

# Package: Rcapture (via r-universe)

September 12, 2024

**Type** Package

**Title** Loglinear Models for Capture-Recapture Experiments

**Version** 1.4-4

**Date** 2022-05-03

**Description** Estimation of abundance and other demographic parameters for closed populations, open populations and the robust design in capture-recapture experiments using loglinear models.

**Imports** stats, graphics, utils

**LazyData** true

**License** GPL-2

**NeedsCompilation** no

**Author** Louis-Paul Rivest [aut, cre], Sophie Baillargeon [aut]

**Maintainer** Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca>

**Repository** CRAN

**Date/Publication** 2022-05-04 13:20:06 UTC

## Contents

Rcapture-package . . . . .	2
BBS2001 . . . . .	7
bunting . . . . .	9
cbird . . . . .	10
closedp . . . . .	11
closedp.bc . . . . .	16
closedp.custom . . . . .	19
closedp.Mtb . . . . .	21
closedpCI . . . . .	22
closedpMS . . . . .	29
descriptive . . . . .	32
duck . . . . .	35
hare . . . . .	36
histpos . . . . .	37

HIV . . . . .	39
ill . . . . .	40
lesbian . . . . .	41
mvole . . . . .	42
openp . . . . .	44
periodhist . . . . .	47
profileCI . . . . .	48
rbvole . . . . .	50
robustd . . . . .	51
uifit . . . . .	54

<b>Index</b>	<b>56</b>
--------------	-----------

---

Rcapture-package	<i>Loglinear Models for Capture-Recapture Experiments</i>
------------------	---

---

## Description

Estimation of abundance and of other demographic parameters for closed populations, open populations and the robust design in capture-recapture experiments using loglinear models.

This package focuses on closed populations. Since version 1.2-0, no new features have been added to open populations and robust design functions.

## Details

Package:	<b>Rcapture</b>
Type:	Package
Version:	1.4-4
Date:	2022-05-03
License:	GPL-2

## SUMMARY OF **Rcapture** CONTENTS

The **Rcapture** package contains nine capture-recapture data sets and the following functions:

- Model fitting functions for
  - > closed populations:
    - `closedp` functions: fit various loglinear models for abundance estimation,
    - `closedpCI` functions: fit one customized loglinear model and calculate a confidence interval for the abundance estimation,
    - `closedpMS.t`: fits various hierarchical loglinear models in a perspective of model selection,
    - `closedp.bc`: performs bias corrections to the abundance estimations from customized loglinear models,
    - `closedp.Mtb`: fits model Mtb, which cannot be fitted by any other function, for abundance estimation;

- > open populations:
  - `openp`: computes various demographic parameters using a loglinear model;
- > robust design:
  - `robustd` functions: compute various demographic parameters and capture probabilities per period using a loglinear model.
- Descriptive statistics functions:
  - `descriptive`: produces basic descriptive statistics for capture-recapture data;
  - `uifit`: produces fit statistics concerning the  $u_i$ , i.e. the numbers of first captures on each capture occasion, for closed population models.
- Data manipulation functions:
  - `histpos` functions: builds a matrix of observable capture histories;
  - `periodhist`: merges capture occasions.

## DESCRIPTION OF DATA SET FORMATS

In capture-recapture experiments, the data collected consist of capture histories for captured units. A capture history is simply a serie of capture indicators for each capture event in the experiment. The capture history of one unit is expressed as a length  $t$  vector  $w = (w_1, \dots, w_t)$ , where  $w_j = 1$  if the unit is captured at the  $j$ th occasion and 0 if not. For closed populations, capture events are named capture occasions, whereas they are named capture periods for open populations.

Capture-recapture data sets are given to **Rcapture** functions through the  $X$  argument.  $X$  must be a numeric matrix. Arguments `dfreq` and `dtype` indicate the format of the matrix. Each have two possible values, meaning that four data set formats are possible with **Rcapture**.

### FORMAT 1 - CAPTURE HISTORY PER UNIT

If `dfreq=FALSE` and `dtype="hist"` (the default),  $X$  has one row per unit captured in the experiment. Each row is an observed capture history. It must contain only zeros and ones; the number one indicates a capture. In this case, the number of columns in the table represents the number of capture occasions in the experiment (noted  $t$ ). Here is an example of a data set of this type for  $t = 2$ :

```
1 1
1 1
1 0
1 0
1 0
1 0
0 1
```

### FORMAT 2 - AGGREGATED CAPTURE HISTORIES

If `dfreq=TRUE` and `dtype="hist"`,  $X$  contains one row per observed capture history followed by its frequency. In that case,  $X$  has  $t+1$  columns. As for format 1, the first  $t$  columns of  $X$ , identifying the capture histories, must contain only zeros and ones. The number one indicates a capture. In this format, the example data set is represented by the following matrix:

```
1 1 2
1 0 4
0 1 1
```

If a possible capture history is not observed, it can appear in  $X$  with a frequency of zero, or it can simply be omitted.

#### FORMAT 3 - NUMBER OF CAPTURES PER UNIT

If `dfreq=FALSE` and `dtype="nbcap"`,  $X$  is a vector with the number of captures for every captured unit. Therefore, this format does not contain complete capture histories. Instead, capture histories are summarized through the number of captures. In this format, the example data set looks like:

```
2 2 1 1 1 1 1
```

#### FORMAT 4 - AGGREGATED NUMBERS OF CAPTURES

If `dfreq=TRUE` and `dtype="nbcap"`,  $X$  is a 2 columns matrix. The first column contains the observed numbers of captures, the second column contains their frequencies. In this format, the example data is:

```
2 2
1 5
```

#### DETAILS ABOUT FORMATS WITH NUMBERS OF CAPTURES

Only few functions have the `dtype` argument. Functions without `dtype` argument accept only a data matrix  $X$  of the form `dtype="hist"`. So the first two formats listed above are the most common.

Formats with `dtype="nbcap"` are used for captures in continuous time (see below). They are also useful to reduce the size of the data set for experiments with a large number of capture occasions  $t$  (often with no units caught a large number of times). For these formats, the number of capture occasions  $t$  cannot be deduced from  $X$  as it can be with `dtype="hist"`. One has no guarantee that the larger number of captures observed is the total number of capture occasions. Therefore, if one gives a data matrix  $X$  with `dtype="nbcap"`, one must also provide  $t$ , the number of capture occasions, as an additional argument.

For now, the data formats with `dtype="nbcap"` are not generalized to the robust design. So `dtype` is not an argument of the `robustd.0` function.

#### CAPTURES IN CONTINUOUS TIME

In some capture-recapture experiments, there is no well defined capture occasions. Captures occur in continuous time. The data set `ill` comes from such an experiment. Bohning and Schon (2005) call this type of capture-recapture data *repeated counting data*. These data sets always have the format `dtype="nbcap"`.

We can estimate abundance for data of this type using the option `t=Inf` with the functions `closedpCI.0` and `closedpCI.0`. The function `descriptive` also accepts `t=Inf`. It modifies the  $y$  coordinate of the exploratory heterogeneity graph.

#### DISTINCTION BETWEEN `.t` and `.0` FUNCTIONS

Capture recapture models for closed population aim at estimating the population size by modelling the probabilities of the different capture histories. The data available to fit these models consist of observed frequencies of capture histories. These frequencies are modeled in Rcapture using loglinear models for frequency tables. For functions with a name ending with `.t` (`closedp.t`, `closedpCI.t`, `closedpMS.t` and `robustd.t`), the observed values of the response variable is the vector of frequencies for every observable capture history. It has length  $2^t - 1$ .

However, for a model without temporal effect (assuming that the probability of capturing a unit does not vary between capture occasions), all the information needed to fit the model can be found in aggregated data. Functions with a name ending with `.0` (`closedp.0`, `closedpCI.0` and `robustd.0`) fit models using as response variable the number of units captured  $i$  times, for  $i = 1, \dots, t$ . It is a vector of length  $t$ , which is much shorter than  $2^t - 1$  for a large  $t$ . Because of an appropriate offset added to the model, the results from a `.0` function match exactly the results from a `.t` function for a corresponding capture-recapture model. Because `.0` functions deal with smaller design matrix, they run faster than `.t` functions, but they cannot fit models with a temporal effect.

## FUNCTIONS USED FOR MODEL FITTING

Most of the **Rcapture** functions use the function `glm` of the **stats** package to fit a loglinear model. However, the function `optim`, again of the **stats**, is also used by three functions:

- `closedpCI.t`, `closedpCI.0` and `closedpMS.t`: when a normal heterogeneous models is requested,
- `closedp.Mtb`: because model `Mtb` does not have a loglinear form.

## ERRORS AND WARNINGS MANAGEMENT IN CLOSED POPULATION FUNCTIONS

If an error occurs while executing a function fitting only one closed population model (`closedpCI` functions, `closedp.bc`, `closedp.Mtb`), the execution is stopped and the error message is printed (usual behavior in R). However, if an error occurs while fitting a model in a call to a `closedp` population function fitting more than one model (`closedp` functions, `closedpMS.t`), the execution of the call is not stopped. Instead, the row in the results table for the problematic model is filled with NA and the error message is stored in an output value. This value is called `glm.err` for `closedp` functions. It is however called `fit.err` for `closedpMS.t` because this function do not always use `glm` to fit the model (as mentionned above, `optim` is used for normal heterogeneous models).

Warning messages while fitting a closed population model, if any, are stored in an output value called `'glm.warn'`, `'fit.warn'` or `'optim.warn'` depending on the function. They are not printed in the console. To inform the user that a warning occurred, the last column of the results table, named `infoFit`, contains a numerical code giving information about errors or warnings encountered. Here is a description of the meaning of the numerical code:

- 0** no error or warning occurred while fitting the model;
- 1** an error occurred while fitting the model;
- 1** a warning indicating that the model fit is questionable occurred (algorithm did not converge, non-positive sigma estimate for a normal heterogeneous model or large asymptotic bias);
- 2** the warning 'design matrix not of full rank' occurred, therefore some model's coefficients are not estimable;
- 3** a warning not of type 1 or 2 occurred (the `glm` warning 'fitted rates numerically 0 occurred' is often encountered with small frequencies, it does not always mean that the model fit is questionable).

The elements in the column `infoFit` can contain more than one number since more than one warning can occur. For example, if `infoFit` takes the value 13 for a model, it means that at least one warning of type 1 and one warning of type 3 have occurred. If more than one warning of the same type are encountered, the number representing the type of warning is not repeated in `infoFit`.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Bohning, D. and Schon, D. (2005) Nonparametric Maximum Likelihood Estimation of Population Size Based on the Counting Distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**(4), 721-737.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**(4), 783-791.
- Cormack, R. M. (1985) Example of the use of glim to analyze capture-recapture studies. In *Lecture Notes in Statistics 29: Statistics in Ornithology*, Morgan, B. et North, P. editors, New York,: Springer-Verlag, 242-274.
- Cormack, R. M. (1989) Loglinear models for capture-recapture. *Biometrics*, **45**, 395-413.
- Cormack, R. M. (1992) Interval estimation for mark-recapture studies of closed populations. *Biometrics*, **48**, 567-576.
- Cormack, R. M. (1993) Variances of mark-recapture estimates. *Biometrics*, **49**, 1188-1193.
- Cormack, R. M. and Jupp, P. E. (1991) Inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika*, **78**(4), 911-916.
- Rivest, L.P. and Levesque, T. (2001) Improved loglinear model estimators of abundance in capture-recapture experiments. *Canadian Journal of Statistics*, **29**, 555-572.
- Rivest, L.P. and Daigle, G. (2004) Loglinear models for the robust design in mark-recapture experiments. *Biometrics*, **60**, 100-107.
- Rivest, L.P. and Baillargeon, S. (2007) Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics*, **63**(4), 999-1006.
- Rivest, L.P. (2008) Why a time effect often has a limited impact on capture-recapture estimates in closed populations. *Canadian Journal of Statistics*, **36**(1), 75-84.

**Examples**

```
# Here is an example on the lesbian data set.

desc <- descriptive(lesbian, dfreq = TRUE)
desc
plot(desc)

# 1612 out of 2185 individuals (74%) appear on one list only.
# The exploratory heterogeneity graph are not quite linear.
# Some heterogeneity in the units capture probabilities
# seem present in the data set.

closedp(lesbian, dfreq = TRUE)

# According to the BIC, the best model is Mth Darroch.
```

```

# Let's see if adding interactions between capture
# histories to the model could improve the model's fit.

closedpMS.t(lesbian, dfreq = TRUE, h = "Darroch")

# According to the BIC, the best heterogeneous Darroch model
# contains the double interactions 12, 13, 14.
# Here is the profile likelihood confidence interval for the
# abundance estimation from this model.

closedpCI.t(lesbian, dfreq = TRUE, mX = "[12,13,14]", h = "Darroch")

#####

# Example to illustrate warnings management in closed population functions.

# Here is a capture-recapture data set one could encounter.

crdata <- cbind(histpos.t(4), c(0,0,3,0,0,0,0,0,0,0,0,1,0,0,2))

# This data set contains 4 capture occasions but only 6 captured units.
# Fitting capture-recapture models on this data set is quite useless.
# The population size should be very close to the sample size.

# Such small frequencies in a capture-recapture data set should
# lead to warnings when fitting a loglinear model on it.

ex <- closedp.t(crdata, dfreq=TRUE)
ex

# Many models produce warnings of type 1 indicating that the model fit
# is questionable. The very large abundance estimation for some models
# are another indicator of questionable model fits.
# Details about the warnings are found in the glm.warn element of the output.

ex$glm.warn

```

BBS2001

*Species Richness Data from the North American Breeding Bird Survey  
(BBS) in 2001*

## Description

This data set contains species richness data from the North American Breeding Bird Survey (BBS) in 2001. The number of capture occasions  $t$  is 50.

## Usage

```
BBS2001
```

## Format

36 by 2 numeric matrix, with the following columns:

nbcap Numbers of captures

freq Observed frequencies for each number of captures

## Details

This data set is presented in Dorazio and Royle (2003). It comes from the North American Breeding Bird Survey (BBS). This survey includes more than 4000 39.4 km survey routes throughout North America. Along each survey route an observer stops at 50 equidistant locations and records the species identity and number of all birds heard or seen within a 3-minute period.

The data in BBS2001 was collected along Route 123, in Maryland, in 2001. In this data set a capture occasion is in fact a location and the frequencies represent the number of species detected at  $i$  locations out of the 50 locations on that route. The frequencies for  $i = 37, \dots, 50$  are null and do not appear in the data set.

## Source

Dorazio, R. M. and Royle, A. J. (2003) Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, **59**, 351–364

## Examples

```
desc <- descriptive(BBS2001, dfreq = TRUE, dtype = "nbcap", t = 50)
plot(desc)
# Note in this plot the convex shape typical of model Mh.

cp <- closedp.0(BBS2001, dfreq = TRUE, dtype = "nbcap", t = 50, t0 = 20)
cp
plot(cp)
# The Gamma estimator has the smallest deviance and all
# its residuals are small.

# Let's calculate a profile confidence interval for the gamma
# estimator:
closedpCI.0(BBS2001, dfreq = TRUE, dtype = "nbcap", t = 50, t0 = 20,
            m = "Mh", h = "Gamma", h.control = list(theta = 3.5))

# One can also calculate a profile confidence interval for Chao's
# lower bound:
closedpCI.0(BBS2001, dfreq = TRUE, dtype = "nbcap", t = 50, t0 = 20,
            m = "Mh", h = "Chao")
```



bunting

*Lazuli Bunting Data***Description**

This data set contains frequencies of capture histories for 8 years of observation of lazuli buntings.

**Usage**

bunting

**Format**

255 by 9 numeric matrix, with the following columns:

p1, p2, p3, p4, p5, p6, p7, p8 Capture histories from eight periods

freq Observed frequencies for each capture history

**Details**

The data come from an eight-year (1973 to 1980) study by Allen W. Stokes of lazuli buntings wintering in Logan, Utah. This data set is analyzed in Burnham and al. (1987) and in Cormack (1993).

Each row of this data set represents an observed capture history followed by its frequency.

**References**

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C. and Pollock, K.H. (1987) Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society Monographs* 5. Bethesda, Maryland.

Cormack, R. M. (1993) The flexibility of GLIM analyses of multiple recapture or resighting data. In J.D. Lebreton and P. North editors: *Marked Individuals in the Study of Bird Population*. Basel, Switzerland: Birkhauser Verlag, 39–49.

**Examples**

```
descriptive(bunting, dfreq = TRUE)
# 1430 birds among the 1681 birds seen (85%) were caught only once.
# This suggests the presence of transient birds at each capture occasion.

op.m1 <- openp(bunting, dfreq = TRUE)
op.m1$model.fit
plot(op.m1)
# The residuals plot shows large residuals for the birds caught twice or
# more while the residuals are small for birds caught once. The Jolly-Seber
```

```

# model does not fit well and the likely presence of transients might
# cause that. Let's remove the birds caught only once from the analysis.

keep2 <- rowSums(histpos.t(8)) > 1
op.m2<- openp(bunting, dfreq = TRUE, keep = keep2)
op.m2$model.fit
# The deviance drop of 94 for 6 degrees of freedom is highly significant.
plot(op.m2)
# The residual plot still shows Pearson residuals larger than 4. We redo
# the analysis without the transient birds and without the large residuals.

keep3p <- residuals(op.m2$glm, type = "pearson") < 4
num3 <- ((1:255)[keep2])[keep3p]
keep3 <- rep(FALSE, 255)
keep3[num3] <- TRUE
op.m3 <- openp(bunting, dfreq = TRUE, keep = keep3)
cbind(op.m2$survivals, op.m3$survivals)
# These changes have little impact on the survival estimates.

# We now test the equality of the survival probabilities and estimate its
# common value phi. Least squares estimates of phi and its standard error:
siginv <- solve(op.m2$cov[8:12, 8:12])
phi <- t(rep(1, 5)) %*% siginv %*% op.m2$survivals[2:6, 1] /
  (t(rep(1, 5)) %*% siginv %*% rep(1, 5))
se <- 1 / sqrt(t(rep(1, 5)) %*% siginv %*% rep(1, 5))
cbind(estimate = phi, stderr = se)
# The chi-square goodness of fit statistic for a constant survival
# and its pvalue are:
chisq4 <- t(op.m2$survivals[2:6, 1] - as.vector(phi)) %*% siginv %*%
  (op.m2$survivals[2:6, 1] - as.vector(phi))
cbind(chi2stat = chisq4, pvalue = 1 - pchisq(chisq4, df = 4))
# The hypothesis of a constant survival is accepted.

```

---

cbird

*Catbird Site Occupancy Data*


---

## Description

This data set contains site occupancy data for catbirds. The number of capture occasions  $t$  is 11.

## Usage

```
cbird
```

## Format

6 by 2 numeric matrix, with the following columns:

nbcap Numbers of captures

freq Observed frequencies for each number of captures

## Details

This data set is presented in Royle (2006). It comes from an experiment in which 50 locations have been visited 11 times. The presence or absence of catbirds was noted at each visit. The experiment aims at estimating the number of locations ( $\leq 50$ ) at which catbirds are present.

In the data set `cbird`, the frequencies represent the number of locations at which `cbirdirbs` have been detected  $i$  times, out of  $t = 11$  visits. The frequencies for  $i = 7, \dots, 11$  are null and do not appear in the data set.

## Source

Royle, A. J. (2006) Site occupancy models with heterogeneous detection probabilities. *Biometrics*, **62**, 97–102

## Examples

```
desc <- descriptive(cbird, dfreq = TRUE, dtype = "nbcap", t = 11)
plot(desc)
# The heterogeneity graph is mildly convex, mostly because of
# 2 sites where catbirds have been seen six times.

closedp.0(cbird, dfreq = TRUE, dtype = "nbcap", t = 11)
# The residuals are OK. The occupancy rate (N/50) estimate
# varies between 40.6% for M0 and 49.4% for Mh Gamma.
# One could also try fitting M0 removing the 2 unusual sites:
closedp.0(cbird, dfreq = TRUE, dtype = "nbcap", t = 11, t0 = 5)
# M0 fits very well and the occupancy rate estimate is 42.6%.
```

---

closedp

*Loglinear Models for Closed Population Capture-Recapture Experiments*

---

## Description

The functions `closedp.t` and `closedp.0` fit various loglinear models for closed populations in capture-recapture experiments. For back compatibility, `closedp.t` is also named `closedp`.

`closedp.t` fits more models than `closedp.0` but for data set with more than 20 capture occasions, the function might fail. However, `closedp.0` works with fairly large data sets (see **Details**).

## Usage

```
closedp(X, dfreq=FALSE, neg=TRUE, ...)
closedp.t(X, dfreq=FALSE, neg=TRUE, ...)

closedp.0(X, dfreq=FALSE, dtype=c("hist","nbcap"), t=NULL, t0=NULL,
          neg=TRUE, ...)

## S3 method for class 'closedp'
```

```

print(x, ...)

## S3 method for class 'closedp'
boxplot(x, main="Boxplots of Pearson Residuals", ...)

## S3 method for class 'closedp'
plot(x, main="Residual plots for some heterogeneity models", ...)

```

### Arguments

<code>x</code>	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
<code>dfreq</code>	A logical. By default FALSE, which means that <code>x</code> has one row per unit. If TRUE, it indicates that the matrix <code>x</code> contains frequencies in its last column.
<code>dtype</code>	A characters string, either "hist" or "nbcap", to specify the type of data. "hist", the default, means that <code>x</code> contains complete observed capture histories. "nbcap" means that <code>x</code> contains numbers of captures (see <a href="#">Rcapture-package</a> for details on data formats).
<code>t</code>	Requested only if <code>dtype="nbcap"</code> . A numeric specifying the total number of capture occasions in the experiment. For <code>closedp.0</code> , the value <code>t=Inf</code> is accepted. It indicates that captures occur in continuous time (see <a href="#">Rcapture-package</a> ).
<code>t0</code>	A numeric. Models are fitted considering only the frequencies of units captured 1 to <code>t0</code> times. By default, if <code>t</code> is not equal to <code>Inf</code> , <code>t0=t</code> . When <code>t=Inf</code> , the default value of <code>t0</code> is the larger number of captures observed.
<code>neg</code>	If this option is set to TRUE, negative eta parameters in Chao's lower bound models are set to zero (see <b>Details</b> ).
<code>...</code>	Further arguments to be passed to <code>glm</code> , <code>print.default</code> , <code>boxplot.default</code> or <code>plot.default</code> .
<code>x</code>	An object, produced by a <code>closedp</code> function, to print or to plot.
<code>main</code>	A main title for the plot.

### Details

`closedp.t` fits models M0, Mt, Mh Chao (LB), Mh Poisson2, Mh Darroch, Mh Gamma3.5, Mth Chao (LB), Mth Poisson2, Mth Darroch, Mth Gamma3.5, Mb and Mbh. `closedp.0` fits only models M0, Mh Chao (LB), Mh Poisson2, Mh Darroch and Mh Gamma3.5. However, `closedp.0` can be used with larger data sets than `closedp.t`. This is explained by the fact that `closedp.t` fits models using the frequencies of the observable capture histories (vector of size  $2^t - 1$ ), whereas `closedp.0` uses the numbers of units captured  $i$  times, for  $i = 1, \dots, t$  (vector of size  $t$ ). See [Rcapture-package](#) for more details about the distinction between `.t` and `.0` functions.

Multinomial profile confidence intervals for the abundance are constructed by `closedpCI.t` and `closedpCI.0`.

To calculate bias corrected abundance estimates, use the `closedp.bc` function.

### CHAO'S LOWER BOUND MODEL

Chao's (or LB) models estimate a lower bound for the abundance, both with a time effect (Mth Chao) and without one (Mh Chao). The estimate obtained under Mh Chao is Chao's (1987) moment estimator. Rivest and Baillargeon (2007) exhibit a loglinear model underlying this estimator and provide a generalization to Mth. For these two models, a small deviance means that there is a heterogeneity in capture probabilities; it does not mean that the lower bound estimate is unbiased. To test whether a certain model for heterogeneity is adequate, one can conduct a likelihood ratio test by subtracting the deviance of Chao's model to the deviance of the heterogeneous model under study. If this heterogeneous model includes a time effect, it must be compared to model Mth Chao. If it does not include a time effect, it must be compared to model Mh Chao. Under the null hypothesis of equivalence between the two models, the difference of deviances follows a chi-square distribution with degrees of freedom equal to the difference between the models' degrees of freedom.

Chao's lower bound models contain  $t - 2$  parameters, called eta parameters, for the heterogeneity. These parameters should theoretically be greater or equal to zero (see Rivest and Baillargeon (2007)). When the argument `neg` is set to `TRUE` (the default), negative eta parameters are set to zero (to do so, columns are removed from the design matrix of the model). Degrees of freedom of Chao's model increase when eta parameters are set to zero.

#### OTHER MODELS FOR HETEROGENEITY

Other models for heterogeneity are defined as follows :

Model	Column for heterogeneity in the design matrix
Poisson2	$2^k - 1$
Darroch	$k^2/2$
Gamma3.5	$-\log(3.5 + k) + \log(3.5)$

where  $k$  is the number of captures. Poisson and Gamma models with alternative to the parameter defaults values 2 and 3.5 can be fitted with the `closedpCI.t` and `closedpCI.0` functions.

Darroch's models for Mh and Mth are considered by Darroch et al. (1993) and Agresti (1994). Poisson and Gamma models are discussed in Rivest and Baillargeon (2007). Poisson models typically yield smaller corrections for heterogeneity than Darroch's model since the capture probabilities are bounded from below under these models. On the other hand, Gamma models can lead to very large estimators of abundance. We suggest considering this estimator only in experiments where very small capture probabilities are likely.

#### PLOT METHODS AND FUNCTIONS

The `boxplot.closedp` function produces boxplots of the Pearson residuals of the fitted loglinear models that converged.

The `plot.closedp` function produces scatterplots of the Pearson residuals in terms of  $f_i$  (number of units captured  $i$  times) for the heterogeneous models Mh Poisson2, Mh Darroch and Mh Gamma3.5 if they converged.

#### Value

`n` The number of captured units.

t	The total number of capture occasions in the data matrix $X$ . When captures occur in continuous time (input argument $t=Inf$ ), this output value is the larger number of captures observed.
t0	For <code>closedp.0</code> only: the value of the argument <code>t0</code> used in the computations.
results	A table containing, for every fitted model: <b>abundance:</b> the estimated population size, <b>stderr:</b> the standard error of the estimated population size, <b>deviance:</b> the model's deviance, <b>df:</b> the number of degrees of freedom, <b>AIC:</b> the Akaike's information criterion, <b>BIC:</b> the bayesian information criterion, <b>infoFit:</b> a numerical code giving information about error or warnings encountered when fitting the model (see <a href="#">Rcapture-package</a> for details).
bias	A vector, the asymptotic bias of the estimated population size for every fitted model.
glm	A list of the 'glm' objects obtained from fitting models.
glm.err	A list of character string vectors. If the <code>glm</code> function generates an error when fitting a model, a copy of the error message is stored in <code>glm.err\$mname</code> where <code>mname</code> is a short name to identify the model. A NULL list element means that no error occurred for the considered model.
glm.warn	A list of character string vectors. If the <code>glm</code> function generates one or more warnings when fitting a model, a copy of these warnings are stored in <code>glm.warn\$mname</code> where <code>mname</code> is a short name to identify the model. A NULL list element means that <code>glm</code> did not produce any warnings for the considered model.
parameters	Capture-recapture parameters estimates. It contains $N$ , the estimated population size, and $p$ or $p_1$ to $p_t$ defined as follows for the different models : <b>M0:</b> the capture probability at any capture occasion, <b>Mt:</b> the capture probabilities for each capture occasion, <b>Mh models:</b> the average probability of capture, <b>Mth models:</b> the average probabilities of capture for each occasion, <b>Mb and Mbh:</b> the probability of first capture at any capture occasion. For models <code>Mb</code> and <code>Mbh</code> , it also contains $c$ , the recapture probability at any capture occasion.
neg.eta	The position of the eta parameters set to zero in the loglinear parameter vector of models <code>MhC</code> and <code>MthC</code> . NULL if no parameters are set to zero.
X	A copy of the data given as input in the function call.
dfreq	A copy of the <code>dfreq</code> argument given in the function call.

### Note

This function uses the `glm` function of the `stats` package.

### Author(s)

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

## References

- Agresti, A. (1994) Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics*, **50**, 494–500.
- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **45**, 427–438.
- Darroch, S.E., Fienberg, G., Glonek, B. and Junker, B. (1993) A three sample multiple capture-recapture approach to the census population estimation with heterogeneous catchability. *Journal of the American Statistical Association*, **88**, 1137–1148.
- Rivest, L.P. and Levesque, T. (2001) Improved loglinear model estimators of abundance in capture-recapture experiments. *Canadian Journal of Statistics*, **29**, 555–572.
- Rivest, L.P. and Baillargeon, S. (2007) Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics*, **63**(4), 999–1006.
- Seber, G.A.F. (1982) *The Estimation of Animal Abundance and Related Parameters*, 2nd edition, New York: Macmillan.

## See Also

[closedpCI.t](#), [closedpCI.0](#), [closedp.bc](#), [closedp.Mtb](#), [closedpMS.t](#), [uifit](#).

## Examples

```
# hare data set
hare.closedp <- closedp.t(hare)
hare.closedp
boxplot(hare.closedp)

# Third primary period of mvole data set
period3 <- mvole[, 11:15]
closedp.t(period3)

# BBS2001 data set
BBS.closedp <- closedp.0(BBS2001, dfreq = TRUE, dtype = "nbcap", t = 50, t0 = 20)
BBS.closedp
plot(BBS.closedp)

### Seber (1982) p.107
# When there is 2 capture occasions, the heterogeneity models cannot be fitted
X <- matrix(c(1,1,167,1,0,781,0,1,254), byrow = TRUE, ncol = 3)
closedp.t(X, dfreq = TRUE)

### Example of captures in continuous time
# Illegal immigrants data set
closedp.0(ill, dtype = "nbcap", dfreq = TRUE, t = Inf)
```

closedp.bc

*Bias Correction for Closed Population Loglinear Models***Description**

This function applies a bias correction to the abundance estimations obtained by closed population models.

**Usage**

```
closedp.bc(X, dfreq=FALSE, dtype=c("hist", "nbcap"), t=NULL, t0=t,
           m=c("M0", "Mt", "Mh", "Mth", "Mb", "Mbh"), h=NULL, h.control=list(), ...)

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

**Arguments**

X	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
dfreq	A logical. By default FALSE, which means that X has one row per unit. If TRUE, it indicates that the matrix X contains frequencies in its last column.
dtype	A characters string, either "hist" or "nbcap", to specify the type of data. "hist", the default, means that X contains complete observed capture histories. "nbcap" means that X contains numbers of captures (see <a href="#">Rcapture-package</a> for details on data formats). If m is "Mt", "Mth", "Mb" or "Mbh", dtype must be "hist".
t	Requested only if dtype="nbcap". A numeric specifying the total number of capture occasions in the experiment.
t0	A numeric used for model M0 or for an Mh model other than Chao's lower bound model : Models are fitted considering only the frequencies of units captured 1 to t0 times. By default t0=t.
m	A character string indicating the model to fit, either "M0"=M0 model, "Mt"=Mt model, "Mh"=Mh model, "Mth"=Mth model, "Mb"=Mb model, "Mbh"=Mbh model.
h	A character string ("LB", "Chao", "Poisson", "Darroch" or "Gamma") or a numerical R function specifying the form of the column(s) for heterogeneity in the design matrix. "LB" and "Chao" represents Chao's lower bound model (the default), "Poisson" represents the function $f(k) = \theta^k - 1$ , "Darroch" represents the function $f(k) = k^2/2$ , and "Gamma" represents the function $f(k) = -\log(\theta + k) + \log(\theta)$ where $k$ is the number of captures and $\theta$ is a parameter specified with the argument theta. If an R function is given, it must be the implementation of any convex mathematical function $f(k)$ (it must have only one argument).



h.control	A list of elements to control the heterogeneous part of the model, if any. For a Poisson or Gamma heterogeneous model:
theta	The value of the parameter $theta$ in $f(k) = theta^k - 1$ for the Poisson model (the default value is 2) and in $f(k) = -\log(theta + k) + \log(theta)$ for the Gamma model (the default value is 3.5).
	For the Chao's lower bound Mth model:
neg	If this option is set to TRUE (the default), negative eta parameters in Chao's lower bound models are set to zero.
...	Further arguments to be passed to <code>glm</code> or <code>print.default</code> .
x	An object, produced by the <code>closedp.bc</code> function, to print.

## Details

For the Mt model:

When  $t=2$ , `closedp.bc` returns the Petersen estimator with Chapman's (1951) bias correction and the bias corrected standard error estimator of Seber (1970) and Wittes (1972). For  $t>2$ , `closedp.bc` implements the bias correction of Rivest and Levesque (2001). The estimate for N and its variance are calculated by solving an estimating equation as proposed in Seber (1982), not by fitting a Poisson regression. This approach works for large values of  $t$ .

For other models:

The bias correction is done through frequency modifications in Poisson regression as described in Rivest and Levesque (2001). The variances calculated with the modified frequencies are less biased than the standard ones, but they can overestimate the mean squared errors, especially when the data is sparse.

This function works with fairly large data set, except if an "Mth" model is requested. In this case, only heterogeneity of the form "LB", "Chao", "Poisson" with  $theta=2$  or "Darroch" is accepted.

## Value

n	The number of captured units
t	The total number of capture occasions in the data matrix X.
t0	For models M0 and Mh only: the value of the argument <code>t0</code> used in the computations.
results	A table containing, for the fitted model: <b>abundance</b> : the estimated population size, <b>stderr</b> : the standard error of the estimated population size, <b>infoFit</b> : a numerical code giving information about error or warnings encountered when fitting the model (see <a href="#">Rcapture-package</a> for details).
glm.warn	Only if the corrected population size estimation was obtained with <code>glm</code> : A vector of character strings. If the <code>glm</code> function generates one or more warnings when fitting the model, a copy of these warnings are stored in <code>glm.warn</code> . NULL if <code>glm</code> did not produce any warnings.
neg.eta	For Chao's lower bound model Mth only: the position of the eta parameters set to zero in the loglinear parameter vector, if any.

**Note**

This function uses the `glm` function of the **stats** package, except for models M<sub>t</sub> and M<sub>h</sub> Chao's lower bound for which exact calculation is performed.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Chapman, D. G. (1951) Some properties of the hypergeometric distribution with applications to zoological sample censuses. *University of California Publications in Statistics*, **1**(7), 131-160.
- Rivest, L.P. and Levesque, T. (2001) Improved loglinear model estimators of abundance in capture-recapture experiments. *Canadian Journal of Statistics*, **29**, 555-572.
- Seber, G.A.F. (1970) The effects of trap response on tag recapture estimates. *Biometrics*, **26**, 13-22.
- Seber, G.A.F. (1982) *The Estimation of Animal Abundance and Related Parameters*, 2nd edition. New York: Macmillan.
- Wittes, J.T. (1972) On the bias and estimated variance of Chapman's two-sample capture-recapture population estimate. *Biometrics*, **28**, 592-597.

**See Also**

[closedp](#)

**Examples**

```
# Third primary period of mvole data set
period3 <- mvole[, 11:15]
closedp.bc(period3, m = "Mh", h = "Darroch")
closedp.bc(period3, m = "Mh", h = "Gamma", h.control = list(theta = 3.5))

# BBS2001 data set
closedp.bc(BBS2001, dfreq = TRUE, dtype = "nbcap", t = 50, t0 = 20,
           m = "Mh", h = "Gamma", h.control = list(theta = 3.5))

# Seber (1982) p.107
# When there are 2 capture occasions, only models M0 and Mt can be fitted
X <- matrix(c(1,1,167,1,0,781,0,1,254), byrow = TRUE, ncol = 3)
closedp.bc(X, dfreq = TRUE, m = "M0")
closedp.bc(X, dfreq = TRUE, m = "Mt")
```

---

closedp.custom	<i>Customization of a Loglinear Model for Closed Populations Capture-Recapture Experiments</i>
----------------	--

---

## Description

As of version 1.2-0 of **Rcapture**, these functions are deprecated, but kept for back compatibility. Please use `closedpCI.t` instead.

The `closedp.mX` function fits a loglinear model given a design matrix `mX`. The `closedp.h` function fits Mh or Mth models for which the form of the column for heterogeneity in the design matrix is determined by the user.

## Usage

```
closedp.mX(X, dfreq=FALSE, mX, mname="Customized model")
```

```
closedp.h(X, dfreq=FALSE, m="Mh", h="Poisson", a=2)
```

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

```
## S3 method for class 'closedp.custom'
boxplot(x, ...)
```

## Arguments

<code>X</code>	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
<code>dfreq</code>	A logical. By default FALSE, which means that <code>X</code> has one row per unit. If TRUE, it indicates that the matrix <code>X</code> contains frequencies in its last column.
<code>mX</code>	The design matrix of the loglinear model. In this matrix, the order of the capture histories is as defined in the <code>histpos.t</code> function.
<code>mname</code>	A character string specifying the name of the customized model.
<code>m</code>	A character string indicating the model to fit, either "Mh"=Mh model or "Mth"=Mth model
<code>h</code>	The character string "Poisson" ( $2^k - 1$ ) or a numerical function taking only one argument specifying the form of the column for heterogeneity in the design matrix.
<code>a</code>	The value of the exponent's base for a Poisson model.
<code>x</code>	An object, produced by the <code>closedp.mX</code> or the <code>closedp.h</code> function, to print or to plot.
<code>...</code>	Further arguments passed to or from other methods.

**Details**

An intercept is added to the model. Therefore, the  $mX$  matrix must not contain a column of ones.

The abundance estimation is calculated as the number of captured units plus the exponential of the intercept. Therefore, these functions are not suited for models with a behavioral effect.

In `closedp.h`, the argument `h` cannot take the value "Chao" or "Darroch" because these models are already fitted by the `closedp` function.

The `boxplot.closedp.custom` function produces a boxplot of the pearson residuals of the customized model.

**Value**

<code>n</code>	The number of captured units
<code>results</code>	A table containing the estimated population size and its standard error, the deviance, the number of degrees of freedom and the Akaike's information criterion.
<code>glm</code>	The 'glm' object obtained from fitting the model.

**Note**

These functions use the `glm` function of the **stats** package.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

Rivest, L.P. and Baillargeon, S. (2007) Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics*, **63**(4), 999–1006.

**See Also**

[closedp](#), [closedp.Mtb](#)

**Examples**

```
# HIV data set
mat <- histpos.t(4)
mX2 <- cbind(mat, mat[, 1] * mat[, 2])
closedp.mX(HIV, dfreq = TRUE, mX = mX2)

# Third primary period of mvole data set
period3 <- mvole[, 11:15]
psi <- function(x) { -log(3.5 + x) + log(3.5) }
closedp.h(period3, h = psi)
```

closedp.Mtb

*Closed Population Capture-Recapture Model Mtb***Description**

This function fits model Mtb for closed populations in capture-recapture experiments.

**Usage**

```
closedp.Mtb(X, dfreq=FALSE, method = "BFGS", ...)
```

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

**Arguments**

X	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
dfreq	A logical. By default FALSE, which means that X has one row per unit. If TRUE, it indicates that the matrix X contains frequencies in its last column.
method	The method to be used by <a href="#">optim</a> . The default is "BFGS".
...	Further arguments to be passed to <a href="#">optim</a> or <a href="#">print.default</a> .
x	An object, produced by the closedp.Mtb function, to print.

**Details**

The Mtb model is non-linear. It is fitted with the [optim](#) function instead of the [glm](#) function. Therefore, the abundance estimate can be unstable.

For the model to be identifiable, the parameters are constrained in the following way:  $\text{logit}(c_i) = \text{logit}(p_i) + b$  for  $i$  in  $2, \dots, l$ .

**Value**

n	The number of captured units
t	The total number of capture occasions in the data matrix X.
results	A table containing, for the fitted model: <b>abundance:</b> the estimated population size, <b>stderr:</b> the standard error of the estimated population size, <b>deviance:</b> the model's deviance, <b>df:</b> the number of degrees of freedom, <b>AIC:</b> the Akaike's information criterion, <b>BIC:</b> the bayesian information criterion, <b>infoFit:</b> a numerical code giving information about error or warnings encountered when fitting the model (see <a href="#">Rcapture-package</a> for details).

optim	The output produced by <code>optim</code> from fitting the model.
optim.warn	A vector of character strings. If the <code>optim</code> function generates one or more warnings when fitting the model, a copy of these warnings are stored in <code>optim.warn</code> . NULL if <code>optim</code> did not produce any warnings.
parMtb	Capture-recapture parameters estimates for model Mtb : the abundance $N$ , $p_1$ to $p_t$ , the probabilities of first capture for each capture occasion, and $c_2$ to $c_t$ , the recapture probabilities for each capture occasion.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.

**See Also**

[closedp](#), [closedpCI.t](#)

**Examples**

```
# hare data set
closedp.Mtb(hare)

## Example producing an unstable estimate
# Fourth primary period of mvole data set
period4 <- mvole[, 16:20]
closedp.Mtb(period4)
```

---

`closedpCI` *Customization of a Loglinear Model and Confidence Interval for Abundance Estimation in Closed Population Capture-Recapture Experiments*

---

**Description**

The `closedpCI.t` and `closedpCI.0` functions fit a loglinear model specified by the user and compute a confidence interval for the abundance estimation. For a normal heterogeneous model, a log-transformed confidence interval (Chao 1987) is produced. For any other model, the multinomial profile likelihood confidence interval (Cormack 1992) is produced.

The model is identified with the argument `m` or `mX`. For heterogeneous models, the form of the heterogeneity is specified with the arguments `h` and `h.control`. If `h` is given with `mX`, heterogeneity is added in `mX`.

These functions extend `closedp.t` and `closedp.0` as they broaden the range of models one can fit and they compute confidence intervals. Unlike the `closedp` functions, it fits only one model at a time.

### Usage

```
closedpCI.t(X, dfreq=FALSE, m=c("M0", "Mt", "Mh", "Mth"), mX=NULL,
            h=NULL, h.control=list(), mname=NULL, alpha=0.05,
            fmaxSupCL=3, ...)

closedpCI.0(X, dfreq=FALSE, dtype=c("hist", "nbcap"), t=NULL, t0=NULL,
            m=c("M0", "Mh"), mX=NULL, h=NULL, h.control=list(),
            mname=NULL, alpha=0.05, fmaxSupCL=3, ...)

plotCI(x.closedpCI, main="Profile Likelihood Confidence Interval", ...)

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

## S3 method for class 'closedpCI'
boxplot(x, main="Boxplots of Pearson Residuals", ...)

## S3 method for class 'closedpCI'
plot(x, main="Scatterplot of Pearson Residuals", ...)
```

### Arguments

<code>X</code>	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
<code>dfreq</code>	A logical. By default FALSE, which means that <code>X</code> has one row per unit. If TRUE, it indicates that the matrix <code>X</code> contains frequencies in its last column.
<code>dtype</code>	A characters string, either "hist" or "nbcap", to specify the type of data. "hist", the default, means that <code>X</code> contains complete observed capture histories. "nbcap" means that <code>X</code> contains numbers of captures (see <a href="#">Rcapture-package</a> for details on data formats).
<code>t</code>	Requested only if <code>dtype="nbcap"</code> . A numeric specifying the total number of capture occasions in the experiment. For <code>closedpCI.0</code> , the value <code>t=Inf</code> is accepted. It indicates that captures occur in continuous time (see <a href="#">Rcapture-package</a> ).
<code>t0</code>	A numeric. Models are fitted considering only the frequencies of units captured 1 to <code>t0</code> times. By default, if <code>t</code> is not equal to <code>Inf</code> , <code>t0=t</code> . When <code>t=Inf</code> , the default value of <code>t0</code> is the larger number of captures observed.
<code>m</code>	A character string indicating the model to fit. For <code>closedpCI.0</code> it can be either "M0"=M0 model or "Mh"=Mh model. For <code>closedpCI.t</code> it can also be "Mt"=Mt model or "Mth"=Mth model.
<code>mX</code>	The design matrix of the loglinear model. By default, the design matrix is built based on the <code>m</code> argument. If a <code>mX</code> argument is given, it must be a matrix (or an object that can be coerced to a matrix by <code>as.matrix</code> ). The order of the capture

	<p>histories in the <code>mX</code> matrix must be as defined in the <code>histpos.t</code> or <code>histpos.0</code> function.</p> <p>For the <code>closedpCI.t</code> function only, <code>mX</code> can also be an object of class "<code>formula</code>" or a hierarchical loglinear model name. The details of a formula specification are given under <b>Details</b> and the details of a hierarchical loglinear model name specification are given in the <code>closedpMS.t</code> documentation.</p> <p>An intercept is always added to the model. Therefore, <code>mX</code> must not contain a column of ones if it is a matrix. Also, if the argument <code>h</code> is not <code>NULL</code>, one or more columns for heterogeneity are added to the design matrix.</p>
<code>h</code>	<p>A character string ("<code>LB</code>", "<code>Chao</code>", "<code>Poisson</code>", "<code>Darroch</code>", "<code>Gamma</code>" or "<code>Normal</code>") or a numerical R function specifying the form of the column(s) for heterogeneity in the design matrix.</p> <p>"<code>LB</code>" and "<code>Chao</code>" represents Chao's lower bound model (the default), "<code>Poisson</code>" represents the function <math>f(k) = \theta^k - 1</math>, "<code>Darroch</code>" represents the function <math>f(k) = k^2/2</math>, and "<code>Gamma</code>" represents the function <math>f(k) = -\log(\theta + k) + \log(\theta)</math> where <math>k</math> is the number of captures and <math>\theta</math> is a parameter specified with the <code>h.control</code> element <code>theta</code>.</p> <p>"<code>Normal</code>" represents heterogeneity modeled with a normal mixing distribution, as presented in Rivest (2011).</p> <p>If an R function is given, it must be the implementation of any convex mathematical function <math>f(k)</math> (it must have only one argument).</p>
<code>h.control</code>	<p>A list of elements to control the heterogeneous part of the model, if any.</p> <p>For a Poisson or Gamma heterogeneous model:</p> <p><code>theta</code> The value of the parameter <math>\theta</math> in <math>f(k) = \theta^k - 1</math> for the Poisson model (the default value is 2) and in <math>f(k) = -\log(\theta + k) + \log(\theta)</math> for the Gamma model (the default value is 3.5).</p> <p>For a Chao's lower bound heterogeneous model:</p> <p><code>neg</code> If this option is set to <code>TRUE</code> (the default), negative eta parameters in Chao's lower bound models are set to zero.</p> <p>For a Normal heterogeneous model:</p> <p><code>initcoef</code> Initial values for the loglinear coefficients given to <code>optim</code>. The default values are taken from Darroch's model.</p> <p><code>initsig</code> Initial value for the sigma parameter of the normal mixing distribution. The default is 0.2.</p> <p><code>method</code> The method to be used by <code>optim</code>. The default is "<code>BFGS</code>".</p>
<code>mname</code>	<p>A character string specifying the name of the customized model. By default, it is derived from the arguments specifying the model.</p>
<code>alpha</code>	<p>A confidence interval with confidence level <math>1-\alpha</math> is constructed. The value of <code>alpha</code> must be between 0 and 1; the default is 0.05.</p>
<code>fmaxSupCL</code>	<p>A numeric (by default 3). The upper end point of the interval to be searched by <code>uniroot</code> to find the upper bound of the multinomial profile likelihood confidence interval (Cormack 1992) is defined by <code>fmaxSupCL</code> <math>\times</math> <math>\hat{N}</math>. In this product, <code>fmaxSupCL</code> is a factor multiplying <math>\hat{N}</math>, the population size estimated by</p>



	Poisson regression. If the upper bound obtained is equal the upper end point $f_{\max\text{SupCL}} \times \hat{N}$ , the element SupCL of the output value CI will begin with the symbol >. In that case, one could try to increase the value of $f_{\max\text{SupCL}}$ to find the upper bound of the multinomial profile likelihood confidence interval instead of a minimum value for this upper bound.
...	Further arguments to be passed to <code>glm</code> , <code>optim</code> , <code>print.default</code> , <code>plot.default</code> or <code>boxplot.default</code> .
x.closedpCI	An object, produced by a <code>closedpCI</code> function, to produce a plot of the multinomial profile likelihood for $N$ .
main	A main title for the plot
x	An object, produced by a <code>closedpCI</code> function, to print or to plot.

## Details

The `closedpCI.t` function fits models using the frequencies of the observable capture histories (vector of size  $2^t - 1$ ), whereas `closedpCI.0` uses the number of units capture  $i$  times, for  $i = 1, \dots, t$  (vector of size  $t$ ). Thus, `closedpCI.0` can be used with data sets larger than those for `closedpCI.t`, but it cannot fit models with a temporal effect. See [Rcapture-package](#) for more details about the distinction between `.t` and `.0` functions.

These functions do not work for closed population models featuring a behavioral effect, such as Mb and Mbh. The abundance estimation is calculated as the number of captured units plus the exponential of the Poisson regression intercept. However, models with a behavioral effect can be fitted with `closedp.t` (Mb and Mbh), `closedp.Mtb` and `closedp.bc`.

## CHAO'S LOWER BOUND MODELS

Chao's lower bound models estimate a lower bound for the abundance. Rivest (2011) presents a generalized loglinear model underlying this estimator. To test whether a certain model for heterogeneity is adequate, one can conduct a likelihood ratio test by subtracting the deviance of a Chao's lower bound model to the deviance of the heterogeneous model under study. The two models should have the same  $m\lambda$  argument. Under the null hypothesis of equivalence between the two models, the difference of deviances follows a chi-square distribution with degrees of freedom equal to the difference between the models' degrees of freedom.

A Chao's lower bound model contains  $t - 2$  parameters, called eta parameters, for the heterogeneity. These parameters should theoretically be greater or equal to zero (see Rivest and Baillargeon (2007)). When the element `neg` of the argument `h.control` is set to TRUE (the default), negative eta parameters are set to zero (to do so, columns are removed from the design matrix of the model). Degrees of freedom of Chao's lower bound model increase when eta parameters are set to zero.

## ARGUMENT $m\lambda$ : FORMULA SPECIFICATION

For the `closedpCI.t` function,  $m\lambda$  can be an object of class "formula". The only accepted variables in this formula are  $c_1$  to  $c_t$ . The variable  $c_i$  represents a capture indicator (1 for a capture, 0 otherwise) for the  $i$ th capture occasions. Also, the formula must not contain a response variable since it is only used to construct the design matrix of the model. For example, if  $t=3$ , the Mt model is fitted if  $m\lambda = \sim .$  or  $m\lambda = \sim c_1 + c_2 + c_3$ . The symbol `.` in this formula is a shortcut for  $c_1 + c_2 + \dots + c_t$ . Formula  $m\lambda$  arguments facilitate the addition of interactions between capture occasions in the model. For example, if  $t=3$ , the Mt model with an interaction between the first and the second

capture occasion is fitted if  $mX = \sim . + c1 : c2$ . See [formula](#) for more details of allowed formulae.

## PLOT METHODS AND FUNCTIONS

The `boxplot.closedpCI` function produces a boxplot of the Pearson residuals of the customized model.

The `plot.closedpCI` function traces the scatterplot of the Pearson residuals in terms of  $f_i$  (number of units captured  $i$  times) for the customized model.

The `plotCI` function produces a plot of the multinomial profile likelihood for  $N$ . The value of  $N$  maximizing the profile likelihood and the bounds of the confidence interval are identified.

## Value

<code>n</code>	The number of captured units
<code>t</code>	The total number of capture occasions in the data matrix $X$ . When captures occur in continuous time (input argument <code>t=Inf</code> ), this output value is the larger number of captures observed.
<code>t0</code>	For <code>closedpCI.0</code> only: the value of the argument <code>t0</code> used in the computations.
<code>results</code>	A table containing, for the fitted model: <b>abundance:</b> the estimated population size, <b>stderr:</b> the standard error of the estimated population size, <b>deviance:</b> the model's deviance, <b>df:</b> the number of degrees of freedom, <b>AIC:</b> the Akaike's information criterion, <b>BIC:</b> the bayesian information criterion, <b>infoFit:</b> a numerical code giving information about error or warnings encountered when fitting the model (see <a href="#">Rcapture-package</a> for details).
<code>bias</code>	Not produced for normal heterogeneous models ( <code>h="Normal"</code> ): The asymptotic bias of the estimated population size.
<code>fit</code>	The 'glm' object obtained from fitting the model except for normal heterogeneous models ( <code>h="Normal"</code> ). These models are not fitted with <code>glm</code> . For them, <code>fit</code> is a list with the following elements: <b>parameters:</b> the matrix of parameters (loglinear coefficients + sigma parameter) estimates with their standard errors, <b>varcov:</b> the estimated variance-covariance matrix of the estimated parameters, <b>y:</b> the y vector used to fit the model, <b>fitted.values:</b> the model fitted values, <b>initparam:</b> the initial values for the parameters (loglinear coefficients + sigma parameter) used by <code>optim</code> , <b>optim:</b> the output produced by <code>optim</code> .
<code>fit.warn</code>	A vector of character strings. If the <code>glm</code> or <code>optim</code> (for normal heterogeneous models) function generates one or more warnings when fitting the model, a copy of these warnings are stored in <code>fit.warn</code> . NULL if no warnings occurred.

neg.eta	For Chao's lower bound models only: the position of the eta parameters set to zero in the loglinear parameter vector, if any.
CI	Not produced for normal heterogeneous models (h="Normal"): A table containing the abundance estimation and its multinomial profile likelihood confidence interval. The last column of the table, named infoCI, contains a numerical code giving information about error or warnings encountered when calculating the confidence interval. Here is a description of the meaning of this numerical code: <b>0</b> no error or warning occurred; <b>-1</b> an error occurred while calculating the multinomial abundance estimation; <b>4</b> a warning occurred while calculating the multinomial abundance estimation; <b>5</b> a warning occurred while calculating the lower bound of the multinomial profile likelihood confidence interval; <b>6</b> a warning occurred while calculating the upper bound of the multinomial profile likelihood confidence interval.
CI.err	Not produced for normal heterogeneous models (h="Normal"): If an error occurred while calculating the multinomial abundance, a copy of the error message is stored in CI.err. NULL if no error occurred.
CI.warn	Not produced for normal heterogeneous models (h="Normal"): If one or more warnings occur while calculating the multinomial abundance estimation or the profile likelihood confidence interval, a copy of these warnings are stored in CI.warn. NULL if no warnings occurred.
alpha	1-the confidence level of the interval.
N.CI	Not produced for normal heterogeneous models (h="Normal"): The x-coordinates for plot.closedpCI.t.
loglik.CI	Not produced for normal heterogeneous models (h="Normal"): The y-coordinates for plot.closedpCI.t.

### Note

For normal heterogeneous models, the closedpCI functions use `optim` from the **stats** package. Otherwise, models are fitted with `glm` and the code to compute the multinomial profile likelihood confidence interval calls the functions `glm`, `optimize` and `uniroot`, all from the **stats** package.

### Author(s)

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

### References

- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**(4), 783–791.
- Cormack, R. M. (1992) Interval estimation for mark-recapture studies of closed populations. *Biometrics*, **48**, 567–576.

Rivest, L.P. (2011) A lower bound model for multiple record systems estimation with heterogeneous catchability. *The International Journal of Biostatistics*, **7**(1), Article 23.

Rivest, L.P. and Baillargeon, S. (2007) Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics*, **63**(4), 999–1006.

### See Also

[closedp](#), [closedp.Mtb](#)

### Examples

```
# hare data set
CI <- closedpCI.t(hare, m = "Mth", h = "Poisson", h.control = list(theta = 2))
CI
plotCI(CI)

# HIV data set
mat <- histpos.t(4)
mX2 <- cbind(mat, mat[, 1] * mat[, 2])
closedpCI.t(HIV, dfreq = TRUE, mX = mX2, mname = "Mt interaction 1,2")
# which can be obtained more conveniently with
closedpCI.t(HIV, dfreq = TRUE, mX = ~ . + c1:c2, mname = "Mt interaction 1,2")

# BBS2001 data set
CI0 <- closedpCI.0(BBS2001, dfreq = TRUE, dtype = "nbcap", t = 50, t0 = 20,
                  m = "Mh", h = "Gamma", h.control = list(theta = 3.5))
CI0
plot(CI0)
plotCI(CI0)

### As an alternative to a gamma model, one can fit a negative Poisson model.
### It is appropriate in experiments where very small capture probabilities
### are likely. It can lead to very large estimators of abundance.

# Third primary period of mvole data set
period3 <- mvole[, 11:15]
psi <- function(x) { 0.5^x - 1 }
closedpCI.t(period3, m = "Mh", h = psi)

### Example of normal heterogeneous models
### diabetes data of Bruno et al. (1994)

histpos <- histpos.t(4)
diabetes <- cbind(histpos, c(58,157,18,104,46,650,12,709,14,20,7,74,8,182,10))
# chosen interaction set I in Rivest (2011)
closedpCI.t(X = diabetes, dfreq = TRUE, mX = ~ . + c1:c3 + c2:c4 + c3:c4,
            h = "Normal", mname = "Mth normal with I")

### Example of captures in continuous time
# Illegal immigrants data set
closedpCI.0(ill, dtype = "nbcap", dfreq = TRUE, t = Inf, m = "Mh", h = "LB")
```

closedpMS

*Hierarchical Loglinear Model Selection in Closed Population  
Capture-Recapture Experiments*

## Description

The `closedpMS.t` function fits every possible hierarchical loglinear models for a given closed population capture-recapture data set, under the constraints set by the given `maxorder` and forced arguments. Parameters for heterogeneity in capture probabilities among units can be added to the models.

The `getAllModels` function lists every possible hierarchical loglinear models for a certain number of capture occasions `t`, under the constraints set by the given `maxorder` and forced arguments.

## Usage

```
closedpMS.t(X, dfreq = FALSE, h = NULL, h.control = list(),
           maxorder = t - 1, forced = 1:t, stopiflong = TRUE, ...)

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

## S3 method for class 'closedpMS'
plot(x, main="Models comparison based on BIC", omitOutliers = TRUE, ...)

getAllModels(t, maxorder = t - 1, forced = 1:t, stopiflong = TRUE)
```

## Arguments

<code>X</code>	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
<code>dfreq</code>	A logical. By default <code>FALSE</code> , which means that <code>X</code> has one row per unit. If <code>TRUE</code> , it indicates that the matrix <code>X</code> contains frequencies in its last column.
<code>h</code>	A character string ("LB", "Chao", "Poisson", "Darroch", "Gamma" or "Normal") or a numerical R function specifying the form of the column(s) for heterogeneity in the design matrix. "LB" and "Chao" represents Chao's lower bound model (the default), "Poisson" represents the function $f(k) = \theta^k - 1$ , "Darroch" represents the function $f(k) = k^2/2$ , and "Gamma" represents the function $f(k) = -\log(\theta + k) + \log(\theta)$ where $k$ is the number of captures and $\theta$ is a parameter specified with the <code>h.control</code> element <code>theta</code> . "Normal" represents heterogeneity modeled with a normal mixing distribution, as presented in Rivest (2011). If an R function is given, it must be the implementation of any convex mathematical function $f(k)$ (it must have only one argument).

h.control	<p>A list of elements to control the heterogeneous part of the model, if any. For a Poisson or Gamma heterogeneous model:</p> <p>theta The value of the parameter <math>theta</math> in <math>f(k) = theta^k - 1</math> for the Poisson model (the default value is 2) and in <math>f(k) = -\log(theta + k) + \log(theta)</math> for the Gamma model (the default value is 3.5).</p> <p>For a Chao's lower bound heterogeneous model:</p> <p>neg If this option is set to TRUE (the default), negative eta parameters in Chao's lower bound models are set to zero.</p> <p>For a Normal heterogeneous model:</p> <p>initsig Initial value for the sigma parameter of the normal mixing distribution. The default is 0.2.</p> <p>method The method to be used by <code>optim</code>. The default is "BFGS".</p>
maxorder	A numeric specifying the higher order accepted for the terms in the models. It can take a value between 1 and $t - 1$ (the default).
forced	A vector of the terms forced in the model (by default every first order terms). NULL if no terms are forced in the model.
stopiflong	A logical indicating whether the function execution should be stopped when the number of possibles models is larger than 10 000, that is when the function might be long to run (default TRUE).
...	Further arguments to be passed to <code>glm</code> , <code>optim</code> or <code>print.default</code> .
x	An object, produced by the <code>closedpMS.t</code> function, to print or to plot.
main	A main title for the plot.
omitOutliers	A logical. If TRUE (the default), models with an outlier abundance or BIC value are removed from the plot. A value is considered an outlier if it is smaller than the first quartile minus 1.5 times the interquartile range, or larger than the third quartile plus 1.5 times the interquartile range.
t	A numeric specifying the number of capture occasions in the experiments. It is deduced from the data set in the <code>closedpMS.t</code> function. It must be an integer between 1 and 9 inclusively.

## Details

### HIERARCHICAL LOGLINEAR MODEL NAME SYNTAX

First, a model's term is written using numbers between 1 and 9 to represent the capture occasions it includes (ex.: 134 represents the three-way interaction  $c_1 : c_3 : c_4$ ). This syntax limits the maximal number of capture occasion to 9. This is not a problem since from 6 capture occasions upwards, the number of hierarchical models becomes very large and difficult to manage.

A hierarchical model name is a list of the model's terms at the top of the hierarchies in the model. These terms are separated by commas, without spaces. They are surrounded by brackets. For example, "[123, 34, 5]" is the name of the model

$\sim 1 + c_1 + c_2 + c_3 + c_4 + c_5 + c_1 : c_2 + c_1 : c_3 + c_2 : c_3 + c_3 : c_4 + c_1 : c_2 : c_3$ .

**Value**

getAllModels returns a character vector with the models names.

closedpMS.t returns a list with the following elements:

n	The number of captured units
t	The total number of capture occasions in the data matrix $X$ . When captures occur in continuous time (input argument <code>t=Inf</code> ), this output value is the larger number of captures observed.
results	A table containing, for every fitted model: abundance the estimated population size, stderr the standard error of the estimated population size, deviance the model's deviance, df the number of degrees of freedom, AIC the Akaike's information criterion, BIC the bayesian information criterion, infoFit a numerical code giving information about error or warnings encountered when fitting the model (see <a href="#">Rcapture-package</a> for details).
bias	A vector, the asymptotic bias of the estimated population size for every fitted model.
fit.err	A list of character string vectors. If an error occurs while fitting a model (with <a href="#">glm</a> or <a href="#">optim</a> ), a copy of the error message is stored in <code>fit.err\$mname</code> where <code>mname</code> is the name of the hierarchical loglinear model. A NULL list element means that no error occurred for the considered model.
fit.warn	A list of character string vectors. If warnings are generated while fitting a model (with <a href="#">glm</a> or <a href="#">optim</a> ), a copy of these warnings are stored in <code>fit.warn\$mname</code> where <code>mname</code> is a the name of the hierarchical model. A NULL list element means that no warnings were generated for the considered model.
neg.eta	For Chao's lower bound models only: the position of the eta parameters set to zero in the loglinear parameter vector, if any.

**Note**

For normal heterogeneous models, `closedpMS.t` uses [optim](#) from the **stats** package. Otherwise, models are fitted with [glm](#), also from the **stats** package.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**(4), 783–791.

Cormack, R. M. (1992) Interval estimation for mark-recapture studies of closed populations. *Biometrics*, **48**, 567–576.

Rivest, L.P. (2011) A lower bound model for multiple record systems estimation with heterogeneous catchability. *The International Journal of Biostatistics*, **7**(1), Article 23.

Rivest, L.P. and Baillargeon, S. (2007) Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics*, **63**(4), 999–1006.

### See Also

[closedp](#), [closedp.Mtb](#)

### Examples

```
# The lesbian data set contains 4 capture occasions.
# By default, closedpMB.t fits the 113 following models:

getAllModels(4)

closedpMS.t(lesbian, dfreq = TRUE)

# We could reduce the number of models by omitting
# those with triple interactions.

closedpMS.t(lesbian, dfreq = TRUE, maxorder = 2)

# Models with heterogeneity fits better.

Darr <- closedpMS.t(lesbian, dfreq = TRUE, h = "Darroch")
Darr

# The plot method allows the visualization of the results
# from models fitted by closedpMS.t().

plot(Darr)

# According to the BIC, the best heterogeneous Darroch model
# for this data set contains the double interactions 12, 13, 14.
# Here is the profile likelihood confidence interval for the
# abundance estimation from this model.

closedpCI.t(lesbian, dfreq = TRUE, mX = "[12,13,14]", h = "Darroch")
```

---

descriptive

*Descriptive Statistics*

---

### Description

This function produces descriptive statistics for capture-recapture data.



**Usage**

```
descriptive(X, dfreq=FALSE, dtype=c("hist", "nbcap"), t=NULL)

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

## S3 method for class 'descriptive'
plot(x, main="Exploratory Heterogeneity Graph", ...)
```

**Arguments**

<code>X</code>	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
<code>dfreq</code>	A logical. By default FALSE, which means that <code>X</code> has one row per unit. If TRUE, it indicates that the matrix <code>X</code> contains frequencies in its last column.
<code>dtype</code>	A characters string, either "hist" or "nbcap", to specify the type of data. "hist", the default, means that <code>X</code> contains complete observed capture histories. "nbcap" means that <code>X</code> contains numbers of captures (see <a href="#">Rcapture-package</a> for details on data formats).
<code>t</code>	Requested only if <code>dtype="nbcap"</code> . A numeric specifying the total number of capture occasions in the experiment. The value <code>t=Inf</code> is accepted. It indicates that captures occur in continuous time. (see <a href="#">Rcapture-package</a> ).
<code>x</code>	An object, produced by the <code>descriptive</code> function, to print or to plot.
<code>main</code>	A main title for the plot
<code>...</code>	Further arguments to be passed to methods (see <code>print.default</code> and <code>plot.default</code> ).

**Details**

The `plot.descriptive` function produces an exploratory heterogeneity graph. In the absence of heterogeneity, the relation(s) presented in the graph should be almost linear. Convex functions indicate heterogeneity (Baillargeon and Rivest (2007)).

**Value**

<code>n</code>	The total number of captured units.
<code>base.freq</code>	A table of basic descriptive statistics. For $i = 1, \dots, t$ : $f_i$ : the number of units captured $i$ times $u_i$ : the number of units captured for the first time on occasion $i$ $v_i$ : the number of units captured for the last time on occasion $i$ $n_i$ : the number of units captured on occasion $i$ If <code>dtype="nbcap"</code> , only the the number of units captured $i$ times, $f_i$ , can be computed.
<code>m.array</code>	Only if <code>dtype="hist"</code> : A $t$ by $t$ matrix containing recapture statistics for units released on occasion $i$ , for $i = 1, \dots, t$ : $n_i$ : the number of units released on occasion $i$ ; $c_2$ : the number of units released on occasion $i$ and recaptured for the first time

on occasion 2;  
 ...  
 ct: the number of units released on occasion  $i$  and recaptured for the first time on occasion  $t$ ;  
 not recapt: the number of units released on occasion  $i$  and never recaptured again.  
 The m.array matrix is used in Mark (see White and Burnham 1999) and MSurge (see Choquet, Reboulet, Pradel, Gimenez, and Lebreton 2004) to fit open population models to the data.

call            The function call (object of class "call").

### Author(s)

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

### References

- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Choquet, R., Reboulet, A.M., Pradel, R., Gimenez, O. and Lebreton, J.D. (2004). M-Surge: New Software Specifically Designed for Multistate Capture-Recapture Models. *Animal Biodiversity and Conservation*, **27**,1, 207–215.
- Hoaglin, D. C. (1980) A Poissonness Plot, *The American Statistician*, **34**, 146–149
- Lindsay, B. G. (1986) Exponential family mixture models (with least-squares estimators). *Annals of Statistics*, **14**, 124–137.
- Rivest, L.P. (2008) Why a time effect often has a limited impact on capture-recapture estimates in closed populations. *Canadian Journal of Statistics*, **36**(1), 75–84.
- White, G. and Burnham, K.P. (1999) Program Mark: Survival Estimation from Populations of Marked Animals. *Bird Study*, **46** (Supplement), 120–138. The software can be downloaded here: <http://www.phidot.org/software/mark/>.

### Examples

```
# hare data set
hare.desc <- descriptive(hare)
hare.desc
plot(hare.desc)

# cbird data set
cbird.desc1 <- descriptive(cbird, dfreq = TRUE, dtype = "nbcap", t = 11)
plot(cbird.desc1)
# To illustrate the option t = Inf.
cbird.desc2 <- descriptive(cbird, dfreq = TRUE, dtype = "nbcap", t = Inf)
plot(cbird.desc2)
# The y coordinate has changed.
```

---

 duck

*Eider Duck Data*


---

### Description

This data set contains capture-recapture data for eider ducks.

### Usage

```
duck
```

### Format

63 by 7 numeric matrix, with the following columns:

p1, p2, p3, p4, p5, p6 Capture histories from six periods

freq Observed frequencies for each capture history

### Details

The data points are extracted from a 25-year study by Coulson (1984). The capture periods are six consecutive years : years 19-24. This data set is analyzed in Cormack (1989).

Each row of this data set represents an observed capture history followed by its frequency.

### Source

Coulson, J. C. (1984) The population dynamics of the Eider Duck *Somateria mollissima* and evidence of extensive non breeding by adults ducks. *Ibis*, **126**, 525–543.

### References

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:[10.18637/jss.v019.i05](https://doi.org/10.18637/jss.v019.i05).

Cormack, R. M. (1989) Loglinear models for capture-recapture. *Biometrics*, **45**, 395–413.

### Examples

```
op.m1 <- openp(duck, dfreq = TRUE)
op.m1$model.fit
# The pvalue of the goodness of fit test based on the deviance is
1 - pchisq(op.m1$model.fit[1, 1], df = 49)
plot(op.m1)
# The residual plot shows a large residual for the 13 individuals
# captured all the times. We redo the analysis without them.

keep2 <- rowSums(histpos.t(6)) != 6
op.m2 <- openp(duck, dfreq = TRUE, keep = keep2)
op.m2$model.fit
```

```

1 - pchisq(op.m2$model.fit[1, 1],df = 48)
# The fit is still not satisfactory.
plot(op.m2)
# The residual plot has the convex shape characteristic of
# heterogeneity in the capture probabilities. We also remove the
# individuals caught at 5 periods out of 6.

keep3 <- rowSums(histpos.t(6)) < 5
op.m3 <- openp(duck, dfreq = TRUE, keep = keep3)
op.m3$model.fit
1 - pchisq(op.m3$model.fit[1, 1], df = 42)
# The fit is better but there is still heterogeneity in the data.

# To investigate whether the capture probabilities are homogeneous,
# one can fit a model with equal capture probabilities.
op.m4 <- openp(duck, dfreq = TRUE, m = "ep", keep = keep3)
op.m4$model.fit
# It gives a much larger deviance; this model is not considered anymore.

# We now investigate models for the growth rate  $N[i+1]/N[i]$  of this
# population using the multivariate normal distribution for the
# abundance estimates. The growth rates and their standard errors are
growth <- op.m3$N[3:5]/op.m3$N[2:4]
partial <- matrix(c(-op.m3$N[3]/op.m3$N[2]^2, 1/op.m3$N[2], 0, 0,
                  0, -op.m3$N[4]/op.m3$N[3]^2, 1/op.m3$N[3], 0,
                  0, 0, -op.m3$N[5]/op.m3$N[4]^2, 1/op.m3$N[4]), 3, 4, byrow = TRUE)
sig <- partial %*% op.m3$cov[9:12, 9:12] %*% t(partial)
cbind(estimate = growth, stderr = sqrt(diag(sig)))
# An estimate for the common growth rate is
siginv <- solve(sig)
growth.e <- t(rep(1, 3)) %*% siginv %*% growth/(t(rep(1, 3)) %*% siginv %*% rep(1, 3))
se <- 1/sqrt(t(rep(1, 3)) %*% siginv %*% rep(1, 3))
cbind(estimate = growth.e, stderr = se)
# A chi-square statistic for testing the equality of the growth rates
# and its pvalue
chisq2 <- t(growth - as.vector(growth.e)) %*% siginv %*% (growth - as.vector(growth.e))
c(chi2stat = chisq2, pvalue = 1 - pchisq(chisq2, df = 2))
# The hypothesis of a common growth rate is rejected

```

---

hare

*Snowshoe Hare Data*


---

### Description

This data set contains capture-recapture data for snowshoe hares.

### Usage

```
hare
```

**Format**

68 by 6 numeric matrix, with the following columns:

c1, c2, c3, c4, c5, c6 Capture histories from the six capture occasions

**Details**

This data set is analyzed in Cormack (1989) and Agresti (1994).

Each row of this data set represents the capture history of one animal.

**References**

Agresti, A. (1994) Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics*, **50**, 494–500.

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.

Cormack, R. M. (1989) Loglinear models for capture-recapture. *Biometrics*, **45**, 395–413.

**Examples**

```
desc <- descriptive(hare)
plot(desc)
# The fi plot shows that the two animals caught on all occasions create
# some heterogeneity in the capture probabilities.

closedp(hare)
# The best fitting model Mth Poisson2(N = 81.1, s.e. = 5.7) has an AIC of 146.
closedpCI.t(hare, m = "Mth", h = "Poisson", h.control = list(theta = 2))

# One can compare the fit of this model with that obtained by removing the
# 2 hares caught 6 times. This can be done by adding a column to the design
# matrix for Mt taking the value 1 for the capture history (1,1,1,1,1,1).
col <- rep(0, 2^6-1)
mat <- histpos.t(6)
col[rowSums(mat) == 6] <- 1
closedpCI.t(hare, mX = cbind(mat, col), mname = "Mt without 111111")
# This gives N = 76.8 (s.e. = 3.9) with an AIC of 146.
```

---

histpos

*Possible Capture Histories*

---

**Description**

histpos.t builds the matrix of observable capture histories in terms of captures and misses for a capture-recapture experiment with t capture occasions.

histpos.0 builds the matrix of observable capture histories in terms of number of captures for each primary period of a robust design with vt capture occasions. For closed populations, vt=t and histpos.0 simply returns t:1.

**Usage**

```
histpos.t(t)
histpos.0(vt)
```

**Arguments**

t	The number of capture occasions.
vt	A vector containing the numbers of capture occasions for each primary sampling period of a robust design. The length of this vector equals the number of primary sampling periods (noted $I$ ).

**Value**

histpos.t gives a  $2^t - 1$  by  $t$  matrix with rows representing capture histories. This matrix contains only zeros and ones.

histpos.0 gives a  $(t_1 + 1) \times (t_2 + 1) \times \dots \times (t_I + 1) - 1$  by  $I$  matrix with rows representing possible capture histories in terms of number of captures. This matrix elements are integers between 0 and  $t_i$  inclusively.

**Note**

The histpos.t function is called by [descriptive](#), [closedp](#), [closedp.bc](#), [closedp.Mtb](#), [openp](#), [robustd.t](#).

The histpos.0 function is called by [robustd.0](#).

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.

**See Also**

[periodhist](#), [descriptive](#)

**Examples**

```
histpos.t(5)
histpos.0(5)

histpos.0(rep(5, 3))
```

HIV

*Epidemiological capture-recapture Data on HIV***Description**

Epidemiological capture-recapture data on HIV from four reporting centers in Rome, Italy.

**Usage**

HIV

**Format**

15 by 5 numeric matrix, with the following columns:

c1, c2, c3, c4 Capture histories from the four capture occasions

freq Observed frequencies for each capture history

**Details**

The capture histories were obtained by linking the records of the four reporting centers.

Each row of this data set represents an observed capture history followed by its frequency.

**Source**

Abeni, D.A., Brancato, G. and Perucci, C. A. (1994) Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection. *Epidemiology*, **5**, 410–414.

**References**

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), [doi:10.18637/jss.v019.i05](https://doi.org/10.18637/jss.v019.i05).

**Examples**

```
desc <- descriptive(HIV, dfreq=TRUE)
desc
# 1774 out of 1896 individuals (94%) appear on one list only.

plot(desc)
# The fi plot is linear showing that heterogeneity is not a problem.
# Models with a time (or list) effect and possible pairwise dependencies
# between lists will be considered.

cp.m1 <- closedpCI.t(HIV, dfreq = TRUE, mX = ~ (c1+c2+c3+c4)^2,
                    mname = "Mt double interaction")
cp.m1
# The model fits well. Let's find out which interactions are important.
summary(cp.m1$fit)$coefficients
```

```
# Eliminating the non significant interactions stepwise shows that only
# the [1,2] interaction is important.

closedpCI.t(HIV, dfreq = TRUE, mX = ~ . + c1:c2, mname = "Mt interaction 1,2")
```

---

 ill

*Illegal immigrants in the Netherlands*


---

### Description

This data set contains the frequency  $f_i$  that an illegal immigrant who is not effectively expelled is apprehended exactly  $i$  times.

### Usage

ill

### Format

6 by 2 numeric matrix, with the following columns:

nbcap Numbers of captures, i.e. number of times an illegal immigrant is apprehended by the police

freq Observed frequencies for each number of captures

### Details

Bohning and Schon (2005) presented this data set as follows:

The illegal immigrant data are from van der Heijden et al. (2003) and refer to count data on illegal immigrants in four large cities in the Netherlands who could not be effectively expelled from the country. Illegal immigrants who are apprehended by the police often cannot be effectively expelled because either they refuse to disclose their nationality or their home country does not cooperate in receiving them back. In such cases they will be asked by the police to leave the country, but it is unlikely that they will abide by this request. The data were collected by the police and date back to the year 1995.

### Source

van der Heijden, P.G.M., Cruyff, M. and van Houwelingen, H.C. (2003) Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statistica Neerlandica*, **57**(3), 289-304.

### References

Bohning, D. and Schon, D. (2005) Nonparametric Maximum Likelihood Estimation of Population Size Based on the Counting Distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**(4), 721-737.



**Examples**

```
sdesc <- descriptive(ill, dtype = "nbcap", dfreq = TRUE, t = Inf)
plot(sdesc)
# A mixture model looks appropriate

closedp.0(ill, dtype = "nbcap", dfreq = TRUE, t = Inf)

# We can try to fit a normal mixture model:
closedpCI.0(ill, dtype = "nbcap", dfreq = TRUE, t = Inf, m = "Mh", h = "Normal")
# We get an estimate similar to the Mh Gamma3.5 estimate.

# Estimates are highly variable and it seems difficult to come up with a
# definitive answer. The lower bound estimate is useful in this context.
closedpCI.0(ill, dtype = "nbcap", dfreq = TRUE, t = Inf, m = "Mh", h = "LB")
# Considering the lower limit of a 95% confidence estimate for the lower bound,
# there should be at least 8 000 illegal immigrants in the Netherlands.
# Less than 25% have been caught.
```

---

lesbian

*Epidemiological capture-recapture Data on the lesbian population*


---

**Description**

Epidemiological capture-recapture data on the lesbian population from four four organisations that serve the lesbian and gay population of Allegheny County, Pennsylvania, United States and maintain large mailing lists.

**Usage**

```
lesbian
```

**Format**

15 by 5 numeric matrix, with the following columns:

A, B, C, D Capture histories from the four organisations mailing lists

freq Observed frequencies for each capture history

**Details**

The capture histories were obtained by matching the names on the four mailing lists.

Each row of this data set represents an observed capture history followed by its frequency.

**Source**

Aaron, D.J., Chang, Y.-F., Markovic, N., LaPorte, R. E. (2003) Estimating the lesbian population: a capture-recapture approach. *Journal of Epidemiology and Community Health* , **57**(3), 207–209.

**Examples**

```

closedp(lesbian, dfreq = TRUE)

# According to the BIC, the best model is Mth Darroch.
# Let's see if adding interactions between capture
# histories to the model could improve the model's fit.

closedpMS.t(lesbian, dfreq = TRUE, h = "Darroch")

# According to the BIC, the best heterogeneous Darroch model
# contains the double interactions 12, 13, 14.

```

---

mvole

---

*Robust Design Data for Adult Male Meadow Voles*


---

**Description**

This data set contains robust design capture history data for adult male meadow voles.

**Usage**

```
mvole
```

**Format**

171 by 30 numeric matrix, with the following columns:

```

c11, c12, c13, c14, c15 Capture histories from the 5 capture occasions within primary period 1
c21, c22, c23, c24, c25 Capture histories from the 5 capture occasions within primary period 2
c31, c32, c33, c34, c35 Capture histories from the 5 capture occasions within primary period 3
c41, c42, c43, c44, c45 Capture histories from the 5 capture occasions within primary period 4
c51, c52, c53, c54, c55 Capture histories from the 5 capture occasions within primary period 5
c61, c62, c63, c64, c65 Capture histories from the 5 capture occasions within primary period 6

```

**Details**

The data set is extracted from Table 19.1 of Williams, Nichols and Conroy (2002). The capture occasions represent five consecutive days of trapping each month from June to December 1981 at Patuxent Wildlife Research Center, Laurel, Maryland.

Each row of this data set represents the capture history of one animal.

**Note**

In this data set, ten animals are in fact not released after capture. These trap deaths are not identified.

## Source

Williams, B.K., Nichols, J.D., and Conroy, M.J. (2002) *Analysis and Management of Animal Populations*, San Diego: Academic Press.

## References

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.

## Examples

```
# This example deals only with the first four primary periods of the data set.
mvole4 <- mvole[, 1:20]

# First, a between primary period Jolly-Seber analysis is obtained.
mvole4.pp <- periodhist(mvole4, vt = rep(5,4))
op.m1 <- openp(mvole4.pp, dfreq = TRUE)
plot(op.m1)
# There is one large residual, removing the corresponding capture history
# from the analysis does not change the results. The model fits well.
keep2 <- residuals(op.m1$glm, type = "pearson") < 4
op.m2 <- openp(mvole4.pp, dfreq = TRUE, keep = keep2)
op.m2$model.fit

# To find a suitable model within each primary period, the function closedp.t
# can be used repeatedly. Heterogeneity is detected in all periods except
# the second one where the data collection was perturbed (the last capture
# occasion doesn't have any new capture and is taken out of the analysis).

# In a robust design, we use Mh models for all primary periods bearing in
# mind the questionable fit in the second one. Since there is no time effect
# within primary periods, we use the function robustd.0 to fit the model.
### The following command might take a few minutes to run.
rd.m1 <- robustd.0(mvole4[, -10], vt = c(5, 4, 5, 5), vm = "Mh", vh = "Chao")
rd.m1$model.fit
rd.m1$emig.fit
# The test for temporary immigration is not significant meaning that capture
# probabilities estimated with the Jolly-Seber model are not different from
# those estimated with the individual closed population models. The
# differences, on the logit scale, of the Jolly-Seber minus the closed
# population models capture probabilities are
rd.m1$emig.param
# Even in period 2, where the closed population model does not fit well, the
# difference on the logit scale is non significant (estimate=.56, s.e.=1.13).

# The following command allows to fit a robust design that does not specify
# any model for the second period.
### The following command might take a few minutes to run.
rd.m3 <- robustd.0(mvole4[, -10], vt = c(5, 4, 5, 5),
                  vm = c("Mh", "none", "Mh", "Mh"), vh = "Chao")

# With Darroch's model, the closed population estimates of the capture
```

```

# probabilities are significantly smaller than those obtained from the
# Jolly-Seber model. This cannot be interpreted as indicating temporary
# immigration. This suggests that Darroch's model is not appropriate within
# primary sessions.

# The smallest AIC is obtained with the Poisson model, with parameter a=1.5
# within sessions.
rd.m4 <- robustd.0(mvle4[, -10], vt = c(5, 4, 5, 5), vm = "Mh",
                  vh = "Poisson", vtheta = 1.5)
# The estimators of the demographic parameters obtained with the robust design
# are similar to those obtained with the Jolly-Seber model applied to the
# between primary period data.
cbind(op.m1$survivals, rd.m4$survivals)
cbind(op.m1$N, rd.m4$N)
cbind(op.m1$birth, rd.m4$birth)
cbind(op.m1$Ntot, rd.m4$Ntot)

```

---

openp

*Loglinear Models for Open Population Capture-Recapture Experiments*


---

### Description

This function computes various demographic parameters using a loglinear model for open populations in capture-recapture experiments.

### Usage

```
openp(X, dfreq=FALSE, m=c("up", "ep"), neg=TRUE, keep=rep(TRUE, 2^I-1))
```

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

```
## S3 method for class 'openp'
plot(x, main="Scatterplot of Pearson Residuals", ...)
```

### Arguments

- |       |  |
|-------|--|
| X     | The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).   |
| dfreq | A logical. By default FALSE, which means that X has one row per unit. If TRUE, it indicates that the matrix X contains frequencies in its last column.   |
| m     | This argument is a character string taking the value "up" (up = unconstrained probabilities) or "ep" (ep = equal probabilities). If m is set to "up" (the default), no constraint is set on the loglinear parameters. Therefore some loglinear and demographic parameters are not estimable. On the other hand, when m takes the value "ep", the capture probabilities are set to a common value; this enables the estimation of all the parameters. |

keep	This option is useful to fit the model on a subset of the possible capture histories. keep is a logical vector of length $2^I - 1$ taking the value TRUE for a history kept and FALSE for a history put aside. In this vector, the order of the capture histories is as defined in the <code>histpos.t</code> function. By default, every capture history is kept.
neg	If this option is set to TRUE, relevant negative gamma parameters are set to zero. This insures that the estimated survival probabilities belong to $[0, 1]$ and that the births are positive.
x	An object, produced by the <code>openp</code> function, to print or to plot.
main	A main title for the plot
...	Further arguments to be passed to methods (see <code>print.default</code> and <code>plot.default</code> ).

### Details

The function `openp` generates statistics to test the presence of a trap effect.

The `plot.openp` function produces a scatterplot of the Pearson residuals of the model versus the frequencies of capture.

If the data matrix  $X$  was obtained through the `periodhist` function, the `dfreq` argument must be set to TRUE.

Standard errors are calculated by linearization.

### Value

n	The number of captured units
model.fit	A table containing the deviance, degrees of freedom and AIC of the fitted model.
trap.fit	A table containing, for the models with an added trap effect, the deviance, degrees of freedom and AIC.
trap.param	The estimated trap effect parameters and their standard errors. For <code>m="up"</code> , the $I - 3$ first rows of <code>trap.param</code> are estimations of the differences $\text{logit}(\text{capture probability after a capture}) - \text{logit}(\text{capture probability after a miss})$ for periods 3 to $I - 1$ . The last row gives a pooled estimate of these differences calculated under the assumption that they are homogeneous.
capture.prob	The estimated capture probabilities per period and their standard errors.
survivals	The estimated survival probabilities between periods and their standard errors.
N	The estimated population sizes per period and their standard errors.
birth	The estimated number of new arrivals in the population between periods and their standard errors.
Ntot	The estimated total number of units who ever inhabited the survey area and its standard error.
glm	The 'glm' object obtained from fitting the loglinear model
loglin.param	The loglinear model parameters estimations and their standard errors, calculated by the <code>glm</code> function.
u.vector	The $U_i$ statistics, useful for the survival probabilities calculation, and their standard errors

v.vector	The $V_i$ statistics, useful for the population sizes estimation, and their standard errors
cov	The covariance matrix of all the demographic parameters estimates.
neg	The position of the gamma parameters set to zero in the loglinear parameter vector.

### Note

If your data contains more than one capture occasion within primary periods, use the [periodhist](#) function to create the input data matrix  $X$  needed by the `openp` function.

This function uses the `glm` function of the **stats** package.

### Author(s)

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

### References

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.

Rivest, L.P. and Daigle, G. (2004) Loglinear models for the robust design in mark-recapture experiments. *Biometrics*, **60**, 100–107.

### See Also

[closedp](#), [periodhist](#), [robustd](#)

### Examples

```
# duck data set
op.m1 <- openp(duck, dfreq = TRUE)
plot(op.m1)

# To remove the capture history 111111.
keep2 <- apply(histpos.t(6), 1, sum) != 6
op.m2 <- openp(duck, dfreq = TRUE, keep = keep2)
op.m2

# To remove the capture histories with 5 captures or more
keep3 <- apply(histpos.t(6), 1, sum) < 5
op.m3 <- openp(duck, dfreq = TRUE, keep = keep3)
op.m3

# mvole data set aggregated per primary period
mvole.op <- periodhist(mvole, vt = rep(5, 6))
openp(mvole.op, dfreq = TRUE)
```

---

periodhist

*Merge capture occasions*

---

### Description

This function produces a reduced matrix of capture histories from a complete one by merging together some capture occasions.

It can also be used to change the format of a capture-recapture data set with complete capture histories: it transforms a data set with one row per captured unit to a data set with one row per capture history followed by its frequency.

### Usage

```
periodhist(X, dfreq=FALSE, vt, drop=TRUE)
```

### Arguments

X	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
dfreq	A logical. By default FALSE, which means that X has one row per unit. If TRUE, it indicates that the matrix X contains frequencies in its last column.
vt	A vector containing the numbers of capture occasions for each pooled capture occasions. The length of this vector equals the number of capture occasions in the reduced matrix (noted $I$ ).
drop	A logical, by default TRUE, meaning that in the output matrix the unobserved capture histories (frequency of 0) are omitted. To keep them, drop must be set to FALSE.

### Details

This function is useful when using an open population model to analyze a robust design data set. It can be used to reduce the data set to one observation per primary period. A one represents a unit caught at least once during the period and a zero a unit never caught.

It is also useful for experiments with a large number of capture occasions but a limited number of catches, especially when there is no capture on some occasions. In such cases, one can pool together some capture occasions with the `periodhist` function.

If a data matrix produced by the `periodhist` function is given as an argument to a **Rcapture** function, the `dfreq` argument must be set to TRUE.

### Value

A  $2^I - 1$  by  $I + 1$  matrix of all the pooled capture histories, with their observed frequencies. If `drop=TRUE` (the default), the number of rows of this matrix is in fact  $2^I - 1$  minus the number of unobserved capture histories.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.

**See Also**

[openp](#)

**Examples**

```
# mvole data set aggregated per primary period
mvole.op <- periodhist(mvole, vt = rep(5, 6))
openp(mvole.op, dfreq = TRUE)
```

---

profileCI

*Profile Likelihood Confidence Interval for Abundance Estimation in Closed Populations Capture-Recapture Experiments*

---

**Description**

As of version 1.2-0 of **Rcapture**, this function is deprecated, but kept for back compatibility. Please use [closedpCI.t](#) instead.

The profileCI function computes the multinomial profile likelihood for the abundance of some closed population capture-recapture models.

**Usage**

```
profileCI(X, dfreq=FALSE, m="M0", h="Chao", a=2, mX=NULL,
          mname="Customized model", neg=TRUE, alpha=0.05)
```

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

**Arguments**

**X** The table of the observed capture histories in one of the two accepted formats. In the default format, it has one row per unit captured in the experiment. In this case, the number of columns in the table represents the number of capture occasions in the experiment (noted  $t$ ). In the alternative format, it contains one row per capture history followed by its frequency. In that case,  $X$  has  $t+1$  columns. The first  $t$  columns of  $X$ , identifying the capture histories, must contain only zeros and ones. The number one indicates a capture.



dfreq	This argument specifies the format of the data matrix $X$ . By default, it is set to "FALSE", which means that $X$ has one row per unit. If it is set to TRUE, then the matrix $X$ contains frequencies in its last column.
m	A character string identifying the model, either "M0"=M0 model, "Mt"=Mt model, "Mh"=Mh model or "Mth"=Mth model.
h	A character string ("Chao" = Chao model, "Poisson" = $2^k - 1$ ), "Darroch" = $(k^2)/2$ or a numerical function taking only one argument specifying the form of the column for heterogeneity in the design matrix.
a	The value of the exponent's base for a Poisson model.
mX	The design matrix of the loglinear model. In this matrix, the order of the capture histories is as defined in the <code>histpos.t</code> function.
mname	A character string specifying the name of the customized model.
neg	If this option is set to TRUE, negative eta parameters in Chao models are set to zero.
alpha	A confidence interval with confidence level $1-\alpha$ is constructed. The value of alpha must be between 0 and 1; the default is 0.05.
x	An object, produced by the <code>profileCI</code> function, to print.
...	Further arguments passed to or from other methods.

### Details

This function does not work for closed population models featuring a behavioral effect, such as Mb and Mbh.

### Value

This function produces a plot of the multinomial profile likelihood for  $N$ . The value of  $N$  maximizing the profile likelihood and the bounds of the confidence interval are identified. It also produces the following objects :

n	The number of captured units
results	A table containing the abundance estimation and its confidence interval.
alpha	$1$ -the confidence level of the interval.

### Author(s)

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

### References

- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Cormack, R. M. (1992) Interval estimation for mark-recapture studies of closed populations. *Biometrics*, **48**, 567–576.

**See Also**

[closedp](#), [closedp.mX](#), [closedp.h](#)

**Examples**

```
# hare data set
profileCI(hare, m = "Mth", h = "Poisson", a = 2)

# HIV data set
mat <- histpos.t(4)
mX2 <- cbind(mat, mat[, 1] * mat[, 2])

profileCI(hare, m = "Mh", h = "Chao")
```

---

rbvole

*Robust Design Data for Red-Back Voles*

---

**Description**

This data set contains robust design capture history data for red-back voles.

**Usage**

```
rbvole
```

**Format**

66 by 19 numeric matrix, with the following columns:

c11, c12, c13 Capture histories from the three capture occasions within primary period 1

c21, c22, c23 Capture histories from the three capture occasions within primary period 2

c31, c32, c33 Capture histories from the three capture occasions within primary period 3

c41, c42, c43 Capture histories from the three capture occasions within primary period 4

c51, c52, c53 Capture histories from the three capture occasions within primary period 5

c61, c62, c63 Capture histories from the three capture occasions within primary period 6

**freq** Observed frequencies for each capture history

**Details**

Data collection was carried out by Etcheverry and al.. The capture occasions represent three consecutive days of trapping in May 1999, July 1999, August 1999, May 2000, July 2000 and August 2000 in the Duchénier conservation area in southeastern Quebec. This data set is analyzed in Rivest and Daigle (2004).

Each row of this data set represents an observed capture history followed by its frequency.

## References

Rivest, L.P. and Daigle, G. (2004) Loglinear models for the robust design in mark-recapture experiments. *Biometrics*, **60**, 100–107.

## Examples

```
# According to Rivest and Daigle (2004), a good robust design model
# for this data set is formed of an Mth Chao model for each period.
# This model can be fitted as follows.

### memory.limit(size = 2000)
### rd <- robustd.t(rbvole, dfreq = TRUE, vt = rep(3, 6), vm = "Mth", vh = "Chao")

# WARNING : Because the data has 18 capture occasions, fitting this
# model uses a lot of memory. Its runtime is several minutes long.
```

---

robustd	<i>Loglinear Models for Capture-Recapture Experiments Carried Out According to a Robust Design</i>
---------	--

---

## Description

These functions compute various demographic parameters and capture probabilities per period using loglinear robust design models in capture-recapture experiments.

`robustd.t` and `robustd.0` fit the model using different response variable. `robustd.t` uses the frequencies of the observable capture histories in terms of capture success or failure for each capture occasions of each primary period. `robustd.0` uses the frequencies of the observable capture histories in terms of number of captures per primary period.

## Usage

```
robustd.t(X, dfreq = FALSE, vt, vm = "M0", vh = list("Chao"),
          vtheta = 2, neg = TRUE)

robustd.0(X, dfreq = FALSE, vt, vm = "M0", vh = list("Chao"),
          vtheta = 2, neg = TRUE)

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

## Arguments

X	The matrix of the observed capture histories (see <a href="#">Rcapture-package</a> for a description of the accepted formats).
dfreq	A logical. By default FALSE, which means that X has one row per unit. If TRUE, it indicates that the matrix X contains frequencies in its last column.

vt	A vector containing the numbers of capture occasions for each primary sampling period. The length of this vector equals the number of primary sampling periods (noted $I$ ).
vm	A vector indicating the closed population model for each primary period. The elements of vm can be "none"=no model, "M0"=M0 model, "Mt"=Mt model, "Mh"=Mh model or "Mth"=Mth model for robustd.t. For robustd.θ, models with time effect are not allowed. So vm can than be "none"=no model, "M0"=M0 model, or "Mh"=Mh model. The 'no model' cannot be selected for the first or the last period. If a single character string is given for vm, the corresponding model is used for all periods. The default value for vm is "M0" for all period.
vh	A list indicating, for each primary period with a heterogeneity model, the form of the columns for heterogeneity in the design matrix. The elements of vh can be "Chao", "Poisson", "Darroch" or any numerical R function created beforehand by the user. "Chao" represents Chao's model, "Poisson" represents the function $f(k) = \theta^k - 1$ , where $k$ is the number of captures, "Darroch" represents the function $f(k) = k^2/2$ , and "Gamma" represents the function $f(k) = -\log(\theta + k) + \log(\theta)$ . If an R function is given, it is the implementation of any convex mathematical function $f(k)$ . It has only one argument. For the Poisson or Gamma model, the parameter theta is specified in the argument vtheta.
vtheta	A vector indicating, for each primary period with a Poisson or Gamma heterogeneity model, the value of the parameter. If vtheta consists of a single number, this number is used for all the periods with a Poisson or Gamma model. The default value for vtheta is 2 for all the periods with a Poisson or Gamma model.
neg	If this option is set to TRUE, negative gamma parameters and negative eta parameters in Chao's models are set to zero. This insures that the estimated survival probabilities belong to $[0, 1]$ and that the births are positive.
x	An object, produced by the robustd.t or the robustd.θ function, to print.
...	Further arguments to be passed to print.default.

## Details

These functions also generate statistics to test the presence of temporary emigration.

The Poisson regression used to fit a robust design model has one entry for each possible capture history, including those that are unobserved. The size of the dependent vector is therefore  $2^{\text{sum}(\text{vt})} - 1$  for robustd.t. Models with a large sum(vt) are hard to fit with robustd.t. robustd.θ uses a more parsimonious coding for the capture histories and can fit larger models.

Standard errors are calculated by linearization.

## Value

n	The number of captured units
models	A vector of length $I$ identifying the closed population models chosen for each period.
model.fit	A table containing the deviance, degrees of freedom and AIC of the fitted model.

emig.fit	A table containing, for the model with an added temporary emigration effect, the deviance, the degrees of freedom and the Akaike's information criterion.
emig.param	The estimated temporary emigration parameters and their standards errors. The $I - 1$ first rows are estimations of the differences $\text{logit}(p^b) - \text{logit}(p^w)$ for periods 2 to $I - 1$ ( $p^b$ represents a between primary period and $p^w$ a within primary period estimate of the capture probability). The last row gives a pooled estimate of these differences calculated under the assumption that they are homogeneous. Negative estimates are associated with a temporary emigration.
capture.prob	The estimated capture probabilities per period and their standard errors.
survivals	The estimated survival probabilities between periods and their standard errors.
N	The estimated population sizes per period and their standard errors.
birth	The estimated number of new arrivals in the population between periods and their standard errors.
Ntot	The estimated total number of units who ever inhabited the survey area and its standard error.
loglin.param	The loglinear model parameters estimations and their standard errors, calculated by the glm function.
u.vector	The $U_i$ statistics, useful for the survival probabilities calculation, and their standard errors
v.vector	The $V_i$ statistics, useful for the population sizes estimation, and their standard errors
cov	The covariance matrix of all the demographic parameters estimates.
neg	The position of the gamma and eta parameters set to zero in the loglinear parameter vector.

**Note**

This function uses the glm function of the **stats** package.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

- Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), doi:10.18637/jss.v019.i05.
- Rivest, L.P. and Daigle, G. (2004) Loglinear models for the robust design in mark-recapture experiments. *Biometrics*, **60**, 100–107.

**See Also**

[closedp](#), [openp](#)

**Examples**

```

# The mvole data set contains a total of 30 capture occasions (the
# tenth capture occasion doesn't have any new capture and is taken
# out of the analysis). This number being large, we can only use
# the robustd.0 function to fit a robust design model.

# Not run :
# robustd.0(mvole[, -10], vt = c(5, 4, rep(5, 4)), vm = "Mh", vh = "Poisson", vtheta = 1.5)
# Should take a few seconds to run.

# Not run:
# robustd.t(mvole[, -10], vt = c(5, 4, rep(5, 4)), vm = 'Mh', vh = 'Poisson', vtheta = 1.5)
# Should fail.

# Considering only the first 3 periods of the data set, we can use the
# robustd.t function to fit a model with a temporal effect.
robustd.t(mvole[, c(1:9, 11:15)], vt = c(5, 4, 5), vm = "Mth", vh = "Poisson", vtheta = 1.5)

```

uifit

*Closed Population Models Fit Statistics Concerning the ui***Description**

This function produces fit statistics concerning the  $u_i$ , i.e. the numbers of first captures on each capture occasion, for closed population models. It also forecasts, for some models, the number of additional units that would be captured if the experiment was continued for five more occasions.

**Usage**

```
uifit(x.closedp)
```

**Arguments**

`x.closedp` An object produced by the `closedp.t` function.

**Value**

`predicted` The observed and predicted values of the  $u_i$  frequencies,  $i = 1, \dots, t$ , for each model in `closedp`. For models M0, Mh Poisson2, Mh Darroch and Mb, predicted values are also extrapolated for  $i = t + 1, \dots, t + 5$ .

`fit.stat` Chi-square fit statistics for each model in `closedp` ( $\sum_{i=1}^t \frac{(\text{observed}_i - \text{predicted}_i)^2}{\text{predicted}_i}$ ).

`day.first.capt` The mean and variance of the day of first capture, calculated with the observed and predicted  $u_i$ , for each model in `closedp.t`.

**Author(s)**

Louis-Paul Rivest <Louis-Paul.Rivest@mat.ulaval.ca> and Sophie Baillargeon

**References**

Baillargeon, S. and Rivest, L.P. (2007) Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, **19**(5), [doi:10.18637/jss.v019.i05](https://doi.org/10.18637/jss.v019.i05).

**See Also**

[closedp](#)

**Examples**

```
# Third primary period of mvole data set
period3 <- mvole[, 11:15]
cp <- closedp(period3)
uifit(cp)
```

# Index

- \* **datasets**
    - BBS2001, 7
    - bunting, 9
    - cbird, 10
    - duck, 35
    - hare, 36
    - HIV, 39
    - ill, 40
    - lesbian, 41
    - mvole, 42
    - rbvole, 50
  - \* **manip**
    - histpos, 37
    - periodhist, 47
  - \* **models**
    - closedp, 11
    - closedp.bc, 16
    - closedp.custom, 19
    - closedp.Mtb, 21
    - closedpCI, 22
    - closedpMS, 29
    - openp, 44
    - profileCI, 48
    - robustd, 51
  - \* **package**
    - Rcapture-package, 2
  - \* **univar**
    - descriptive, 32
    - uifit, 54
- BBS2001, 7
- boxplot.closedp (closedp), 11
- boxplot.closedp.custom (closedp.custom), 19
- boxplot.closedpCI (closedpCI), 22
- boxplot.default, 12, 25
- bunting, 9
- cbird, 10
- closedp, 2, 5, 11, 18, 20, 22, 23, 28, 32, 38, 46, 50, 53, 55
- closedp.0, 5, 23
- closedp.bc, 2, 5, 12, 15, 16, 25, 38
- closedp.custom, 19
- closedp.h, 50
- closedp.h (closedp.custom), 19
- closedp.Mtb, 2, 5, 15, 20, 21, 25, 28, 32, 38
- closedp.mX, 50
- closedp.mX (closedp.custom), 19
- closedp.t, 4, 23, 25
- closedpCI, 2, 5, 22
- closedpCI.0, 4, 5, 12, 13, 15
- closedpCI.t, 4, 5, 12, 13, 15, 22, 48
- closedpMS, 29
- closedpMS.t, 2, 4, 5, 15, 24
- descriptive, 3, 4, 32, 38
- duck, 35
- formula, 24–26
- getAllModels (closedpMS), 29
- glm, 5, 12, 14, 17, 18, 25–27, 30, 31
- hare, 36
- histpos, 3, 37
- HIV, 39
- ill, 4, 40
- lesbian, 41
- mvole, 42
- openp, 3, 38, 44, 48, 53
- optim, 5, 21, 22, 24–27, 30, 31
- optimize, 27
- periodhist, 3, 38, 46, 47
- plot.closedp (closedp), 11



plot.closedpCI (closedpCI), 22  
plot.closedpMS (closedpMS), 29  
plot.default, 12, 25  
plot.descriptive (descriptive), 32  
plot.openp (openp), 44  
plotCI (closedpCI), 22  
print.closedp (closedp), 11  
print.closedp.bc (closedp.bc), 16  
print.closedp.custom (closedp.custom),  
19  
print.closedp.Mtb (closedp.Mtb), 21  
print.closedpCI (closedpCI), 22  
print.closedpMS (closedpMS), 29  
print.default, 12, 17, 21, 25, 30  
print.descriptive (descriptive), 32  
print.openp (openp), 44  
print.profileCI (profileCI), 48  
print.robustd (robustd), 51  
profileCI, 48  
  
rbvole, 50  
Rcapture (Rcapture-package), 2  
Rcapture-package, 2  
robustd, 3, 46, 51  
robustd.0, 5, 38  
robustd.t, 4, 38  
  
uifit, 3, 15, 54  
uniroot, 24, 27