Package ‘maxlike’

August 18, 2025

Type Package

Title Model Species Distributions by Estimating the Probability of Occurrence Using Presence-Only Data

Version 0.1-12

Date 2025-7-29

Depends R (>= 2.10.0), raster

Suggests dismo

Author Richard Chandler [aut],
Andy Royle [aut],
Roeland Kindt [cre, aut] (ORCID:
<`https://orcid.org/0000-0002-7672-0712`>)

Maintainer Roeland Kindt <`RoelandCEKindt@gmail.com`>

Description Provides a likelihood-based approach to modeling species distributions using presence-only data. In contrast to the popular software program MAXENT, this approach yields estimates of the probability of occurrence, which is a natural descriptor of a species' distribution.

License GPL (>= 3)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2025-08-18 10:10:02 UTC

Contents

maxlike-package 2
carw 2
MaungaWhau 3
maxlike 4

Index **9**

`maxlike-package` *Model species distributions by estimating the probability of occurrence using presence-only data.*

Description

A species' distribution can be characterized by the probability that it occurs at some location in space. Estimating occurrence probability can be easily accomplished using presence-absence data, but often researchers only have presence locations and environmental data for the study area. MAXENT is a popular software program for modeling species distributions, but it does not estimate the probability of occurrence. Rather, it returns various indices that are not easy to interpret (see Royle et al. 2012). Package "maxlike" provides a simple likelihood-based alternative.

Note

All presence-only models require a random sample of data from locations where the species is present. Unfortunately, random sampling is not a feature of most presence-only datasets, and bias in the estimated probability surface should be expected in such cases. This assumption can be greatly relaxed if one has presence-absence data, which will always contain more information about a species' distribution.

References

J. A. Royle, R. B. Chandler, C. Yackulic and J. D. Nichols. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods in Ecology and Evolution 3:545–554. doi: 10.1111/j.2041-210X.2011.00182.x.

See Also

`maxlike`, `raster`

`carw` *The Carolina Wren data used by Royle et al. (2012)*

Description

These data come from the North American Breeding Bird Survey. They include presence-only and presence-absence data for all 25 square-kilometer covering the contiguous United States. See Royle et al. (2012) for details.

Usage

```
data(carw)
```

Format

The format is: List of 5 $ raster.data:'data.frame': 31980 obs. of 6 variables:

$ pa.data :'data.frame': 4615 obs. of 9 variables:

$ xy1 :'data.frame': 12082 obs. of 2 variables:

$ ext : num [1:4] -3043880 3106120 114746 3364746

$ dim : num [1:2] 130 246

Details

The component `raster.data` contains the spatially-referenced covariate data that can be coverted to a `raster` object. `pa.data` is a data.frame of presence-absence data. `xy1` is a data.frame of coordinates of routes where Carolina Wrens were detected. The other components are attibutes necessary for formatting `raster.data` as a `raster` object.

References

Royle, J.A., R.B. Chandler, C. Yackulic and J. D. Nichols. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods in Ecology and Evolution. doi: 10.1111/j.2041-210X.2011.00182.x

Examples

```
data(carw)

# Convert data frame to a list of rasters
rl <- lapply(carw.data$raster.data, function(x) {
   m <- matrix(x, nrow=carw.data$dim[1], ncol=carw.data$dim[2],
               byrow=TRUE)
   r <- raster(m)
   extent(r) <- carw.data$ext
   r
})

# Stack and name them
rs <- stack(rl[[1]], rl[[2]], rl[[3]], rl[[4]], rl[[5]], rl[[6]])
names(rs) <- names(carw.data$raster.data)

plot(rs)
```

`MaungaWhau` *Fake data for the Maunga Whau volcano*

Description

Based upon volcano in the datasets package.

Usage

```
data(MaungaWhau)
```

Format

The format is: List of 3 $ elev : num [1:87, 1:61] -1.17 -1.13 -1.09 -1.05 -1.01 ...

$ precip: num [1:87, 1:61] -2.27 -2.25 -2.24 -2.22 -2.2 ...

$ xy : num [1:1000, 1:2] 38.5 42.5 30.5 25.5 16.5 36.5 22.5 56.5 49.5 18.5 ...

Details

No coordinate system attributes are included

Examples

```
data(MaungaWhau)
elev <- raster(MaungaWhau$elev, xmn=0, xmx=61, ymn=0, ymx=87)
precip <- raster(MaungaWhau$precip, xmn=0, xmx=61, ymn=0, ymx=87)
```

`maxlike` *Model occurrence probability using presence-only data*

Description

This function estimates the probability of occurrence using presence-only data and spatially-referenced covariates. Species distribution maps can be created by plotting the expected values of occurrence probability. The model is described by Royle et al. (2012).

Usage

```
  maxlike(formula, rasters, points, x=NULL, z=NULL,
          link=c("logit", "cloglog"),
          starts, hessian = TRUE, fixed, removeDuplicates=FALSE,
          savedata=FALSE, na.action = "na.omit", ...)
```

Arguments

`formula`	A right-hand side `formula` describing the model. At least 1 continuous covariate must be present in the formula.
`rasters`	The spatially-referenced covariate data formatted as a ‘raster stack’ created by the `stack` function in the `raster-package`. It’s a good idea to standardize these by subtracting the mean and dividing by the standard deviation. This will make it easier for `optim` to find the maximum-likelihood estimates.
`points`	A `matrix` or `data.frame` with the X and Y coordinates of the presence locations.

`x`	A `matrix` or `data.frame` with the explanatory data for presence locations. In case data is provided for `x` and `z`, arguments `rasters` and `points` will be ignored
`z`	A `matrix` or `data.frame` with the explanatory data for background locations. In case data is provided for `x` and `z`, arguments `rasters` and `points` will be ignored
`link`	The link function. Either `"logit"` (the default) or `"cloglog"`.
`starts`	Starting values for the parameters. This should be a vector with as many elements as there are parameters. By default, all starting values are 0, which should be adequate if covariates are standardized.
`hessian`	Logical. Should the hessian be computed and the variance-covariance matrix returned?
`fixed`	Optional vector for fixing parameters. It must be of length equal to the number of parameters in the model. If an element of `fixed` is NA, then the parameter is estimated, otherwise if it is a real number, the parameter is fixed at this value.
`removeDuplicates`	Logical. Should duplicate points be removed? Defaults to FALSE, but note that the MAXENT default is TRUE.
`savedata`	Should the raster data be saved with the fitted model? Defaults to FALSE in order to reduce the size of the returned object. If you wish to make predictions, it is safer to set this to TRUE, otherwise the raster data are searched for in the working directory, and thus may not be the data used to fit the model.
`na.action`	See `options` for choices
`...`	Additional arguments passed to `optim`

Details

`points` and `rasters` should the same coordinate system. The program does not check this so it is up to the user.

Value

A list with 8 components

`Est`	data.frame containing the parameter estimates (Ests) and standard errors (SE).
`vcov`	variance-covariance matrix
`AIC`	AIC
`call`	the original call
`pts.removed`	The points removed due to missing values
`pix.removed`	The pixels removed due to missing values
`optim`	The object returned by `optim`
`not.fixed`	A logical vector indicating if a parameter was estimated or fixed.
`link`	The link function

Warnings

Maximizing the log-likelihood function is achieved using the `optim` function, which can fail to find the global optima if sensible starting values are not supplied. The default starting values are `rep(0, npars)`, which will often be adequate if the covariates have been standardized. Standardizing covariates is thus recommended. Even when covariates are standardized, it is always a good idea to try various starting values to see if the log-likelihood can be increased. When fitting models with many parameters, good starting values can be found by fitting simpler models first.

Note

In general it is very hard to obtain a random sample of presence points, which is a requirement of both the Royle et al. (2012) method and of MAXENT. This is one of many reasons why presence-absence data are preferable to presence-only data. When presence-absence data are available, they can be modeled using functions such as `glm`. Creating species distribution maps from `glm` is easily accomplished using the `predict` method.

The MAXENT software assumes that species prevalence is known a priori. If the user does not specify a value for prevalence, prevalence is set to 0.5. MAXENT predictions of occurrence probability are highly sensitive to this setting. In contrast, `maxlike` directly estimates prevalence.

Another weakness of models for presence-only data is that they do not allow one to model detection probability, which is typically less than one in field conditions. If detection probability is affected by the same covariates that affect occurrence probability, then bias is inevitable. The R package **unmarked** (Fiske and Chandler 2011) offers numerous methods for jointly modeling both occurrence and detection probability when detection/non-detection data are available.

References

Royle, J.A., R.B. Chandler, C. Yackulic and J. D. Nichols. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods in Ecology and Evolution. doi: 10.1111/j.2041-210X.2011.00182.x

Fiske, I. and R.B. Chandler. 2011. unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. Journal of Statistical Software 43(10).

See Also

`maxlike-package`, `raster`, `carw`

Examples

```
## Not run:

# Carolina Wren data used in Royle et. al (2012)
data(carw)

# Covert data.frame to a list of rasters
rl <- lapply(carw.data$raster.data, function(x) {
   m <- matrix(x, nrow=carw.data$dim[1], ncol=carw.data$dim[2], byrow=TRUE)
   r <- raster(m)
   extent(r) <- carw.data$ext
   r
```

```
})

# Create a raster stack and add layer names
rs <- stack(rl[[1]], rl[[2]], rl[[3]], rl[[4]], rl[[5]], rl[[6]])
names(rs) <- names(carw.data$raster.data)

plot(rs)


# Fit a model
fm <- maxlike(~pcMix + I(pcMix^2) + pcDec + I(pcDec^2)+ pcCon +
    I(pcCon^2) + pcGr + I(pcGr^2) +
    Lat + I(Lat^2) + Lon + I(Lon^2), rs, carw.data$xy1,
    method="BFGS", removeDuplicates=TRUE, savedata=TRUE)

summary(fm)
confint(fm)
AIC(fm)
logLik(fm)


# Produce species distribution map (ie, expected probability of occurrence)
psi.hat <- predict(fm) # Will warn if savedata=FALSE
plot(psi.hat)
points(carw.data$xy1, pch=16, cex=0.1)



# MAXENT sets "default prevalence" to an arbitrary value, 0.5.
# We could do something similar by fixing the intercept at logit(0.5)=0.
# However, it seems more appropriate to estimate this parameter.

# fm.fix <- update(fm, fixed=c(0, rep(NA,length(coef(fm))-1)))

# Predict data.frame
presenceData <- as.data.frame(extract(rs, carw.data$xy1))
presenceData <- presenceData[complete.cases(presenceData), ]
presence.predictions <- predict(fm, newdata=presenceData)
summary(presence.predictions)

# Calibrate with data.frames
PresenceUniqueCells <- unique(cellFromXY(rs, xy=carw.data$xy1))
PresenceUnique <- xyFromCell(rs, PresenceUniqueCells)
presenceData <- as.data.frame(extract(rs, PresenceUnique))
library(dismo)
background <- randomPoints(rs, n=ncell(rs), extf=1.00)
backgroundData <- as.data.frame(extract(rs, y=background))
backgroundData <- backgroundData[complete.cases(backgroundData), ]
fm2 <- maxlike(~pcMix + I(pcMix^2) + pcDec + I(pcDec^2)+ pcCon +
    I(pcCon^2) + pcGr + I(pcGr^2) +
    Lat + I(Lat^2) + Lon + I(Lon^2),
    rasters=NULL, points=NULL,
    x=presenceData, z=backgroundData,
```

```
    method="BFGS", removeDuplicates=TRUE, savedata=TRUE)

summary(fm2)

fm2$rasters <- rs
psi.hat2 <- predict(fm2)


# Simulation example

set.seed(131)
x1 <- sort(rnorm(100))
x1 <- raster(outer(x1, x1), xmn=0, xmx=100, ymn=0, ymx=100)

x2 <- raster(matrix(runif(1e4), 100, 100), 0, 100, 0, 100)

# Code factors as dummy variables.
# Note, using asFactor(x3) will not help
x3 <- raster(matrix(c(0,1), 100, 100), 0, 100, 0, 100)

logit.psi <- -1 + 1*x1 + 0*x2
psi <- exp(logit.psi)/(1+exp(logit.psi))
plot(psi)

r <- stack(x1, x2, x3)
names(r) <- c("x1", "x2", "x3")
plot(r)

pa <- matrix(NA, 100, 100)
pa[] <- rbinom(1e4, 1, as.matrix(psi))
str(pa)
table(pa)

pa <- raster(pa, 0, 100, 0, 100)
plot(pa)

xy <- xyFromCell(pa, sample(Which(pa==1, cells=TRUE), 1000))

plot(x1)
points(xy)

fm2 <- maxlike(~x1 + x2 + x3, r, xy)

summary(fm2)
confint(fm2)
AIC(fm2)
logLik(fm2)


## End(Not run)
```

Index

∗ **datasets**
  carw, 2
  MaungaWhau, 3
∗ **package**
  maxlike-package, 2

carw, 2, *6*

data.frame, *4, 5*

formula, *4*

glm, *6*

matrix, *4, 5*
MaungaWhau, 3
maxlike, *2*, 4
maxlike-package, 2

optim, *4–6*
options, *5*

predict, *6*

raster, *2, 3*, *6*

stack, *4*

volcano, *3*