Journal of Hydrology (NZ) 41(2): 197-202, 2002

A comment on modelling extremes: Links between Multi-Component Extreme Value and General Extreme Value distributions

Martin A.J. van Montfort and Bram van Putten[1]
Department of Mathematics and Statistics, Wageningen University
Dreijenlaan 4, 6703 HA Wageningen, The Netherlands
[1] *Bram.vanPutten@mat.dpw.wau.nl*

Abstract

Revfeim (1991) reduced the number of parameters of the Multi-Component Extreme Value distribution with m components from $2m$ to 4, where one of the four parameters ($C_{\mu\mu}$) quantifies the variation between the expectations of the Exponential size distributions. A well-known method for determining Jenkinson's General Extreme Value (GEV) distribution in the Peak Over Threshold approach is using a Compound distribution of the maximum of a Poisson number (or count) of sizes with a Pareto distribution, where the Pareto distribution is a generalisation of the Exponential distribution. An Exponential distribution with fixed expectation μ can be generalised to a Pareto distribution by replacing μ by a 2-parameter Gamma distribution, in which these two parameters are functions of the expectation and variation of the random size. The 3 parameters of the GEV distribution can then be interpreted in terms of an expected count, an expected size and the variation of the sizes.

Introduction

Several methods are used to model the distribution of maxima, including hydrological maxima. One of these is to think in terms of the maximum Y of N event sizes with N a random number with support $\{0,1,2, \ldots\}$; these events have a random size X. In this so-called Peak Over Threshold approach, the number of events and the event sizes are assumed to be stochastically independent. In the simplest case, N has a Poisson distribution with expectation ρ, denoted by $N \propto$ Poisson (ρ), and X has an Exponential distribution with expectation μ, denoted by $X \propto \text{Exp}(\mu)$. Then for the compound distribution of Y, denoted by C.P.Exp:

$$
\begin{aligned}
Pr\{Y \leq y\} &= \sum_{n=0}^{\infty} e^{-\rho}\rho^{n} / n!(1-e^{-y/\mu})^{n} \\
&= \exp(-\rho e^{-y/\mu}), \text{ or} \\
&= \exp\{-\exp(-\frac{y-\mu\ln(\rho)}{\mu})\}, \ (y \geq 0)
\end{aligned}
$$

and zero for $y < 0$. Note $Pr\ (Y \leq 0) = Pr\ (Y = 0) = \exp\ (-r) = Pr\ (N = 0)$. In the case $N = 0$ the maximum size is defined to be the lower bound of the support of X. For non-negative y this is the well-known EV1 or Gumbel distribution with location parameter (= mode) $\mu \ln(\rho)$ and scale parameter μ.

Two distributional choices have been made: a Poisson count distribution and an Exponential size distribution. Nature can be more complicated. The count distribution can be generalised from the Poisson distribution to a Negative Binomial distribution, by adding one more parameter. Analogously, the size distribution can be generalised to the so-called (Generalised) Pareto distribution (with tails that may be less or more heavy when compared to those of the Exponential distribution). Van Montfort and Otten (1991) dealt with detecting statistical support for a non-Poisson distribution of the count and/or a non-Exponential distribution for the sizes, so increasing the number of parameters from 2 to 4.

C.P.Pareto coincides on its support with Jenkinson's General Extreme Value (GEV) distribution, introduced by Jenkinson (1955). C.P.Pareto has a point probability on zero equal to exp (-ρ), whilst GEV is a continuous distribution. The cumulative distribution function of the GEV is obtained by replacing $e^{-y/\mu}$ in the EV1 formula by $(1-e^{-\theta y/\mu})/\theta$ with an infinite upper bound of the support for $\theta < 0$; this shows the generalisation of the Exponential to the Pareto distribution.

Other generalisations of C.P.Exp to extreme-value distributions with more than 2 parameters can be found in the hydrologic literature, see e.g. Revfeim (1989, 1991, 2001).

Revfeim considered the annual maximum on a specific point as the maximum of the m independent monthly or seasonal maxima. For the 'superstar' Y of the maxima of, say, m seasons $(X_1, \ldots X_m)$ with distribution C.P.Exp (ρ_i, μ_i), $(i = 1, \ldots, m)$ one gets, in the case of independent maxima,

$$
Pr\{Y \leq y\} = \prod_{i=1}^{m} \exp(-\rho_i e^{-y/\mu_i}),
$$

so

$$
-\ln(Pr) = \sum \rho_i e^{-y/\mu_i}
$$

A different problem in practise leads to the same formulae, namely where Y is the superstar of m independent maxima over the same time period, at m points within a climatologically homogeneous area. The assumption that ρ is constant seems practical for, say, precipitation amounts, but the stochastic independence of these maxima needs special attention insofar as stochastic dependence is influencing the result.

Revfeim reduces the number of parameters from $2m$ *to* 4 by a numerical approximation using series expansion. This reduction of the Multi-Component Extreme Value distribution is close to the Two-Component Extreme Value distribution (which also has 4 parameters). The 4 parameters ρ, μ, $C_{\mu\mu}$, and $C_{\rho\mu}$ are defined by

$$\rho=\Sigma\rho_i/m;\ \mu=\Sigma\mu_i/m;\ C_{\mu\mu}=\Sigma(\mu_j/\mu-1)^2/m \text{ and } C_{\rho\mu}=\Sigma(\mu_j/\mu-1)(\rho_i/\rho-1)/m.$$

Revfeim gets (with $\tilde{y}=y/\mu$)

$$-\ln(Pr\{Y\leq y\})\approx(m\rho)\exp(-\tilde{y})\{1+(\tilde{y}^2/2-\tilde{y})C_{\mu\mu}+\tilde{y}C_{\rho\upsilon}\}$$

For $C_{\rho\mu}=0$ one gets a 3-parameter distribution with parameters $\rho,\mu,C_{\mu\mu}$; this happens at no correlation of (ρ_i,μ_i), $(i=1,\ldots,m)$ or at no variation in ρ and/or μ.

Here we want to focus on variation in μ at $\rho_1=\ldots=\rho_m$ $(=\rho)$, so on a 3-parameter distribution with parameters $\rho,\mu,C_{\mu\mu}$ with as a special case the C.P.Exp for $C_{\mu\mu}=0$.

This 3-parameter model could be useful for the maximum over an area with constant ρ and variation of μ over that area, and thus useful for climatological applications. The idea of a varying μ at a fixed ρ gives rise to the question how to generalise C.P.Exp (ρ,μ) to Jenkinson's GEV by replacing μ by a random variable; this question is the same as how to generalise the Exp-distribution to get a generalised Pareto distribution. It turns out that replacing μ by a two-parameter Gamma distribution does the right job.

Generalised Pareto distribution as a gamma-mixture of exponentials

The results discussed in this section can also be found in Johnson & Kotz (1970) referencing Harris (1968), but do not seem to be in use in hydrology. The probability density function of the Exponential distribution with expectation $1/\lambda$ (denoted above by μ) reads

$$f_{Exp}(x)=\begin{cases}\lambda e^{-\lambda x}, & 0<x<\infty\\ 0 & , x<0\end{cases}.$$

The probability density function of the Gamma distribution with parameters a and $1/k$ reads

$$f_{Gamma}(\lambda)=\begin{cases}k^a\lambda^{a-1}e^{-k\lambda}/\Gamma(a), & 0<\lambda<\infty\\ 0 & ,\lambda<0\end{cases}$$

For the probability density function of the compound distribution of Exp_λ where λ is a random variate with a Gamma distribution one gets

$$f_{compound}(y)=\int_0^\infty f_{Exp}(y|\lambda)f_{Gamma}(\lambda)d\lambda\,.$$

For $y<0$ one gets $f_{compound}(y)=0$.

For $y>0$ one gets $f_{compound}(y)=\int_0^\infty \lambda e^{-\lambda y}k^a\lambda^{a-1}e^{-k\lambda}/\Gamma(a)\,d\lambda$

$$=\ldots=a\,k^a\,/(y+k)^{a+1}$$

So $f_{compound}(y)=f_{Pareto}(y)$, where the support is $(0,\infty)$ and the cumulative distribution function *cdf* of the Pareto distribution on its support reads

$cdf_{Pareto}(y)=1-(k/(y+k))^a$.

Note that:

$E(Y)=k/(a-1)$ for $a>1$; $\mathrm{var}(Y)=ak^2/((a-1)^2\,(a-2))$ for $a>2$, and $CV_Y^2=a/(a-2)$.

Here we applied the Gamma distribution to generalise the Exponential distribution to a Pareto distribution. Solving the problem of which distribution has to be plugged in for this generalisation needs further algebra and results in the Gamma distribution.

Concluding remarks

Revfeim's 4-parameter Extreme-value distribution contributes to the choice of Jenkinson's GEV as a physically relevant model for hydrologic maxima in areas that are reasonably homogeneous in climate.

Revfeim's 4-parameter Extreme-value distribution is the result of a distributionally relevant reduction of the 2m parameters of the MCEV to only 4 parameters, with only minor changes in the cumulative distribution function. A different way is to start with choosing m=2, where MCEV is denoted by 2CEV or TCEV. This TCEV entered quantitative hydrology in the 1980s (see Rossi *et al.*,1984; Beran *et al.*, 1986) and was succesfully used afterwards, see e.g. Connell and Pearson (2001). Beran *et al.* compared TCEV with other Extreme-value distributions used in quantitative hydrology by plotting the distributions on Gumbel paper: paper on which the Gumbel distribution (=EV1) is represented by a straight line. All non-Gumbel models for extremes are represented by lines curved like the GEV, with an infinite upper bound (θ< 0).

Van Montfort (2000) showed that this GEV, MCEV (including TCEV), extremes with an EV1-distribution after a power transformation

$$\left[x \rightarrow \left(x^{\lambda}-1\right)/\lambda \text{ for } \lambda<1 \text{ with } \lim_{\lambda-->0}\left(x^{\lambda}-1\right)/\lambda=\ln(x)\right]$$

all show the same curvature in the EV1 plot, as do some other commonly used distributions of extremes with more than two parameters.

This means that for the purpose of description the GEV does a good job, but that some types of distributions of extremes more accurately reflect the real hydrological situation, e.g. they are correctly modelled by TCEV or Revfeim's approximation of the MCEV.

Proper interpretation of the parameters also contributes to sound regionalisation of maxima.

References

Beran, M.; Hosking, J.R.M.; Arnell, N. 1986: Comment on "Two Component ExtremeValue Distribution for Flood Frequency Analysis" by Fabio Rossi, Mauro Fiorentino, and Pasquale Versace. *Water Resources Research 22*: 263-266.

Connell, R.J.; Pearson, C.P. 2001: Two-Component Extreme Value Distribution Applied to Canterbury Annual Maximum Flood Peaks. *Journal of Hydrology New Zealand 40*: 105-127.

Harris, C.M. 1968: The Pareto distribution as a queue service discipline. *Operations Research 16*: 307-313.

Jenkinson, A.F. 1955: The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly Journal of the Royal Meteorological Society 81*: 158-171.

Johnson, N.L.; Kotz, S. 1970: Distributions in statistics: continuous univariate distributions – 1. A. Wiley - Interscience publication. Chapter 19: Pareto Distribution.

Revfeim, K.J.A. 1989: A framework for interpreting rainfall models. *In*: Sansom, J (*ed.*) Proceedings Fourth International Meeting on Statistical Climatology, 27–31 March 1989, Rotorua, New Zealand: 233-237.

Revfeim, K.J.A. 1991: Annual maxima and totals of seasonally varying processes. *Stochastic Hydrology and Hydraulics 5*: 147–153.

Revfeim, K.J.A. 2001: Extremes of observations bounded by limits on the size or number of events. *In*: Handout at 8 IMSC, Lüneburg, Germany, 12-16 March 2001; unpublished paper, 11 p.

Rossi, F.; M. Fiorentino, M.; Versace, P. 1984: Two component extreme value distributions for flood frequency analysis. *Water Resources Research 20:* 847-856.

Van Montfort, M.A.J.; Otten, A. 1991: The first and second e of the extreme value distribution EV1. *Stochastic Hydrology and Hydraulics 5*: 69-76.

Van Montfort, M.A.J. 2000: EV1-plot of some distributions of extremes. Topic 4 in *Some topics relating to maxima in hydrology*, Technical Note 00-05, Subdepartment of Mathematics, Wageningen University: 6 p.

Manuscript received: 23 August 2001; accepted for publication: 17 May 2002.