

Package: MRCCC (via r-universe)

October 8, 2026

Type Package

Title Bayesian Mendelian Randomization for Causal Cell-Cell Communication

Version 0.99.4

Description Implements MR-CCC, a Bayesian Mendelian randomization framework for inferring causal ligand-receptor signalling between cell types from population-scale single-cell expression and genotype data. For each ligand-receptor-pathway triplet, cis-eQTL genotypes instrument ligand expression in the sender cell type and receptor expression in the receiver cell type, and a working model with a ligand-by-receptor interaction is fitted to downstream pathway activity by a blocked Gibbs sampler. A spike-and-slab prior on the causal block yields a posterior inclusion probability that serves as the communication score, and the retained draws support credible intervals, sign probabilities, the receptor level at which the ligand effect reverses sign, and Bayesian false discovery rate control across many triplets.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.5.0)

Imports Rcpp, SummarizedExperiment, ggplot2 (>= 3.4.0), stats, methods, withr (>= 2.5.0)

LinkingTo Rcpp, RcppArmadillo

Suggests testthat (>= 3.0.0), knitr, rmarkdown, BiocStyle

VignetteBuilder knitr

biocViews Software, StatisticalMethod, Bayesian, GeneExpression, SingleCell, Genetics, SystemsBiology, GeneRegulation

URL <https://github.com/bitansa/MRCCC>

BugReports <https://github.com/bitansa/MRCCC/issues>

Roxygen list(markdown = TRUE)

Config/testthat/edition 3
NeedsCompilation yes
Config/roxygen2/version 8.1.0
Config/pak/sysreqs zlib1g-dev
Repository <https://biocstaging.r-universe.dev>
Date/Publication 2026-10-08 20:03:06 UTC
RemoteUrl <https://github.com/BiocStaging/MRCCC>
RemoteRef HEAD
RemoteSha 8078793179259416e2e7643070057fad472410b

Contents

bayes_fdr	2
bfdr_threshold_set	3
credible_intervals	4
mr_ccc	5
mr_ccc_gibbs	9
mrccc_hyperparameters	11
plot.mrccc_fit	13
plot_effect_curve	13
print.mrccc_fit	15
se_row	16
sign_probabilities	17
sign_reversal	18
simulate_mrccc	19
summary.mrccc_fit	21
Index	23

bayes_fdr	<i>Bayesian false discovery rate of nested PIP-ranked sets</i>
-----------	--

Description

For a vector of posterior inclusion probabilities from many ligand-receptor-pathway triplets, computes the Bayesian false discovery rate (bFDR) of every nested set obtained by declaring the top- k triplets by PIP, for $k = 1, \dots, m$.

Usage

bayes_fdr(pip)

Arguments

pip	Numeric vector of posterior inclusion probabilities in $[0, 1]$, one per triplet. Missing values are not allowed.
-----	--

Details

For a declared set S , the posterior expected number of false discoveries is $\sum_{j \in S} (1 - \text{PIP}_j)$, so the Bayesian false discovery rate is

$$\text{bFDR}(S) = |S|^{-1} \sum_{j \in S} (1 - \text{PIP}_j).$$

Ranking the triplets by decreasing PIP and accumulating gives, for each possible cut-off, the bFDR incurred by declaring everything at or above it. Because the PIPs are posterior probabilities, this controls the expected proportion of false discoveries without a separate multiplicity correction. The value returned for triplet j is the bFDR of the smallest nested set that contains j ; it is non-decreasing in the rank of j . Ties in PIP are broken by input order.

Value

A numeric vector of the same length as `pip`, in the order of the input, giving the bFDR of the nested set that ends at each triplet.

References

Newton MA, Noueiry A, Sarkar D, Ahlquist P (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* 5(2):155-176.

See Also

[bfdr_threshold_set\(\)](#), [mr_ccc\(\)](#)

Examples

```
pip <- c(0.9, 0.8, 0.5)
bayes_fdr(pip)
# Unsorted input is handled; results are returned in input order.
bayes_fdr(c(0.5, 0.9, 0.8))
```

<code>bfdr_threshold_set</code>	<i>Largest PIP-ranked set with Bayesian FDR at most alpha</i>
---------------------------------	---

Description

Selects the largest nested set of triplets, ranked by decreasing posterior inclusion probability, whose Bayesian false discovery rate does not exceed `alpha`. The rule is monotone: if the top- k set qualifies, so does every smaller top- k' set. If no set qualifies, nothing is selected.

Usage

```
bfdr_threshold_set(pip, alpha)
```

Arguments

<code>pip</code>	Numeric vector of posterior inclusion probabilities in $[0, 1]$, one per triplet.
<code>alpha</code>	Target Bayesian false discovery rate in $(0, 1)$.

Value

A logical vector of the same length as `pip`, in input order, that is TRUE for the selected triplets.

See Also

[bayes_fdr\(\)](#)

Examples

```
pip <- c(0.9, 0.8, 0.5, 0.2)
bayes_fdr(pip)
bfdr_threshold_set(pip, alpha = 0.2)
bfdr_threshold_set(pip, alpha = 0.05)
```

<code>credible_intervals</code>	<i>Credible intervals for the MR-CCC effect sizes</i>
---------------------------------	---

Description

Computes equal-tailed posterior credible intervals for β_X , β_{XZ} and β_Z from the pooled retained draws of an MR-CCC fit.

Usage

```
credible_intervals(fit, level = 0.95, standardized = TRUE)
```

Arguments

<code>fit</code>	An object of class <code>mrccc_fit</code> returned by mr_ccc() .
<code>level</code>	Credible level in $(0, 1)$; the default 0.95 gives the 2.5 and 97.5 percent posterior quantiles.
<code>standardized</code>	Logical; if TRUE (the default) the draws are rescaled to the standardised scale (per standard deviation of X and Z , in standard deviations of Y) before the quantiles are taken. If FALSE, the intervals are on the centred input scale.

Details

The intervals are marginal posterior quantiles of each coefficient. Because the prior on (β_X, β_{XZ}) is a spike-and-slab mixture, the marginal posterior of each of these two coefficients is itself a mixture of a component concentrated near zero (draws with $\gamma = 0$) and a slab component (draws with $\gamma = 1$). An interval that covers zero therefore reflects the posterior weight of the spike as much as the uncertainty within the slab; see [sign_probabilities\(\)](#) for a decomposition.

Value

A data frame with columns parameter (beta_X, beta_XZ, beta_Z), lower, median and upper.

See Also

[mr_ccc\(\)](#), [sign_probabilities\(\)](#)

Examples

```
sim <- simulate_mrccc(n = 150, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
             n_iter = 600, burn_in = 100, seed = 1)
credible_intervals(fit)
credible_intervals(fit, level = 0.9, standardized = FALSE)
```

mr_ccc

Fit the MR-CCC model to one ligand-receptor-pathway triplet

Description

Fits the Bayesian Mendelian randomization model for causal cell-cell communication to a single triplet consisting of ligand expression in a sender cell type, receptor expression in a receiver cell type, and pathway activity in the receiver cell type, using cis-eQTL genotypes as instruments. The function validates and centres the inputs, runs one or more Gibbs chains through [mr_ccc_gibbs\(\)](#), pools the retained draws, and computes the posterior inclusion probability (PIP) together with its Monte Carlo standard error and convergence diagnostics.

Usage

```
mr_ccc(
  X,
  Z,
  Y,
  G,
  H,
  V,
  n_iter = 20000,
  burn_in = 2000,
  thin = 1,
  n_chains = 4,
  init_scale = 1,
  pip_threshold = 0.5,
  hyper = mrccc_hyperparameters(),
  seed = NULL,
  verbose = FALSE
)
```

Arguments

X	Ligand expression in the sender cell type: a numeric vector of length n or an $n \times 1$ matrix, one value per donor. A SummarizedExperiment with exactly one row (gene) is accepted directly; its first assay row is used. <code>se_row()</code> selects a gene from a larger SummarizedExperiment.
Z	Receptor expression in the receiver cell type, same format as X (including a single-row SummarizedExperiment).
Y	Pathway activity in the receiver cell type, same format as X (including a single-row SummarizedExperiment).
G	Numeric matrix with n rows: cis-eQTL genotypes that instrument the ligand (at least one column).
H	Numeric matrix with n rows: cis-eQTL genotypes that instrument the receptor (at least one column).
V	Numeric matrix with n rows of shared donor covariates (for example genotype principal components, age or sex). Required, with at least one column; omitting it or supplying NULL is an error. When no covariates are available, pass a single non-constant column such as age. Columns are centred internally.
n_iter	Total number of Gibbs iterations per chain.
burn_in	Number of initial iterations discarded per chain. Must be smaller than n_iter.
thin	Thinning interval: every thin-th post-burn-in iteration is retained.
n_chains	Number of independent chains, pooled for all posterior summaries. With two or more chains the Gelman-Rubin statistic is reported and the chains start from dispersed values. Pooling C chains of n_iter iterations gives roughly the Monte Carlo precision of one chain of $C * (n_iter - burn_in)$ retained draws, because per-chain standard errors combine in quadrature; each chain repeats the burn-in, which is the only extra cost.
init_scale	Standard deviation of the normal distribution from which β_X and β_{XZ} are initialised when n_chains > 1. With a single chain the sampler starts deterministically at zero.
pip_threshold	Threshold in (0, 1) that defines a discovery ($PIP > pip_threshold$). The default of 0.5 corresponds to the median probability model. Used only for the borderline warning and for the summary methods.
hyper	Prior hyperparameters, as returned by <code>mrccc_hyperparameters()</code> .
seed	Optional integer; if supplied, the chains are run under this seed, which makes the whole multi-chain fit reproducible as a unit, and the caller's random number state is restored afterwards. The default NULL uses the current random number stream, so a caller can equally call <code>set.seed()</code> beforehand and omit this argument.
verbose	Logical; if TRUE, each chain prints a progress line every 1000 iterations.

Details

Model. With $X^* = E[X \mid G, V]$ and $Z^* = E[Z \mid H, V]$ the genetically predicted ligand and receptor levels, the outcome model is

$$Y = \mu + \gamma(\beta_X X^* + \beta_{XZ} X^* Z^*) + \beta_Z Z^* + \alpha_Y^\top V + \varepsilon_Y,$$

where $\gamma \in \{0, 1\}$ is a spike-and-slab inclusion indicator for the causal block (β_X, β_{XZ}) . The indicator is implemented through the prior variance of that block (a continuous spike-and-slab: under $\gamma = 0$ the prior variance is multiplied by `nu1`, 10^{-4} by default), so that under the spike the two coefficients are concentrated tightly at zero rather than fixed at zero; at this value of `nu1` the two formulations are numerically indistinguishable. The posterior inclusion probability $\text{PIP} = P(\gamma = 1 \mid \text{data})$ is the communication score. The receptor-modulated ligand effect is $\beta_X + \beta_{XZ}z$, which changes sign at $\tau = -\beta_X/\beta_{XZ}$; see `sign_reversal()`.

Centring. The model requires X , Z and Y , and every column of G , H and V , to have mean zero: neither first stage carries an intercept, so uncentred instruments or covariates would force each first-stage fit through the origin. All six inputs are centred internally. The standard deviations of the original X , Z and Y are stored in the `scale` element so that effects can be reported on a standardised scale (per standard deviation of X and Z , in standard deviations of Y).

Priors. The first-stage blocks $\pi_X, \alpha_X, \pi_Z, \alpha_Z$ and the covariate block α_Y carry Zellner g-priors on their observed designs G , H and V . The second-stage blocks are scaled by the size of their regressors, but with that scale fixed:

$$(\beta_X, \beta_{XZ}) \mid \gamma, \sigma_Y^2 \sim N_2(0, g s_\gamma \sigma_Y^2 D_0^{-1}), \quad \beta_Z \mid \sigma_Y^2 \sim N(0, g \sigma_Y^2 / d_Z),$$

where $D_0 = \text{diag}(\|\hat{X}^*\|^2, \|\hat{X}^* \circ \hat{Z}^*\|^2)$ and $d_Z = \|\hat{Z}^*\|^2$ are computed once from the least-squares first stage and held fixed, and s_γ is 1 in the slab and `nu1` in the spike. Equivalently, the causal coefficients carry independent normal priors on the scale of the standardised regressors. The scale is fixed rather than recomputed from the current draws because X^* and Z^* are functions of the first-stage parameters; a prior whose covariance moved with them would enter their full conditionals, and the sampler would no longer be a Gibbs sampler for the stated model. The values used are returned in `settings$prior_scale`. Unless overridden in `hyper`, every scale g is set to $\min(n, 100)$.

Instrument strength. The first-stage partial F statistic of each instrument block, after the covariates, is computed by ordinary least squares and returned in `instruments`. When either F is below 10 a message is issued. The threshold is the conventional two-stage least-squares heuristic and is reported for orientation: the one-sample Bayesian model fitted here does not require it, but posterior inclusion probabilities are less decisive under weak instruments and longer chains may be needed.

Chains and pooling. Chain c starts with $\gamma = 1$ when c is odd and $\gamma = 0$ when c is even, and, when `n_chains` > 1, with (β_X, β_{XZ}) drawn from a normal distribution with mean zero and standard deviation `init_scale`. Retained draws are pooled across chains for all posterior summaries.

Monte Carlo error of the PIP. The PIP is the mean of an autocorrelated binary chain, so its Monte Carlo standard error (MCSE) is estimated by non-overlapping batch means within each chain and combined across chains in quadrature: $\text{MCSE} = \sqrt{\sum_c \text{MCSE}_c^2 / C}$. Effective sample sizes add across chains. Batch means are never applied to the concatenated chains, because a batch straddling the join between two chains started at different values would inflate the variance estimate. If $|\text{PIP} - t| < 2 \text{MCSE}$, where t is `pip_threshold`, the discovery status of the triplet is not resolved at the current chain length and a warning is issued.

Convergence diagnostics. The Gelman-Rubin statistic $\hat{R}$ is reported for $\beta_X, \beta_{XZ}, \gamma$ and the joint log-likelihood. It compares the variance between chains with the variance within them, so it requires `n_chains` >= 2 and is NA for a single chain; it is also only informative when the chains start from dispersed values, which `init_scale` > 0 and the alternating γ start provide. Values below 1.01 are conventionally taken as consistent with convergence; a warning is issued when any $\hat{R}$ exceeds 1.01,

and a message when only one chain is run. Because a useful $\hat{R}$ is the reason to run several chains at all, the default `n_chains` is 4.

Units. The g-priors are invariant to the scale of the inputs, but the inverse-gamma priors on the residual variances are not: their defaults assume that X , Z and Y are on a scale of roughly unit variance. Inputs on very different scales should be standardised first.

Value

An object of class `mrccc_fit`: a list with elements

`pip` Posterior inclusion probability, pooled over chains.

`pip_mcse` Monte Carlo standard error of `pip`.

`pip_ess` Effective sample size of the pooled γ chain.

`estimates` Data frame with columns `parameter` (`beta_X`, `beta_XZ`, `beta_Z`), `raw_mean` (posterior mean on the centred input scale) and `std_mean` (posterior mean on the standardised scale).

`draws` List of pooled numeric vectors `beta_X`, `beta_XZ`, `beta_Z`, `gamma`, `mu` and `loglik`, the first five on the raw (centred) scale, each of length `n_keep * n_chains`. `loglik` is the joint log-likelihood of the three equations at each retained draw and is useful for trace plots.

`scale` List with `sd_X`, `sd_Z`, `sd_Y` (standard deviations of the original inputs), `mean_X`, `mean_Z`, `mean_Y` (the subtracted means) and `Z_range_std` (range of the centred receptor expression in standard-deviation units).

`diagnostics` Data frame with columns `parameter` (`beta_X`, `beta_XZ`, `gamma`, `loglik`) and `rhat`.

`instruments` Data frame with one row per exposure and columns `exposure` (X or Z), `n_instruments`, `F_statistic` (first-stage partial F of the instrument block after the covariates) and `weak` (TRUE when `F_statistic` < 10).

`settings` List with `n_iter`, `burn_in`, `thin`, `n_chains`, `n_keep` (retained draws per chain), `pip_threshold`, `hyper`, `g` (the g-prior scale used) and `prior_scale` (the fixed second-stage prior scale `d_X`, `d_XZ`, `d_Z`).

`call` The matched call.

References

Sarkar B, Ni Y (2026). MR-CCC: Bayesian Mendelian Randomization for Causal Cell-Cell Communication. [arXiv:2604.23917](https://arxiv.org/abs/2604.23917).

See Also

[summary.mrccc_fit\(\)](#), [credible_intervals\(\)](#), [sign_probabilities\(\)](#), [sign_reversal\(\)](#), [plot_effect_curve\(\)](#), [bayes_fdr\(\)](#), [simulate_mrccc\(\)](#), [mrccc_hyperparameters\(\)](#).

Examples

```
sim <- simulate_mrccc(n = 150, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
              n_iter = 600, burn_in = 100, seed = 1)
fit
fit$pip
```

```
fit$estimates
```

```
mr_ccc_gibbs
```

Gibbs sampler for the MR-CCC model

Description

Runs the blocked Gibbs sampler for a single ligand-receptor-pathway triplet and returns posterior means together with the retained draws of the scalar parameters of inferential interest. This is the low-level engine called by `mr_ccc()`; most users should call `mr_ccc()`, which validates and centres the inputs, runs several chains, and computes convergence diagnostics.

Usage

```
mr_ccc_gibbs(
  X,
  Z,
  Y,
  G,
  H,
  V,
  n_iter = 20000L,
  burn_in = 2000L,
  thin = 1L,
  a_sigma = 3,
  b_sigma = 2,
  a_rho = 3,
  b_rho = 1,
  nu1 = 1e-04,
  gG = 100,
  gV = 100,
  gH = 100,
  gZ = 100,
  gBeta = 100,
  ridge = 1e-08,
  init_gamma = 1L,
  init_scale = 0,
  verbose = FALSE
)
```

Arguments

X	Numeric matrix of dimension n x 1: centred ligand expression in the sender cell type.
Z	Numeric matrix of dimension n x 1: centred receptor expression in the receiver cell type.

Y	Numeric matrix of dimension n x 1: centred pathway activity in the receiver cell type.
G	Numeric matrix of dimension n x pG: column-centred sender cis-eQTL genotypes (instruments for X).
H	Numeric matrix of dimension n x pH: column-centred receiver cis-eQTL genotypes (instruments for Z).
V	Numeric matrix of dimension n x pV, pV >= 1: column-centred shared covariates. The first stages have no intercept, so all three design matrices must be centred, as <code>mr_ccc()</code> does.
n_iter	Total number of Gibbs iterations.
burn_in	Number of initial iterations discarded.
thin	Thinning interval; every thin-th post-burn-in iteration is retained.
a_sigma, b_sigma	Inverse-gamma shape and scale for the three residual variances.
a_rho, b_rho	Beta shape parameters for the prior inclusion probability rho.
nu1	Spike variance multiplier; small values enforce a near-zero spike.
gG, gV, gH, gZ, gBeta	Prior scale factors for the first-stage sender effects, the covariate effects, the first-stage receiver effects, the receptor main effect, and the causal block (Beta_X, Beta_XZ). The first three are Zellner g-priors on the observed designs G, V and H; the last two scale a fixed diagonal computed once from the least-squares first stage (see Details).
ridge	Diagonal ridge added before every matrix inversion.
init_gamma	Starting value (0 or 1) of the inclusion indicator.
init_scale	If greater than zero, Beta_X and Beta_XZ start from independent normal draws with standard deviation init_scale.
verbose	Logical; if TRUE, a progress line is printed every 1000 iterations.

Details

The priors on the second-stage coefficients are

$$(\beta_X, \beta_{XZ}) \mid \gamma, \sigma_Y^2 \sim N_2(0, g_\beta s_\gamma \sigma_Y^2 D_0^{-1}), \quad \beta_Z \mid \sigma_Y^2 \sim N(0, g_Z \sigma_Y^2 / d_Z),$$

with $D_0 = \text{diag}(\|\hat{X}^*\|^2, \|\hat{X}^* \circ \hat{Z}^*\|^2)$ and $d_Z = \|\hat{Z}^*\|^2$ computed once from least squares of X on [G V] and of Z on [H V] and held fixed, and $s_\gamma = 1$ in the slab and nu1 in the spike. Because the projected regressors X^* and Z^* are functions of the first-stage parameters, a prior whose covariance was recomputed from the current draws would enter the full conditionals of those parameters; fixing the scale keeps every update an exact conjugate step. The values used are returned in `prior_scale`.

Value

A named list with posterior means (Pi_X_mean, Alpha_X_mean, Pi_Z_mean, Alpha_Z_mean, Beta_X_mean, Beta_XZ_mean, Beta_Z_mean, Alpha_Y_mean, mu_mean, sigma_X_sq_mean, sigma_Z_sq_mean, sigma_Y_sq_mean, gamma_mean, rho_mean), the number of retained draws n_keep, the retained

draws Beta_X_draws, Beta_XZ_draws, Beta_Z_draws, gamma_draws, mu_draws and loglik_draws (each a numeric vector of length n_keep), and prior_scale, a named numeric vector giving d_X, d_XZ and d_Z. loglik_draws is the joint log-likelihood of the three equations evaluated at each retained draw, excluding prior terms.

See Also

[mr_ccc\(\)](#) for the user-facing interface.

Examples

```
sim <- simulate_mrccc(n = 120, seed = 1)
ctr <- function(v) matrix(v - mean(v), ncol = 1)
ctr_cols <- function(M) scale(M, center = TRUE, scale = FALSE)
out <- mr_ccc_gibbs(ctr(sim$X), ctr(sim$Z), ctr(sim$Y),
                   ctr_cols(sim$G), ctr_cols(sim$H), ctr_cols(sim$V),
                   n_iter = 400, burn_in = 100, thin = 1)

out$gamma_mean
out$prior_scale
length(out$Beta_XZ_draws) == out$n_keep
```

mrccc_hyperparameters *Hyperparameters of the MR-CCC prior*

Description

Constructs and validates the list of prior hyperparameters used by [mr_ccc\(\)](#). The defaults are those used throughout the MR-CCC analyses and are weakly informative for inputs on a scale of roughly unit variance: with several hundred donors the likelihood then dominates the inverse-gamma and beta priors. The inverse-gamma priors are not invariant to the units of X, Z and Y, so inputs on very different scales should be standardised before fitting.

Usage

```
mrccc_hyperparameters(
  a_sigma = 3,
  b_sigma = 2,
  a_rho = 3,
  b_rho = 1,
  nu1 = 1e-04,
  g = NULL,
  ridge = 1e-08
)
```

Arguments

a_sigma, b_sigma	Shape and scale of the inverse-gamma prior on the three residual variances $\sigma_X^2, \sigma_Z^2, \sigma_Y^2$. Both must be positive.
a_rho, b_rho	Shape parameters of the beta prior on the prior inclusion probability ρ . Both must be positive.
nu1	Spike variance multiplier in $(0, 1)$. The spike component of the prior on (β_X, β_{XZ}) has the slab covariance scaled by nu1, so small values enforce a near-zero spike.
g	Common g-prior scale applied to every coefficient block (first-stage sender effects, first-stage receiver effects, covariate effects, receptor main effect, and the causal block). If NULL (the default), <code>mr_ccc()</code> sets $g = \min(n, 100)$, where n is the number of donors. A positive number overrides this rule.
ridge	Non-negative diagonal ridge added before every matrix inversion for numerical stability.

Details

Each first-stage block θ with observed design D (the genotype matrices G , H or the covariates V) has the Zellner g-prior $\theta \sim N(0, g \sigma^2 (D^\top D)^{-1})$. The second-stage blocks (β_X, β_{XZ}) and β_Z use the same scale g but against a fixed diagonal computed once from the least-squares first stage, because their regressors X^* and Z^* are themselves parameters of the model; see the *Priors* section of `mr_ccc()`. The default $g = \min(n, 100)$ follows the unit-information convention for small samples and caps the prior dispersion for large ones. The spike-and-slab prior on the causal block uses scale g in the slab and $\nu_1 g$ in the spike.

Value

A named list of class `mrccc_hyper` with elements `a_sigma`, `b_sigma`, `a_rho`, `b_rho`, `nu1`, `g` and `ridge`.

See Also

`mr_ccc()`

Examples

```
hyper <- mrccc_hyperparameters()
str(hyper)

# A more diffuse slab and a fixed g-prior scale
mrccc_hyperparameters(nu1 = 1e-3, g = 50)
```

plot.mrccc_fit	<i>Plot the receptor-modulated ligand effect of an MR-CCC fit</i>
----------------	---

Description

Method for `plot()` that dispatches to `plot_effect_curve()`.

Usage

```
## S3 method for class 'mrccc_fit'
plot(x, Z_observed = NULL, ...)
```

Arguments

<code>x</code>	An object of class <code>mrccc_fit</code> returned by <code>mr_ccc()</code> .
<code>Z_observed</code>	Optional numeric vector of observed receptor expression (on the original scale) used to set the plotting range. If <code>NULL</code> , the range stored in the fit is used.
<code>...</code>	Further arguments passed to <code>plot_effect_curve()</code> .

Value

A ggplot object.

See Also

`plot_effect_curve()`

Examples

```
sim <- simulate_mrccc(n = 150, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
              n_iter = 600, burn_in = 100, seed = 1)
plot(fit)
```

plot_effect_curve	<i>Plot the receptor-modulated ligand effect with a credible band</i>
-------------------	---

Description

Draws the posterior mean of the receptor-modulated ligand effect $\text{effect}(z) = \beta_X + \beta_{XZ}z$ over the observed range of receptor expression, together with a pointwise credible band computed from paired posterior draws.

Usage

```
plot_effect_curve(
  fit,
  Z_observed = NULL,
  level = 0.95,
  n_draws = 2000,
  standardized = TRUE
)
```

Arguments

<code>fit</code>	An object of class <code>mrccc_fit</code> returned by <code>mr_ccc()</code> .
<code>Z_observed</code>	Optional numeric vector of observed receptor expression on the original (un-centred) scale, typically the <code>Z</code> passed to <code>mr_ccc()</code> . It determines the horizontal range of the curve. If <code>NULL</code> , the range stored in the fit is used.
<code>level</code>	Credible level in $(0, 1)$ of the pointwise band.
<code>n_draws</code>	Maximum number of paired draws used for the band. A systematic subsample of the pooled draws of this size is taken; the default of 2000 is ample for the quantiles of the band.
<code>standardized</code>	Logical; if <code>TRUE</code> (the default) the effect is shown on the standardised scale (standard deviations of Y per standard deviation of X) against receptor expression in standard-deviation units relative to its mean. If <code>FALSE</code> , the effect is on the centred input scale against centred receptor expression.

Details

The band is built from paired draws of (β_X, β_{XZ}) , for the same reason as the sign-reversal threshold in `sign_reversal()`: the curve depends on two correlated parameters at once, so combining their marginal intervals would misstate the uncertainty. At each grid point z , the band is the equal-tailed interval of $\beta_X^{(s)} + \beta_{XZ}^{(s)}z$ across the subsampled draws. A horizontal reference line marks an effect of zero; where the band crosses it, the direction of the ligand effect is uncertain.

The draws include both spike ($\gamma = 0$) and slab ($\gamma = 1$) iterations, so the band reflects the full spike-and- slab posterior.

Value

A `ggplot` object that can be further modified.

See Also

`mr_ccc()`, `sign_reversal()`, `plot.mrccc_fit()`

Examples

```
sim <- simulate_mrccc(n = 150, beta_X = 0.3, beta_XZ = -0.3, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
  n_iter = 600, burn_in = 100, seed = 1)
p <- plot_effect_curve(fit, Z_observed = sim$Z)
```

p

print.mrccc_fit	<i>Print an MR-CCC fit</i>
-----------------	----------------------------

Description

Prints a compact description of an `mrccc_fit` object: the chain configuration, the posterior inclusion probability with its Monte Carlo standard error, and the standardised posterior means of the three effect sizes.

Usage

```
## S3 method for class 'mrccc_fit'
print(x, digits = 3L, ...)
```

Arguments

x	An object of class <code>mrccc_fit</code> returned by mr_ccc() .
digits	Number of significant digits used for printing.
...	Further arguments; ignored.

Value

x, invisibly.

See Also

[mr_ccc\(\)](#), [summary.mrccc_fit\(\)](#)

Examples

```
sim <- simulate_mrccc(n = 150, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
              n_iter = 600, burn_in = 100, seed = 1)
print(fit)
```

se_row	<i>Extract one gene from a SummarizedExperiment as a donor-level vector</i>
--------	---

Description

Extracts a single gene (row) from a `SummarizedExperiment` and returns it as a plain named numeric vector suitable for the X, Z or Y argument of `mr_ccc()`. The expected layout is donor-level (pseudo-bulk) expression with genes as rows and donors (samples) as columns, so the returned vector holds one value per donor, named by the column names of the object.

Usage

```
se_row(se, gene, assay = 1L)
```

Arguments

se	A <code>SummarizedExperiment</code> with genes as rows and donors (samples) as columns.
gene	A single character string: the name of the gene to extract. Must match one of <code>rownames(se)</code> .
assay	The assay to read from: either a single positive integer index or the name of an assay of se. Defaults to the first assay.

Value

A named numeric vector of length `ncol(se)` with names `colnames(se)`, holding the expression of gene in the selected assay.

See Also

[mr_ccc\(\)](#)

Examples

```
sim <- simulate_mrccc(n = 60, seed = 1)
m <- rbind(LIG1 = as.numeric(sim$X), REC1 = as.numeric(sim$Z))
colnames(m) <- paste0("donor", seq_len(ncol(m)))
se <- SummarizedExperiment::SummarizedExperiment(
  assays = list(logcounts = m)
)
x <- se_row(se, "LIG1")
head(x)
```

sign_probabilities	<i>Posterior sign probabilities of the interaction effect</i>
--------------------	---

Description

Computes the posterior probability that the ligand-by-receptor interaction coefficient β_{XZ} is positive and that it is negative, both marginally and conditionally on the causal block being included ($\gamma = 1$).

Usage

```
sign_probabilities(fit)
```

Arguments

`fit` An object of class `mrccc_fit` returned by `mr_ccc()`.

Details

The prior on (β_X, β_{XZ}) is a spike-and-slab mixture, so the marginal posterior of β_{XZ} is a mixture of two components: draws taken while $\gamma = 0$, which are concentrated near zero and carry no directional information, and draws taken while $\gamma = 1$, which come from the slab. The two summaries therefore answer different questions and both are informative.

- The **marginal** probabilities `p_pos_marginal` and `p_neg_marginal` use every retained draw. They combine the evidence that there is an effect at all with the evidence about its direction. By the law of total probability, $P(\beta_{XZ} > 0) = \text{PIP} \cdot P(\beta_{XZ} > 0 \mid \gamma = 1) + (1 - \text{PIP}) \cdot P(\beta_{XZ} > 0 \mid \gamma = 0)$, and the last factor is close to one half because the spike is symmetric about zero.
- The **conditional** probabilities `p_pos_conditional` and `p_neg_conditional` use only the draws with $\gamma = 1$. They describe the direction of the interaction given that the ligand acts on the pathway, and sum to one (up to draws exactly at zero).

A triplet with a high PIP and a conditional probability near one has a well-determined direction; a triplet with a modest PIP can still have a decisive conditional direction, in which case the marginal probability is limited mainly by the inclusion uncertainty. When no retained draw has $\gamma = 1$, the conditional columns are NA.

Value

A one-row data frame with columns `pip`, `p_pos_marginal`, `p_neg_marginal`, `p_pos_conditional`, `p_neg_conditional` and `n_conditional` (the number of pooled draws with $\gamma = 1$).

See Also

`mr_ccc()`, `credible_intervals()`, `sign_reversal()`

Examples

```
sim <- simulate_mrccc(n = 150, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
             n_iter = 600, burn_in = 100, seed = 1)
sign_probabilities(fit)
```

sign_reversal

Receptor level at which the ligand effect reverses sign

Description

Computes the posterior distribution of the sign-reversal threshold $\tau = -\beta_X/\beta_{XZ}$, the receptor expression level at which the receptor-modulated ligand effect $\beta_X + \beta_{XZ}z$ changes sign, from paired posterior draws of an MR-CCC fit.

Usage

```
sign_reversal(fit, Z_observed = NULL, level = 0.95)
```

Arguments

fit	An object of class <code>mrccc_fit</code> returned by <code>mr_ccc()</code> .
Z_observed	Optional numeric vector of observed receptor expression on the original (un-centred) scale, typically the <code>Z</code> passed to <code>mr_ccc()</code> . If supplied, the posterior probability that τ falls inside the observed range is also returned. If <code>NULL</code> , the range stored in the fit is used.
level	Credible level in $(0, 1)$ for the reported interval.

Details

τ is a ratio of two parameters, so its posterior can only be obtained from paired draws of (β_X, β_{XZ}) ; combining the two marginal summaries would misstate the uncertainty because the two coefficients are strongly correlated a posteriori. Draws in which the standardised interaction is negligible, $|\beta_{XZ}^{(s)}| = |\beta_{XZ}| \text{sd}(X) \text{sd}(Z) / \text{sd}(Y) \leq 10^{-8}$, are dropped, since τ diverges there. Applying the cutoff on the standardised scale makes it independent of the units of X , Z and Y . All remaining draws are used, including those from the spike ($\gamma = 0$); for a fit with a PIP well below one, the threshold is therefore best interpreted together with the PIP. Note that the model's interaction acts on the genetically predicted receptor level Z^* , whereas `p_in_range` refers to the range of the observed receptor expression.

Scale. `mr_ccc()` centres Z before fitting, so on the raw scale $\tau_{\text{raw}} = -\beta_X/\beta_{XZ}$ is expressed in the original units of Z relative to the mean receptor expression. The function reports $\tau = \tau_{\text{raw}}/\text{sd}(Z)$, in standard-deviation units of Z relative to its mean, which is the scale on which the standardised coefficients are defined: $-\beta_X^{(s)}/\beta_{XZ}^{(s)} = -(\beta_X/\beta_{XZ})/\text{sd}(Z)$. A value of $\tau = 0$ therefore corresponds to a donor with average receptor expression, and $\tau = 1$ to a donor one

standard deviation above the average. The observed range is expressed on the same scale, $(Z - \bar{Z})/\text{sd}(Z)$.

Whether the reversal is realised in the population is determined by `p_in_range`, the posterior probability that τ lies inside the observed standardised receptor range, rather than by τ itself: a threshold far outside the observed range describes a direction change that no donor exhibits. The posterior of a ratio is heavy-tailed, so the median and quantiles are reported rather than a mean.

Value

A list with elements

`tau_median` Posterior median of τ in standard-deviation units of Z .

`tau_lower`, `tau_upper` Equal-tailed credible bounds at `level`.

`level` The credible level.

`n_draws` Number of paired draws used.

`Z_range_std` The observed receptor range in standard-deviation units.

`p_in_range` Posterior probability that τ lies inside `Z_range_std`.

`tau_draws` The paired draws of τ (standardised scale).

All numeric summaries are NA when no usable paired draw is available.

See Also

[mr_ccc\(\)](#), [plot_effect_curve\(\)](#), [sign_probabilities\(\)](#)

Examples

```
sim <- simulate_mrccc(n = 150, beta_X = 0.3, beta_XZ = -0.3, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
              n_iter = 600, burn_in = 100, seed = 1)
sr <- sign_reversal(fit, Z_observed = sim$Z)
c(sr$tau_median, sr$tau_lower, sr$tau_upper, sr$p_in_range)
```

`simulate_mrccc`

Simulate one ligand-receptor-pathway triplet from the MR-CCC working model

Description

Generates donor-level data for a single triplet under the MR-CCC data generating process, including cis-eQTL instruments for the ligand and the receptor, observed covariates, and an unmeasured confounder that loads on the ligand, the receptor and the pathway. The output can be passed directly to [mr_ccc\(\)](#).

Usage

```
simulate_mrccc(
  n = 300,
  pG = 5,
  pH = 5,
  pV = 3,
  beta_X = 0.3,
  beta_XZ = 0.3,
  beta_Z = 0.5,
  pi_val = 0.5,
  conf_strength = 0.7,
  seed = NULL
)
```

Arguments

n	Number of donors.
pG	Number of sender instruments (columns of G).
pH	Number of receiver instruments (columns of H).
pV	Number of observed covariates (columns of V); at least one.
beta_X	True ligand main effect on the pathway.
beta_XZ	True ligand-by-receptor interaction effect.
beta_Z	True receptor main effect on the pathway.
pi_val	Common first-stage effect of every instrument on its target gene (instrument strength).
conf_strength	Loading of the unmeasured confounder U on X , Z and Y . Zero removes confounding.
seed	Optional integer; if supplied, the data are simulated under this seed and the caller's random number state is restored afterwards. The default NULL uses the current random number stream.

Details

With $G_{ij}, H_{ij}, V_{ij}, U_i \sim N(0, 1)$ independently, the model is

$$X_i = G_i^\top \pi_X + V_i^\top \alpha_X + cU_i + \varepsilon_{Xi},$$

$$Z_i = H_i^\top \pi_Z + V_i^\top \alpha_Z + cU_i + \varepsilon_{Zi},$$

$$Y_i = \beta_X X_i + \beta_Z Z_i + \beta_{XZ} X_i Z_i + V_i^\top \alpha_Y + cU_i + \varepsilon_{Yi},$$

where c is `conf_strength`, every element of π_X and π_Z equals `pi_val`, every element of α_X , α_Z and α_Y equals 0.3, and the residuals are standard normal. The confounder U induces a spurious association between the exposures and the outcome that a naive regression absorbs into the causal coefficients; the instruments allow MR-CCC to separate the causal signal from it. Setting `beta_XZ = 0` gives a null triplet.

Value

A list with elements X, Z, Y (numeric matrices of dimension $n \times 1$), G, H, V (numeric matrices with n rows), U (the unmeasured confounder, for reference only) and truth, a list of the generating parameters beta_X, beta_XZ, beta_Z, pi_val, conf_strength and tau (the true sign-reversal threshold $-\beta_X/\beta_{XZ}$ in raw units of Z, or NA when beta_XZ is zero).

See Also

[mr_ccc\(\)](#)

Examples

```
sim <- simulate_mrccc(n = 200, seed = 42)
str(sim$truth)
dim(sim$G)

# A null triplet
null_sim <- simulate_mrccc(n = 200, beta_X = 0, beta_XZ = 0, seed = 42)
```

summary.mrccc_fit	<i>Summarise an MR-CCC fit</i>
-------------------	--------------------------------

Description

Prints the posterior inclusion probability with its Monte Carlo standard error and effective sample size, the standardised posterior means and credible intervals of the three effect sizes, the convergence diagnostics, and a note when the discovery status is not resolved at the current chain length. The assembled table is returned invisibly.

Usage

```
## S3 method for class 'mrccc_fit'
summary(object, level = 0.95, digits = 3L, ...)
```

Arguments

object	An object of class mrccc_fit returned by mr_ccc() .
level	Credible level for the reported intervals.
digits	Number of significant digits used for printing.
...	Further arguments; ignored.

Details

The discovery status is considered unresolved when $|\text{PIP} - t| < 2 \text{ MCSE}$, where t is the discovery threshold stored in the fit; in that case a note recommends a longer run or additional chains. The Gelman-Rubin column is NA unless the fit used at least two chains.

Value

Invisibly, a data frame with one row per quantity (beta_X, beta_XZ, beta_Z, pip) and columns raw_mean, std_mean, std_lower, std_upper and rhat. For the pip row the raw and standardised means both equal the PIP, the interval columns are NA, and rhat refers to the γ chain.

See Also

[mr_ccc\(\)](#), [credible_intervals\(\)](#)

Examples

```
sim <- simulate_mrccc(n = 150, seed = 1)
fit <- mr_ccc(sim$X, sim$Z, sim$Y, sim$G, sim$H, sim$V,
              n_iter = 600, burn_in = 100, seed = 1)
tab <- summary(fit)
tab
```

Index

bayes_fdr, [2](#)
bayes_fdr(), [4](#), [8](#)
bfdr_threshold_set, [3](#)
bfdr_threshold_set(), [3](#)

credible_intervals, [4](#)
credible_intervals(), [8](#), [17](#), [22](#)

mr_ccc, [5](#)
mr_ccc(), [3–5](#), [9–19](#), [21](#), [22](#)
mr_ccc_gibbs, [9](#)
mr_ccc_gibbs(), [5](#)
mrccc_hyperparameters, [11](#)
mrccc_hyperparameters(), [6](#), [8](#)

plot(), [13](#)
plot.mrccc_fit, [13](#)
plot.mrccc_fit(), [14](#)
plot_effect_curve, [13](#)
plot_effect_curve(), [8](#), [13](#), [19](#)
print.mrccc_fit, [15](#)

se_row, [16](#)
se_row(), [6](#)
set.seed(), [6](#)
sign_probabilities, [17](#)
sign_probabilities(), [4](#), [5](#), [8](#), [19](#)
sign_reversal, [18](#)
sign_reversal(), [7](#), [8](#), [14](#), [17](#)
simulate_mrccc, [19](#)
simulate_mrccc(), [8](#)
summary.mrccc_fit, [21](#)
summary.mrccc_fit(), [8](#), [15](#)