R program: modselrLOS

Version: 0.0.0.1 (under developing)
Input packages: eva, lmomco, Rsolnp
Date: 2025-09-20
Authors: Yire Shin (shinyire87@gmail.com) and Jeong-soo Park (jsparkstat@gmail.com)
Description:

Two algorithms to select a good r-largest order statistics (rLOS) model with an appropriate r are implemented. The first one is the algorithm Survival. The second is the algorithm r_median. The first algorithm selects the best model first, then determines an appropriate r for the fixed model. The second algorithm determines an appropriate r, then selects the best model for the fixed r. The pool of models for rLOS are rK4D, rK3D, rGEV, rGLO, rLD, rGGD, and rGD.

Contents:

sel.rmod.surv (main program)
sel.rmod.rmed (main.program)
rk4d.fit.park
r3kd.fit.park
rglo.fit.park
rggd.fit.park
rgev.fit.park
multi.eEdtest.park
one_optr
find_optr
com.aic
Pohang (rainfall data)

References:

1. Bader B, Yan J, Zhang XB (2017) Automated selection of r for the r largest order statistics approach with adjustment for sequential testing. Statistics and Computing, 27(6): 1435-1451.
2. Busababodhin P, Seenoi P, Shin Y, and Park J-S (2025) Model selection for r-largest order statistics with hydrological applications. Submitted manuscript.

3. Shin Y, Park J-S (2023) Modeling climate extremes using the four-parameter kappa distribution for r-largest order statistics, Weath Clim Extrem 39 100533. Revision at arXiv.2007.12031

Main program-1: sel.rmod.surv

---

sel.rmod.surv = function(xdat, maxr=NULL, h.fix=NULL, mscrit="AIC", num_inits=20,
show=TRUE, sig.ed=0.05, qq=c(.98,.99,.995,.998), true.para= NULL)

Usage: The algorithm Survival selects the best model first, then determines an appropriate r for the fixed model.

Arguments:

xdat: Data of matrix with R number of columns for order statistics and n rows for independent observations.
maxr: R represents the maximum, predetermined number of the top order. Default is that R <- ncol(xdat).
h.fix: Use when the user have any prior set for the parameter h.
mscrit: Criterion for selcting a model based on the likelihood. Default is "AIC". There are options for "AICc" and "BIC".
num_inits: Number of trials in finding the MLE using numerical optimization routine.
show: Logical. If it is TRUE, it shows the progress of eliminating the models.
sig.ed: Significance level for the ED test. Default is 0.05.
qq: Probability values. These values are used for computing $T = 1/(1 - qq)$ year return levels.
true.para: Useful for simulation study when the user know the true parameter already.

Values

model names: Model names in the pool of models.
surv.model: Name of the best model selected.
opt_r: Optimal r for the selected model, which was obtained by the ED test for the best model.
theta.best.model: Parameter estimates of the best model.
rt: Return level estimates corresponding to the input qq.

Examples
sel.rmod.surv(data)

Main program-2: sel.rmod.rmed

---

sel.rmod.rmed = function(xdat, maxr=NULL, choose="median", h.fix=NULL, num_inits=20, sig.ed=0.05, qq=c(.98,.99,.995,.998), true.para= NULL)

Usage: The algorithm r_median determines an appropriate r, then selects the best model for the fixed r.

Arguments:

xdat: Data of matrix with R number of columns for order statistics and n rows for independent observations.
maxr: R represents the maximum, predetermined number of the top order. Default is that R <- ncol(xdat).
choose: 'median' or 'average' of some rs, to determine an appropriate r. Default is 'median'.
h.fix: Use when the user have any prior set for the parameter h.
num_inits: Number of trials in finding the MLE using numerical optimization routine.
sig.ed: Significance level for the ED test. Default is 0.05.
qq: Probability values. These values are used for computing $T = 1/(1 - qq)$ year return levels.
true.para: Useful for simulation study when the user know the true parameter already.

Values

model names: Model names in the pool of models.
best.model: Name of the best model selected.
opt_r: The rs obtained by the ED test for the pool of models.
rhat: The r obtained by median or average of rs from the ED test for the pool of models.
theta.best.model: Parameter estimates of the best model.
rt: Return level estimates corresponding to the input qq.

Examples
sel.rmod.rmed(data)