Package: vandalico (via r-universe)

September 7, 2024

Type Package

Title Evaluation of Presence-Absence Models

Version 0.1.0

Date 2024-09-06

Maintainer Alberto Jiménez-Valverde

<alberto.jimenez.valverde@gmail.com>

Description Collection of functions to evaluate presence-absence
models. It comprises functions to adjust discrimination
statistics for the representativeness effect through
case-weighting, along with functions for visualizing the
outcomes. Originally outlined in: Jiménez-Valverde (2022) The
uniform AUC: dealing with the representativeness effect in
presence-absence models. Methods Ecol. Evol, 13, 1224-1236.

License GPL-3

Imports ROCR (>= 1.0-7), stats, graphics

Depends R (>= 3.5.0)

RoxygenNote 7.1.0

Encoding UTF-8

NeedsCompilation no

Author Alberto Jiménez-Valverde [aut, cre]
(<<https://orcid.org/0000-0001-9962-2106>>)

Date/Publication 2024-09-06 16:20:02 UTC

Repository https://ajimenezvalverde.r-universe.dev

RemoteUrl https://github.com/cran/vandalico

RemoteRef HEAD

RemoteSha 1c054c5e15c5f3291ecfa64e18a1a489b76d3751

Contents

AUCuniform   *Calculates the uniform* AUC *and uniform* Se*

Description

This function computes the uniform *AUC* (*uAUC*) and uniform *Se** (*uSe**) following Jiménez-Valverde (2022).

Usage

```
AUCuniform(
  mat,
  rep = 100,
  by = 0.1,
  deleteBins = NULL,
  plot = FALSE,
  plot.adds = FALSE
)
```

Arguments

mat	A matrix with two columns. The first column must contain the suitability values (i.e., the classification rule); the second column must contain the presences and absences.
rep	Number of sampling replications. By default, rep = 100.
by	Size of the suitability intervals (i.e., bins). By default, by = 0.1.
deleteBins	A vector (e.g., from 1 to 10 if by = 0.1) with the bins that have to be excluded (1 for [0,0.1), 10 for [0.9,1]) from the resampling procedure (trimming); NULL by default.
plot	Logical. Indicates whether or not the observed ROC curve is plotted.
plot.adds	Logical. Indicates whether or not the negative diagonal and the point of equivalence are added to the observed ROC plot.

Details

This function performs the stratified weighted bootstrap to calculate the uniform *AUC* (*uAUC*) and uniform *Se** (*uSe**) as suggested in Jiménez-Valverde (2022). A warning message will be shown if the sample size of any bin is zero. Another warning message will be shown if the sample size of any bin is lower than 15. In such case, trimming should be considered. The *AUC* (non-uniform) is estimated non-parametrically (Bamber 1975). *Se** is calculated by selecting the point that minimizes the absolute difference between sensitivity and specificity and by doing the mean of those values (Jiménez-Valverde 2020).

Value

A list with the following elements:

AUC: the *AUC* value (non-uniform), a numeric value between 0 and 1.

Se: the *Se** value (non-uniform), a numeric value between 0 and 1.

bins: a table with the sample size of each bin.

suit.sim: a matrix with the bootstrapped suitability values.

sp.sim: a matrix with the bootstrapped presence-absence data.

uAUC: a numeric vector with the (*uAUC*) values for each replication.

uAUC.95CI: a numeric vector with the sample (*uAUC*) quantiles corresponding to the probabilities 0.025, 0.5 and 0.975.

uSe: a numeric vector with the (*uSe**) values for each replication.

uSe.95CI: a numeric vector with the sample (*uSe**) quantiles corresponding to the probabilities 0.025, 0.5 and 0.975.

References

Bamber, D. (1975). The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph. *J. Math. Psychol.*, 12, 387-415.

Jiménez-Valverde, A. (2020). Sample size for the evaluation of presence-absence models. *Ecol. Indic.*, 114, 106289.

Jiménez-Valverde, A. (2022). The uniform AUC: dealing with the representativeness effect in presence-absence models. *Methods Ecol. Evol.*, 13, 1224-1236.

Examples

```
suit<-rbeta(100, 2, 2) #Generate suitability values
random<-runif(100)
sp<-ifelse(random < suit, 1, 0) #Generate presence-absence data
result<-AUCuniform(cbind(suit, sp), plot = TRUE, plot.adds = TRUE)
result$uAUC.95CI[2] #Get the uAUC
```

AUCuniform_trap	*Calculates the uniform* AUC *and uniform* Se* *by the trapezoidal method.*

Description

This function computes the uniform *AUC* (*uAUC*) and uniform *Se** (*uSe**) using the weighted trapezoidal method instead of the weighted bootstrapping method used in `AUCuniform` and originally proposed in Jiménez-Valverde (2022). This procedure reduces bias and improves the coverage of confidence intervals (Jiménez-Valverde 2024). Additionally, the weights vector associated to each case can be customized. See Jiménez-Valverde (2024) for details.

Usage

```
AUCuniform_trap(
  mat,
  by = 0.1,
  deleteBins = NULL,
  w = NULL,
  plot = FALSE,
  plot.compare = FALSE,
  plot.adds = FALSE
)
```

Arguments

mat	A matrix with two columns. The first column must contain the suitability values (i.e., the classification rule); the second column must contain the presences and absences.
by	Size of the suitability intervals (i.e., bins). By default, by = 0.1.
deleteBins	A vector (e.g., from 1 to 10 if by = 0.1) with the bins that have to be excluded (1 for [0,0.1), 10 for [0.9,1]) from the calculation of the uniform statistics; NULL by default.
w	A vector with the weights associated to each case. If NULL (default), then the uniform *AUC* (*uAUC*) and uniform *Se** (*uSe**) are calculated.
plot	Logical. Indicates whether or not the observed ROC curve is plotted (gray dots).
plot.compare	Logical. Indicates whether or not the weighed ROC curve is plotted (black line).
plot.adds	Logical. Indicates whether or not the negative diagonal and the points of equivalence (weighted and unweighted) are added to the ROC plot.

Details

This function calculates the uniform *AUC* (*uAUC*) and uniform *Se** (*uSe**) using the weighted trapezoidal method as suggested in Jiménez-Valverde (2024). A warning message will be shown if the sample size of any bin is zero. Another warning message will be shown if the sample size of any bin

is lower than 15. In such case, trimming should be considered using deleteBins (Jiménez-Valverde 2022). Alternatively, the weights associated to each case can be fully customized with the w parameter (Jiménez-Valverde 2024). In this case, no warnings regarding sample size issues are raised, and deleteBins is not used. The *AUC* (non-uniform, unweighted) is estimated non-parametrically by the trapezoidal rule, which is equivalent to the Wilcoxon-based estimation (Hanley & McNeil 1982) used in AUCuniform. *Se** is calculated as in AUCuniform.

Value

A list with the following elements:

AUC: the *AUC* value (non-uniform, unweighted), a numeric value between 0 and 1.

Se: the *Se** value (non-uniform, unweighted), a numeric value between 0 and 1.

bins: a table with the sample size of each bin (only if w = NULL).

uAUC: the uniform *AUC* value (only if w = NULL).

uSe: the uniform *Se** value (only if w = NULL).

wAUC: the weighted *AUC* estimated with the vector w

wSe: the weighted *Se** estimated with the vector w

References

Hanley, J. A. & McNeil, B. J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*., 143, 29-36.

Jiménez-Valverde, A. (2022). The uniform AUC: dealing with the representativeness effect in presence-absence models. *Methods Ecol. Evol.*, 13, 1224-1236.

Jiménez-Valverde, A. (2024). Improving the uniform AUC (uAUC): towards a case-by-case weighting evaluation in species distribution models. *In preparation*.

Examples

```
suit<-rbeta(100, 2, 2) #Generate suitability values
random<-runif(100)
sp<-ifelse(random < suit, 1, 0) #Generate presence-absence data
result<-AUCuniform_trap(cbind(suit, sp), plot = TRUE, plot.compare = TRUE)
result$AUC #Get the AUC
result$uAUC #Get the uAUC. Note how it is closer to the reference value of
            #0.83 since the suitability values are simulated to be
            #well-calibrated (see Jimenez-Valverde 2022).
```

CALplot   *Calibration graph*

Description

A function to plot a calibration graph.

Usage

```
CALplot(mat, by = 0.1)
```

Arguments

mat	A matrix with two columns. The first column must contain the suitability values (i.e., the classification rule); the second column must contain the presences and absences.
by	Size of the suitability intervals (bins). By default, by = 0.1.

Details

Dots for bins with 15 or more cases are shown in solid black; dots for bins with less than 15 cases are shown empty (see Jiménez-Valverde et al. 2013). This way, by plotting the calibration graph before running AUCuniform, one can get a glimpse of how reliable *uAUC* or *uSe** can be expected to be.

Value

This function returns a calibration plot

References

Jiménez-Valverde, A., Acevedo, P., Barbosa, A. M., Lobo, J. M. & Real, R. (2013). Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecol. Biogeogr.*, 22, 508-516.

Examples

```
suit<-rbeta(100, 2, 2) #Generate suitability values
random<-runif(100)
sp<-ifelse(random < suit,1 , 0) #Generate presence-absence data
CALplot(cbind(suit, sp))
```

HSgraph   *Suitability values distribution graph*

Description

A function to visualize the distribution of the suitability values associated to presences, absences, and all cases together.

Usage

```
HSgraph(mat, breaks = 10, hist.total = TRUE)
```

Arguments

mat	A matrix with two columns. The first column must contain the suitability values (i.e., the classification rule); the second column must contain the presences and absences.
breaks	Number of cells for the total histogram. By default, breaks = 10.
hist.total	Logical. Indicates whether or not the distribution of suitability values for all the cases together is graphed.

Details

In blue, the distribution of the suitability values associated to presences. In red, the distribution of the suitability values associated to absences. This graph helps to understand why the *AUC* (or *Se**) is greater, equal to, or less than the *uAUC* (or *uSe**) (see Jiménez-Valverde 2022).

Value

This function returns a multiple histogram.

References

Jiménez-Valverde, A. (2022). The uniform AUC: dealing with the representativeness effect in presence-absence models. *Methods Ecol. Evol.*, 13, 1224-1236.

Examples

```
suit<-rbeta(100, 2, 2) #Generate suitability values
random<-runif(100)
sp<-ifelse(random < suit, 1 , 0) #Generate presence-absence data
HSgraph(cbind(suit, sp), breaks = 20, hist.total = TRUE)
```

Index