Package ‘EMGCR’

November 18, 2025

Type Package

Title Fit a Mixture Cure Rate Model with Custom Link Function

Version 0.2.0

Maintainer Jalmar M. F. Carrasco `<carrasco.jalmar@ufba.br>`

Description Tools to fit Mixture Cure Rate models via the Expectation-Maximization (EM) algorithm, allowing for flexible link functions in the cure component and various survival distributions in the latency part. The package supports user-specified link functions, includes methods for parameter estimation and model diagnostics, and provides residual analysis tailored for cure models. The classical theory methods used are described in Berkson, J. and Gage, R. P. (1952) <doi:10.2307/2281318>, Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) <`https://www.jstor.org/stable/2984875`>, Bazán, J., Torres-Avilés, F., Suzuki, A. and Louzada, F. (2017)<doi:10.1002/asmb.2215>.

License GPL-3

Encoding UTF-8

Depends R (>= 4.2.0)

Imports survival, Formula, knitr, actuar, flexsurv, tibble, ggplot2

RoxygenNote 7.3.2

LazyData true

NeedsCompilation no

Author Chaeyeon Yoo [aut],
Dipak K. Dey [aut],
Victor H. Lachos [aut],
Jalmar M. F. Carrasco [aut, cre]

Repository CRAN

Date/Publication 2025-11-18 09:50:08 UTC

Contents

`liver` *Liver Cancer Data*

Description

A sample of 2,766 patients diagnosed with liver cancer between 2012 and 2016, whose cancer grades were well identified. Available individual-level covariates include age at diagnosis, sex, pathological grade of the liver cancer, number of relapses, and median household income.

Usage

```
data(liver)
```

Format

A data frame with 2,766 observations and 10 variables:

ID Unique patient identifier

age Age as a factor (e.g., age group)

ageNumeric Age as a numeric variable

grade Pathological grade of liver cancer (I, II, III, IV)

medh Median household income as factor (possibly grouped)

medhNumeric Median household income as numeric

relapse Number of relapses after first diagnosis

sex Sex of patient: 1 = male, 0 = female

status Event indicator: 1 = death, 0 = censored

time Survival time in months

Details

The grade of the disease is categorized into four levels:

1. Grade I – Well differentiated
2. Grade II – Moderately differentiated
3. Grade III – Poorly differentiated
4. Grade IV – Undifferentiated/anaplastic

Examples

```
data(liver)
head(liver)
```

`liver2` *Liver Cancer Data 2*

Description

A sample of 1,736 patients who have been diagnosed with liver cancer between 2012 and 2016, whose cancer grades are well identified. Available individual-level covariates include age at diagnosis, sex, race, grade of liver cancer, median household income, time to treatment, tumor size, radiation indicator, and chemotherapy indicator. The grade of diseas is categorized into four levels: well-differentiated (Grade I), moderately differentiated (Grade II), poorly differentiated (Grade III), and undifferentiated/anaplastic (Grade IV).

Usage

```
data(liver2)
```

Format

A data frame with 1,736 observations and 10 variables:

ID Unique patient identifier

time Survival time in months

status censored = 0, dead due to liver cancer = 1

sex Sex of patient: Male=1, Female=0

age Scaled age of diagnosis

medh Scaled median household income of the subject

race While = 1, Other = 0

grade Pathological grade of liver cancer

chemo chemotherapy = 1, non = 0

radio radiation = 1, non = 0

Examples

```
data(liver2)
head(liver2)
```

`MCRfit` *Fit a Mixture Cure Rate (MCR) Survival Model*

Description

Fits a cure rate model using a flexible link function and a variety of survival distributions. The model accounts for a cured fraction through a logistic-type link and estimates the model via an EM-like algorithm.

Usage

```
MCRfit(
  formula,
  data,
  dist = "weibull",
  link = "logit",
  tau = 1,
  maxit = 1000,
  tol = 1e-05
)
```

Arguments

`formula`	A two-part formula of the form `Surv(time, status) ~ x | w`, where `x` are covariates for the survival part, and `w` are covariates for the cure fraction.
`data`	A data frame containing the variables in the model.
`dist`	A character string indicating the baseline distribution. Supported values are `"weibull"`, `"exponential"`, `"rayleigh"`, `"lognormal"`, `"loglogistic"`, and `"invgauss"`.
`link`	A character string specifying the link function for the cure fraction. Options are `"logit"`, `"probit"`, `"plogit"`, `"rplogit"`, and `"cauchit"`.
`tau`	A numeric value used when `link` = `"plogit"` or `"rplogit"`. Defaults to 1.
`maxit`	Maximum number of iterations for the EM-like algorithm. Defaults to 1000.
`tol`	Convergence tolerance. Defaults to 1e-5.

Value

An object of class `"MCR"`, which is a list containing:

`coefficients`	Estimated regression coefficients for the survival part.
`coefficients_cure`	Estimated coefficients for the cure part.
`scale`	Estimated scale parameter of the baseline distribution.
`loglik`	Final log-likelihood value.

`n`	Number of observations used in the model.
`deleted`	Number of incomplete cases removed before fitting.
`ep`	Estimated standard errors.
`iter`	Number of iterations used for convergence.
`dist`	Distribution used.
`link`	Link function used.
`tau`	Tau parameter used (if applicable).

Examples

```
require(EMGCR)

data(liver2)
names(liver2)
liver2$sex <- factor(liver2$sex)
liver2$grade <- factor(liver2$grade)
liver2$radio <- factor(liver2$radio)
liver2$chemo <- factor(liver2$chemo)
str(liver2)
model <- MCRfit(
  survival::Surv(time, status) ~ age + sex + grade + radio + chemo |
    age + medh + grade + radio + chemo,
  dist = "loglogistic",
  link = "plogit",
  tau = 0.15,
  data = liver2
)
model
```

`plot.MCR` *Plot multiple MCR model fits against Kaplan-Meier curve*

Description

Plot multiple MCR model fits against Kaplan-Meier curve

Usage

```
## S3 method for class 'MCR'
plot(...)
```

Arguments

`...`	One or more fitted MCR objects from `MCRfit()`.

Value

A ggplot object with Kaplan-Meier and survival curves for each model.

`qqMCR` *QQ-Plot of Residuals for MCR Model*

Description

Produces a Q-Q plot of residuals from a Mixture Cure Rate (MCR) model fitted via `MCRfit`. Optionally, a simulation envelope can be included for Cox-Snell residuals.

Usage

```
qqMCR(
  object,
  type = c("cox-snell", "quantile"),
  envelope = FALSE,
  nsim = 100,
  censor = NULL,
  ...
)
```

Arguments

`object`	An object of class `MCR`, typically returned by `MCRfit`.
`type`	Character. Type of residual to use in the QQ-plot. Options are `"cox-snell"` or `"quantile"`. Defaults to `"cox-snell"`.
`envelope`	Logical. Whether to add a simulation envelope to the QQ-plot. Default is `FALSE`.
`nsim`	Integer. Number of simulations used to construct the envelope. Default is `100`.
`censor`	Logical vector or NULL. Censoring indicator used when simulating data for the envelope. Required only when `envelope = TRUE` and `type = "cox-snell"`.
`...`	Additional arguments (currently ignored).

Details

The function generates QQ-plots of either Cox-Snell or quantile residuals. When `envelope = TRUE` and `type = "cox-snell"`, a simulation envelope is added using Monte Carlo replications.

Value

A QQ-plot is produced as a side effect. Nothing is returned.

See Also

`MCRfit`, `residuals.MCR`

Examples

```
data(liver)
fit <- MCRfit(survival::Surv(time, status) ~ age + medh + relapse + grade | sex + grade,
              data = liver, dist = "weibull", link = "logit")
qqMCR(fit, type = "quantile", envelope = TRUE, nsim = 50, censor = liver$status)
```

`residuals.MCR` *Compute residuals for MCR model*

Description

This function computes Global Cox-Snell and randomized quantile residuals for objects of class `MCR`.

Usage

```
## S3 method for class 'MCR'
residuals(object, type = c("cox-snell", "quantile"), ...)
```

Arguments

`object`	An object of class `MCR`, typically returned from `MCRfit`.
`type`	Type of residual.
`...`	Additional arguments (not used).

Value

A numeric vector of residuals.

Examples

```
data(liver)
names(liver)

model <- MCRfit(
 survival::Surv(time, status) ~ age + medh + relapse + grade | sex + age + medh + grade,
 data = liver
)
summary(residuals(model,type="quantile"))
```

`rMCM` *Generate Random Samples for Mixture Cure Rate (MCR) Model*

Description

Simulates survival data from a mixture cure rate model with covariates and user-defined link and latency distributions. Censoring is applied randomly.

Usage

```
rMCM(
  n,
  x,
  w,
  censor,
  alpha,
  beta,
  eta,
  dist = "weibull",
  link = "logit",
  tau = 1
)
```

Arguments

`n`	Integer. Number of observations to simulate.
`x`	Matrix or numeric. Covariate matrix for the latency component (must include intercept if needed).
`w`	Matrix or numeric. Covariate matrix for the cure component (no intercept assumed).
`censor`	Numeric. Maximum censoring time (uniformly distributed).
`alpha`	Numeric. Shape parameter for the survival distribution.
`beta`	Numeric vector. Coefficients for the latency part.
`eta`	Numeric vector. Coefficients for the cure part.
`dist`	Character. Distribution for the latency part. Options: `"weibull"`, `"lognormal"`, `"loglogistic"`, `"invgauss"`, `"exponential"`, `"rayleigh"`.
`link`	Character. Link function for cure component. Options: `"logit"`, `"probit"`,`"plogit"` ,`"rplogit"`, `"cauchit"`.
`tau`	A numeric value used when `link` = `"plogit"` or `"rplogit"`. Defaults to 1.

Value

A list with elements:

time Observed (possibly censored) survival time.

status Event indicator (1 = event, 0 = censored).

x Covariate matrix for the latency component.

w Covariate matrix for the cure component.

pCcensur Percentage of cured individuals.

pUCcensur Percentage of censored cases among the uncured.

Examples

```
# Example: Simulating survival data using the inverse Gaussian distribution
library(EMGCR)

n <- 500
beta <- c(1, -1, -2)
eta <- c(0.5, -0.5)
alpha <- 1.5

p <- length(beta)
q <- length(eta)

set.seed(10)
X <- matrix(rnorm(n*(p-1),0,1),n,p-1)
X <- cbind(1,X)

set.seed(20)
W <- matrix(runif(n*q,-1,1),n,q)
W <- scale(W)

max_censoring <- 10

set.seed(1234)
sim_data <- rMCM(n=n, x = X, w = W,
                 censor = max_censoring,
                 beta = beta, eta = eta,
                 alpha = alpha,
                 link = "logit", dist = "invgauss", tau = 1)

names(sim_data)
head(sim_data)
attributes(sim_data)
attr(sim_data, "pCcensur")
attr(sim_data, "pUCcensur")
```

Index