Australasian Marketing Journal 25 (2017) 288–293

Contents lists available at ScienceDirect

Australasian Marketing Journal

journal homepage: www.elsevier.com/locate/amj

If the model fits, use it: Methods and benchmarks for evaluating NBD-Dirichlet goodness-of-fit

Carl Driesener [a,*], Melissa Banelis [b], Cam Rungie [a]

[a] Ehrenberg Bass Institute, University of South Australia, Adelaide, South Australia, Australia
[b] The University of Adelaide, Adelaide, South Australia, Australia

ARTICLE INFO

Article history:
Received 9 March 2017
Revised 13 September 2017
Accepted 25 October 2017
Available online 23 November 2017

Keywords:
Dirichlet model
Assessing fit
AAE
MAPE
Choice models
Non-purchase related choice

ABSTRACT

The Dirichlet model is an empirical generalization describing and predicting repeated choice amongst a set of competitive alternatives. With the advent of big data, there are many new potential applications for this model. Its developers emphasized one goodness-of-fit statistic, and subsequent researchers have used this along with others. There is, however, no consensus in the literature regarding which measures to use or, more importantly, benchmarks. This paper proposes a suite of six goodness-of-fit statistics developed from the literature to assess the fit of the model and develops two new measures that account for category specific factors enabling the development of benchmarks. It also provides appropriate benchmarks for all statistics derived from 54 FMCG categories in the UK.

CHINESE ABSTRACT

Dirichlet模型是描述和预测在一组竞争性替代品中重复多次选择的经验泛化模型。随着大数据的到来，这种模型有了很多新的应用潜力。它的开发者强调了一个拟合度统计量，随后的研究者也随之使用该方法。然而，对于选择什么量度以及更重要的选择什么基准，各文献中并没有共识。本文提出了一套从文献中开发的六种拟合度统计量，以评估模型的拟合度，并提出两个新的量度，考虑类别特定因素，便于研发基准。本文还为从英国54种消费品类别衍生的所有统计量提供了适当的基准。

© 2017 Australian and New Zealand Marketing Academy. Published by Elsevier Ltd. All rights reserved.

1. Introduction

An unintended consequence of the information technology revolution is the enormous volume of data now being collected about human behaviour in general and choice in particular. Facebook and similar websites, for example, collect data on people's social behaviour and networks. Consequently, researchers now have access to a wide range and volume of behavioural data in areas previously measurable only as a result of small experiments or ad hoc projects, if at all. There are models in marketing that have successfully described large data sets and predicted complex human behaviour in the past, and these same models show considerable promise for wider application in the era of big data.

Marketers know something about human behaviour, at least in the context of competitive consumer (and some B2B) markets (Keng et al., 1998; Sharp et al., 2002; Uncles and Ehrenberg, 1990). Buying behaviour in these markets can be characterized as choosing from a (limited) repertoire of a larger set of essentially similar options in an apparently random manner with relatively fixed propensities (at least in the short to medium term; Goodhardt et al., 1984). Consumers are generally disinterested in what they are choosing in the sense that the choice occupies a small fraction of their thoughts; there are generally other things the individuals would rather be thinking about or doing. Furthermore that choice occurs in the context of a stream of other activities (Sharp, 2010). Understanding the nature of buying behaviour has implications for marketing activities undertaken to attract, retain and extract value from buyers.

The Dirichlet model (Goodhardt et al., 1984) was developed to model buying behaviour in competitive markets, and has been applied to a significant number of markets with considerable success (Ehrenberg et al., 2004), and is generally considered to be one of the few empirical generalizations in marketing (Bass, 1993). It may be used as a comparison or benchmark for other modelling approaches (Kalwani et al., 1994), and is a useful starting point in

* Corresponding author.
E-mail address: Carl.Driesener@unisa.edu.au (C. Driesener).

https://doi.org/10.1016/j.ausmj.2017.10.003

understanding behaviour more generally in areas other than the typical buying context.

When applied to behavioural data, the model is a method for assessing the nature of the behaviour: where the model fits, the behaviour is consistent with normal buying behaviour, and therefore may be derived from the same heuristics or mental processes. Where the model does not fit, presumably different heuristics and mental processes may be in play, potentially requiring a different marketing approach.

Of course it is necessary to determine if the model predictions are congruent with the observed behaviour; that is, does the model fit the data? As with any model, the Dirichlet will not provide an exact match between its predictions, known as the 'model estimates' and the data, or the 'observed' values. The questions that are not clearly answered in the literature, however, are (i) how to measure the goodness-of-fit between the model and the data, and (ii) how to evaluate the fit and, more specifically, what benchmarks should be used with the goodness-of-fit statistics.

We review the limited literature that specifically examines the fit of the model and describe the methods in use for evaluating fit. We discuss the strengths and weaknesses of the various methods and develop new approaches to rectify identified problems. Finally we recommend a suite of eight tests of fit and provide benchmarks for all of them based on 54 FMCG categories in the UK.

2. Evaluating the fit of the Dirichlet model

Researchers usually make a qualitative evaluation of fit of the Dirichlet model by comparing the observed values and model estimates of brand performance metrics including penetration and purchase frequency (Dalal et al., 1984; Goodhardt et al., 1984; Uncles and Kwok, 2009; Uncles et al., 2010; Wagner and Taudes, 1986). If the observed values and model estimates are similar, the model is considered to fit (Ehrenberg, 1988) but there is not a clear consensus for what a good fit is for any single application of the model. This is not an unusual situation. In the stochastic modelling literature where the evaluations often emphasize statistical tests, valid and generalized goodness-of-fit statistics are often absent (Herniter, 1971; Jeuland et al., 1980; Marsh et al., 2004; Zufryden, 1977, 1978).

Mathematically speaking the preferable approach is to compare the Dirichlet model's distributions to the observed distributions using statistical techniques (e.g. Kemp and Zufryden in Goodhardt et al., 1984; Jeuland et al., 1980; Wagner and Taudes, 1986). This, though, has the significant limitation of requiring access to disaggregated, individual level data, which is often not available to the researcher. Furthermore while there are statistical tests for the univariate negative binomial distribution and bi-variate beta binomial distribution, there are no formal tests of fit for the multivariate Dirichlet multinomial distribution (Dalal et al., 1984), hence other avenues must be pursued.

The natural approach is to assess the fit of the model's estimates with the observed data. The standard method of fit assessment for the Dirichlet model is based on individual brand deviations calculated as the difference between observed values and model estimates (Fader and Schmittlein, 1993; Goodhardt et al., 1984; Scriven and Bound, 2004; Wright, 1999; Wright et al., 2002). An assessment of fit, however, is not determined by how well the model estimates match the observed metrics for a single brand, but rather the match over all brands (Fader and Schmittlein, 1993); that is how well the model fits the category as a whole (Scriven and Bound, 2004).

Dirichlet modellers use a range of goodness-of-fit statistics in evaluations where fit is quantified beyond just 'eye balling' the data. These tend to be comparisons of the average of the observed and model estimates of metrics (as shown in Table 1) and correlations of the observed and model estimates (Ehrenberg, 1988; Goodhardt et al., 1984; Uncles and Ellis, 1989; Wright et al., 2002). More recently authors have used the average absolute error (AAE) and/or a mean absolute percentage error (MAPE) as an assessment of fit (e.g. Uncles and Kwok, 2009). In the marketing literature historically the AAE has been referred to as the mean absolute deviation (MAD). While these measures are useful in the evaluation of fit, if used in isolation they may lead to incorrect conclusions: a single measure of fit may suggest that the model is a good fit when it is not. It is for this reason that most authors use more than one method of evaluating goodness-of-fit (e.g. Wright et al., 2002).

Table 1
Dirichlet model fit – Toothpaste UK 2000.

Leading brands (and market shares)		Penetration (%)		Purchase frequency	
		O	T	O	T
Category	(100%)	86	86	6.1	6.1
Colgate	(28%)	43	46	3.4	3.2
Aquafresh	(12%)	23	24	2.8	2.7
Sensodyne	(7%)	11	14	3.1	2.5
Macleans Standard	(6%)	14	13	2.4	2.5
Tesco	(5%)	9	10	2.9	2.5
Asda Protect	(4%)	8	9	2.8	2.5
Boots	(1%)	4	3	2.0	2.4
Morrisons	(1%)	3	3	2.6	2.4
Thera-Med	(1%)	2	2	2.2	2.4
Macleans Sensitive	(1%)	2	2	2.7	2.4
All other brands*	(10%)	22	19	2.3	2.6
Average brand		10	10	2.5	2.5

* 'All other brands' is the aggregate of any existing 'all other brands', 'all other private label brands', and brands with less than 1% market share. This results in a single 'brand' with a market share of 10%. This aggregate is comprised of 43 small brands.

3. Comparison of averages (AVE)

The first of the methods examined is the comparison of the average of the observed and model estimates of the values for each brand performance measure (hereafter AVE). The use of averages to compare these two sets of metrics is not as simplistic as it may first appear. Such comparisons provide an evaluation of any aggregate bias between the observed values and model estimates (i.e. the residuals, Ehrenberg and Bound, 1993). If the model is a good approximation of the data the residuals should be randomly distributed rather than showing any consistent bias (Ehrenberg, 1975; Ehrenberg and Bound, 1993). Thus the comparison of the averages of model estimates and observed values provides a good evaluation of overall bias. If the model estimated values are consistently larger than the observed, the average of the model estimates will be larger than that of the observed data, and *vice versa*.

$$AAE = \frac{\sum_{j=1}^{g} O_j - \sum_{j=1}^{g} T_j}{g} \quad (1)$$

Equation 1 provides the formula for the comparison of averages method for g brands where O and T are the observed values and model estimates of the metric respectively (Scriven and Bound, 2004; Sharp and Driesener, 2000; Uncles and Kwok, 2009; Wright, 1999; Wright et al., 2002). To illustrate this method, Table 1 shows observed values (O) and model estimates (T) for the toothpaste category in the UK, along with their averages.

This data is indicative of the behavioural metrics in competitive markets; there is some variation between many brands' observed values and model estimates (e.g. Sensodyne). In this example, however, the difference between the average observed value and model estimates is very close for both penetration and purchase frequency. In fact the difference is AVE = 0. Therefore the model overall is unbiased despite the individual brand variations. It is not, however,