Package 'NBDdirichlet'

May 7, 2026

Type Package

Title NBD-Dirichlet Model of Consumer Buying Behavior for Marketing Research

Version 1.4

Date 2022-05-26

Author Feiming Chen

Maintainer Feiming Chen <feimingchen@yahoo.com>

URL https://ani.stat.fsu.edu/~fchen/statistics/R-package-NBDdirichlet/how-to-use-Dirichlet-marketing-model.html

Description The Dirichlet (aka NBD-Dirichlet) model describes the purchase incidence and brand choice of consumer products. We estimate the model and summarize various theoretical quantities of interest to marketing researchers. Also provides functions for making tables that compare observed and theoretical statistics.

License GPL (>= 3)

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-29 14:00:02 UTC

Contents

NBDdirichlet-package 2
dirichlet 3
plot.dirichlet 6
print.dirichlet 8
summary.dirichlet 9

Index **13**

NBDdirichlet-package *NBD-Dirichlet of Model Consumer Buying Behavior*

Description

The Dirichlet (aka NBD-Dirichlet) model describes the probability distributions of the consumer purchase incidences and brand choices. We estimate the model and calculate various theoretical quantities of interest to marketing researchers.

Author(s)

Feiming Chen

References

The Dirichlet: A Comprehensive Model of Buying Behavior. G.J. Goodhardt, A.S.C. Ehrenberg, C. Chatfield. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 147, No. 5 (1984), pp. 621-655

Repeat-Buying:Facts,Theory and Applications, 2nd edn. A.S.C. Ehrenberg, 1988, London, Charles Griffin, ISBN 0 85264 287 3

Book Review: Repeat-Buying:Facts,Theory and Applications by A.S.C. Ehrenberg. Norman Pigden. *The Statistician*, Vol. 40, No. 3, Special Issue (1991), pp. 349-350

See Also

dirichlet, print.dirichlet, summary.dirichlet, plot.dirichlet

Examples

```
cat.pen <- 0.56 # Category Penetration
cat.buyrate <- 2.6 # Category Buyer's Average Purchase Rate in a given period.
brand.share <- c(0.25, 0.19, 0.1, 0.1, 0.09, 0.08, 0.03, 0.02) # Brands' Market Share
brand.pen.obs <- c(0.2,0.17,0.09,0.08,0.08,0.07,0.03,0.02) # Brand Penetration
brand.name <- c("Colgate DC", "Macleans","Close Up","Signal","ultrabrite",
"Gibbs SR","Boots Priv. Label","Sainsbury Priv. Lab.")

dobj <- dirichlet(cat.pen, cat.buyrate, brand.share, brand.pen.obs, brand.name)
print(dobj)
summary(dobj)
# plot(dobj)
```

`dirichlet` *Estimation of the Dirichlet model*

Description

Given consumer purchase summary data, it estimates the parameters of the Dirichlet model, which describes the consumer repeat-buying behavior of branded products. It also returns with several probability functions for users to calculate various theoretical quantities.

Usage

```
dirichlet(cat.pen, cat.buyrate, brand.share, brand.pen.obs,
brand.name = NA, cat.pur.var = NA, nstar = 50,
max.S = 30, max.K = 30, check = F)
```

Arguments

`cat.pen`	Product category penetration, which is the observed proportion of category buyers over a specific time period.
`cat.buyrate`	Category buyers' average purchase rate in a given period. This is derived as the total number of category purchase occasions divided by the total number of category buyers during a time period.
`brand.share`	A vector of brand market share. We typically define it as the proportions of purchase occasions that belong to different brands during the time period.
`brand.pen.obs`	A vector of observed brand penetration, which is the proportion of buyers for each brand during the time period.
`brand.name`	A character vector of the brand names. If not given (default), use "B1", "B2", etc.
`cat.pur.var`	The observed variance of category purchase rates across individuals. It is used for the method of moment estimation of the parameter K in the Dirichlet model. If it is not given (default), then estimate K by "mean and zeros"(see reference).
`nstar`	Maximum number of category purchases in the time period considered in the calculation. Any number larger than `nstar` is assumed to have occurred with probability zero. By default, it is 50. For higher frequently purchased category and longer study time period, it is necessary to increase `nstar` to the level (say, 100, 300, etc.) where $\sum_{n=1}^{nstar} P(n) > 0.9999$. We did not use the truncation procedure (suggested by the reference authors) in order to simplify coding.
`max.S`	Upper bound for the model parameter S in the optimization procedure to solve for S. Default to 30.
`max.K`	Upper bound for the model parameter K in the optimization procedure to solve for K. Default to 30.
`check`	A logical value. If T, print some diagnostic information. Defaul to F.

Details

The Dirichlet model and its estimation can be found in the reference paper. It is found to fit and reproduce the patterns of repeat buying of branded products quite well. Specifically, the dirichlet model is a mixture of distributions at four levels:

1. Each consumer's purchase incidences in a product category follow the Poisson process.
2. The purchase rates of the category by different consumers follow a Gamma distribution.
3. Each consumer's choices among the available brands follow a multinomial distribution, and
4. These brand choice probabilities follow a multivariate Beta or "Dirichlet" distribution across different consumers.

There are three structural parameters to be estimated:

M Mean purchase rate of the category.

K Measures the diversity of the overal category purchase frequency across consumers (smaller K implies more diversity).

S Measures the diversity of the brand purchase propensity across consumers (smaller S implies more diversity).

To estimate M and K, we use the observed category penetration (`cat.pen`) and purchase rate (`cat.buyrate`). To estimate S, we use additionally the observed brand penetrations (`brand.pen.obs`) and brand market shares (`brand.share`). Note however once these three parameters are estimated, only the brand market shares are needed by the Dirichlet model to compute various repeat-buying theoretical statistics.

The estimated parameters, along with several probability functions that can access the object data, are passed back in a list, which is assigned a "dirichlet" class attribute. The result can be used by the `print.dirichlet`, `summary.dirichlet`, and `plot.dirichlet` method.

The study period (where we report the model result) is assumed to be 4 times of the observation period (input data). So if we use quarterly data, the model output is annulized. This multiple (4) can be changed using the member function `period.set`.

Value

A list with the following components:

`M`	Estimated Dirichlet model parameter: mean purchase rate of the category.
`K`	Estimated Dirichlet model parameter: it measures the diversity of the overal category purchase frequency (smaller K implies more diversity).
`S`	Estimated Dirichlet model parameter: it measures the diversity of the brand purchase propensity (smaller S implies more diversity).
`nbrand`	Number of brands being considered in the produt category.
`nstar`	Input parameter: Maximum number of category purchases considered.
`cat.pen`	Input parameter: Category penetration in a given time period.
`cat.buyrate`	Input parameter: Category buyers' average purchase rate in a given time period.
`brand.share`	Input parameter: A vector of brand market share.

brand.pen.obs	Input parameter: A vector of observed brand penetration.
brand.name	Input parameter: A character vector of the brand names.
check	A logical flag that indicates whether to print the intermediate information in the model estimation. Default to F.
error	A logical flag that indicates if nstar is too small to sufficently cover the support of category purchase probabilities (calculated by Pn, see below). If it is returned T, then nstar should be increased and the Dirichlet model be re-estimated.
period.set	A member function of the "dirichlet" class object with one required parameter (t), which can be any positive real number. It resets the study time period to be t times of the assumed base time period in the sample.
period.print	A member function of the "dirichlet" class object with no parameter. It indicates the current time period by printing the multiple t of the base time period.
p.rj.n	A member function of the "dirichlet" class object with three required parameters (r_j, n, j). It calculates the conditional probability of buying brand j for exactly r_j times given that the consumer has made n category purchases.
Pn	A member function of the "dirichlet" class object with one required parameter (n). It calculates the probability that a consumer has made n category purchases in the study time period.
brand.pen	A member function of the "dirichlet" class object with one required and one optional parameter (j, limit=c(0:nstar)). It calculates the probability that a consumer makes at least one purchase of the brand j (theoretical penetration) in the study time period. The optional vector limit enumerates the exact frequencies that a consumer will be buying in the category and is used to index the summation of the probabilities of *not* buying brand j given those category purchases in limit.
brand.buyrate	A member function of the "dirichlet" class object with one required and one optional parameter (j, limit=c(0:nstar)). It calculates the expected number of brand j purchases given that the consumer is a buyer of the brand j in the time period (theoretical brand buying rate). The limit parameter has the same meaning as that in the function brand.pen.
wp	A member function of the "dirichlet" class object with one required and one optional parameter (j, limit=c(0:nstar)). It calculates the expected number of the product category purchases given that the consumer is a buyer of the brand j in the time period (theoretical category buying rate for brand buyer). The limit parameter has the same meaning as that in the function brand.pen.

Author(s)

Feiming Chen

References

The Dirichlet: A Comprehensive Model of Buying Behavior. G.J. Goodhardt, A.S.C. Ehrenberg, C. Chatfield. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 147, No. 5 (1984), pp. 621-655

See Also

print.dirichlet, summary.dirichlet, plot.dirichlet, NBDdirichlet-package

Examples

```
# The following data comes from the example in section 3 of
# the reference paper.  They are Toothpaste purchase data in UK
# in 1st quarter of 1973 from the AGB panel (5240 static panelists).

cat.pen <- 0.56 # Category Penetration
cat.buyrate <- 2.6 # Category Buyer's Average Purchase Rate in a given period.
brand.share <- c(0.25, 0.19, 0.1, 0.1, 0.09, 0.08, 0.03, 0.02) # Brands' Market Share
brand.pen.obs <- c(0.2,0.17,0.09,0.08,0.08,0.07,0.03,0.02) # Brand Penetration
brand.name <- c("Colgate DC", "Macleans","Close Up","Signal","ultrabrite",
"Gibbs SR","Boots Priv. Label","Sainsbury Priv. Lab.")

dobj <- dirichlet(cat.pen, cat.buyrate, brand.share, brand.pen.obs, brand.name)
print(dobj)
```

---

plot.dirichlet *Plot of theoretical penetration growth and buying rate growth*

Description

This function plots a 'dirichlet' object. It is a method for the generic function plot for objects of the class 'dirichlet'. It plots the theoretical penetration growth and buying rate growth across multiple brands according to the Dirichlet model over a specified time sequence.

Usage

```
## S3 method for class 'dirichlet'
plot(x, t = 4, brand = 1:x$nbrand, incr = 1, result = NULL,...)
```

Arguments

x	An object of "dirichlet" class.
t	Maximum of the projection time period, which is specified as a multiple of the base time period. For example, if the base time period is quarterly, then t=4 would mean annually.
brand	A vector specifying the subset of brands to be ploted.
incr	Increment for the time sequence that starts at 0. Its unit is one base time period. Can be a fractional number such as 0.1.
result	A list returned from the previous run of the plot.dirichlet. It is used to avoid repeating the computation when incr is changed.
...	Other parameters passing to the generic function.

Details

A time sequence will be made from 0 up to `t` with increment `incr`, against each component of which the theoretical penetration and brand buying rate will be plotted. Each plotted point represents the *cumulated* penetration or buying rate from time 0 to its current time point (expressed as the multiple of the base time period).

Value

A list with two components:

`pen`	A matrix with the penetration values. Its number of rows is the number of brands, and its number of columns is the length of the time sequence used for plotting the X coordinates of the points.
`buy`	A matrix with the buying rate values. Its dimension is the same as that of `pen`.

Author(s)

Feiming Chen

References

The Dirichlet: A Comprehensive Model of Buying Behavior. G.J. Goodhardt, A.S.C. Ehrenberg, C. Chatfield. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 147, No. 5 (1984), pp. 621-655

See Also

`dirichlet`, `summary.dirichlet`, `print.dirichlet`, `NBDdirichlet-package`

Examples

```
cat.pen <- 0.56 # Category Penetration
cat.buyrate <- 2.6 # Category Buyer's Average Purchase Rate in a given period.
brand.share <- c(0.25, 0.19, 0.1, 0.1, 0.09, 0.08, 0.03, 0.02) # Brands' Market Share
brand.pen.obs <- c(0.2,0.17,0.09,0.08,0.08,0.07,0.03,0.02) # Brand Penetration
brand.name <- c("Colgate DC", "Macleans","Close Up","Signal","ultrabrite",
"Gibbs SR","Boots Priv. Label","Sainsbury Priv. Lab.")

dobj <- dirichlet(cat.pen, cat.buyrate, brand.share, brand.pen.obs, brand.name)
plot(dobj)
```

print.dirichlet *Print Dirichlet model information*

Description

This function prints a 'dirichlet' object. It is a method for the generic function `print` of class 'dirichlet'. It prints descriptive information on the product category, brand, and the estimated Dirichlet model parameters.

Usage

```
## S3 method for class 'dirichlet'
print(x,...)
```

Arguments

x	An object of "dirichlet" class.
...	Other parameters passing to the generic function

Details

The following information is printed:

- Number of brands in the category
- Brand list
- Brands' market shares.
- Brands' penetrations.
- The multiple of the base time period that indicates the study time period, and the corresponding M value.
- Category penetration and category buying rate.
- Estimated dirichlet model parameters: M (for base period), K, and S.

Value

NULL

Author(s)

Feiming Chen

References

The Dirichlet: A Comprehensive Model of Buying Behavior. G.J. Goodhardt, A.S.C. Ehrenberg, C. Chatfield. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 147, No. 5 (1984), pp. 621-655

See Also

dirichlet, summary.dirichlet, plot.dirichlet, NBDdirichlet-package

Examples

```
cat.pen <- 0.56 # Category Penetration
cat.buyrate <- 2.6 # Category Buyer's Average Purchase Rate in a given period.
brand.share <- c(0.25, 0.19, 0.1, 0.1, 0.09, 0.08, 0.03, 0.02) # Brands' Market Share
brand.pen.obs <- c(0.2,0.17,0.09,0.08,0.08,0.07,0.03,0.02) # Brand Penetration
brand.name <- c("Colgate DC", "Macleans","Close Up","Signal","ultrabrite",
"Gibbs SR","Boots Priv. Label","Sainsbury Priv. Lab.")

dobj <- dirichlet(cat.pen, cat.buyrate, brand.share, brand.pen.obs, brand.name)
print(dobj)  # YOU WILL SEE THE FOLLOWING:

# Number of Brands in the Category = 8
# Brand List : Colgate DC : Macleans : Close Up : Signal : ultrabrite :
#              Gibbs SR : Boots Priv. Label : Sainsbury Priv. Lab.
# Brands' Market Shares: 0.25 0.19 0.1 0.1 0.09 0.08 0.03 0.02
# Brands' Penetration:   0.2 0.17 0.09 0.08 0.08 0.07 0.03 0.02
# Multiple of Base Time Period: 1 , Current M = 1.456

# Channel Penetration = 0.56 , with Shopping Rate = 2.6
# Estimated Dirichlet Model Parameters:
# NBD: M = 1.46 ,  K = 0.78 ;  Dirichlet: S = 1.55
```

---

summary.dirichlet *Theoretical summary statistics from the Dirichlet model.*

Description

This function summarizes a 'dirichlet' object. It is a method for the generic function summary of class 'dirichlet'. It calculate four types of theoretical summary statistics, which can be compared with the corresponding observed statistics.

Usage

```
## S3 method for class 'dirichlet'
summary(object, t = 1, type = c("buy", "freq", "heavy", "dup"),
digits = 2, freq.cutoff = 5, heavy.limit = 1:6, dup.brand = 1, ...)
```

Arguments

object	An object of "dirichlet" class.
t	Multiple of the base time period. For example, if the assumed base time period is quarterly, then t=4 would mean annually. Default to one.
type	A character vector that specifies which types of theoretical statistics (during the time period indicated by t) are to be calculated. Four character strings can be listed:

buy Theoretical brand penetration, buying rate, and the buying rate of the category per brand buyer.

freq Distribution of the number of brand purchases.

heavy Theoretical brand penetration and buying rate among category buyers with a specific frequency range.

dup Brand duplication (proportion of buyers of a particular brand also buying other brand).

`digits` Number of decimal digits to control the rounding precision of the reported statistics. Default to 2.

`freq.cutoff` For the `type="freq"` table, where to cut off and lump the tail of the frequency distribution.

`heavy.limit` For the `type="heavy"` table, a vector containing the specific purchase frequency range of the category buyers. The upper-bound of the frequency is `nstar`.

`dup.brand` For the `type="dup"` table, the focal brand. Defaul to the first brand in the brand list.

`...` Other parameters passing to the generic function.

Details

The output corresponds to the theoretical portion of the Table 3, 4, 5, 6 in the reference paper. We also have another set of functions (available upon request) that put observed and theoretical statistics together for making tables that resemble those in the reference.

Let P_n be the probability of a consumer buying the product category n times. Then P_n has a Negative Binomial Distribution (NBD). Let $p(r_j|n)$ be the probability of making r_j purchases of brand j, gien that n purchases of the category having been make ($r_j \leq n$). Then $p(r_j|n)$ has a Beta-Binomial distribution.

The theoretical brand penetration b is then

$$b = 1 - \sum_{n=0} P_n p(0|n)$$

The theoretical brand buying rate w is

$$w = \frac{\sum_{n=1}\{P_n \sum_{r=1}^{n} rp(r|n)\}}{b}$$

and the category buying rate per brand buyer w_P is

$$w_P = \frac{\sum_{n=1}\{nP_n[1 - p(0|n)]\}}{b}$$

The brand purchase frequency distribution is

$$f(r) = \sum_{n \geq r} P_n p(r|n)$$

The brand penetration given a specific category purchase frequency range $R = \{i_1, i_2, i_3, \ldots\}$ is

$$1 - \frac{\sum_{n \in R} P(n)p(0|n)}{\sum_{n \in R} P(n)}$$

The brand buying rate given a specific category purchase frequency range $R = \{i_1, i_2, i_3, \ldots\}$ is

$$\frac{\sum_{n \in R} P(n)[\sum_{r=1}^{n} rp(r|n)]}{\sum_{n \in R} P(n)[1 - p(0|n)]}$$

To calculate the brand duplication measure, we first get the penetration $b_{(j+k)}$ of the "composite" brand of two brands j and k as:

$$b_{(j+k)} = 1 - \sum_{n} P_n p_k(0|n) p_j(0|n)$$

Then the theoretical proportion b_{jk} of the population buying both brands at least once is:

$$b_{jk} = b_j + b_k - b_{(j+k)}$$

and the brand duplication $b_{j/k}$ (where brand k is the focal brand) is:

$$b_{j/k} = b_{jk}/b_k$$

Value

A list with those components that are specified by the input `type` parameter.

`buy`	A data frame with three components: `pen.brand` (Theoretical brand penetration), `pur.brand` (buying rate of the brand), `pur.cat` (buying rate of the category per brand buyer).
`freq`	A matrix that lists the distribution of brand purchases. The number of rows is the number of brands.
`heavy`	A matrix with two columns. The first column (`Penetration`) is the theoretical brand penetration among category buyers with a specific frequency range. The second column (`Avg Purchase Freq`) is the brand buying rate of those category buyers. The number of rows is the number of brands.
`dup`	A vector with dimension as the number of brands. It reports the brand duplication (proportion of buyers of a particular brand also buying other brand) of the focal brand (`dup.brand`).

Author(s)

Feiming Chen

References

The Dirichlet: A Comprehensive Model of Buying Behavior. G.J. Goodhardt, A.S.C. Ehrenberg, C. Chatfield. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 147, No. 5 (1984), pp. 621-655

See Also

`dirichlet`, `print.dirichlet`, `plot.dirichlet`, `NBDdirichlet-package`

Examples

```
cat.pen <- 0.56 # Category Penetration
cat.buyrate <- 2.6 # Category Buyer's Average Purchase Rate in a given period.
brand.share <- c(0.25, 0.19, 0.1, 0.1, 0.09, 0.08, 0.03, 0.02) # Brands' Market Share
brand.pen.obs <- c(0.2,0.17,0.09,0.08,0.08,0.07,0.03,0.02) # Brand Penetration
brand.name <- c("Colgate DC", "Macleans","Close Up","Signal","ultrabrite",
"Gibbs SR","Boots Priv. Label","Sainsbury Priv. Lab.")

dobj <- dirichlet(cat.pen, cat.buyrate, brand.share, brand.pen.obs, brand.name)

## Not run: summary(dobj)
summary(dobj, t=4, type="freq")
summary(dobj, t=4, type="heavy", heavy.limit=c(7:50))
summary(dobj, t=1, type="dup", dup.brand=2)
```

Index

∗ **methods**
plot.dirichlet, 6
print.dirichlet, 8
summary.dirichlet, 9
∗ **models**
dirichlet, 3
∗ **package**
NBDdirichlet-package, 2

dirichlet, *2*, 3, *7*, *9*, *11*

NBDdirichlet-package, 2

plot.dirichlet, *2*, *4*, *6*, 6, *9*, *11*
print.dirichlet, *2*, *4*, *6, 7*, 8, *11*

summary.dirichlet, *2*, *4*, *6, 7*, *9*, 9