evd: Extreme Value Distributions

by Alec Stephenson

Extreme value distributions arise as the limiting distributions of normalized maxima. They are often used to model extreme behaviour; for example, joint flooding at various coastal locations.

evd contains simulation, distribution, quantile and density functions for univariate and multivariate parametric extreme value distributions.

It also provides functions that calculate maximum likelihood estimates for univariate and bivariate models.

A user's guide is included in the package. It can also be downloaded directly (in postscript or pdf) from `http://www.maths.lancs.ac.uk/~stephena/`.

Introduction

Let $X_1, \ldots, X_m$ be *iid* random variables with distribution function F. Let $M_m = \max\{X_1, \ldots, X_m\}$. Suppose there exists normalizing sequences a_m and b_m such that $a_m > 0$ and as $m \to \infty$

$$\Pr(Z_m \le z) = [F(a_m z + b_m)]^m \to G(z)$$

for $z \in \mathbb{R}$, where G is a non-degenerate distribution function and $Z_m = (M_m - b_m)/a_m$ is a sequence of normalized maxima. It follows that the distribution function G is generalized extreme value, namely

$$G(z) = \exp\left[-\{1 + \xi\,(z-\mu)\,/\sigma\}_+^{-1/\xi}\right],$$

where (μ, σ, ξ) are location, scale and shape parameters, $\sigma > 0$ and $h_+ = \max(h, 0)$. The case $\xi = 0$ is defined by continuity.

Multivariate extreme value distributions arise in a similar fashion (see Kotz and Nadarajah (2000) for details). In particular, any bivariate extreme value distribution can be expressed as

$$G(z_1, z_2) = \exp\left\{-(y_1 + y_2) A\left(\frac{y_1}{y_1 + y_2}\right)\right\},$$

where

$$y_j = y_j(z_j) = \{1 + \xi_j(z_j - \mu_j)/\sigma_j\}_+^{-1/\xi_j}$$

for $j = 1, 2$. The dependence function A characterizes the dependence structure of G. $A(\cdot)$ is a convex function on $[0, 1]$ with $A(0) = A(1) = 1$ and $\max(\omega, 1 - \omega) \le A(\omega) \le 1$ for all $0 \le \omega \le 1$. The jth univariate marginal distribution is generalized extreme value, with parameters (μ_j, σ_j, ξ_j).

Parametric models for the dependence function are commonly used for inference. The logistic model appears to be the most widely used. The corresponding distribution function is

$$G(z_1, z_2; \alpha) = \exp\left\{-(y_1^{1/\alpha} + y_2^{1/\alpha})^\alpha\right\}, \qquad (1)$$

where the dependence parameter $\alpha \in (0, 1]$. Independence is obtained when $\alpha = 1$. Complete dependence is obtained as $\alpha \downarrow 0$. Non-parametric estimators of A also exist, most of which are based on the estimator of Pickands (1981).

Features

- Simulation, distribution, quantile, density and fitting functions for the generalized extreme value and related models. This includes models such as $[F(\cdot)]^m$ for a given integer m and distribution function F, which enable e.g. simulation of block maxima.

- Simulation, distribution, density and fitting functions for eight parametric bivariate extreme value models. Non-parametric estimates of the dependence function can also be calculated and plotted.

- Simulation and distribution functions for two parametric multivariate extreme value models.

- Linear models for the generalized extreme value location parameter(s) can be implemented within maximum likelihood estimation. (This incorporates the forms of non-stationary most often used in the literature.)

- All fitting functions allow any of the parameters to be held fixed, so that nested models can easily be compared.

- Model diagnostics and profile deviances can be calculated/plotted using `plot`, `anova`, `profile` and `profile2d`.

Application

The `sealevel` data frame (Coles and Tawn, 1990) is included in the package. It has two columns containing annual sea level maxima from 1912 to 1992 at Dover and Harwich, two sites on the coast of Britain. There are 39 missing maxima in total; nine at Dover and thirty at Harwich.

The maxima on both margins appear to be increasing with time. The following snippet fits the logistic model (1) with simple linear trend terms on each marginal location parameter.

```
data(sealevel) ; sl <- sealevel
tt <- (1912:1992 - 1950)/100
lg <- fbvlog(sl, nsloc1 = tt, nsloc2 = tt)}
```

The fitted model contains maximum likelihood estimates for $(\gamma_1, \beta_1, \sigma_1, \xi_1, \gamma_2, \beta_2, \sigma_2, \xi_2, \alpha)$ where, for the ith observation, the marginal location parameters are

$$\mu_j(i) = \gamma_j + \beta_j t_i$$

for $j = 1, 2$ and t_i = `tt[i]`. The significance of the trend parameters can be tested using the following analysis of deviance.

```
lg2 <- fbvlog(sl)
anova(lg, lg2)
```

This yields a p-value of about 10^{-6} for the hypothesis $\beta_1 = \beta_2 = 0$.

More complex models for the marginal location parameters can be fitted. Further tests suggest that a quadratic trend could be implemented for the Harwich maxima, but we retain the model `lg` for further analysis. The profile deviance of the dependence parameter α from (1), which corresponds to element `"dep"` in `lg$estimate`, can be produced using `profile`, as shown below.

```
pr <- profile(lg, "dep", xmax = 1)
plot(pr)
```

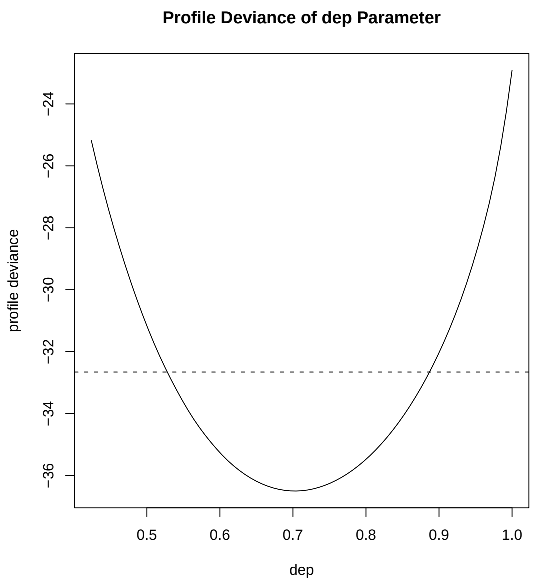

Figure 1: Profile deviance for α.

The horizontal line on the plot represents the 95% confidence interval $(0.53, 0.89)$, which can be calculated explicitly by `pcint(pr)`.

Diagnostic plots for the dependence structure and for the generalized extreme value margins can be produced as follows. The diagnostic plots for the dependence structure include Figure 2, which compares the fitted estimate of the dependence function A to the non-parametric estimator of Capéraà et al. (1997).

```
plot(lg)
plot(lg, mar = 1)
plot(lg, mar = 2)
```

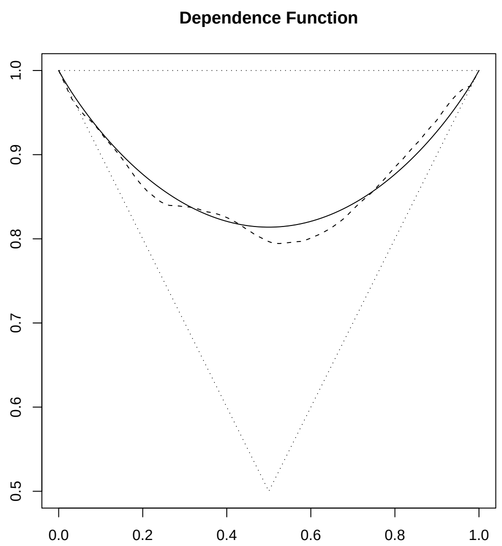

Figure 2: Estimates for the dependence function $A(\cdot)$; the logistic model (solid line), and the non-parametric estimator of Capéraà et al. (1997) (dashed line). The dotted border represents the constraint $\max(\omega, 1-\omega) \le A(\omega) \le 1$ for all $0 \le \omega \le 1$.

Alternative parametric models for the dependence structure can be fitted in a similar manner. The logistic model has the lowest deviance (evaluated at the maximum likelihood estimates) amongst those models included in the package that contain one dependence parameter. The models that contain two dependence parameters produce similar fits compared to the logistic. The models that contain three dependence parameters produce unrealistic fits, as they contain a near singular component. Amongst all models included in the package, the logistic is seen to give the best fit under a number of widely used criteria.

Bibliography

P. Capéraà, A.-L. Fougères, and C. Genest. A non-parametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84:567–577, 1997. 32

S. G. Coles and J. A. Tawn. Statistics of coastal flood prevention. *Phil. Trans. R. Soc. Lond., A*, 332:457–476, 1990. 31

S. Kotz and S. Nadarajah. *Extreme Value Distributions*. Imperial College Press, London, 2000. 31

J. Pickands. Multivariate extreme value distributions. *Proc. 43rd Sess. Int. Statist. Inst.*, 49:859–878, 1981. 31

Alec Stephenson
Lancaster University, UK
`a.stephenson@lancaster.ac.uk`