

Package ‘RMT4DS’

November 4, 2022

Title Computation of Random Matrix Models

Version 0.0.1

Authors Xiucan Ding <xiucaiding89@gmail.com>, Yichen Hu <ethhu@ucdavis.edu>

Maintainer Xiucan Ding <xiucaiding89@gmail.com>

Description Functions working with general Marchenko-Pastur distribution and Tracy-Widom, limits and distributions of eigenvalues and generalized components of spiked covariance matrices, estimation of all population eigenvalues of spiked covariance matrix model, test of population covariance matrix, matrix denoising for signal-plus-noise model.

License ‘use_mit_license()’, ‘use_gpl3_license()’ or friends to pick a license

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.1

Repository CRAN, Github

Collate 'Limits.R'
'CovEst.R'
'CovTest.R'
'GeneralMP.R'
'GeneralWishartMax.R'
'SignalPlusNoise.R'

R topics documented:

CovEst	2
CovTest	3
GeneralMPLaw	4
GeneralWM	5
Limits	7
SignalNoise	8

Index	10
--------------	-----------

Description

Estimation of the eigenvalues of population covariance matrix given samples.

Usage

```
MPEst(X, n=nrow(X), k=1, num=NULL, penalty=FALSE, n_spike=0)
MomentEst(X, n=nrow(X), k=1, n_spike=0)
```

Arguments

X	n by p data matrix.
n	sample size.
k	repeated times in estimation. If k>1, estimation will be the average.
num	numbers of mass points chosen in estimation.
penalty	whether to implement L-1 penalty in inverting Marchenko-Pastur law
n_spike	number of spikes in population spectral.

Details

Given $E(X) = 0$ and $Cov(X) = \Sigma$ with Σ unknown and fourth moment of X exists, we want to estimate spectrum of Σ from sample covariance matrix $X'X/n$.

MPEst estimates spectrum by inverting Marchenko-Pastur law while MomentEst estimates spectrum by estimating the moment of population spectral density.

Those two functions give estimates of the eigenvalues by d and estimates of spectral density by xs and cdf.

Value

MPEst and MomentEst give estimation of the spectrum of population covariance matrix and corresponding spectral density.

Author(s)

Xiucui Ding, Yichen Hu

References

- [1] El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. The Annals of Statistics, 36(6), 2757-2790.
- [2] Kong, W., & Valiant, G. (2017). Spectrum estimation from samples. The Annals of Statistics, 45(5), 2218-2247.

Examples

```
library(MASS)
n = 500
p = 250
X = mvnrm(n, rep(0,p), diag(c(rep(2,p/2),rep(1,p/2))))
MPEst(X, n)$d
MomentEst(X, n)$d
```

CovTest

*High-dimensional Covariance Test***Description**

Test of given population covariance matrix, test of equal covariance of two or more samples.

Usage

```
OneSampleCovTest(X, mean=NULL, S=NULL)
TwoSampleCovTest(X1, X2, mean=NULL)
MultiSampleCovTest(..., input=NULL)
```

Arguments

X, X1, X2	input samples in the form n by p where p is the dimension.
mean	population mean of samples. If it is missing, sample mean will be used.
S	covariance matrix to be tested. If it is missing, test of identity covariance will be performed.
...	any samples to be tested.
input	list of samples to be tested. Please choose either ... or input as input form.

Value

OneSampleCovTest tests given covariance matrix of one sample,
 TwoSampleCovTest tests equal covariance matrices of two samples,
 MultiSampleCovTest tests equal covariance matrices of multiple samples.

Author(s)

Xiucui Ding, Yichen Hu

Source

Maximal likelihood tests fail in high-dimensional settings, so corrections are made. Note all tests are one-sided. Large statistics indicate violation of null hypothesis.

References

[1] Zheng, S., Bai, Z., & Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics*, 43(2), 546-591.

Examples

```
library(MASS)
n = 500
p = 100
S1 = diag(rep(1,p))
S2 = diag(sample(c(1,4),p,replace=TRUE))
OneSampleCovTest(mvrnorm(n,rep(0,p),S2), S=S1)
TwoSampleCovTest(mvrnorm(n,rep(0,p),S1), mvrnorm(n,rep(0,p),S2))
MultiSampleCovTest(mvrnorm(n,rep(0,p),S1), mvrnorm(n,rep(0,p),S2))
```

GeneralMPLaw

General Marchenko-Pastur Distribution

Description

Density, distribution function, quantile function and random generation for the general Marchenko-Pastur distribution, the limiting distribution of empirical spectral measure for large Wishart matrices.

Usage

```
qgmp(p, ndf=NULL, pdim=NULL, svr=ndf/pdim, eigens=NULL, lower.tail=TRUE, log.p=FALSE, m=500)
rgmp(n, ndf=NULL, pdim=NULL, svr=ndf/pdim, eigens=NULL, m=500)
pgmp(q, ndf=NULL, pdim=NULL, svr=ndf/pdim, eigens=NULL, lower.tail=TRUE,
     log.p=FALSE, m=500)
dgmp(x, ndf=NULL, pdim=NULL, svr=ndf/pdim, eigens=NULL, log.p=FALSE, m=500)
```

Arguments

<code>x, q</code>	vector of quantiles.
<code>p</code>	vector of probabilities.
<code>m</code>	number of points used in estimating density.
<code>ndf</code>	the number of degrees of freedom for the Wishart matrix.
<code>pdim</code>	the number of dimensions (variables) for the Wishart matrix.
<code>svr</code>	samples to variables ratio; the number of degrees of freedom per dimension.
<code>log, log.p</code>	logical; if TRUE, probabilities <code>p</code> are given as $\log(p)$.
<code>lower.tail</code>	logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.
<code>eigens</code>	input eigenvalues of population covariance matrix.

Details

Those functions work only for non-spiked part.

To achieve high accuracy of estimation, `eigens` should be large, like larger than 500.

In general Marchenko Pastur distributions, the support of density is the union of one or more intervals.

Value

dgmp gives the density,
 pgmp gives the distribution function,
 qgmp gives the quantile function,
 rgmp generates random deviates,

Author(s)

Xiucui Ding, Yichen Hu

Source

If eigens is missing, functions from package RMTstat will be used to compute classical Marchenko-Pastur distribution.

References

- [1] Knowles, A., & Yin, J. (2017). Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, 169(1), 257-352.
- [2] Bai, Z., & Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106, 167-177.
- [3] Ding, X. (2021). Spiked sample covariance matrices with possibly multiple bulk components. *Random Matrices: Theory and Applications*, 10(01), 2150014.
- [4] Ding, X., & Trogdon, T. (2021). A Riemann–Hilbert approach to the perturbation theory for orthogonal polynomials: Applications to numerical linear algebra and random matrix theory. *arXiv preprint arXiv:2112.12354*.

Examples

```
N = 1000
M = 300
d = c(rep(3.8,M/3),rep(1.25,M/3),rep(0.25,M/3))
qgmp(0.5, ndf=N, pdim=M, eigens=d)
pgmp(3, ndf=N, pdim=M, eigens=d)
dgmp(2, ndf=N, pdim=M, eigens=d)
rgmp(2, ndf=N, pdim=M, eigens=d)
```

Description

Density, distribution function, quantile function and random generation for the maximum eigenvalue from a general non-spiked Wishart matrix (sample covariance matrix) with ndf degrees of freedom, pdim dimensions, and order parameter beta.

Usage

```
dWishartMax(x, eigens, ndf, pdim, beta, log = FALSE)
pWishartMax(q, eigens, ndf, pdim, beta, lower.tail = TRUE, log.p = FALSE)
qWishartMax(p, eigens, ndf, pdim, beta, lower.tail = TRUE, log.p = FALSE)
rWishartMax(n, eigens, ndf, pdim, beta)
```

Arguments

<code>x, q</code>	vector of quantiles.
<code>p</code>	vector of probabilities.
<code>n</code>	number of observations.
<code>eigens</code>	eigenvalues of population covariance matrix.
<code>ndf</code>	the number of degrees of freedom for the Wishart matrix
<code>pdim</code>	the number of dimensions (variables) for the Wishart matrix
<code>beta</code>	the order parameter. 1 for real Wishart and 2 for complex Wishart.
<code>log, log.p</code>	logical; if TRUE, probabilities p are given as log(p).
<code>lower.tail</code>	logical; if TRUE (default), probabilities are $P[X \leq x]$, otherwise, $P[X > x]$.

Details

A real Wishart matrix is equal in distribution to $X^T X/n$, where X are $n \times p$ real matrix with elements of mean zero and covariance matrix Σ . A complex Wishart matrix is equal in distribution to $X^* X/n$, where both real and imaginary part of X are $n \times p$ complex matrices with elements of mean zero and covariance matrix $\Sigma/2$. `eigens` are eigenvalues of Σ . These functions give the limiting distribution of the largest eigenvalue from the such a matrix when `ndf` and `pdim` both tend to infinity.

Value

`dWishartMax` gives the density,
`pWishartMax` gives the distribution function,
`qWishartMax` gives the quantile function,
`rWishartMax` generates random deviates.

Author(s)

Xiucan Ding, Yichen Hu

References

- [1] El Karoui, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 35(2), 663-714.
- [2] Lee, J. O., & Schnelli, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *The Annals of Applied Probability*, 26(6), 3786-3839.

Examples

```

n = 500
p = 100
eigens = c(rep(2,p/2), rep(1, p/2))
beta = 2
rWishartMax(5, eigens, n, p, beta=beta)
qWishartMax(0.5, eigens, n, p, beta=beta)
pWishartMax(3.5, eigens, n, p, beta=beta)
dWishartMax(3.5, eigens, n, p, beta=beta)

```

Limits

*Limits in High-dimensional Sample Covariance***Description**

Some limits of eigenvalues and eigenvectors in high-dimensional sample covariance.

Usage

```

MP_vector_dist(k, v, ndf=NULL, pdim, svr=ndf/pdim, cov=NULL)
cov_spike(spikes, eigens, ndf, svr)
quadratic(k, cov, svr, spikes, type=1)

```

Arguments

k	k-th eigenvector. In MP_vector_dist, k can be a serie.
v	vector to be projected on.
ndf	the number of degrees of freedom for the Wishart matrix.
pdim	the number of dimensions (variables) for the Wishart matrix.
svr	samples to variables ratio; the number of degrees of freedom per dimension.
cov	population covariace matrix. If it is null, it will be regarded as identity.
eigens	input eigenvalues of population covariance matrix without spikes.
spikes	spikes in population covariance matrix.
type	transformation of eigenvalues. n for n-th power. 0 for logarithm.

Details

In MP_vector_dist, the variance computed is for $\sqrt{\text{pdim}} u_k^T v$, where u_k is the k-th eigenvector.
 Note in quadratic, k should be within the spikes.

Value

MP_vector_dist gives asymptotic variance of projection of eigenvectors of non-spiked Wishart matrix,
 cov_spike gives spikes in sample covariance matrix and their asymptotic variance.
 quadratic gives mean of certain quadratic forms of k-th sample eigenvector of spiked models.
 Note k should be within the spikes.

Author(s)

Xiucan Ding, Yichen Hu

References

- [1] Knowles, A., & Yin, J. (2017). Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, 169(1), 257-352.
- [2] Jolliffe, I. (2005). Principal component analysis. *Encyclopedia of statistics in behavioral science*.

Examples

```
k = 1
n = 200
p = 100
v = runif(p)
v = v/sqrt(sum(v^2))
MP_vector_dist(k,v,n,p,cov=diag(p))
cov_spike(c(10),rep(1,p),n,n/p)
quadratic(k,diag(p),n/p,c(30))
```

SignalNoise

Signal-Plus-Noise Models

Description

Estimation of signals, rank of signals.

Usage

```
StepWiseSVD(Y, threshold=NULL, B=1000, level=0.02, methods='kmeans',
  u_threshold=NULL, v_threshold=NULL, sparse=TRUE)
ScreeNot(Y, r1)
GetRank(Y, r1, type=c("1","2"), level=0.1, B=500)
signal_value(d, svr)
signal_vector(k1, k2, d1, d2, svr, left=TRUE)
```

Arguments

Y	matrix to be denoised.
B	repeat time of simulations.
threshold	threshold used in determining rank of signal.
level	significance level in determining ranks.
method	methods used in determining sparse structure.
u_threshold, v_threshold	thresholds used in determining sparse structure if kmeans is not used.
sparse	whether signals have sparse structure.
r1	upper bound of rank.
type	type of test.

k1, k2	k-th eigenvector.
d1, d2	eigenvalues of corresponding signal matrix
left	whether to use left singular vectors.
svr	ndf/ndim of Y.

Details

StepWiseSVD works well in sparse setting and requires i.i.d normal noise and a lot simulation time. ScreeNot is to pick the best TSVD result so works well in general setting.

Value

StepWiseSVD performs step-wise SVD to denoise and returns decomposed strcuture,
 ScreeNot performs ScreeNot to denoise and returns decomposed strcuture,
 GetRank gives rank of signals.
 signal_value gives corrected signal eigenvalue from SVD result,
 signal_vector gives limiting inner product between signal vector and corresponding signal-plus-noise vector.

Author(s)

Xiucan Ding, Yichen Hu

References

- [1] Ding, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli*, 26(1), 387-417.
- [2] Donoho, D. L., Gavish, M., & Romanov, E. (2020). Screenot: Exact mse-optimal singular value thresholding in correlated noise. *arXiv preprint arXiv:2009.12297*.
- [3] Ding, X., & Yang, F. (2022). Tracy-Widom distribution for heterogeneous Gram matrices with applications in signal detection. *IEEE Transactions on Information Theory*, vol. 68, no. 10, pp. 6682-6715.

Index

- * **distribution**
 - GeneralMPLaw, [4](#)
 - GeneralWM, [5](#)
- * **estimation**
 - CovEst, [2](#)
- * **limit, distribution**
 - Limits, [7](#)
- * **limits, estimation**
 - SignalNoise, [8](#)
- * **test**
 - CovTest, [3](#)

cov_spike (Limits), [7](#)
CovarianceTest (CovTest), [3](#)
CovEst, [2](#)
CovTest, [3](#)

dgmp (GeneralMPLaw), [4](#)
dWishartMax (GeneralWM), [5](#)

GeneralMPLaw, [4](#)
GeneralWM, [5](#)
GetRank (SignalNoise), [8](#)

Limits, [7](#)

MomentEst (CovEst), [2](#)
MP_vector_dist (Limits), [7](#)
MPEst (CovEst), [2](#)
MultiSampleCovTest (CovTest), [3](#)

OneSampleCovTest (CovTest), [3](#)

pgmp (GeneralMPLaw), [4](#)
pWishartMax (GeneralWM), [5](#)

qgmp (GeneralMPLaw), [4](#)
quadratic (Limits), [7](#)
qWishartMax (GeneralWM), [5](#)

rgmp (GeneralMPLaw), [4](#)
rWishartMax (GeneralWM), [5](#)

ScreeNot (SignalNoise), [8](#)
signal_values (SignalNoise), [8](#)
signal_vector (SignalNoise), [8](#)
SignalNoise, [8](#)
StepWiseSVD (SignalNoise), [8](#)
TwoSampleCovTest (CovTest), [3](#)