

Package: EMGCR (via r-universe)

October 9, 2026

Type Package

Title Mixture Cure Rate Models with Flexible Link Functions via the EM Algorithm

Version 0.3.0

Description Fits mixture cure rate models by the Expectation-Maximization (EM) algorithm. The incidence component (the probability of being uncured) accepts the logit, probit, cauchit, power logit and reversed power logit link functions, and the latency component accepts the exponential, Rayleigh, Weibull, log-normal, log-logistic and inverse Gaussian distributions. The package provides parameter estimates with standard errors, simulation of data from the model, and diagnostic tools based on residuals and simulated envelopes. The methods build on Berkson and Gage (1952) <doi:10.2307/2281318>, Dempster, Laird and Rubin (1977) <doi:10.1111/j.2517-6161.1977.tb01600.x> and Bazán, Torres-Avilés, Suzuki and Louzada (2017) <doi:10.1002/asmb.2215>.

URL <https://github.com/carrascojalmar/EMGCR>

BugReports <https://github.com/carrascojalmar/EMGCR/issues>

License GPL-3

Encoding UTF-8

Language en-US

Depends R (>= 4.2.0)

Imports survival, Formula, actuar, flexsurv, tibble, ggplot2 (>= 3.4.0), stats, graphics

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

LazyData true

NeedsCompilation no

Author Chaeyeon Yoo [aut], Dipak K. Dey [aut], Victor H. Lachos [aut],
Jalmar M. F. Carrasco [aut, cre]
Maintainer Jalmar M. F. Carrasco <carrasco.jalmar@ufba.br>
Config/pak/sysreqs libicu-dev
Repository <https://carrascojalmar.r-universe.dev>
Date/Publication 2026-10-06 20:40:44 UTC
RemoteUrl <https://github.com/carrascojalmar/emgcr>
RemoteRef HEAD
RemoteSha 2969a08e75e13972e55530efb27d48655748142d

Contents

liver	2
MCRfit	3
plot.MCR	5
qqMCR	6
residuals.MCR	7
rMCM	8
Index	11

liver	<i>Liver Cancer Data</i>
-------	--------------------------

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, radiation indicator, and chemotherapy indicator. The grade of disease is categorized into four levels: well-differentiated (Grade I), moderately differentiated (Grade II), poorly differentiated (Grade III), and undifferentiated/anaplastic (Grade IV).

Usage

data(liver)

Format

- A data frame with 1,736 observations and 10 variables:
- ID** Unique patient identifier
 - time** Survival time in months
 - status** Event indicator: death due to liver cancer = 1, censored = 0
 - sex** Sex of patient: male = 1, female = 0

age Scaled age of diagnosis
medh Scaled median household income of the subject
race Race: white = 1, other = 0
grade Pathological grade of liver cancer (1 = I, 2 = II, 3 = III, 4 = IV)
chemo Chemotherapy: yes = 1, no = 0
radio Radiation: yes = 1, no = 0

Examples

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

MCRfit

Fit a Mixture Cure Rate (MCR) Survival Model

Description

Fits a mixture cure rate model by the Expectation-Maximization (EM) algorithm, with a flexible link function for the incidence (cure) component and a choice of survival distributions for the latency component.

Usage

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

Arguments

formula	A two-part formula of the form $\text{Surv}(\text{time}, \text{status}) \sim x \mid w$, where x are co-variates for the survival part, and w are covariates for the incidence part (the probability of being uncured).
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 probability of being uncured. 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.

Details

In the latency component, the covariates enter through the scale parameter $\lambda_i = \exp(x_i^\top \beta)$, as in an accelerated failure time model (the same convention as [survreg](#)). A positive coefficient therefore means longer survival times for uncured individuals. With shape parameter α , the survival functions of the uncured are:

- exponential: $S(t) = \exp(-t/\lambda)$;
- Rayleigh: $S(t) = \exp\{-(t/\lambda)^2\}$;
- Weibull: $S(t) = \exp\{-(t/\lambda)^\alpha\}$;
- log-normal: $\log T \sim N(\log \lambda, \alpha^2)$;
- log-logistic: $S(t) = 1/\{1 + (t/\lambda)^\alpha\}$;
- inverse Gaussian: mean λ and shape α .

Value

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

<code>coefficients</code>	Estimated regression coefficients for the survival part.
<code>coefficients_cure</code>	Estimated coefficients for the incidence part, that is, for the probability of being uncured.
<code>scale</code>	Estimated shape parameter α of the latency distribution (fixed at 1 for the exponential and 2 for the Rayleigh).
<code>loglik</code>	Final log-likelihood value.
<code>n</code>	Number of observations used in the model.
<code>deleted</code>	Number of incomplete cases removed before fitting.
<code>ep</code>	Estimated standard errors.
<code>iter</code>	Number of EM iterations.
<code>convergence</code>	Logical; TRUE if the EM algorithm converged within <code>maxit</code> iterations.
<code>dist</code>	Distribution used.
<code>link</code>	Link function used.
<code>tau</code>	Tau parameter used (if applicable).

Examples

```
require(EMGCR)

data(liver)
names(liver)
liver$sex <- factor(liver$sex)
liver$grade <- factor(liver$grade)
liver$radio <- factor(liver$radio)
liver$chemo <- factor(liver$chemo)
str(liver)
model <- MCRfit(
```

```

survival::Surv(time, status) ~ age + sex + grade + radio + chemo |
  age + medh + grade + radio + chemo,
dist = "loglogistic",
link = "plogit",
tau = 0.15,
data = liver
)
summary(model)

```

plot.MCR

Compare fitted MCR models with the Kaplan-Meier curve

Description

Plots the Kaplan-Meier estimate of the survival function together with the population survival function implied by one or more fitted mixture cure rate models.

Usage

```

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

```

Arguments

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

Details

For each model, the fitted population survival function is averaged over the observations,

$$\hat{S}_{pop}(t) = \frac{1}{n} \sum_{i=1}^n \{1 - \hat{\theta}_i + \hat{\theta}_i \hat{S}(t \mid x_i)\},$$

where $\hat{\theta}_i$ is the estimated probability of being uncured and $\hat{S}(t \mid x_i)$ is the estimated survival function of the uncured. The Kaplan-Meier curve (dashed) is computed from the data of the first model, so all models should be fitted to the same data. Curves are coloured by latency distribution.

Value

A ggplot object.

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)
liver$sex <- factor(liver$sex)
liver$grade <- factor(liver$grade)
liver$radio <- factor(liver$radio)
liver$chemo <- factor(liver$chemo)

fit <- MCRfit(
  survival::Surv(time, status) ~ age + sex + grade + radio + chemo |
    age + medh + grade + radio + chemo,
  dist = "loglogistic", link = "plogit", tau = 0.15,
  data = liver
)
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 of class "MCRresiduals". Printing it shows the first six residuals and a summary; use `as.vector()` to drop the class.

Examples

```

data(liver)
liver$sex <- factor(liver$sex)
liver$grade <- factor(liver$grade)
liver$radio <- factor(liver$radio)
liver$chemo <- factor(liver$chemo)

```

```

model <- MCRfit(
  survival::Surv(time, status) ~ age + sex + grade + radio + chemo |
    age + medh + grade + radio + chemo,
  dist = "loglogistic", link = "plogit", tau = 0.15,
  data = liver
)
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, a chosen link function for the incidence part and a chosen latency distribution. Censoring times are drawn from a uniform distribution on $(0, \text{censor})$.

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. The first column must be the intercept (a column of ones).
w	Matrix or numeric. Covariate matrix for the incidence part (the probability of being uncured). Include a column of ones if an intercept is wanted.
censor	Numeric. Maximum censoring time (uniformly distributed).
alpha	Numeric. Shape parameter α of the latency distribution. Ignored for the exponential and Rayleigh distributions.
beta	Numeric vector. Coefficients for the latency part, which enter through the scale parameter $\lambda = \exp(x^\top \beta)$ (see MCRfit for details).
eta	Numeric vector. Coefficients for the incidence part.

dist	Character. Distribution for the latency part. Options: "weibull", "lognormal", "loglogistic", "invgauss", "exponential", "rayleigh".
link	Character. Link function for the probability of being uncured. Options: "logit", "probit", "plogit", "rplogit", "cauchit".
tau	A numeric value used when link = "plogit" or "rplogit". Defaults to 1.

Value

A **tibble** with columns:

time Observed (possibly censored) survival time.

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

x1, x2, ... Columns of x without the intercept.

w1, w2, ... Columns of w .

It also has two attributes: "pCcensur", the percentage of cured individuals, and "pUCcensur", the percentage of censored observations 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

`liver`, [2](#)

`MCRfit`, [3](#), [6–8](#)

`plot.MCR`, [5](#)

`qqMCR`, [6](#)

`residuals.MCR`, [6](#), [7](#)

`rMCM`, [8](#)

`survreg`, [4](#)

`tibble`, [9](#)