# Package 'profileScoreDist'

April 10, 2025

**Type** Package

**Title** Profile score distributions

**Version** 1.35.0

**Date** 2015-12-09

**Author** Paal O. Westermark

**Maintainer** Paal O. Westermark <pal-olof.westermark@charite.de>

**Description** Regularization and score distributions for position count
matrices.

**License** MIT + file LICENSE

**biocViews** Software, GeneRegulation, StatisticalMethod

**LazyData** TRUE

**LinkingTo** Rcpp

**Imports** Rcpp, BiocGenerics, methods, graphics

**Depends** R(>= 3.3)

**Suggests** BiocStyle, knitr, MotifDb

**VignetteBuilder** knitr

**RoxygenNote** 5.0.1

**NeedsCompilation** yes

**git_url** https://git.bioconductor.org/packages/profileScoreDist

**git_branch** devel

**git_last_commit** f562e3f

**git_last_commit_date** 2024-10-29

**Repository** Bioconductor 3.21

**Date/Publication** 2025-04-09

# Contents

---

backgroundDist *Background distribution.*

---

### Description

`backgroundDist` returns the background distribution of a profile object.

### Usage

```
backgroundDist(x)
```

### Arguments

x A ProfileDist object.

### Details

This is a generic function.

### Value

The background distribution vector.

### Examples

```
anObject <- ProfileDist()
backgroundDist(anObject)
```

---

| computeScoreDist | *Compute exact position weight/count matrix score distribution.* |

---

## Description

Computes the discretisized score distribution of a position count matrix (PCM) or a position weight matrix (PWM), using the method described by Rahmann et al.

## Usage

```
computeScoreDist(motif, gc, granularity = 0.01, unit = "nat")
```

## Arguments

| | |
|---|---|
| motif | A matrix representing a PCM or PWM; each column a position and each row a base corresponding to A, C, G, T. This order is assumed, unless the rows are correspondingly named in a different order. |
| gc | A scalar giving the GC fraction to assume. |
| granularity | The granularity of the discretization, defaults to 0.01. |
| unit | The logarithm unit of the score computed from the PCM or PWM, can be "nat" (default, natural logarithm), "bit" (base 2), or "dit" (base 10). |

## Value

a ProfileDist object

## References

Rahmann, S., Mueller, T., and Vingron, M. (2003). On the power of profiles for transcription factor binding site detection. Stat Appl Genet Mol Biol 2, Article7.

## Examples

```
data(INR)
thedist <- computeScoreDist(regularizeMatrix(INR), 0.5)
plotDist(thedist)
```

---

INR                              *The position count matrix for INR.*

---

**Description**

The position count matrix for the initiator (INR) core promoter element. This matrix was obtained from the JASPAR public domain database, but was originally published by P. Bucher (1990); in that publication (and elsewhere) it was termed Cap signal.

**Usage**

    INR

**Format**

A matrix with named rows corresponding to the counts for each of the four nucleotides.

**Value**

The position count matrix for INR.

**Source**

http://jaspar.genereg.net

**References**

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of Molecular Biology 212, 563???-578.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.-Y., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Research 42, D142–D147.

---

plotDist                         *Plot background and signal distributions.*

---

**Description**

`plotDist` creates a rudimentary plot of signals and backgrounds.

**Usage**

    plotDist(x)

## Arguments

x                A ProfileDist object.

## Details

This is a generic function.

## Value

The scores vector.

## Examples

```
data(INR)
thedist <- computeScoreDist(regularizeMatrix(INR), 0.5)
plotDist(thedist)
```

---

ProfileDist-class        *ProfileDist*

---

## Description

This class represents signal and background score distributions for a profile.

## Usage

```
## S4 method for signature 'ProfileDist'
show(object)

## S4 method for signature 'ProfileDist'
score(x)

## S4 method for signature 'ProfileDist'
signalDist(x)

## S4 method for signature 'ProfileDist'
backgroundDist(x)

## S4 method for signature 'ProfileDist'
plotDist(x)
```

## Arguments

object         A ProfileDist object for the show method.

x                A ProfileDist object.

**Value**

A ProfileDist object.

**Methods (by generic)**

- `show`: Shows useful information
- `score`: Accessor for the scores
- `signalDist`: Accessor for the signal distribution
- `backgroundDist`: Accessor for the background distribution
- `plotDist`: Simple plot method for signal and background distributions

**Slots**

`f` Signal distribution

`g` Background distribution

`Scores` Scores for the distributions

**Constructor**

`ProfileDist(f=numeric, g=numeric, Scores=numeric)`

---

| regularizeMatrix | *Careful regularization (pseudocount addition) to a position count matrix.* |
|---|---|

---

**Description**

Carries out the regularization suggested by Rahmann et al. This lets each column in the regularized matrix be a linear combination of the column in the non-regularized matrix and rho, the overall base distribution of all positions. The weighting of the linear combination is determined by the parameter E in a non-trivial way, see Rahmann et al. for more information. A default value E=1.5 usually works well.

**Usage**

```
regularizeMatrix(motif, E = 1.5)
```

**Arguments**

| | |
|---|---|
| `motif` | A position count matrix; each column a position and each row a base corresponding to A, C, G, T. This order is assumed, unless the rows are correspondingly named in a different order. |
| `E` | Weighting parameter between 0 and 3 for the regularization. |

## Value

The regularized matrix

## References

Rahmann, S., Mueller, T., and Vingron, M. (2003). On the power of profiles for transcription factor binding site detection. Stat Appl Genet Mol Biol 2, Article7.

## Examples

```
data(INR)
regularizeMatrix(INR)
```

---

scoreDistCutoffs          *False discovery rate and power for PWM Score distributions.*

---

## Description

Computes score cutoffs for a PWM or a PCM, given distributions as calculated with `computeScoreDist()`. Cutoffs can be computed for a given false discovery rate (FDR), for a given false negative rate (FNR), and the optimal tradeoff between the two, in the sense that $c \times FDR = FNR$ for some $c$ that the user may choose.

## Usage

```
scoreDistCutoffs(scoreDist, n, m = 1, c = 1, cutoff = 0.01)
```

## Arguments

| | |
|---|---|
| scoreDist | A ProfileDist object, as computed by `computeScoreDist()` |
| n | The number of scores considered for the given PWM. If one sequence is considered and a score is computed for all overlapping windows of the same length as the PWM, this will be the length of the sequence, minus the PWM length plus 1. If scanning a sequence and its reverse complement too, this number must be further multiplied by two. The number forms the basis for the FDR, since this is a multiple testing problem. |
| m | The number of true positives assumed for computing the FNR. |
| c | A factor expressing how much more important the FDR is compared to the FNR, when computing the tradeoff cutoff that considers both FDR and FNR. See Rahmann et al. for details. |
| cutoff | The FDR and FNR considered, typically 0.01 or 0.05. |

## Value

a list with elements:

**cutoffa** Score cutoff for FDR=cutoff

**cutoffb** Score cutoff for FNR=cutoff

**cutoffopt** Score cutoff for c*FDR = FNR

## References

Rahmann, S., Mueller, T., and Vingron, M. (2003). On the power of profiles for transcription factor binding site detection. Stat Appl Genet Mol Biol 2, Article7.

## Examples

```
data(INR)
thedist <- computeScoreDist(regularizeMatrix(INR), 0.5)
scoreDistCutoffs(thedist, n=2000, cutoff=0.05)
```

---

| signalDist | *Signal distribution.* |
|---|---|

---

## Description

signalDist returns the signal distribution of a profile object.

## Usage

```
signalDist(x)
```

## Arguments

x                         A ProfileDist object.

## Details

This is a generic function.

## Value

The signal distribution vector.

## Examples

```
anObject <- ProfileDist()
backgroundDist(anObject)
```

# Index