Package ‘nlshrink’

May 9, 2026

Type Package

Title Non-Linear Shrinkage Estimation of Population Eigenvalues and Covariance Matrices

Version 1.0.1

Date 2016-04-11

Depends R (>= 3.2.3)

Description Non-linear shrinkage estimation of population eigenvalues and covariance matrices, based on publications by Ledoit and Wolf (2004, 2015, 2016).

License GPL-3

Imports stats (>= 3.2.3), MASS (>= 7.3-45), nloptr (>= 1.0.4), graphics (>= 3.2.3)

RoxygenNote 5.0.1

ByteCompile TRUE

NeedsCompilation no

Author Pratik Ramprasad [aut, cre]

Maintainer Pratik Ramprasad <pratik.ramprasad@gmail.com>

Repository CRAN

Date/Publication 2016-04-12 00:55:21

Contents

ESD *Compute the empirical spectral distribution (ESD) for a set of population eigenvalues*

Description

The Marcenko Pastur (MP) law relates the limiting distribution of the sample eigenvalues to that of the population eigenvalues. In the finite-dimensional case, the population spectral distribution (PSD) can be represented as a sum of point masses, and the empirical spectral distribution (ESD) can be obtained by solving the discretized MP equation. Theoretical and implementation details in the references.

Usage

```
ESD(tau, n)
```

Arguments

`tau`	(Required) A non-negative numeric vector of population eigenvalues.
`n`	(Required) A positive integer representing the number of datapoints of a hypothetical data matrix with dimension `c(n, p = length(tau))`.

Value

A named numeric vector of containing points of the ESD. The names give the corresponding points on the x axis.

References

- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. Journal of Multivariate Analysis, 139(2)
- Ledoit, O. and Wolf, M. (2016). Numerical Implementation of the QuEST function. arXiv:1601.05870 [stat.CO]

Examples

```
tau_ESD <- ESD(tau = rep(1,200), n = 300)
plot(names(tau_ESD), tau_ESD, ylab="F(x)", xlab="x")
```

lambda_estimate — *Generate sample eigenvalues from population eigenvalues*

Description

The Marcenko Pastur (MP) law relates the limiting distribution of the sample eigenvalues to that of the population eigenvalues. In the finite-dimensional case, the population spectral distribution (PSD) can be represented as a sum of point masses, and the empirical spectral distribution (ESD) can be obtained by solving the discretized MP equation. The QuEST function(see references), uses the quantile function of the ESD to compute the sample eigenvalues for any given ratio $c = p/n \in (0, \infty)$.

Usage

```
lambda_estimate(tau, n)
```

Arguments

`tau`	(Required) A non-negative numeric vector of population eigenvalues.
`n`	(Required) A positive integer representing the number of datapoints of a hypothetical data matrix with dimension `c(n, p = length(tau))`.

Value

A numeric vector of the same length as `tau`, containing the sample eigenvalue estimates, sorted in ascending order.

References

- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. Journal of Multivariate Analysis, 139(2)
- Ledoit, O. and Wolf, M. (2016). Numerical Implementation of the QuEST function. arXiv:1601.05870 [stat.CO]

Examples

```
lambda_estimate(tau = rep(1,200), n = 300)
```

`linshrink` *Linear-shrinkage estimator of population eigenvalues.*

Description

`linshrink` estimates the population eigenvalues from the sample eigenvalues by shrinking each sample eigenvalue towards the global mean based on a shrinkage factor. Details in referenced publications.

Usage

```
linshrink(X, k = 0)
```

Arguments

`X`	A data matrix.
`k`	(Optional) Non-negative integer less than `ncol(X)`. If `k == 0` (default), `X` is assumed to contain 1 class, which will be centered. If `k >= 1`, `X` is assumed to contain `k` classes, each of which has already been centered.

Value

A numeric vector of length `ncol(X)`, containing the population eigenvalue estimates sorted in ascending order.

References

- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2)
- Ledoit, O. and Wolf, M. (2016). Numerical Implementation of the QuEST function. arXiv:1601.05870 [stat.CO]

Examples

```
linshrink(X = matrix(rnorm(1e4, mean = 5), nrow = 100, ncol = 100)) # 1 class; will be centered
linshrink(X = matrix(rnorm(1e4), nrow = 100, ncol = 100), k = 1) # 1 class; no centering
```

linshrink_cov — *Linear-shrinkage estimator of population covariance matrix.*

Description

The linear shrinkage estimator of the population covariance matrix is computed by shrinking the sample covariance matrix towards the identity matrix based on a shrinkage factor. Note that the eigenvalues of the population covariance matrix estimate are not the same as the linear shrinkage estimates of population eigenvalues. Details in referenced publication.

Usage

```
linshrink_cov(X, k = 0)
```

Arguments

`X`	A data matrix.
`k`	(Optional) Non-negative integer less than `ncol(X)`. If `k == 0` (default), `X` is assumed to contain 1 class, which will be centered. If `k >= 1`, `X` is assumed to contain `k` classes, each of which has already been centered.

Value

Population covariance matrix estimate. A square positive semi-definite matrix of dimension `ncol(X)`.

References

- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2)

Examples

```
linshrink_cov(X = matrix(rnorm(1e4, mean = 5), nrow = 100, ncol = 100)) # 1 class; will be centered
linshrink_cov(X = matrix(rnorm(1e4), nrow = 100, ncol = 100), k = 1) # 1 class; no centering
```

nlshrink — *Package*

Description

A package for estimating population eigenvalues and covariance matrices, based on publications by Ledoit and Wolf (2004, 2012, 2015, 2016).

Details

A common assumption in statistics is that for a data matrix X of dimension $n \times p$, the number of predictor variables (p) vanishes relative to the number of datapoints (n) as $n \to \infty$. However, in modern datasets, it is often the case that p is comparable to or greater than n. In this scenario, a more appropriate asymptotic framework is to assume that the ratio $c := p/n$ approaches a finite positive value as $n, p \to \infty$. In this case, the sample covariance matrix S is no longer a consistent estimator of the population covariance matrix Σ. Similarly, the sample eigenvalues deviate substantially from the population eigenvalues. This package contains implementations of Ledoit and Wolf's linear and non-linear shrinkage population eigenvalue and covariance estimation methods, based on their 2016 publication and the accompanying MATLAB code. Theoretical and implementation details of these methods can be found in the following publications:

- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2)
- Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. Annals of Statistics, 40(2).
- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. Journal of Multivariate Analysis, 139(2).
- Ledoit, O. and Wolf, M. (2016). Numerical Implementation of the QuEST function. arXiv:1601.05870 [stat.CO].

`nlshrink_cov` *Non-linear shrinkage estimator of population covariance matrix.*

Description

`nlshrink_cov` calls `tau_estimate` to estimate the population eigenvalues. Note that the eigenvalues of the estimated population covariance matrix are not the same as the non-linear shrinkage estimates of the population eigenvalues. Theoretical and implementation details in references.

Usage

```
nlshrink_cov(X, k = 0, method = "nlminb", control = list())
```

Arguments

`X`	A data matrix.
`k`	(Optional) Non-negative integer less than `ncol(X)`. If `k == 0` (default), `X` is assumed to contain 1 class, which will be centered. If `k >= 1`, `X` is assumed to contain `k` classes, each of which has already been centered.
`method`	(Optional) The optimization routine called in `tau_estimate`. Choices are `nlminb` (default) and `nloptr`.
`control`	(Optional) A list of control parameters. Must correspond to the selected optimization method. See `nlminb`, `nloptr` for details.

Value

A numeric positive semi-definite matrix of dimension `ncol(X)`.

References

- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. Journal of Multivariate Analysis, 139(2)
- Ledoit, O. and Wolf, M. (2016). Numerical Implementation of the QuEST function. arXiv:1601.05870 [stat.CO]

Examples

```
# generate matrix of uniform random variates
X <- matrix(sapply(1:20, function(b) runif(50, max=b)), nrow = 50, ncol = 20)
Sigma <- diag((1:20)^2/12) # true population covariance matrix
nlshrink_X <- nlshrink_cov(X, k=0) # compute non-linear shrinkage estimate
linshrink_X <- linshrink_cov(X, k=0) # compute linear shrinkage estimate
S <- cov(X) # sample covariance matrix

# compare accuracy of estimators (sum of squared elementwise Euclidean distance)
sum((S-Sigma)^2)
sum((nlshrink_X - Sigma)^2)
sum((linshrink_X - Sigma)^2)
```

`nlshrink_demo` *Demonstration of non-linear shrinkage estimator of population eigenvalues*

Description

This is a demonstration of the non-linear shrinkage method for estimating population eigenvalues. The inputted population eigenvalues are used to simulate a matrix of multivariate normal random variates with covariance matrix `diag(tau)`. This data matrix is then used to estimate the input population eigenvalues using the non-linear shrinkage method. The output plot shows the comparison of the various estimators (sample eigenvalues, linear shrinkage, non-linear shrinkage) to the true population eigenvalues.

Usage

```
nlshrink_demo(tau = NULL, n = 300, p = 300, method = "nlminb",
  control = list())
```

Arguments

`tau`	(Optional) Input population eigenvalues. Non-negative Numeric vector of length `p`.
`n`	(Optional) Number of rows in simulated data matrix (Default 300).
`p`	(Optional) Number of columns in simulated data matrix (Default 300).
`method`	(Optional) The optimization routine called in `tau_estimate`. Choices are `nlminb` (default) and `nloptr`.
`control`	(Optional) A list of control parameters. Must correspond to the selected optimization method. See `nlminb`, `nloptr` for details.

NOTE

`nlminb` is usually robust and accurate, but does not allow equality constraints, so the sum of the estimated population eigenvalues is in general not equal to the sum of the sample eigenvalues. `nloptr` enforces an equality constraint to preserve the trace, but is substantially slower than `nlminb`. The default optimizer used for `nloptr` is the Augmented Lagrangian method with local optimization using LBFGS. These can be modified using the control parameter. `Rcgmin` does not enforce equality constraints, but may be more efficient for certain higher dimensional problems. The ideal optimization routine depends on the underlying structure of the population eigenvalues.

References

- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. Journal of Multivariate Analysis, 139(2)
- Ledoit, O. and Wolf, M. (2016). Numerical Implementation of the QuEST function. arXiv:1601.05870 [stat.CO]

`tau_estimate` *Non-linear shrinkage estimator of population eigenvalues.*

Description

The population eigenvalue estimates are computed by numerically inverting the QuEST function (see references). The starting point is the linear shrinkage estimate of the population eigenvalues, computed using `linshrink`.

Usage

```
tau_estimate(X, k = 0, method = "nlminb", control = list())
```

Arguments

X	A data matrix.
k	(Optional) Non-negative integer less than ncol(X). If k == 0 (default), X is assumed to contain 1 class, which will be centered. If k >= 1, X is assumed to contain k classes, each of which has already been centered.
method	(Optional) The optimization routine used. Choices are nlminb (default) and nloptr.
control	(Optional) A list of control parameters. Must correspond to the selected optimization method. See nlminb, nloptr for details.

Value

A numeric vector of length ncol(X), containing the population eigenvalue estimates, sorted in ascending order.

NOTE

nlminb is usually robust and accurate, but does not allow equality constraints, so, in general, the sum of the estimated population eigenvalues is not equal to the sum of the sample eigenvalues. nloptr enforces an equality constraint to preserve the trace, but is substantially slower than nlminb. The default optimizer used for nloptr is the Augmented Lagrangian method with local optimization using LBFGS. These can be modified using the control parameter.

References

- Ledoit, O. and Wolf, M. (2015). Spectrum estimation: a unified framework for covariance matrix estimation and PCA in large dimensions. Journal of Multivariate Analysis, 139(2)
- Ledoit, O. and Wolf, M. (2016). Numerical Implementation of the QuEST function. arXiv:1601.05870 [stat.CO]

Examples

```
tau_estimate(X = matrix(rnorm(1e3, mean = 5), nrow = 50, ncol = 20))
```

Index