Estimate.ED.R

Marco Del Giudice (2020). Version 1. Contact: marcodg@unm.edu

A simple R function to compute estimates of effective dimensionality (ED), either from raw data or from a correlation/covariance matrix. The function returns four estimators, both uncorrected and corrected for small-sample bias if enough information is available. Disattenuated estimates can also be computed if desired, by supplying a vector of reliability coefficients. For more information see Del Giudice (2020).

ED estimators: The n_1 index is suitable as a general-purpose estimator of ED (Cangelosi & Goriely 2007, Roy & Vetterli 2007, Gnedenko & Yelnik 2016); n_2 is more conservative and can be used as a lower bound estimate of dimensionality (Fraedrich et al. 1995, Bretherton et al. 1999, Pirkl et al. 2012). The estimators n_C (Cheverud 2001, Wagner et al. 2008) and n_∞ (Kirkpatrick et al. 2009) are not recommended, and are only reported for comparison with the published literature. Whereas n_C typically overestimates ED, n_∞ is extremely conservative and insensitive to changes in the correlation structure.

Small-sample correction: if raw data are supplied, the function uses the `nlshrink` package (Ramprasad 2016) to perform nonlinear shrinkage of the eigenvalues (Ledoit & Wolf 2012, 2015). If only the correlation or covariance matrix is supplied, the function applies the correction method by Mestre (2008). If the corrected ED is lower than the uncorrected ED (which may happen when the uncorrected ED is very close to the number of variables), the uncorrected value is returned in place of the corrected one. IMPORTANT NOTE: implementing Mestre's method is tricky; sometimes, the `uniroot.all` function may fail to find the correct mu values. When applying the small-sample correction to a correlation/covariance matrix, please check the results for unexpected or implausible values.

Disattenuation: to correct for measurement error, disattenuated estimates of ED can be obtained by supplying a vector of reliability coefficients (e.g., Cronbach's α, McDonald's ω_t or ω_h). For details see Del Giudice (2020).

Note: If the matrix supplied by the user is indefinite (or becomes indefinite after disattenuation), Higham's (2002) method is automatically applied to replace it with the nearest positive definite matrix (`Matrix` package by Bates & Maechler, 2019). A warning message is displayed in the output summary.

Bates, D., & Maechler, M. (2019). Matrix v. 1.2-17. URL: https://CRAN.R-project.org/package=Matrix

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., & Bladé, I. (1999). The effective number of spatial degrees of freedom of a time-varying field. Journal of Climate, 12, 1990-2009.

Cangelosi, R., & Goriely, A. (2007). Component retention in principal component analysis with application to cDNA microarray data. Biology Direct, 2, 2.

Cheverud, J. M. (2001). A simple correction for multiple comparisons in interval mapping genome scans. Heredity, 87, 52-58.

Del Giudice, M. (2020). Effective dimensionality: A tutorial. Multivariate Behavioral Research. https://doi.org/10.1080/00273171.2020.1743631

Fraedrich, K., Ziehmann, C., & Sielmann, F. (1995). Estimates of spatial degrees of freedom. Journal of Climate, 8, 361-369.

Gnedenko, B., & Yelnik, I. (2016). Minimum entropy as a measure of effective dimensionality. SSRN, doi: 10.2139/ssrn.2767549.

Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. IMA journal of Numerical Analysis, 22, 329-343.

Kirkpatrick, M. (2009). Patterns of quantitative genetic variation in multiple dimensions. Genetica, 136, 271-284.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. Annals of Statistics, 40, 1024-1060.

Ledoit, O., & Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. Journal of Multivariate Analysis, 139, 360-384.

Pirkl, R. J., Remley, K. A., & Patané, C. S. L. (2012). Reverberation chamber measurement correlation. IEEE Transactions on Electromagnetic Compatibility, 54, 533-545.

Ramprasad, P. (2016). nlshrink v. 1.0.1. URL: https://CRAN.R-project.org/package=nlshrink

Roy, O., & Vetterli, M. (2007). The effective rank: A measure of effective dimensionality. In M. Domański, R. Stasiński, & M. Bartkowiak (Eds.), 15th European Signal Processing Conference (pp. 606-610). Poznań, Poland: PTETiS.

Function `estimate.ED`

```
estimate.ED (x, sample.size = NULL, rel.values = NULL, cov.mat = FALSE,
small.sample.c = FALSE, round.digits = 2, print.summary = TRUE)
```

Arguments

`x`	data frame or correlation/covariance matrix
`sample.size`	sample size (optional: required for small-sample bias correction from a correlation/covariance matrix)
`rel.values`	vector of reliability coefficients (optional: only for disattenuation)
`cov.mat`	if TRUE, ED estimators are computed from the covariance matrix; if FALSE (default), from the correlation matrix
`small.sample.c`	if TRUE (default), the function returns estimators corrected for small-sample bias
`round.digits`	rounding digits for the output (default is 2)
`print.summary`	if TRUE (default), the function prints a summary of the analysis before returning the results

Value

returns a list object containing some or all of the following:

`n1`	ED estimator based on the Shannon entropy H_1
`n2`	ED estimator based on the quadratic entropy H_2
`nInf`	ED estimator based on the min-entropy H_∞; extremely conservative, not recommended
`nC`	ED estimator by Cheverud (2001); not recommended
`n1.c`	n_1 corrected for small-sample bias
`n2.c`	n_2 corrected for small-sample bias
`nInf.c`	n_∞ corrected for small-sample bias
`nC.c`	n_C corrected for small-sample bias