SR (1993). *Maximum Likelihood Estimation - Logic and Practice*. California, Sage Publications Inc.

Goodhardt GJ; Ehrenberg ASC et al. (1984). The Dirichlet: A Comprehensive Model of Buying Behaviour. *Journal of the Royal Statistical Society, Series A (General)* ***147***(Part 5): 621-655.

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 et al. (1997). *Discrete Multivariate Distributions*. New York, John Wiley & Sons, Inc.

Johnson NL; Kotz S et al. (1993). *Univariate Discrete Distributions*. 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***,** 547-551.

Kalwani MU & Morrison DG (1980). Sample Size Requirements for Zero-Order Models. *Journal of Marketing Research, XVII,* 221-227.

Kearns Z (2000). *Dirichlet No Solver Software*. New Zealand, Massey University.

Lilien GL & Rangaswamy A (2003). *Marketing Engineering.* Upper Saddle River, New Jersey, Prentice Hall.

Rungie C; Laurent G et al. (2003). *Modelling Long-run Dynamic Markets.* Adelaide, Australia, School of Marketing, University of South Australia.

Rungie CM; Goodhardt G et al. (2003). *Calculation of Brand Performance Measures from the Parameters of the Dirichlet Model.* Adelaide, School of Marketing, University of South Australia.

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

Rungie CM; Laurent G et al. (2003). *Modeling Long-Run Dynamic Markets*. Adelaide, School of Marketing, University of South Australia.

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

Uncles MD; Ehrenberg ASC et al. (1995). Patterns of Buyer Behavior: Regularities, Models and Extensions. *Marketing Science, 14*(3, 2 of 2): G71-G78.

Cam Rungie is a Senior Lecturer in the School of Marketing University of South Australia.