Package 'touchard'

May 31, 2019

Type Package

Title Touchard Model and Regression

Description Tools for analyzing count data with the Touchard model (Matsushita et al., 2018, Comm Stat Th Meth <doi:10.1080/03610926.2018.1444177>). It includes univariate estimation (ML and MM) and regression tools developed by Andrade et al. (submitted).

Version 2.0.1

Date 2019-05-31

Author Bernardo Andrade and Sandro Oliveira

Maintainer Bernardo Andrade <bbandrade@unb.br>

License GPL-2

Depends R (>= 3.2.0)

Imports MASS, nleqslv, numDeriv, plotrix

NeedsCompilation no

Repository CRAN

Date/Publication 2019-05-31 12:40:03 UTC

R topics documented:

touchard-package *Touchard Model and Regression*

Description

Analysis of count data with the Touchard model including: d-*, p-*, q-* and r-* functions; inference for univariate analysis (maximum likelihood and method of moment estimation); regression modeling and graphical assessment of goodness of fit (rootogram and touchardness plot).

Details

The DESCRIPTION file:

Package:	touchard
Type:	Package
Title:	Touchard Model and Regression
Description:	Tools for analyzing count data with the Touchard model (Matsushita et al., 2018, Comm Stat Th Meth <doi:10.
Version:	2.0.1
Date:	2019-05-31
Author:	Bernardo Andrade and Sandro Oliveira
Maintainer:	Bernardo Andrade <bbandrade@unb.br>
License:	GPL-2
Depends:	R (>= 3.2.0)
Imports:	MASS, nleqslv, numDeriv, plotrix

Index of help topics:

```
Accidents               Trafic Accidents Data
Crabs                   Crabs Data
Epilepsy                Epilepsy Data
Touchard                The Touchard Distribution
rgram                   Tukey's (hanging, standing, suspended)
                        Rootogram for Assessing a Touchard Model Fit
score.delta             Score Test for Poisson vs Touchard
touchard-package        Touchard Model and Regression
toufit                  Touchard Estimation
touplot                 Touchardness Plot
toureg                  Touchard Regression
```

Author(s)

Bernardo Andrade and Sandro Oliveira

Maintainer: Bernardo Andrade <bbandrade@unb.br>

References

Matsushita RY, Pianto D, Andrade BB, Cancado A, Silva S (2018) The Touchard distribution, *Communications in Statistics - Theory and Methods*, <doi:10.1080/03610926.2018.1444177>

Andrade, BB; Matsushita, RY; Oliveira, SB (submitted) Analyzing Count Data with the Touchard Model. *available upon request*.

`Accidents` *Trafic Accidents Data*

Description

Data on daily traffic accidents.

Usage

```
data("Accidents")
```

Format

A data frame with 1096 observations on 6 variables.

`Year` year (2011, 12, 13).

`Month` integer, 1–12.

`WeekDay` integer, 1–7.

`Season` factor (Fall, Spring, Summer, Winter).

`Holiday` integer (0=no, 1=yes).

`Y` integer, observed count of accidents.

Source

<https://data.ny.gov> Open Data NY: State of New York

Examples

```
data(Accidents)
summary(Accidents)
```

`Crabs` *Crabs Data*

Description

Horseshoe crab data in Table 4.3.2 (Agresti, Categ. Data Anal., 3nd edition) by courtesy of Jane Brockmann, Zoology Department, University of Florida.

Usage

```
data("Crabs")
```

Format

A data frame with 173 observations on the following 6 variables.

`crab` crab id

`y` number of satellites

`weight` in kg

`width` carapace width in cm

`color` has values 1-4 with 1=light

`spine` spine condition

Source

<http://www.stat.ufl.edu/~aa/cda/data.html>

References

Agresti, A. (2013). *Categorical Data Analysis*. Wiley, 3d ed.

Brockmann, H. J. (1996), Satellite Male Groups in Horseshoe Crabs, Limulus polyphemus. *Ethology*, 102: 1-21.

Examples

```
data(Crabs)
dim(Crabs)
head(Crabs)
```

Epilepsy *Epilepsy Data*

Description

Data on Epileptic Seizures.

Usage

```
data("Epilepsy")
```

Format

A data frame with 351 observations on counts of epileptic seizures.

`seizures` numeric vector of length 351.

Source

see reference below

References

Bhati D, Sastry DVS, Qadri PZM (2015). A New Generalized Poisson-Lindley Distribution: Applications and Properties. *Austrian Journal of Statistics*, 4, 35-51.

Examples

```
data(Epilepsy)
seizures
table(seizures)
```

rgram *Tukey's (hanging, standing, suspended) Rootogram for Assessing a Touchard Model Fit*

Description

The Hanging Rootogram is variation of the histogram with the vertical axis showing the square root of the frequencies hanging from (Touchard) fitted values so that the discrepancies are visualized against a straight line (the axis) rather than against a curve.

Usage

```
rgram(object, xlim = NULL, ylim = NULL, xlab = "Count", ylab = NULL, main = NULL,
breaks = NULL, border = "black", fill = "lightgray", col = "blue",
         lwd = 2, pch = 19, lty = 1,  axes = TRUE, width = NULL, plot = TRUE,
        style = c("hanging", "standing", "suspended"), scale = c("sqrt", "raw"),
         max = NULL, ...)
```

Arguments

`object`	either a `toufit` or a `toureg` object.
`breaks`	numeric. Breaks for the bar intervals.
`style`	character. Style of rootogram (see below).
`scale`	character. Scale of vertical axis, raw frequencies or their square roots; `style="standing"` with `scale="raw"` results in the usual histogram with the fitted curve superimposed.
`plot`	logical. Should the plot be displayed?
`width`	numeric. Widths of the histogram bars.
`main`	character. Title for the plot.
`xlab`	character. Label for the x axis.
`ylab`	character. Label for the y axis.
`xlim, ylim, border, fill, col, lwd, pch, lty, axes`	graphical parameters.
`max`	the largest count value used when using the `rgram` with `toureg` objects. If `NULL`, it sets to `max(1.5 * max(count), 20L)`.
`...`	further graphical parameters passed to the plotting function.

Details

The rootogram is a visual tool for comparing the empirical distribution and fitted values (here from a Touchard model). The square-root scale de-emphasizes outlying values and right skewness (common for count data). The hanging (from the fitted values) style allows discrepancies to be visualized against a straight line (the axis).

The code has been largely based on more general rootogram functions: `rootogram` in package **vcd** and `countreg::rootogram` available through R-forge.

Value

Returns invisibly a data frame with quantities used in plotting.

Author(s)

Bernardo Andrade and Sandro Oliveira

References

Friendly M, Meyer D (2015). *Discrete Data Analysis with R*. Chapman and Hall.

Kleiber C, Zeileis A (2016). Visualizing Count Data Regressions Using Rootograms. *The American Statistician*, **70**(3), 296–303. doi: 10.1080/00031305.2016.1173590.

Tukey JW (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading.

See Also

`rootogram`, `touplot`

Examples

```
data(Epilepsy)
seiz <- seizures
fm <- toufit( seiz )
rgram(fm)

data(Accidents)
acc <- within(subset(Accidents, subset=Year==2013), {
FriSat <- ifelse(WeekDay > 5, 1, 0)
Spring <- ifelse(Season == "Spring", 1, 0)
})
fmTraff <- toureg( Y ~ FriSat + Spring, data = acc )
rgram(fmTraff)
```

`score.delta` *Score Test for Poisson vs Touchard*

Description

Score test for the extra (with respect to the Poisson) parameter in the Touchard model (Null: $\delta = 0$) performed without the need to fit the Touchard model. Analogous likelihood-ratio and Wald tests (which require fitting the full model) are available in the output of `toureg`.

Usage

```
score.delta(x, freq = NULL, max = 50, data)
```

Arguments

`x`	for univariate data: either an object of class `table` with the observed counts as classifying factors *or* the observed counts (raw data) *or* the unique values of the observed counts in which case the observed frequencies must be given in argument `freq`; for regression data: an object of class `"formula"` with the symbolic description of the model to be fitted.

`freq`	unnecessary (and ignored) if `x` is a table or the raw data or a formula; must be provided if `x` is the unique values of the observed counts, in the same order.
`max`	number of Poisson terms used in calculations of necessary moments
`data`	data frame containing the variables in the model.

Value

A list with elements

`stat`	the value the chi-squared test statistic.
`pval`	the p-value for the test (from chi^2(df=1)).

Author(s)

Bernardo Andrade and Sandro Oliveira

References

Andrade, BB; Matsushita, RY; Oliveira, SB (submitted) Analyzing Count Data with the Touchard Model. *available upon request*.

Examples

```
data(Epilepsy)
score.delta(seizures)

data(Crabs)
score.delta( y ~ weight + color, data = Crabs )
```

`Touchard` *The Touchard Distribution*

Description

Density, normalizing constant, distribution function, quantile function and random number generation for the Touchard distribution with Poisson-like parameter equal to `lambda` and shape/dispersion parameter equal to `delta`.

Usage

```
dtouch(x, lambda, delta, N=NULL, eps=sqrt(.Machine$double.eps), log = FALSE)
ptouch(x, lambda, delta, N=NULL, eps=sqrt(.Machine$double.eps))
qtouch(p, lambda, delta, N=NULL, eps=sqrt(.Machine$double.eps))
rtouch(n, lambda, delta, N=NULL, eps=sqrt(.Machine$double.eps))
tau(lambda, delta, N=NULL, eps=sqrt(.Machine$double.eps))
```

Arguments

`x`	vector of quantiles
`p`	vector of probabilities.
`n`	number of observations.
`lambda`	Poisson-like (location) parameter which corresponds to the mean of the distribution when `delta` = 0
`delta`	shape/dispersion parameter which produces unequal dispersion (var x mean) when different from zero and mild zero excess compared to the Poisson distribution
`N`	number of terms in the computation (series) of the normalizing constant. If `NULL` a recursion formula is used and iterated until the specified relative error is reached.
`eps`	relative error in the computation (series) of the normalizing constant. Only used if `N=NULL`. See reference for details.
`log`	logical; if TRUE, probability p is given as log(p).

Details

The Touchard distribution with parameters λ and δ has density

$$f(x) = \frac{\lambda^x (x+1)^\delta}{x! \tau(\lambda, \delta)}$$

for $y = 0, 1, 2, ...$, $\lambda > 0$ and δ real.

Value

`dtouch` gives the density, `ptouch` gives the distribution function, `qnorm` gives the quantile function, and `rtouch` generates random deviates.

`rtouch` uses the inverse transform method. The length of the result is determined by `n` and is the maximum of the lengths of the numerical arguments for the other functions. The numerical arguments other than `n` are recycled to the length of the result.

`qtouch` uses an initial approximation based on the Cornish-Fisher expansion followed by a search in the appropriate direction.

`tau` gives the value of the normalizing constant in the Touchard density.

Author(s)

Bernardo Andrade and Sandro Oliveira

References

Matsushita RY, Pianto D, Andrade BB, Cancado A, Silva S (2018) The Touchard distribution, *Communications in Statistics - Theory and Methods*, <doi:10.1080/03610926.2018.1444177>

See Also

rgram, touplot

Examples

```
for(N in c(2, 5, 10, 20, 50)) print( tau(lambda=7, delta=-1, N) )
tau(lambda=7, delta=-1)
dtouch(0:10, lambda=7, delta=-1)
ptouch(0:10, lambda=7, delta=-1)
qtouch(c(.1,.25,.5,.75,.9), lambda=7, delta=-1)
rtouch(10, lambda=7, delta=-1)
```

---

toufit *Touchard Estimation*

Description

Maximum-likelihood and method-of-moments estimation of the Touchard model.

Usage

```
toufit(x, freq = NULL, start, method = c("ml", "mm", "gmm"),
    rc = FALSE, trunc.at.zero = FALSE)
```

Arguments

x	either an object of class table with the observed counts as classifying factors *or* the observed counts (raw data) *or* the unique values of the observed counts in which case the observed frequencies must be given in argument freq; see example below for the three formats.
freq	unnecessary (and ignored) if x is a table or the raw data; must be provided if x is the unique values of the observed counts, in the same order.
start	starting values to be used by optimization/nonlinear solving algorithms. If missing, the default value is explained in the reference below.
method	one of 'ml' (maximum likelihood), 'mm' (method of moments based on the usual first and second moment conditions) or 'gmm' (generalized method of moments based on the first two moments and an extra condition on the first moment of *log(Y+1)*.
rc	TRUE for right-censored data. Not implemented for methods 'mm' and 'gmm'.
trunc.at.zero	TRUE for model truncated at zero. Not implemented for methods 'mm' and 'gmm'.

Details

For method 'ml' the Touchard likelihood is numerically maximized via optim() with 'method=L-BFGS-B'. For methods 'mm' and 'gmm', the system of moment conditions is solved by calling nleqslv and numDeriv.

Value

A list with the following:

fit	a list with point estimates, standard errors and variance matrix of estimates
aic	Akaike's information criterion, -2*log-likelihood + 2*p
bic	Schwarz's Bayesian criterion, -2*log-likelihood + log(n)*p
test	a data frame with likelihood ratio and Wald test results for the Null: delta = 0 (Poisson)
method	estimation method used ('ml', 'mm' or 'gmm')
data	a list with elements x and freq

Author(s)

Bernardo Andrade and Sandro Oliveira

References

Matsushita RY, Pianto D, Andrade BB, Cancado A, Silva S (2018) The Touchard distribution, *Communications in Statistics - Theory and Methods*, <doi:10.1080/03610926.2018.1444177>

Andrade, BB; Matsushita, RY; Oliveira, SB (submitted) Analyzing Count Data with the Touchard Model. *available upon request*.

See Also

optim , rgram , Touchard , touplot

Examples

```
data(Epilepsy)
# 'seizures' is vector of raw data (351 counts)
toufit( x = seizures, method = 'ml' )
toufit( x = seizures, method = 'mm' )
toufit( x = seizures, method = 'gmm' )

# suppose data were recorded as table object
TAB <- table(seizures)
TAB
toufit(TAB)  # same as above (method = 'ml')

# suppose data were recorded as data.frame
DF <- data.frame( y = as.numeric(names(TAB)), fr = as.numeric(TAB) )
DF
toufit(x = DF$y, freq = DF$fr) # same as above (method = 'ml')
```

`touplot` *Touchardness Plot*

Description

Touchardness Plot: diagnostic distribution plot for the Touchard model.

Usage

```
touplot(x, freq = NULL, plot = TRUE, conf.level = 0.95,
                        ylab = "Count Metameter", xlab = "Count",
                        main = "Touchardness Plot", ...)
```

Arguments

`x`	either a vector of counts or a 1-way table of frequencies of counts.
`freq`	unnecessary (and ignored) if `x` is a table or the raw data; must be provided if `x` is the unique values of the observed counts, in the same order.
`plot`	logical. Should the plot be displayed?
`conf.level`	numeric in (0,1). Confidence level for confidence intervals.
`main`	character. Title for the plot.
`xlab`	character. Label for the x axis.
`ylab`	character. Label for the y axis.
`...`	other parameters to be passed through to plotting functions.

Details

Plots the number of occurrences (counts) against the count metameter of the Touchard distribution. Circles are the observed count metameters and the filled points show the confidence interval (dashed lines) centers. Estimate of λ based on the fitted line is shown on top margin along with the MLE (`toufit`). If the Touchard model fits the data well, the plot should show a straight line and the two estimates of λ should be close.

Value

Returns invisibly a data frame containing the counts (`y`), frequencies (`freq`), count metameter (`metameter`), the CI center (`CIcenter`) and the CI margin (`CImargin`).

Author(s)

Bernardo Andrade and Sandro Oliveira

References

D. C. Hoaglin, F. Mosteller & J. W. Tukey (eds.), *Exploring Data Tables, Trends and Shapes*, chapter 9. John Wiley & Sons, New York.

M. Friendly & D. Meyer (2015), *Discrete Data Analysis with R*. Chapman and Hall.

See Also

rgram

Examples

```
data(Epilepsy)
touplot(seizures)
touplot(seizures, plot=FALSE)
```

---

toureg *Touchard Regression*

Description

Touchard Regression via either maximum likelihood or quasi-likelihood.

Usage

```
toureg(formula, data, x = FALSE, y = FALSE, start.beta, start.delta,
        parscale = rep.int(1, length(start.beta)+1),  maxit, abstol = -Inf,
        reltol = 1e-6, etol = 1e-6, gtol = 1e-4,
        N=100, eps=1e-6, dm = 10, regress = c("mu", "lambda"),
        method = c("BFGS", "CG", "Nelder-Mead", "glm", "qp1", "qp2"),  ... )
```

Arguments

`formula` an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.

`data` an optional data frame containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which `toureg` is called.

`start.beta, start.delta`
starting values for the parameters in the linear predictor. If missing, the default values explained in the reference below are used.

`parscale, maxit, abstol, reltol`
arguments passed to `control` in `optim`. `reltol` is alos used as a relative tolerance for termination of iteratively weighted least squares (`method="glm"`) if the relative absolute change in the function (**f** = loglikelihood) falls below `reltol`*(`reltol` + |**f**|).

`gtol` iteratively weighted least squares (`method="glm"`) stops if largest component (in magnitude) of gradient is less than this value.

`etol` iteratively weighted least squares (`method="glm"`) stops when the last relative step length is sufficiently small, i.e. below `etol`*(`etol` + ||**b**||_2), where **b** is current state of the minimizer.

`dm`	non-zero scalar: with `method` set to `"glm"` or `"qp2"` the estimated **delta** is obtained by solving some nonlinear equation with root-finding started in the interval `start.delta` +/- `dm`.
`regress`	whether regression is based on log(`"mu"` or log(`"lambda"`.
`method`	optimization method for maximization of loglikelihood: (i) Broyden-Fletcher-Goldfarb-Shanno, Conjugate Gradient or Nelder-Mead as implemented in `optim`; (ii) iteratively weighted least squares (given **delta**) combined with optimization over **delta** (given the regression coefficients) as in GLM-type models; or (iii) quasi-Poisson-Touchard (QPT) method with two variants: `"qp1"` assumes variance = **mu-delta** and `"qp2"` assumes the exact Touchard variance (see also `dm`).
`N, eps`	arguments passed to `tau`.
`x, y`	logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value.
`...`	not used.

Details

Touchard regression with either $log(\mu)$ or $log(\lambda)$ modeled linearly on the predictors as described in Andrade et al (submitted). Estimation can be performed by maximum likelihood via `optim` with three available methods (`'BFGS', 'CG', 'Nelder-Mead'`) and analytical gradients. Default starting values for the coefficients are obtained from Poisson GLM. Default starting value for δ is obtained by regressing the metameter on the sufficient statistics Y and log(Y+1). Standard errors are obtained from the diagonal of inverse of observed Fisher information as reported at the final iteration.

Estimation may also be performed by combination of iteratively weighted least squares and maximization over δ given current estimate of β. Details are given Andrade et al (submitted).

Finally, estimation can be performed by Poisson Quasi-MLE (or Poisson pseudo-MLE): the estimator for is β is the same as in the Poisson model (which can be thought of as simply a motivation to the first-order condition defining the estimator); the variance is specified independently without restriction of equidispersion. Two specifications are available: (i) a linear specification variance = **mu-delta** which corresponds to an approximation to the Touchard variance and (ii) the exact Touchard variance, both allowing for under- and over-dispersion. Details are given Andrade et al (submitted).

Extractor functions for fitted model objects (of class `"toureg"`): `print`, `summary`, `plot`, `residuals`, `predict`, `cooks.distance`, `hatvalues` and `gleverage`.

`toureg` returns an object of class `"toureg"`, a list with components as described below.

Value

`call`	the original function call.
`coefficients`	named vector of estimated regression coefficients.
`convergence`	integer code from `otim` indicating either successful completion or faulty termination.

`data`	the data provided in the function call.
`delta`	named vector (of length one) of estimated delta parameter.
`df`	residual degrees of freedom in the fitted model.
`fitted.values`	a vector of fitted values of lambda.
`formula`	the formula provided in the function call.
`lambda`	vector of fitted values of lambda.
`loglik`	log-likelihood of the fitted model or pseudo-log-likelihood in case os `method` set to `"glm"`, `"qp1"` or `"qp2"`.
`method`	method used.
`mu`	vector of fitted means.
`residuals`	vector of raw residuals (y - mu).
`se`	standard errors of estimated parameters.
`start.beta, start.delta`	the starting values for the parameters passed to the optimizations routines.
`w`	weights in the (projection) hat matrix analogous to GLMs.
`terms`	the 'terms' object used.
`var`	vector of fitted variances.
`vcov`	covariance matrix of estimates.
`x`	if requested, the model matrix.
`y`	if requested, the response vector.

Author(s)

Bernardo Andrade and Sandro Oliveira

References

Andrade, BB; Matsushita, RY; Oliveira, SB (submitted) Analyzing Count Data with the Touchard Model in R. *available upon request*.

See Also

`glm`, `formula`

Examples

```
### Horseshoe crab data used by several textbook sources
data(Crabs)

### Model Fitting (with different methods) and Plotting
summary( fm <- toureg(y ~ weight + color, data=Crabs) )
# same as
# summary( fm <- toureg(y ~ weight + color, data=Crabs, regress='lambda', method='BFGS') )
```

```
# other methods based on log(mu):
# summary( fm2 <- toureg(y ~ weight + color, data=Crabs, regress='mu', method='glm') )
# summary( fm3 <- toureg(y ~ weight + color, data=Crabs, regress='mu', method='qp1') )


plot(fm)
plot(fm , which = 1)
rgram(fm)

### Diagnostics
plot(hvalues(fm))
plot(gleverage(fm))
plot(cooks.dist(fm))


sum(residuals(fm,'response')^2)
sum(residuals(fm,'pearson')^2)
sum(residuals(fm,'deviance')^2)



### Predicted values for 'newdata' ###

# Predicted mean values (on the scale of the response variable, i.e. \hat{\mu}):
predict(fm, newdata=data.frame(weight=c(5,6), color=c(2,4)), type="response", se.fit=TRUE)
# Predicted values of lambda:
predict(fm, newdata=data.frame(weight=c(5,6), color=c(2,4)), type="lambda", se.fit=TRUE)
# Predicted values of the linear predictor  x'beta, SEs not yet available:
predict(fm, newdata=data.frame(weight=c(5,6), color=c(2,4)), type="linpred")
# Predicted variances, i.e. \hat{\sigma}^2, SEs not yet available:
predict(fm, newdata=data.frame(weight=c(5,6), color=c(2,4)), type="variance")
```

Index