Package 'BimodalIndex'

September 24, 2026

Version 1.1.13

Date 2026-09-24

Title The Bimodality Index

Depends R (>= 3.0)

Imports oompaBase (>= 3.0.1), mclust

Suggests oompaData

Description Defines the functions used to compute the bimodal index as defined by Wang et al. (2009) <https://pmc.ncbi.nlm.nih.gov/articles/PMC2730180/>, <doi:10.4137/CIN.S2846>.

License Apache License (== 2.0)

LazyLoad yes

biocViews Microarray

URL http://silicovore.com/OOMPA/classpred.html

NeedsCompilation no

Author Kevin R. Coombes [aut, cre]

Maintainer Kevin R. Coombes <krc@silicovore.com>

Repository CRAN

Date/Publication 2026-09-24 20:50:02 UTC

Contents

bimodalIndex 2

Index **4**

`bimodalIndex` *The Bimodality Index*

Description

The "bimodality index" is a continuous measure of the extent to which a set of (univariate) data fits a two-component mixture model. The score is larger if the two components are balanced in size or if the separation between the two modes is larger.

Usage

```
bimodalIndex(dataset, verbose=TRUE)
```

Arguments

`dataset`	A matrix or data.frame, usually with columns representing samples and rows representing genes or proteins.
`verbose`	A logical value; should the function output a stream of information while it is working?

Details

Identifying genes with bimodal expression patterns from large-scale expression profiling data is an important analytical task, which is often addressed using model-based clustering. That technique commonly uses the Bayesian information criterion (BIC) or the Akaike information criterion (AIC) for model selection. In practice, however, BIC and AIC appear to be overly sensitive and may lead to the identification of bimodally expressed genes that are unreliable or not clinically useful. We propose using a novel criterion, the bimodality index, not only to identify but also to rank meaningful and reliable bimodal patterns.

We model the data as a mixture

$$\pi N(\mu_1, \sigma) + (1 - \pi) N(\mu_2, \sigma)$$

of two normal components with a common standard deviation. We define the standardized distance between the two means to be

$$\delta = \frac{|\mu_1 - \mu_2|}{\sigma}.$$

We then define the bimodality index as

$$BI = \delta \sqrt{\pi(1 - \pi)}.$$

The bimodality index can be computed by first using either a mixture model-based algorithm such as Mclust or by using Markov chain Monte Carlo (MCMC) techniques to estimate the model parameters. In this package, we rely on the Mclust implementation.

In the paper by Wang et al. referenced below, we provide a statistical justification for the definition of the bimodality index, based on considerations of power and sample size. Theoretical considerations suggest that, in the limit over the number of samples, a bimodality index of 1.1 or greater is

likely to indicate a "useful" bimodal pattern of expression. Higher cutoffs are needed when there are relatively few samples, and can be chosen by simulating from the null distribution. We carried out simulation studies and applied the method to real data from a lung cancer gene expression profiling study. Our findings suggest that BIC behaves like a lax cutoff based on the bimodality index (much smaller than 1), and that the bimodality index provides an objective measure to identify and rank meaningful and reliable bimodal patterns from large-scale gene expression datasets.

Value

Returns a data frame containing six columns, with the rows corresponding to the rows of the original data set. The columns contain the four parameters from the normal mixture model (`mu1`, `mu2`, `sigma`, and `pi`) along with the standardized distance `delta` and the bimodal index `BI`.

Author(s)

Kevin R. Coombes <krc@silicovore.com>

References

Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR.
The bimodality index: A criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data.
Cancer Informatics, 2009 Aug 5; 7:199–216.

Examples

```
library(oompaData)
data(lungData)
bi <- bimodalIndex(lung.dataset, verbose=FALSE)
summary(bi)
```

Index

∗ **models**
bimodalIndex, 2
∗ **multivariate**
bimodalIndex, 2

bimodalIndex, 2

Mclust, *2*