

# Package ‘caviarpd’

September 6, 2024

**Type** Package

**Title** Cluster Analysis via Random Partition Distributions

**Version** 0.3.13

**Description** Cluster analysis is performed using pairwise distance information and a random partition distribution. The method is implemented for two random partition distributions. It draws samples and then obtains and plots clustering estimates. An implementation of a selection algorithm is provided for the mass parameter of the partition distribution. Since pairwise distances are the principal input to this procedure, it is most comparable to the hierarchical and k-medoids clustering methods. The method is Dahl, Andros, Carter (2022+) <[doi:10.1002/sam.11602](https://doi.org/10.1002/sam.11602)>.

**License** MIT + file LICENSE | Apache License 2.0

**URL** <https://github.com/dbdahl/caviarpd>

**BugReports** <https://github.com/dbdahl/caviarpd/issues>

**Depends** R (>= 4.2.0)

**Suggests** salso (>= 0.3.0)

**SystemRequirements** Cargo (Rust's package manager), rustc (>= 1.66.1)

**Encoding** UTF-8

**RoxygenNote** 7.3.1

**NeedsCompilation** yes

**Author** David B. Dahl [aut, cre] (<<https://orcid.org/0000-0002-8173-1547>>),  
R. Jacob Andros [aut] (<<https://orcid.org/0000-0002-1289-385X>>),  
J. Brandon Carter [aut] (<<https://orcid.org/0000-0003-1687-0564>>),  
Alex Crichton [ctb] (Rust crates: cfg-if, proc-macro2),  
Brendan Zabaraszkas [ctb] (Rust crate: approx),  
David B. Dahl [ctb] (Rust crates: dahl-partition, dahl-salso, epa, roxido, roxido\_macro),  
David Tolnay [ctb] (Rust crates: proc-macro2, quote, syn, unicode-ident),  
Jim Turner [ctb] (Rust crate: ndarray),

Jorge Aparicio [ctb] (Rust crate: libm),  
 Josh Stone [ctb] (Rust crate: autocfg),  
 Mikhail Vorotilov [ctb] (Rust crate: roots),  
 R. Janis Goldschmidt [ctb] (Rust crate: matrixmultiply),  
 Sean McArthur [ctb] (Rust crate: num\_cpus),  
 Stefan Lankes [ctb] (Rust crate: hermit-abi),  
 The Cranelift Project Developers [ctb] (Rust crate: wasi),  
 The CryptoCorrosion Contributors [ctb] (Rust crates: ppv-lite86,  
 rand\_chacha),  
 The Rand Project Developers [ctb] (Rust crates: getrandom, rand,  
 rand\_chacha, rand\_core, rand\_distr, rand\_pcg),  
 The Rust Project Developers [ctb] (Rust crates: libc, num-complex,  
 num-integer, num-traits, rand, rand\_chacha, rand\_core),  
 Ulrik Sverdrup ``bluss" [ctb] (Rust crate: ndarray),  
 bluss [ctb] (Rust crates: matrixmultiply, rawpointer)

**Maintainer** David B. Dahl <dahl@stat.byu.edu>

**Repository** CRAN

**Date/Publication** 2024-09-06 16:00:12 UTC

## Contents

caviarpd . . . . .	2
<b>Index</b>	<b>5</b>

---

caviarpd	<i>Cluster Analysis via Random Partition Distributions</i>
----------	--

---

## Description

Returns a clustering estimate given pairwise distances using the CaviarPD method.

## Usage

```
caviarpd(
  distance,
  nClusters,
  mass = NULL,
  nSamples = 200,
  gridLength = 5,
  loss = "binder",
  temperature = 100,
  similarity = c("exponential", "reciprocal")[1],
  maxNClusters = 0,
  nRuns = 4,
  nCores = nRuns
)
```

**Arguments**

distance	An object of class 'dist' or a pairwise distance matrix.
nClusters	A numeric vector that specifies the range for the number of clusters to consider in the search for a clustering estimate.
mass	The mass value to use for sampling. If NULL, the mass value is found by inverting values from nClusters.
nSamples	The number of samples drawn per candidate estimate.
gridLength	The number of candidate estimates to consider. The final estimate is obtained from $nSamples \times gridLength$ total samples.
loss	The SALSO method (Dahl, Johnson, Müller, 2021) tries to minimize this expected loss when searching the partition space for an optimal estimate. This must be either "binder" or "VI".
temperature	A positive number that accentuates or dampens distance between observations.
similarity	Either "exponential" or "reciprocal" to indicate the desired similarity function.
maxNClusters	The maximum number of clusters that can be considered by the SALSO method.
nRuns	The number of runs of the SALSO algorithm.
nCores	The number of CPU cores to use. A value of zero indicates to use all cores on the system.

**Details**

A range for the number of clusters to be considered is supplied using the nClusters argument.

**Value**

A object of class `salso.estimate`, which provides a clustering estimate (a vector of cluster labels) that can be displayed and plotted.

**References**

D. B. Dahl, J. Andros, J. B. Carter (2023), Cluster Analysis via Random Partition Distributions, *Statistical Analysis and Data Mining*, doi:10.1002/sam.11602.

D. B. Dahl, D. J. Johnson, and P. Müller (2022), Search Algorithms and Loss Functions for Bayesian Clustering, *Journal of Computational and Graphical Statistics*, 31(4), 1189-1201, doi:10.1080/10618600.2022.2069779.

**Examples**

```
# To reduce load on CRAN servers, limit the number of samples, grid length, and CPU cores.
set.seed(34)
iris.dis <- dist(iris[,-5])
est <- caviarpd(distance=iris.dis, nClusters=c(2,4), nSamples=20, nCores=1)
if ( require("salso") ) {
  summ <- summary(est, orderingMethod=2)
  plot(summ, type="heatmap")
}
```

```
    plot(summ, type="mds")  
  }
```

# Index

caviarpd, [2](#)