Package 'exposureEM'

September 21, 2026

Type Package

Title Combined-Exposure Models by EM and Marquardt Optimization

Version 0.3.0

Description Fits general two-component combined-exposure models for binary event histories when the event setting is not observed.
The observed binary event is represented as the union of two latent component-specific binary events. Known exposure proportions enter as offsets. Parameters can be estimated by expectation-maximization, direct Marquardt-damped Newton-Raphson maximization of the observed likelihood, or a hybrid that uses several expectation-maximization iterations before direct optimization. Uncertainty is estimated with Louis' formula for the expectation-maximization estimator and the inverse observed Hessian for direct and hybrid fits. Complementary log-log, logit, and log component links are available for all three estimation methods.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.1.0)

Imports stats, utils

NeedsCompilation no

Author Wenjing Meng [aut, cre]

Maintainer Wenjing Meng <w2meng@ucsd.edu>

Repository CRAN

Date/Publication 2026-09-21 21:00:07 UTC

Contents

`component_contrast` *Contrast between the two exposure components*

Description

Computes `exp(beta_component_1 - beta_component_0)` with a Wald confidence interval. Its interpretation is a rate ratio for `cloglog`, an odds ratio for `logit`, and a risk ratio for `log`.

Usage

```
component_contrast(object, level = 0.95)
```

Arguments

`object`	A fitted `exposure_em_fit` object.
`level`	Confidence level.

Value

A one-row data frame containing the log contrast, standard error, exponentiated estimate, and confidence limits.

`exposure_em` *Fit a two-component exposure model by EM*

Description

Fits the paper's combined-exposure model. For observation `i`, the observed event is `Y_i = I(Y_0i + Y_1i >= 1)`, where the two latent component events are conditionally independent. The two design rows share all covariates and have separate component-specific intercepts. Exposure proportions enter as fixed offsets.

Usage

```
exposure_em(
  formula,
  data,
  exposure = c("a0", "a1"),
  link = c("cloglog", "logit", "log"),
  start = NULL,
  control = exposure_em_control(),
  na.action = stats::na.omit,
  contrasts = NULL
)
```

Arguments

formula	A two-sided formula. The response must be binary. The right side specifies covariates shared by both latent components; the usual formula intercept is replaced by two component-specific intercepts.
data	A data frame.
exposure	Character vector naming the component 0 and component 1 exposure columns, in that order.
link	Component inverse link: "cloglog" (the primary censored-Poisson model), "logit", or "log".
start	Optional full numeric starting vector ordered as the component 0 intercept, component 1 intercept, then expanded formula columns. A named vector is reordered to the package's coefficient names.
control	A list returned by exposure_em_control(), or a named partial list overriding its defaults.
na.action	Missing-data action, normally stats::na.omit, stats::na.exclude, or stats::na.fail.
contrasts	Optional contrasts passed to stats::model.matrix().

Details

The outer criterion selects a preliminary stopping check. Termination also requires an absolute observed likelihood change below tol and score' information^{-1} score / length(score) <= tol, with unmodified observed information having eigenvalues above 1e-10. The maximum absolute unmodified observed Newton step must also be at most sqrt(tol) to avoid treating a vanishing boundary score as a finite root. The M-step keeps E-step expectations fixed and checks Q ascent. Formula offsets are rejected: the two exposure columns supply the model's offsets. Both response values must occur, and the active component design must have full column rank. Constant cloglog exposure proportions are rejected because the separate intercepts are not identified. The rank check alone does not establish observed-model identification in other designs. For na.exclude, in-sample prediction, fitted-value, and residual methods restore excluded rows as NA; stored arrays and posterior summaries contain only analyzed rows. Model-based covariance assumes independent observations. The component contrast compares equal exposures (or the underlying rates, odds, or risks per unit exposure), not unequal observed component exposures.

Value

An object of class `exposure_em_fit`. It contains `coefficients`, `vcov`, the unmodified observed `information`, `logLik`, `aic`, the final observed `score` and `newton_decrement`, `converged`, `convergence`, iteration counts, and an optional `trace`. Analyzed-row probabilities, posterior expectations, residuals, exposures, and the model frame are also stored. A singular or non-positive information matrix produces an unavailable (`NA`) covariance and a warning; a convergence flag does not prove a global maximum or exclude weak identification.

Examples

```
set.seed(42)
n <- 500
a1 <- c(rep(0, 100), stats::rbeta(400, 2, 12))
a0 <- 1 - a1
age <- stats::rnorm(n)
lambda0 <- a0 * exp(-3.2 + 0.15 * age)
lambda1 <- a1 * exp(-1.4 + 0.15 * age)
y <- stats::rbinom(n, 1, -expm1(-(lambda0 + lambda1)))
dat <- data.frame(y, a0, a1, age)
fit <- exposure_em(y ~ age, dat, exposure = c("a0", "a1"))
coef(fit)
component_contrast(fit)
```

---

exposure_em_control *Control parameters for exposure-component EM fitting*

Description

Control parameters for exposure-component EM fitting

Usage

```
exposure_em_control(
  tol = 1e-08,
  maxit = 10000L,
  mstep_tol = 1e-08,
  mstep_maxit = 1000L,
  criterion = c("parameter", "loglik", "both"),
  step_factor = 0.5,
  min_step = 2^-20,
  ridge_initial = 0,
  ridge_multiplier = 10,
  ridge_max = 1e+08,
  exposure_tol = sqrt(.Machine$double.eps),
  check_exposure_sum = TRUE,
  keep_trace = TRUE,
  trace = FALSE
)
```

Arguments

`tol`	Outer EM convergence tolerance.
`maxit`	Maximum number of EM iterations.
`mstep_tol`	Convergence tolerance for the Newton-Raphson M-step.
`mstep_maxit`	Maximum number of inner Newton-Raphson iterations.
`criterion`	Outer convergence criterion: parameter change, observed log-likelihood change, or both. The observed likelihood and curvature-scaled score safeguards described in `exposure_em()` always apply.
`step_factor`	Multiplicative factor used during step-halving.
`min_step`	Smallest accepted fraction of a Newton step.
`ridge_initial`	First diagonal ridge used if the information matrix is not positive definite. A value of zero first attempts an undamped step.
`ridge_multiplier`	Factor by which the ridge is increased.
`ridge_max`	Maximum relative diagonal ridge.
`exposure_tol`	Tolerance used to check that the two exposure proportions sum to one.
`check_exposure_sum`	Whether to require exposure proportions to sum to one. Set to `FALSE` only when the exposure proportions intentionally do not sum to one; each component must still lie in `[0, 1]`.
`keep_trace`	Whether to retain the EM iteration history.
`trace`	Whether to print one concise line per EM iteration.

Value

A named list consumed by `exposure_em()`.

`exposure_fit` *Fit a two-component exposure model using EM, Marquardt, or both*

Description

Fits the same two-component combined-exposure likelihood through one of three computational paths. `method = "em"` uses Bernoulli-indicator EM. `method = "marquardt"` directly maximizes the observed-data likelihood with Marquardt-stabilized Newton-Raphson. `method = "hybrid"` first performs a limited number of EM outer iterations and then passes those coefficients to the direct optimizer.

Usage

```
exposure_fit(
  formula,
  data,
  exposure = c("a0", "a1"),
  link = c("cloglog", "logit", "log"),
  method = c("em", "marquardt", "hybrid"),
  start = NULL,
  hybrid_em_iterations = 10L,
  em_control = exposure_em_control(),
  marquardt_control = exposure_marquardt_control(),
  na.action = stats::na.omit,
  contrasts = NULL
)
```

Arguments

`formula`	A two-sided formula. The response must be binary. Covariates on the right side have shared effects in both exposure components.
`data`	A data frame.
`exposure`	Character vector naming component 0 and component 1 exposure columns, in that order.
`link`	Component inverse link: `"cloglog"`, `"logit"`, or `"log"`, available with every estimation method.
`method`	Computational method: `"em"`, `"marquardt"`, or `"hybrid"`.
`start`	Optional coefficient starting vector ordered as the two component intercepts followed by expanded formula columns. Named starts are matched to the package coefficient names.
`hybrid_em_iterations`	Maximum number of complete EM outer iterations before switching to direct Marquardt optimization. Switch earlier if EM converges or stops with invalid observed curvature.
`em_control`	EM settings from `exposure_em_control()`, or a named list overriding selected defaults.
`marquardt_control`	Direct-optimization settings from `exposure_marquardt_control()`, or a named list overriding defaults.
`na.action`	Missing-data action, normally stats::na.omit, stats::na.exclude, or stats::na.fail.
`contrasts`	Optional contrasts passed to `stats::model.matrix()`.

Details

All three methods support complementary log-log, logit, and log component links. Hybrid fitting uses the selected link in both phases.

Direct termination requires the sum of squared coefficient changes, absolute likelihood change, and the unmodified Newton decrement per coefficient all to be at most `tol`. At an exactly zero score the

proposed change is zero. Diagonal inflation uses `0.99 * abs(diag(H)) + 0.01 * abs(sum(diag(H)))`, with a machine-precision floor for flat diagonals. Backtracking enforces Armijo sufficient increase up to a rounding allowance of `1e-12 * (1 + abs(logLik))`. Final covariance never uses inflation. Log-link steps must keep every component probability below one; invalid proposals are rejected during backtracking. For logit, proposals whose component probabilities round to one are also rejected numerically. These are interior optimizers: reaching a probability boundary does not certify a regular finite maximum or valid Wald inference. Hybrid fitting switches after at most `hybrid_em_iterations`, or earlier if EM terminates. Each phase respects its own `keep_trace` setting; the combined trace contains both phases when both are retained.

Value

An object inheriting from `exposure_fit` and `exposure_em_fit`, with the fields described in `exposure_em()`. Direct and hybrid fits additionally retain `em_trace` and `marquardt_trace` when requested. `iterations` is the sum of completed EM and accepted direct updates; `em_iterations` and `marquardt_iterations` give the separate counts.

Examples

```
set.seed(12)
n <- 500
a1 <- ifelse(stats::runif(n) < 0.2, 0, stats::rbeta(n, 2, 15))
a0 <- 1 - a1
x <- stats::rnorm(n)
p <- -expm1(-(a0 * exp(-3.2 + 0.2 * x) +
               a1 * exp(-1.5 + 0.2 * x)))
dat <- data.frame(y = stats::rbinom(n, 1, p), a0, a1, x)

fit_em <- exposure_fit(y ~ x, dat, method = "em")
fit_direct <- exposure_fit(y ~ x, dat, method = "marquardt")
fit_hybrid <- exposure_fit(
  y ~ x, dat, method = "hybrid", hybrid_em_iterations = 5
)
component_contrast(fit_hybrid)
fit_logit <- exposure_fit(y ~ x, dat, link = "logit", method = "marquardt")
fit_log <- exposure_fit(y ~ x, dat, link = "log", method = "hybrid")
```

`exposure_marquardt_control`

Control parameters for direct Marquardt fitting

Description

These settings control direct maximization of the observed-data likelihood. They are used by `method = "marquardt"` and by the direct phase of `method = "hybrid"` in `exposure_fit()`.

Usage

```
exposure_marquardt_control(
  tol = 1e-08,
  maxit = 200L,
  step_factor = 0.5,
  min_step = 1e-04,
  armijo = 1e-04,
  inflation_initial = 0.01,
  inflation_multiplier = 10,
  inflation_attempts = 20L,
  pd_tol = 1e-10,
  keep_trace = TRUE,
  trace = FALSE
)
```

Arguments

`tol`	Convergence tolerance applied to the sum of squared parameter changes, absolute likelihood change, and curvature-scaled score measure.
`maxit`	Maximum number of direct Marquardt iterations.
`step_factor`	Multiplicative reduction used during backtracking.
`min_step`	Smallest fraction of a proposed Marquardt step.
`armijo`	Small nonnegative constant used in the sufficient-increase check. Set it to zero to require only a nondecreasing likelihood.
`inflation_initial`	First Marquardt diagonal-inflation multiplier used when the unmodified negative Hessian is not positive definite.
`inflation_multiplier`	Factor used to increase the inflation multiplier.
`inflation_attempts`	Maximum number of diagonal-inflation attempts per iteration.
`pd_tol`	Smallest eigenvalue accepted as positive definite.
`keep_trace`	Whether to retain the direct-iteration history.
`trace`	Whether to print one concise line per direct iteration.

Value

A named list consumed by `exposure_fit()`.

posterior_components *Posterior latent-component probabilities*

Description

Returns the converged E-step quantities. `component_0` and `component_1` are the posterior expectations of the two latent indicators; because both events can occur, they are not mutually exclusive and can sum to more than one.

Usage

```
posterior_components(object, joint = TRUE)
```

Arguments

`object`	A fitted `exposure_em_fit` object.
`joint`	Whether to include the four mutually exclusive joint latent states `(Y0,Y1) = (0,0), (1,0), (0,1), (1,`

Value

A data frame with one row per analyzed observation.

predict.exposure_em_fit *Predict from a fitted exposure model*

Description

Predict from a fitted exposure model

Usage

```
## S3 method for class 'exposure_em_fit'
predict(object, newdata = NULL, type = c("response", "component", "link"), ...)
```

Arguments

`object`	A fitted `exposure_em_fit` object.
`newdata`	Optional data frame containing covariates and both exposure columns. If omitted, fitted values are returned.
`type`	`"response"` for the probability of any observed event, `"component"` for the two marginal component probabilities, or `"link"` for the two component linear predictors including log-exposure offsets.
`...`	Unused.

Value

A numeric vector for `type = "response"`, otherwise a two-column matrix.

Index