# VGAM Family Functions for Extreme Value Data

T. W. Yee

November 20, 2007

Beta Version 0.7-4



Department of Statistics,
University of Auckland,
New Zealand
yee@stat.auckland.ac.nz
http://www.stat.auckland.ac.nz/∼yee

## Contents

[Important note: This document and code is not yet finished, but should be completed one day . . . ]

# 1 Introduction

This document describes in more detail about VGAM family functions for extreme value theory. It has become quite a large field in its own right. Many of VGAM's features come from `glm()` and `gam()` so that readers unfamiliar with these functions are referred to Chambers and Hastie (1993). Additionally, the VGAM *User Manual* should be consulted for general instructions about the software. In this document the terms *quantiles*, *percentiles* and *centiles* will be used interchangeably, and note that, for example, a 0.5 quantile is equivalent to percentile $= 50$, which is the median. VGAM family functions use the argument `percentile` to input quantiles, so it should be assigned values between 0 and 100. This article can be thought of as a supplement to Yee and Stephenson (2007).

# 2 Models for Extreme Value Data

Suppose we have data $y_i$, $i = 1, \ldots, n$, assumed to be a random sample from some distribution with distribution function $F$. Extreme value theory is the branch of statistics concerned with inferences on the *tail* of $F$. This contrasts with almost every other branch of statistics where the main concern is in the central part of the distribution. Extreme value theory has important applications in many fields such as environmental science (sea-levels, wind speeds), reliability modelling (weakest-link-type models), finance (e.g., an insurance company is at risk of bankrupcy from large claims) and sport science. There is now a substantial general literature in the area; some starts include Beirlant et al. (1996), Castillo et al. (2005), Coles (2001), Embrechts et al. (1997), Finkenstadt and Rootzen (2003), Kotz and Nadarajah (2000), Reiss and Thomas (2007), and Smith (2003).

## 2.1 Classical Theory

Let $M_n = \max(Y_1, \ldots, Y_n)$ where $Y_i$ are i.i.d. from a continuous cumulative distribution function $F$. Suppose we can find normalizing constants $a_n > 0$ and $b_n$ such that

$$P\left(\frac{M_n - b_n}{a_n} \leq y\right) \longrightarrow G(y) \tag{1}$$

as $n \to \infty$, where $G$ is some proper distribution function. Then $G$ is necessarily one of three possible types of limiting distribution functions. These have been called the Gumbel type, Fréchet type and Weibull type. In the past, controversy would often be present because practitioners chose one of these types to model their data—without much justification.

### 2.1.1 GEV

Nowadays, it is realized that it is more convenient to consider the *generalized extreme value* (GEV) distribution, which holds the three types as special cases. The GEV cumulative distribution function can be written

$$G(y; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, \quad \sigma > 0, \quad -\infty < \mu < \infty, \tag{2}$$

$1 + \xi(y-\mu)/\sigma > 0$, where $x_+ = \max(x, 0)$. The $\mu$, $\sigma$ and $\xi$ are known as the *location*, *scale* and *shape* parameters respectively. The cases $\xi > 0$, $\xi < 0$ and $\xi = 0$ correspond to the Fréchet, Weibull and Gumbel types respectively.

It can be noted that the Gumbel (or Type I) distribution accommodates many commonly-used distributions such as the normal, lognormal, logistic, gamma, exponential and Weibull. However, the rate of convergence of (1) varies enormously, for example, the standard exponential converges quickly but for the standard normal it is extremely slow. The Gumbel cumulative distribution function is

$$G(y) = \exp\left\{-\exp\left[-\frac{y-\mu}{\sigma}\right]\right\}, \qquad -\infty < y < \infty.$$

For the GEV distribution, the $k$th moment about the mean exists if $\xi < k^{-1}$. Provided they exist, the mean and variance are given by $\mu + \sigma\{\Gamma(1-\xi)-1\}/\xi$ and $\sigma^2\{\Gamma(1-2\xi) - \Gamma^2(1-\xi)\}/\xi^2$ respectively, where $\Gamma$ is the gamma function. When $\xi = 0$ then $E(Y) = \mu + \sigma\gamma$ where $\gamma \approx 0.5722$ is Euler's constant.

VGAM fits the GEV distribution by the family function `gev()`. It handles a multivariate response too (see Section 2.1.2). However, for univariate responses, `egev()` is recommended instead because it is faster. Both have

$$\boldsymbol{\eta} = \left(\mu, \log\sigma, \log\left(\xi + \frac{1}{2}\right)\right)^T \tag{3}$$

as default. For parametric models, they provide maximum likelihood estimates (MLEs). Why the link function for $\xi$ in (3)? Answer: Smith (1985) established that when $\xi > -0.5$, the maximum likelihood estimators are completely regular, and the usual asymptotic properties apply. The $\frac{1}{2}$ in the above formula is controlled by the argument `Offset`. Although for $\xi < -0.5$ the usual asymptotic properties do not apply, the MLE generally exists and is superefficient for $-1 < \xi < -0.5$, so it is "better" than normal. The family function `gev()` allows for $-1 < \xi$ by setting `Offset=1`. When $\xi < -1$ the MLE generally does not exist as it effectively becomes

a two parameter problem. To bound $\xi$ between two values use `lshape="elogit"` and the `rshape` argument.

In terms of quantiles,

$$y_p \;=\; \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], \tag{4}$$

where $G(y_p) = 1 - p$. In extreme value terminology, $y_p$ is the *return level* associated with the return period $1/p$. In the argument `percentile` in `gev()`, the user can specify values of $p$ for the fitted values (4).

### 2.1.2 GEV: The $r$-Largest Order Statistics

Suppose now that instead of recording the maximum value we we record the most extreme $r_i$ values (at a fixed value of $x_i$). Thus the type of data can be written $(x_i, \boldsymbol{y}_i)^T$, $i = 1, \ldots, n$, where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ir_i})^T$, $y_{i1} \geq y_{i2} \geq \cdots \geq y_{ir_i}$. We call this *block* data.

Given $x_i$, the data (not just the extremes) are assumed to be i.i.d. realizations from some distribution with continuous distribution function $F$. Here are some examples of this type of data:

1. Table 1 provides data where, for each year between $x = 1931$ and 1981, the 10 highest sea levels in Venice were measured (except for 1935 where the 6 highest were recorded). Thus $r_i = 10$ for all $i$ except for one of them.

2. The top 10 runners in each age group in a school are used to estimate the 99 percentile of running speed as a function of age.

3. The 10 most intelligent children in each age group in a large school are tested with the same IQ test. Fixing the definition of "gifted" as being within the top 1%, the data helps determine the cut-off score for that particular IQ test for each age group in order to screen for gifted children.

The `gev()` family function can handle $r_i > 1$ data. It assumes that the maximum $y_{i1}$ has a GEV distribution. If the $r_i$ are not all equal then the response must be entered as a matrix that is padded with `NA`s.

Table 1: Subset of the Venice sea levels data. For each year from 1931 to 1981 the 10 highest daily sea levels (cm) are recorded (data for only the 6 highest are available for 1935).

| $t$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1931 | 103 | 99 | 98 | 96 | 94 | 89 | 86 | 85 | 84 | 79 |
| 1932 | 78 | 78 | 74 | 73 | 73 | 72 | 71 | 70 | 70 | 69 |
| 1933 | 121 | 113 | 106 | 105 | 102 | 89 | 89 | 88 | 86 | 85 |
| 1934 | 116 | 113 | 91 | 91 | 91 | 89 | 88 | 88 | 86 | 81 |
| 1935 | 115 | 107 | 105 | 101 | 93 | 91 | NA | NA | NA | NA |
| 1936 | 147 | 106 | 93 | 90 | 87 | 87 | 87 | 84 | 82 | 81 |
| 1937 | 119 | 107 | 107 | 106 | 105 | 102 | 98 | 95 | 94 | 94 |
| ⋮ | | | | ⋮ | | | | ⋮ | | |
| 1979 | 166 | 140 | 131 | 130 | 122 | 118 | 116 | 115 | 115 | 112 |
| 1980 | 134 | 114 | 111 | 109 | 107 | 106 | 104 | 103 | 102 | 99 |
| 1981 | 138 | 136 | 130 | 128 | 119 | 110 | 107 | 104 | 104 | 104 |

### 2.1.3 The Block-Gumbel Model

Suppose the maxima is Gumbel (GEV with $\xi = 0$) and let $Y_{(1)}, \ldots, Y_{(r)}$ be the $r$ largest observations, such that $Y_{(1)} \geq \cdots \geq Y_{(r)}$. Given that $\xi = 0$, the joint distribution of

$$\left( \frac{Y_{(1)} - b_n}{a_n}, \ldots, \frac{Y_{(r)} - b_n}{a_n} \right)^T$$

has, for large $n$, a limiting distribution, having density

$$f(y_{(1)}, \ldots, y_{(r)}; \mu, \sigma) = \sigma^{-r} \exp \left\{ - \exp \left( - \frac{y_{(r)} - \mu}{\sigma} \right) - \sum_{j=1}^{r} \left( \frac{y_{(j)} - \mu}{\sigma} \right) \right\},$$

for $y_{(1)} \geq \cdots \geq y_{(r)}$. Upon taking logarithms, one can treat this as an approximate log-likelihood.

Smith (1986) used the results of Weissman (1978) to derive quantiles allowing $\mu$ to be linear in $x$. Rosen and Cohen (1996) extended Smith (1986) to allow for smoothing splines—this is the roughness penalty framework that VGAM naturally operates in. A VGAM family function called `gumbel()` has been written to implement this model (formerly it was called `gumbel.block()`). It should be noted that $\boldsymbol{\eta}(x) = (\mu(x), \log \sigma(x))^T$ is the default so $M = 2$, and that the likelihood used is only an approximate likelihood. When $r_i = 1$ the Gumbel distribution has $E(Y) = \mu$ and $\mathrm{Var}(Y) = \pi^2 \sigma^2 / 6$.

Extreme quantiles for the *block-Gumbel* model can be calculated as follows. If the $y_{i1}, \ldots, y_{ir_i}$ are the $r_i$ largest observations from a population of size $R_i$ at $x_i$ then a large $\alpha = 100(1 - c_i/R_i)\%$ percentile of $F$ can be estimated by

$$\widehat{\mu}_i - \widehat{\sigma}_i \, \log c_i \; . \tag{5}$$

For example, for the Venice data, $R_i = 365$ (if *all* the data were collected there would be one observation per day of the year resulting in 365 observations) and so a 99 percentile is obtained from $\hat{\mu}_i - \hat{\sigma}_i \log(3.65)$. When $R_i$ is missing (which is the default, signified by the argument `R=NA`) then the $\alpha\%$ percentile of $F$ can be estimated using $c_i = -\log(\alpha/100)$ in (5).

The *median predicted value* (MPV) for a particular year is the value for which the maximum of that year has an even chance of exceeding. It corresponds to $c_i = \log(\log(2)) \approx -0.673$ in (5).

From a practical point of view, one weakness of the block-Gumbel model is that one often does not have sufficient data to verify the assumption that $\xi = 0$.

Note that Tawn (1988) considered the block-GEV model, which is more general, and that `gev()` handles this model. Also note that `egumbel()` implements the Gumbel distribution but only handles a univariate response; however it is faster than `gumbel()` on such data.

### 2.1.4 GPD

A second distribution that is important in extreme value theory is the *generalized Pareto* distribution (GPD). This is a common alternative approach to extreme value statistics based on the idea of exceedances over high thresholds. The idea is to pick a high threshold value $u$ (if optimally it becomes a difficult problem) and to study all the *exceedances* of $u$, i.e., values of $Y$ greater than $u$. In extreme value terminology, $Y - u$ are the *excesses*.

Suppose $Y$ has cumulative distribution function $F$, and let $Y^* = Y - u$ given $Y > u$. Then

$$P(Y^* \leq y^*) = P(Y \leq u + y^* | Y > u) = \frac{F(u + y^*) - F(u)}{1 - F(u)}.$$

The cumulative distribution function of the GPD can be written

$$G(y;\mu,\sigma,\xi) \;=\; 1-\left(1+\xi\,\frac{y-\mu}{\sigma}\right)_+^{-1/\xi}, \quad \sigma>0. \tag{6}$$

The parameters $\mu$, $\sigma$ and $\xi$ are known as the *location*, *scale* and *shape* parameters respectively. The function $1-G$ is known as the *survivor function*, and the limit $\xi \to 0$ gives the shifted exponential as a special case, i.e., $G(y) = 1-\exp\{-(y-\mu)/\sigma\}$. The support is $y>\mu$ for $\xi>0$, and $\mu<y<\mu-\sigma/\xi$ for $\xi<0$, i.e., has a finite upper endpoint. In the GPD, if $\xi>0$ then it is long-tailed. If $\xi=0$ then it has an exponential distribution with mean $\sigma$.

It can be shown that

$$E(Y-u|Y>u>0) \;=\; \frac{\sigma+\xi\,u}{1-\xi}, \tag{7}$$

which holds for any $u>\mu$, provided $\xi<1$. This gives a simple diagnostic for threshold selection: the residual mean life should be linear in $u$ at levels for which the model is valid. This suggests producing an empirical plot of the residual life plot and looking for linearity. It involves plotting the sample mean of exceedances of $u$ versus $u$. This is known as a *mean life residual plot* or a *mean excess plot*. It is implemented in VGAM with `meplot()`.

Like the GEV, the GPD has three different special cases depending on the sign of $\xi$. Also, the GPD has mean and variance given by $\mu+\sigma/\{1-\xi\}$ and $\sigma^2/\{(1-2\xi)(1-\xi)^2\}$ provided $\xi<1$ and $\xi<\frac{1}{2}$ respectively. The mean is returned as the fitted value if `percentile=NULL`. VGAM fits the GPD by the family function `gpd()`, which accepts $\mu$ as known input and internally operates on the excesses $y-\mu$. Note that the working weight matrices $\mathbf{W}_i$ in the VGLM/VGAM algorithm are positive-definite only if $\xi>-\frac{1}{2}$, and this is ensured with the default `lshape` argument.

The fitted values of `gpd()` are percentiles obtained from (6):

$$y_p \;=\; \mu+\frac{\sigma}{\xi}\left[(1-p)^{-\xi}-1\right], \tag{8}$$

where $G(y_p)=1-p$. If $\xi=0$ then

$$y_p \;=\; \mu-\sigma\,\log(1-p). \tag{9}$$

In terms of regularity, the GPD is very similar to the GEV. Smith (1985) showed that for $\xi>-\frac{1}{2}$ the information matrix is finite and the classical asymptotic theory of maximum likelihood estimators is applicable, while for $\xi\leq-\frac{1}{2}$ the problem is nonregular and special procedures are needed.

## 2.2 Software and Resources for Extreme Value Data

There are a number of software implementations for extreme value data, of which many run on PCs, see, e.g., Xtremes in Reiss and Thomas (2007), Stephenson and Gilleland (2006). In S-PLUS there is the EVIS library, currently available at
`http://www.math.ethz.ch/~mcneil/software.html`. S. Coles has code for his book at
`http://www.maths.bris.ac.uk/~masgc`. Also, there is the R package **evd** at
`http://www.maths.lancs.ac.uk/~stephena`. See also `http://www.xtremes.math.uni-siegen.de/database.htm`.

## 2.3 General Numerical Details and Summary

Here are some miscellaneous notes.

1. The observed information matrix (OIM) is the Hessian of the negative log-likelihood; its inverse provides a good approximation (in most cases) to the variance-covariance matrix of the estimators. Currently, none of the extremes VGAM family functions use the OIM. This would be Newton-Raphson.

   The expected information matrix (EIM) is the expected value of the same Hessian, evaluated at the true parameters. Currently, many of the extremes VGAM family functions use the EIM, especially the `egev()` and `egumbel()`. This is Fisher scoring.

   For some models the expected information is harder to compute than the observed information (e.g., negative binomial, Gumbel).

2. For `gev()`, VGAM may fail because of the range restriction in (2). If this problem occurs at initialization then the error may be irrecoverable. After the first iteration, with `vglm()` half-stepping avoids this problem, but `vgam()` does not implement half-stepping. If smoothing is required and a range restriction problem occurs try using `vglm()` with regression splines.

3. As well as Rosen and Cohen (1996), other workers have applied penalized likelihood for extreme data, e.g., Pauli and Coles (2001), Chavez-Demoulin and Davison (2005). VGAM provides a neat theoretical and software framework that handles these. It uses iteratively reweighted least squares (IRLS) to obtain maximum likelihood estimates in the parametric case, and obtains penalized likelihood estimates in the nonparametric case. See also Hall and Tajvidi (2000).

4. VGAM family functions in this section are tagged as `"vextremes"` in the `vfamily` slot of the family function. This is accessed by methods functions such as `qtplot.vglm()` and then dispatched accordingly.

5. Smith (2000) discussed Bayesian methods and probability weighted moments as alternative estimation methods for extreme value data.

To summarize, VGAM currently provides the family functions for extreme value data listed in Table 2. The functions `egev()` and `egumbel()` only handle $r_i = 1$ and use expected second derivative matrices.

Table 2: Central S functions for extreme value data. The LHS column are VGAM family functions, the middle are plotting functions. There are many other VGAM family functions associated with extreme value analysis—see other documentation elsewhere.

| | |
|---|---|
| `gev` | `qtplot` |
| `gpd` | `rlplot` |
| `gumbel` | `guplot` |
| `egumbel` | `meplot` |
| `egev` | |

## 2.4 Other Topics

### 2.4.1 Input

For the block-Gumbel model, if $r_i > 1$ then the response in `vglm()`/`vgam()` must be a $n \times \max(r_i)$ matrix, and padded with `NA`s if the $r_i$ are unequal. If there are `NA`s then one must use `na.action=na.pass`. See Section 2.5.3 for an example. However, if there are `NA`s in the model matrix then an error will occur. See the next section for more details about missing values in general.

### 2.4.2 Missing Values

Since `NA`'s are used to 'pad' an input matrix, it is useful to describe how R and S-PLUS handles missing values in general.

The treatment of missing values is controlled by the argument `na.action`. Currently it may be assigned one of the functions given in Table 3, though users can write their own `NA`-handling function. The function `na.exclude()` is similar to `na.omit()` except it pads out the output from generic functions `resid()`, `predict()`, `fitted()` with `NA`s so that the length of the output matches the length of the original data. In contrast, `na.omit()` causes these generic functions to return output with length equal to the number of complete cases at fitting time. S-PLUS's `lm()`, `glm()` have `na.fail()` as default, whereas R's default is `na.omit()`. VGAM's default follows these two defaults. S-PLUS's `gam()` treats missing values quite uniquely and `vgam()` does not follow such behaviour.

Table 3: Functions that `na.action` can be assigned.

| | |
|---|---|
| `na.fail()` | An error message occurs if there are any `NA`s |
| `na.pass()` | The `NA`s are treated the same as ordinary data so that the model frame remains unchanged |
| `na.omit()` | Any row with an `NA` in `x` or `y` is deleted |
| `na.exclude()` | Same as `na.omit` |

### 2.4.3 Quantile Plotting

With models with quantiles, there is a generic function called `qtplot()` which computes and plots quantiles. There is a methods function `qtplot.vextremes()` which is available for block-Gumbel models. See Section 2.5.3 for an example.

## 2.5 Examples

### 2.5.1 GPD

Here is analysis of some rain data using the GPD.

```
> myth = 300
> (fit = vglm(rainfall ~ 1, gpd(threshold = myth, perc = c(50,
+     90, 99)), subset = rainfall > myth))

Call:
vglm(formula = rainfall ~ 1, family = gpd(threshold = myth, perc = c(50,
```

```
    90, 99)), subset = rainfall > myth)

Coefficients:
(Intercept):1 (Intercept):2
    4.3095194    -0.3791078

Degrees of Freedom: 304 Total; 302 Residual
Log-likelihood: -835.0867

> (v = vcov(fit))

              (Intercept):1 (Intercept):2
(Intercept):1    0.01558197   -0.01138651
(Intercept):2   -0.01138651    0.01970718

> diag(v)^0.5

(Intercept):1 (Intercept):2
    0.1248278     0.1403823

> fitted(fit)[1:2, ]

        50%      90%      99%
38 355.0158 513.4578 839.8831
67 355.0158 513.4578 839.8831
```

### 2.5.2 GEV

Here is analysis of some annual maximum temperatures (in °F) data collected at Oxford from 1901 to 1980 using the GEV distribution.

```
> data(oxtemp)
> (fit = vglm(maxtemp ~ 1, egev, data = oxtemp))

Call:
vglm(formula = maxtemp ~ 1, family = egev, data = oxtemp)

Coefficients:
(Intercept):1 (Intercept):2 (Intercept):3
    83.838343      1.449252     -1.547308

Degrees of Freedom: 240 Total; 237 Residual
Log-likelihood: -228.8965

> coef(fit, mat = TRUE)

            location log(scale) logoff(shape, list(offset=0.5))
(Intercept) 83.83834   1.449252                       -1.547308

> Coef(fit)

  location      scale      shape
83.8383434  4.2599271 -0.2871798

> (v = vcov(fit))

              (Intercept):1 (Intercept):2 (Intercept):3
(Intercept):1  0.2668816559 -0.0009631106   -0.04925200
(Intercept):2 -0.0009631106  0.0072372680   -0.01357609
(Intercept):3 -0.0492519963 -0.0135760893    0.07533999

> diag(v)^0.5

(Intercept):1 (Intercept):2 (Intercept):3
   0.51660590    0.08507213    0.27448131

> fitted(fit)[1:2, ]

       95%      99%
1 92.35074 94.71365
2 92.35074 94.71365
```

### 2.5.3 The Block-Gumbel Model

The dataframe `venice` has the 10 highest sea levels (in cm) for each year for the years $x =$ 1931 to 1981; see Table 1. Notice in 1935 that only the top 6 values are available.

```
> data(venice)
> fit = vglm(cbind(r1, r2, r3, r4, r5) ~ year, gumbel(R = 365,
+     mpv = TRUE, zero = 2, lscale = "identity"), venice)
> coef(fit, mat = TRUE)

               location    scale
(Intercept) -780.2947033 12.75869
year           0.4583112  0.00000
```

These results agree with Smith (1986).

A preliminary VGAM fitted to all the data is

```
> y = as.matrix(venice[, paste("r", 1:10, sep = "")])
> fit1 = vgam(y ~ s(year, df = 3), gumbel(R = 365, mpv = TRUE),
+     data = venice, na.action = na.pass)
> green = if (is.R()) "green4" else 5
> red = if (is.R()) "red" else 2
> blue = if (is.R()) "blue" else 4
> darkgreen = if (is.R()) "darkgreen" else 5
> if (!is.R()) ps.options(colors = ps.colors.rgb[c("black",
+     "red", "green", "blue", "dark green", "DarkSeaGreen"),
+     ])
```

Then Figure 1 was produced by

```
> par(mfrow = c(2, 1), mar = c(5, 4, 0.2, 1) + 0.1, xpd = TRUE)
> plot(fit1, se = TRUE, lcol = blue, scol = green, lty = 1,
+     lwd = 2, slwd = 2, slty = if (is.R()) "dashed" else 3)
```

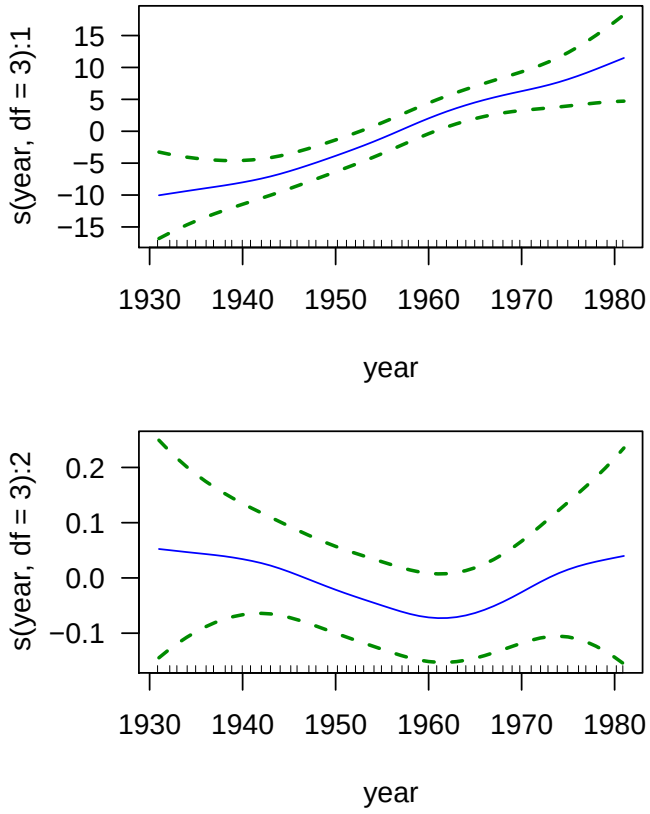


Figure 1: VGAM fitted to the Venice sea level data (`fit1`).

It appears that the first function, $\mu$, is linear and the second, $\sigma$, may be constant. Let's fit such a model with

```
> fit2 = vglm(y ~ year, gumbel(R = 365, mpv = TRUE, zero = 2),
+     venice, na.action = na.pass)
> fitted(fit2)[1:4, ]

       95%      99%      MPV
1 67.78093 88.78784 110.4709
2 68.26374 89.27065 110.9537
3 68.74655 89.75346 111.4365
4 69.22936 90.23627 111.9193
```

Then the quantile plot in Figure 2 is produced by

```
> par(mfrow = c(1, 1), bty = "l", mar = c(4, 4, 0.2, 3) +
+     0.1, xpd = TRUE)
> qtplot(fit2, mpv = TRUE, lcol = c(1, 2, 5), tcol = c(1,
+     2, 5), lwd = 2, pcol = blue, tadj = 0.1)
```

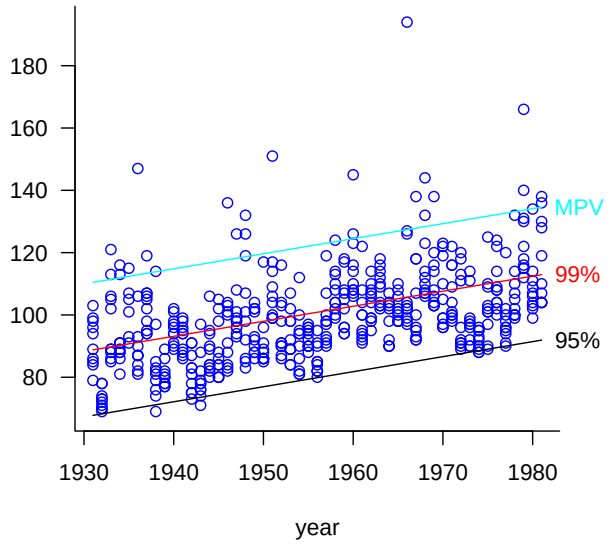


Figure 2: Quantile plot. Venice sea level data (`fit2`).

Clearly, it appears that the response is increasing over time, and that a linear model appears to do well. Now

```
> print(summary(fit2))

Call:
vglm(formula = y ~ year, family = gumbel(R = 365, mpv = TRUE,
    zero = 2), data = venice, na.action = na.pass)

Pearson Residuals:
               Min       1Q   Median      3Q    Max
location   -2.1386 -0.87125 -0.29681 0.74714 3.0077
log(scale) -1.6890 -1.01673 -0.59232 0.31620 4.5688

Coefficients:
                   Value Std. Error t value
(Intercept):1 -826.61784  77.396426 -10.680
(Intercept):2    2.56897   0.042416  60.566
year             0.48281   0.039564  12.203

Number of linear predictors:  2

Names of linear predictors: location, log(scale)

Dispersion Parameter for gumbel family:   1

Log-likelihood: -1086.177 on 99 degrees of freedom
```

```
Number of Iterations: 5
```

so that all the linear coefficients are significant.

The rest of the analysis follows Rosen and Cohen (1996), but allows for the missing values. We'll use `fit1`. Following (5),

```
> par(mfrow = c(1, 1), mar = c(3, 4, 0.2, 1) + 0.1, xpd = TRUE)
> year = venice$year
> matplot(year, y, ylab = "sea level (cm)", type = "n")
> matpoints(year, y, pch = "*", col = blue)
> lines(year, fitted(fit1)[, "99%"], lwd = 2, col = red)
```

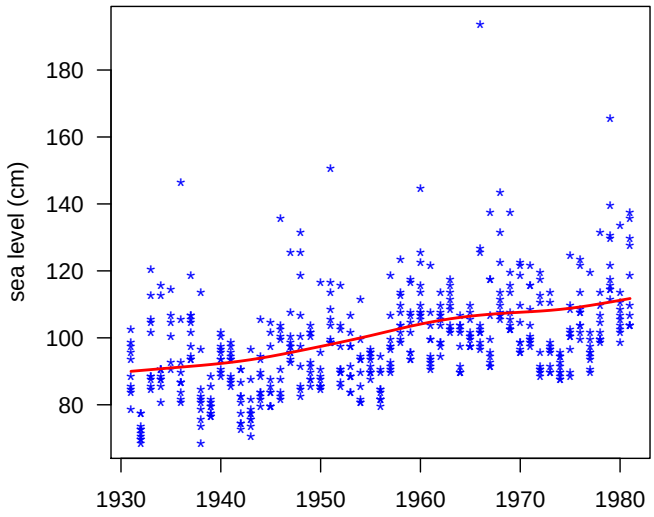


Figure 3: Venice data overlaid with the fitted 99 percentile of `fit1`.

produces Figure 3: 99 percentiles of the distribution. That is, for any particular year, we should expect $99\% \times 365 \approx 361$ observations below the line, or equivalently, 4 observations above the line. It is seen that there is a general increase in extreme sea levels over time (or that Venice is sinking).

To check this, Figure 4 was produced by

```
> par(mfrow = c(1, 1), mar = c(3, 4, 0.2, 1) + 0.1, xpd = TRUE,
+     lwd = 2)
> plot(year, y[, 4], ylab = "sea level", type = "n")
> points(year, y[, 4], pch = "4", col = blue)
> lines(year, fitted(fit1)[, "99%"], lty = 1, col = red)
> lines(smooth.spline(year, y[, 4], df = 4), col = darkgreen,
+     lty = 3)
```

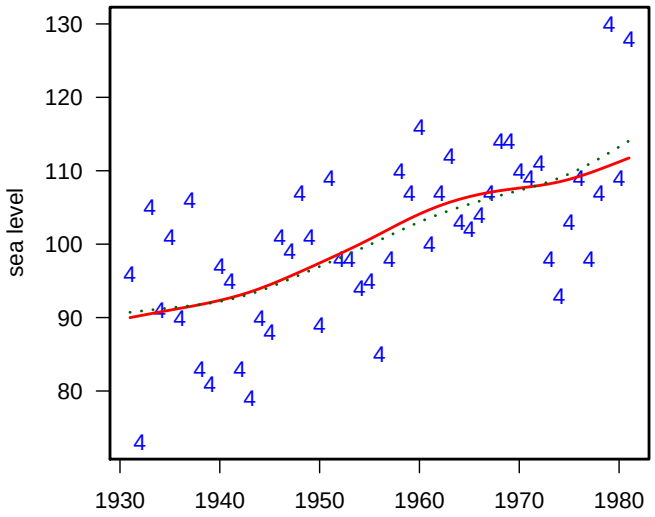


Figure 4: Comparison between the fitted 99 percentile of `fit1` (solid line) and a cubic spline fitted to the 99 percentile data values (4th highest sea level each year; dashed line).

This plot compares a cubic spline fitted to the fourth order statistic ($4/365 \approx 1\%$) values with the fitted 99 percentile values of the block-Gumbel model (same as Figure 3). Although both have approximately the same amount of smoothing, the cubic spline is less wiggly. However, the overall results are very similar.

Finally, Figure 5 plots the *median predicted value*. It was produced by

```
> par(mfrow = c(1, 1), mar = c(3, 4, 0.2, 1) + 0.1, xpd = TRUE,
+     lwd = 2)
> plot(year, y[, 1], ylab = "sea level", type = "n")
> points(year, y[, 1], pch = "1", col = blue)
> lines(year, fitted(fit1)[, "MPV"], lty = 1, col = red)
> lines(smooth.spline(year, y[, 1], df = fit1@nl.df[1] +
+     2), col = darkgreen, lty = 3)
```

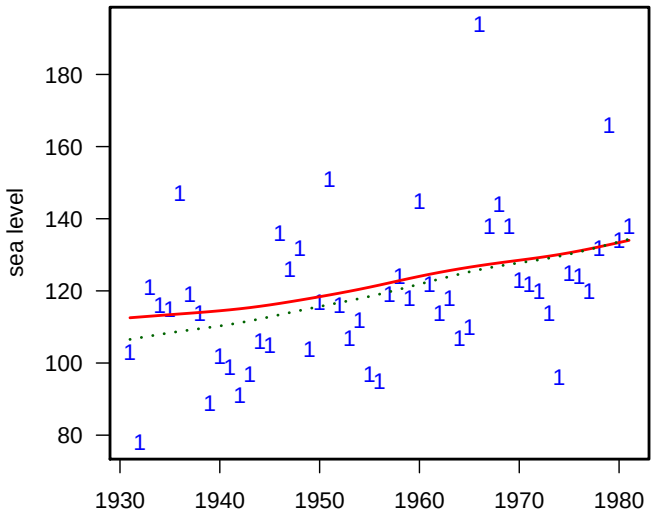


Figure 5: Fitted median predicted value of the Venice sea level data from `fit1`. The points are the highest sea levels for each year. The MPV of `fit1` is the solid line and a spline fitted to these data is the dashed line.

The MPV for a particular year is the value for which the maximum of that year has an even chance of exceeding. It is evident from this plot too that the sea level is increasing over time.

# Exercises

1. Show that (7) holds for the GPD.

2. Write some S functions including `residplot.vextremes()` to plot residuals for models for extreme data. It should produce a histogram, with the probability density function of a standard normal overlaid for comparison. It should plot a QQ-plot of the residuals too.

3. Show for the block-Gumbel model that

$$\begin{aligned}
\frac{\partial \ell_i}{\partial \mu_i} &= \sigma_i^{-1}\left\{r_i - \exp\left(\frac{y_{ir_i}-\mu_i}{\sigma_i}\right)\right\},\\
\frac{\partial \ell_i}{\partial \sigma_i} &= \left\{\sum_{k=1}^{r_i}\frac{y_{ik}-\mu_i}{\sigma_i^2} - \frac{r_i}{\sigma_i} - \frac{y_{ir_i}-\mu_i}{\sigma_i^2}\,e^{-(y_{ir_i}-\mu_i)/\sigma_i}\right\},\\
-E\left[\frac{\partial^2 \ell_i}{\partial \mu_i^2}\right] &= \frac{r_i}{\sigma_i^2},\\
-E\left[\frac{\partial^2 \ell_i}{\partial \sigma_i^2}\right] &= \sigma_i^{-2}\left\{2\left(1+r_i\,\psi(r_i)\right) - 2\sum_{k=1}^{r_i-1}\psi(k) + r_i\left(\psi'(r_i)+\psi^2(r_i)-1\right)\right\},\\
-E\left[\frac{\partial^2 \ell_i}{\partial \mu_i\,\partial \sigma_i}\right] &= \frac{-(r_i\,\psi(r_i)+1)}{\sigma_i^2}.
\end{aligned}$$

4. Consider the GEV distribution for a univariate response $y_i$, $i = 1, \ldots, n$. Write down the log-likelihood, and show that, as $\xi_i$ tends to zero,

$$\frac{\partial \ell_i}{\partial \mu} = \left(1 - e^{-z_i}\right)/\sigma_i, \tag{10}$$

$$\frac{\partial \ell_i}{\partial \sigma} = \left[z_i\left(1 - e^{-z_i}\right) - 1\right]/\sigma_i, \tag{11}$$

$$\frac{\partial \ell_i}{\partial \xi} = z_i\left[\frac{z_i}{2}\left(1 - e^{-z_i}\right) - 1\right], \tag{12}$$

where $z_i = (y_i - \mu_i)/\sigma_i$.

Now consider the expected information matrix, as given by Prescott and Walden (1980). Let $p = (1+\xi)^2\Gamma(1+2\xi)$ and $q = \Gamma(2+\xi)\{\psi(1+\xi) + (1+\xi)/\xi\}$ where $\psi$ is the digamma function. Let $\gamma \approx 0.5772\cdots$ be Euler's constant. Then Prescott and Walden (1980) showed that (the subscript $i$ will be dropped)

$$-E\left[\frac{\partial^2 \ell}{\partial \mu^2}\right] = \frac{p}{\sigma^2}, \tag{13}$$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma^2}\right] = \frac{1 - 2\Gamma(2+\xi) + p}{\sigma^2\xi^2}, \tag{14}$$

$$-E\left[\frac{\partial^2 \ell}{\partial \xi^2}\right] = \frac{1}{\xi^2}\left\{\frac{\pi^2}{6} + \left(1 - \gamma + \frac{1}{\xi}\right)^2 - \frac{2q}{\xi} + \frac{p}{\xi^2}\right\}, \tag{15}$$

$$-E\left[\frac{\partial^2 \ell}{\partial \mu\,\partial \sigma}\right] = \frac{-1}{\sigma^2\xi}\left\{p - \Gamma(2+\xi)\right\}, \tag{16}$$

$$-E\left[\frac{\partial^2 \ell}{\partial \mu\,\partial \xi}\right] = \frac{-1}{\sigma\xi}\left(q - \frac{p}{\xi}\right), \tag{17}$$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma\,\partial \xi}\right] = \frac{-1}{\sigma\xi^2}\left\{1 - \gamma + \frac{1 - \Gamma(2+\xi)}{\xi} - q + \frac{p}{\xi}\right\}. \tag{18}$$

Calculate the limit of these expressions as $\xi$ tends to zero. Note that the expected second derivative with respect to $\xi$ does not depend on the other two parameters; calculate its value to about 2 decimal places (numerically, as it is the most difficult and probably intractable).

Note: show that $\Gamma'(x) = \psi(x)\Gamma(x)$ and $\Gamma''(x) = \Gamma(x)\left\{\psi'(x) + \psi^2(x)\right\}$.

5. Consider the GPD distribution for a univariate response $y_i$, $i = 1, \ldots, n$. Write down the log-likelihood, and show that, as $\xi_i$ tends to zero,

$$\frac{\partial \ell_i}{\partial \sigma_i} = \frac{1}{\sigma_i}\left[\frac{y_i}{\sigma_i} - 1\right], \tag{19}$$

$$\frac{\partial \ell_i}{\partial \xi_i} = \frac{y_i}{\sigma_i}\left[\frac{y_i}{2\sigma_i} - 1\right]. \tag{20}$$

## ACKNOWLEDGEMENTS

The author wishes to thank Stuart Coles for providing helpful manuscripts of material, as well as the Rain and Oxford data. Also thanks to Alec Stephenson and O. Rosen for helpful input and comments.

# References

Beirlant, J., Teugels, J. L., Vynckier, P., 1996. Practical Analysis of Extreme Values. Leuven University Press, Leuven, Belgium.

Castillo, E., Hadi, A. S., Balakrishnan, N., Sarabia, J. M., 2005. Extreme Value and Related Models with Applications in Engineering and Science. Wiley, Hoboken, N.J., USA.

Chambers, J. M., Hastie, T. J. (Eds.), 1993. Statistical Models in S. Chapman & Hall, New York.

Chavez-Demoulin, V., Davison, A. C., 2005. Generalized additive modelling of sample extremes. Applied Statistics 54, 207–222.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer-Verlag, New York.

Finkenstadt, B., Rootzen, H. (Eds.), 2003. Extreme Values in Finance, Telecommunications and the Environment. Chapman & Hall/CRC, Boca Raton, FL.

Kotz, S., Nadarajah, S., 2000. Extreme Value Distributions: Theory and Applications. Imperial College Press, London.

Pauli, F., Coles, S., 2001. Penalized likelihood inference in extreme value analyses. Journal of Applied Statistics 28, 547–560.

Prescott, P., Walden, A. T., 1980. Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. Biometrika 67, 723–724.

Reiss, R.-D., Thomas, M., 2007. Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields, 3rd Edition. Birkhäuser, Basel, Switzerland.

Rosen, O., Cohen, A., 1996. Extreme percentile regression. In: Härdle, W., Schimek, M. G. (Eds.), Statistical Theory and Computational Aspects of Smoothing: Proceedings of the COMPSTAT '94 Satellite Meeting held in Semmering, Austria, 27–28 August 1994. Physica-Verlag, Heidelberg, pp. 200–214.

Smith, R. L., 1985. Maximum likelihood estimation in a class of nonregular cases. Biometrika 72, 67–90.

Smith, R. L., 1986. Extreme value theory based on the $r$ largest annual events. Journal of Hydrology 86, 27–43.

Smith, R. L., 2000. Extreme value statistics in meteorology and the environment. `http://www.unc.edu/depts/statistics/faculty/rsmith.html` .

Smith, R. L., 2003. Statistics of extremes, with applications in environment, insurance and finance. In: Finkenstadt and Rootzen (2003), pp. 1–78.

Stephenson, A. G., Gilleland, E., 2006. Software for the analysis of extreme events: The current state and future directions. Extremes 8, 87–109.

Tawn, J. A., 1988. An extreme-value theory model for dependent observations. Journal of Hydrology 101, 227–250.

Weissman, I., 1978. Estimation of parameters and large quantiles based on the $k$ largest observations. Journal of the American Statistical Association 73, 812–815.

Yee, T. W., Stephenson, A. G., 2007. Vector generalized linear and additive extreme value models. Extremes 10, 1–19.