

Package ‘Entropic.Scree’

September 15, 2026

Title Information-Theoretic Dimensionality Estimation

Version 1.0.1

Copyright Terrence J. Lee-St. John (Enli)

Description An information-theoretic diagnostic technique for estimating the intrinsic dimensionality of tabular datasets. Evaluates shared probability mass via a transformed mutual information metric. Aims to extract the Intrinsic Generative Rank (r) and structural topology. For full methodological details, see the preprint by Lee-St. John (2026) <<https://zenodo.org/records/22028087>>.

URL <https://zenodo.org/records/22028087>

License Apache License 2.0

Encoding UTF-8

Imports data.table, ggplot2, infotheo, parallel, patchwork, Rcpp

LinkingTo Rcpp

Config/roxygen2/version 8.1.0

NeedsCompilation yes

Author Terrence J. Lee-St. John [aut, cre]

Maintainer Terrence J. Lee-St. John <terry@enli.com.au>

Repository CRAN

Date/Publication 2026-09-15 14:00:02 UTC

Contents

Entropic.Scree	2
Update.Entropic.Scree	4
Index	6

Description

An information-theoretic diagnostic technique for estimating the intrinsic dimensionality of tabular datasets. Evaluates shared probability mass via a transformed mutual information metric. Aims to extract the Intrinsic Generative Rank (r) and structural topology. For full methodological details, see the preprint at <https://zenodo.org/records/22028087>.

Usage

```
Entropic.Scree(
  data,
  low_entropy_thresh = 0.05,
  num_bins = NULL,
  bin_multiplier = 1,
  num_cores = parallel::detectCores() - 2,
  interactive_mode = TRUE,
  purge_constants = TRUE,
  check_collinearity = TRUE,
  triple_tap_window = 20,
  fwer_alpha = 0.01,
  extract_eigenvectors = FALSE,
  extract_bipolar_modules = FALSE,
  bipolar_top_n = 0.2,
  return_processed_data = FALSE
)
```

Arguments

<code>data</code>	A <code>data.table</code> containing the dataset to evaluate. Base <code>data.frames</code> are not supported.
<code>low_entropy_thresh</code>	Numeric. Minimum marginal entropy threshold. Default is 0.05.
<code>num_bins</code>	Integer. Number of bins for discretization. Default is NULL (auto-calculated).
<code>bin_multiplier</code>	Numeric. Multiplier for the automatic bin calculation. Default is 1.0.
<code>num_cores</code>	Integer. Number of CPU cores for OpenMP parallelization.
<code>interactive_mode</code>	Logical. If TRUE, displays plot and prompts user to confirm ranks.
<code>purge_constants</code>	Logical. If TRUE, automatically removes zero-variance variables.
<code>check_collinearity</code>	Logical. If TRUE, purges perfectly collinear variables.

triple_tap_window	Integer. Window size for the triple-tap heuristic. Default is 20.
fwer_alpha	Numeric. Family-wise error rate alpha threshold. Default is 0.01.
extract_eigenvectors	Logical. If TRUE, extracts and returns eigenvectors.
extract_bipolar_modules	Logical. If TRUE, forces eigenvector extraction to build topological poles.
bipolar_top_n	Numeric. Proportion of top variables to include in poles.
return_processed_data	Logical. If TRUE, includes the processed data.table in the output.

Value

A list of class "entropic_scree" containing the following components:

eigenvalues	A numeric vector of extracted eigenvalues.
similarity_matrix	The double-centered mutual information matrix.
retained_features	A character vector of variables that passed entropy thresholds.
bin_distributions	A table of bin counts for the discretized data.
R_eff	Total Unique Probabilistic Volume scalar.
K_log_gap	Integer estimate for the Observed Generative Rank.
K_triple_tap	Integer estimate for the Extended Signal Tail.
triple_tap_multiplier	The local dynamic t-multiplier used in Engine B.
K_roots	The finalized integer representing the Intrinsic Generative Rank.
K_extended	The finalized integer representing the Extended Signal Tail.
top_of_bulk	Index marking the boundary of idiosyncratic variance.
total_signal_volume	Total shared signal volume scalar.
unique_signal_volume	Unique signal volume scalar.
redundant_signal_volume	Redundant signal volume scalar.
idiosyncratic_volume	Idiosyncratic informational variance scalar.
AIG	Average Informational Gravity variable equivalent.
FSIG_final	A numeric vector of Factor-Specific Informational Gravity.
structural_topology_profile	A numeric vector of relative factor gravities.
FSIG_extended_bulk	Extended FSIG metric utilizing the macro bulk bound.

FSIG_extended_kaiser Extended FSIG metric utilizing the Kaiser bound.

eigenvectors A matrix of extracted eigenvectors, if requested.

bipolar_modules A structured list of topological extraction poles, if requested.

processed_data The pre-processed data.table, if requested.

Examples

```
# Generate a small random dataset
dummy_data <- data.table::data.table(
  V1 = rnorm(50),
  V2 = rnorm(50),
  V3 = rnorm(50)
)

# Run the estimation (interactive_mode MUST be FALSE for automated checks)
res <- Entropic.Scree(dummy_data, interactive_mode = FALSE, num_cores = 1)
```

Update.Entropic.Scree *Update Entropic Scree Rank Estimates*

Description

A companion function to modify the rank estimates of a previously evaluated Entropic Scree object. This allows users to recalculate informational gravity and structural composition without re-running the computationally expensive C++ mutual information engine.

Usage

```
Update.Entropic.Scree(
  scree_obj,
  new_K_roots = NULL,
  new_K_extended = NULL,
  bipolar_top_n = 0.2,
  interactive_mode = TRUE
)
```

Arguments

scree_obj A valid output list from Entropic.Scree().

new_K_roots Integer. The new Observed Generative Rank. If NULL, retains original.

new_K_extended Integer. The new Extended Signal Tail Rank. If NULL, retains original.

bipolar_top_n Numeric. Proportion of top variables to include in newly extracted poles.

interactive_mode Logical. If TRUE, prints updated metric dashboards to the console.

Value

A mutated list of class "entropic_scree" containing the updated structural topology profiles, signal volumes, and informational gravity metrics based on the new rank thresholds.

Index

Entropic.Scree, [2](#)

Update.Entropic.Scree, [4](#)