UNIVERSITY OF
COPENHAGEN

Advanced Methods in Applied Statistics

Generalized extreme value distribution

Application in Climate Sciences

Authors: Tinus Blæsbjerg
Gabriel Guyader

Copenhagen, March 5, 2025

1 Introduction

Extreme Value Theory (EVT) is an area of statistics that deals with the low probability values located at the tails of probability density functions (PDF). It is used in different research fields to study extremes and this write-up directly addresses how it is used in Climate Science by looking at the paper entitled "Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions" Huang et al. 2016.

2 Review of theory

2.1 Fischer-Tippet-Gnedenko Theorem

Suppose you have multiple sequences of independent and identically distributed random variables $(X_1, X_2, ...X_n)$ with block length n, drawn from a cumulative distribution function (CDF) F. When $n \to \infty$, $M_n = max(X_1, X_2, ...X_n)$ can also go to infinity therefore M_n needs normalization. The Fischer-Tippet-Gnedenko theorem (Fisher and Tippett 1928; Gnedenko 1943) then says that for specific normalization constants $a_n > 0$ and b_n, then if $\frac{M_n - b_n}{a_n}$ converges in F to a nondegenerate distribution function as $n \to \infty$:

$$\lim_{n\to\infty} P(\frac{M_n - b_n}{a_n} \leq x) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \tag{1}$$

then G(x) is the CDF of a GEV distribution. This means for recurring events with sufficiently large enough block length n one could find multiple maximums $[M_{n1}, M_{n2}, ...M_{nj}]$ and fit it to a GEV distribution. An example could be annual maximum precipitation in an area. The GEV distributions can also be used to look at extreme minimums instead.

2.2 Generalized extreme value distribution

This section is primarily based on the work of Coles et al. (2001). Three parameters specify the GEV distribution:

- The **location** parameter μ is related to the mode. A greater μ value will give a greater mode for the distribution (and vice-versa). For a distribution where $\xi = 0$, μ is the mode.

- The **scale** parameter σ provides information about the spread of values. The higher the scale parameter, the more spread out the values will be.

- The **shape** parameter ξ determines the behaviour of the tail's distribution. If $\xi = 0$, the tail is open towards both $+\infty$ and $-\infty$, and if $\xi > 0$ (resp. $\xi < 0$) the tail is open toward $-\infty$ (resp. $+\infty$). There are three families of GEV distribution (cf. Figure1) depending on the shape parameter: **Gumbel** ($\xi = 0$), **Fréchet** ($\xi > 0$), and **Weibull** ($\xi < 0$).

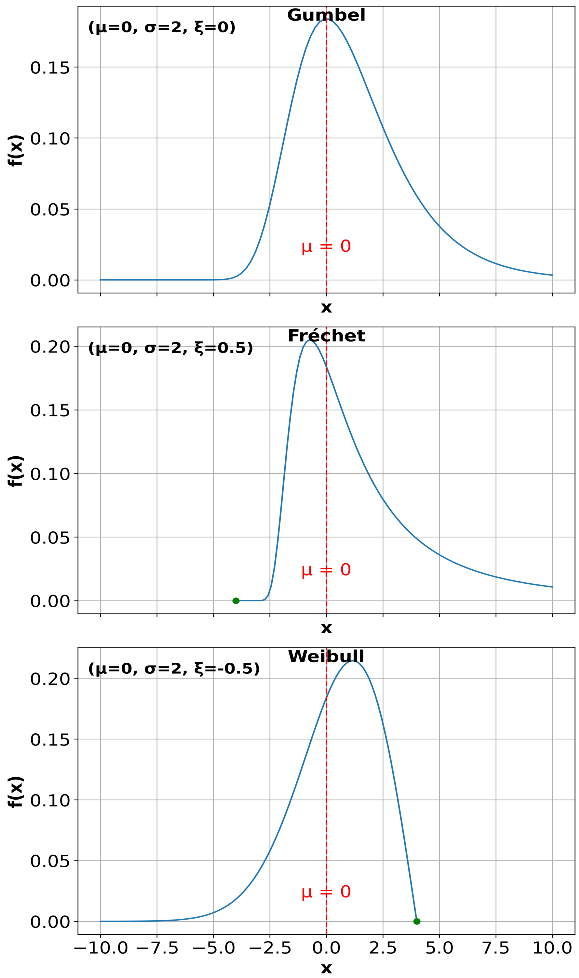

Figure 1: The three families of a GEV distribution.

The support of a GEV distribution depends on the shape parameter and is defined as :

- $[\mu - \frac{\sigma}{\xi}, +\infty[$ if $\xi > 0$.
- $]-\infty, +\infty[$ if $\xi = 0$.
- $]-\infty, \mu - \frac{\sigma}{\xi}]$ if $\xi < 0$.

The probability density function (PDF) of a GEV distribution is given by :

$$\frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}, \tag{2}$$

where

$$t(x) = \begin{cases} 1 + \xi \left(\frac{x-\mu}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ e^{\frac{x-\mu}{\sigma}} & \text{if } \xi = 0. \end{cases}$$

2.3 Return level

An important aspect of GEV is the return level which looks at the quantiles of the fitted GEV distribution. The return level x_r is the value in the GEV distribution that is on average exceeded once per r years[1]. The

[1] We use years because it is common in Climate Science to look at annual maximums. However the choice of time unit depends on the block length n, so one could use months or centuries instead.

probability of exceeding x_r in any given year is given by $P(x_r < X) = \frac{1}{r}$. The probability of not exceeding x_r is $P(X \leq x_r) = 1 - 1/r$ and can be found by using the CDF of the underlying GEV distribution:

$$G(x_r) = 1 - 1/r. \tag{3}$$

One can now find the return year of a specific maximum by inverting the $G(x_r)$ function and solving for x_r. Likewise, the return period of an observed extreme can be estimated by solving for r (Huang et al. 2016). We haven't presented the cumulative distribution functions of the GEV distributions in this write-up but they can be looked up when solving for x_r and r.

3 Application in Climate Sciences: Temperature extremes

Extreme value theory assumes that measurements and events follow **stationary conditions** which means that underlying conditions do not change in time. This is however not true for a changing climate because of an increasing concentration of CO_2. To account for this the most commonly used nonstationary GEV model assumes that location and scale parameters change linearly in time and the shape parameter is time-invariant (Kharin and Zwiers 2005). Huang et al. (2016) investigate how the parameters (μ, σ, ξ) change in a changing climate and how it affects the GEV distribution and the return level (cf. Figure 2).

In the paper, Huang et al. consider the annual maxima of daily maximum temperatures and the annual minima of daily minimum temperatures across the contiguous United States. As it is difficult to evaluate changes in extremes with short observational records, the data are derived from global climate models that were run over 1,000 years under pre-industrial, 700 ppm, and 1,400 ppm CO2 concentrations.

3.1 Paper results

Huang et al. find the annual maxima and then fit these values to a GEV distribution to examine changes in the fit parameters (μ, σ, ξ) caused by different CO_2 concentrations. They fit the temperature extremes to a Weibull distribution because it has an upper bound restricting results that are not physically possible. They find that μ increases for both warm and cold extremes all over the United States because of an overall increase in mean temperature due to more CO_2. The change in σ and ξ is more complex but does show a decrease in σ for cold extremes inland and generally shows that a linear and time-invariant change is not sufficient for σ and ξ.

Additionally, they assess how the change in the parameters impacts the return levels of extreme events (cf. Equation (3)). Here they find generally all across the United States that for warm extremes the return level value is mostly affected by a change in μ, which corresponds to a uniform increase in the return level (cf.

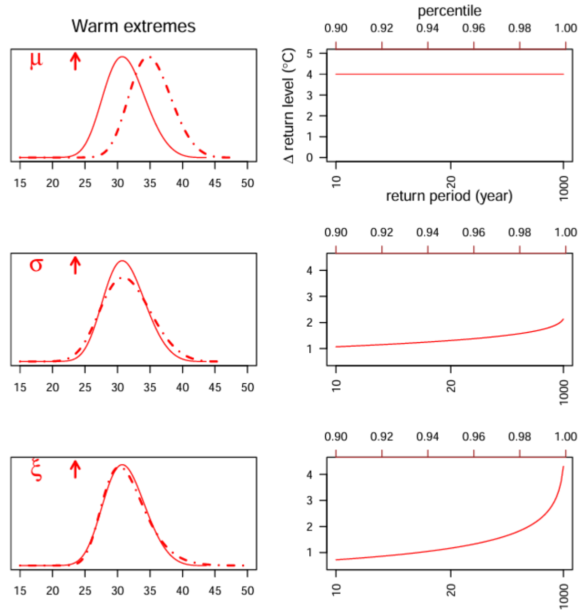

Figure 2: Shows what the effect of changing the parameters of the GEV distribution has on the return level. In the left column are the GEV distribution PDFs (we assume the x-axis is temperature, it was not stated in the figure of the paper), and in the right is the effect on the return level. The dashed line is the GEV after an increase in the parameter shown. *They say in the paper that the extreme temperatures follow the Weibull GEV, so we assume the GEVs seen are Weibull.* Figure 3 from Huang et al. (2016)

.

Figure 2). For cold extremes, they find it is more geographically dependent but in general, all three μ, σ, and ξ parameters show changes. A change in σ and ξ affects the return level values for high return periods close to 1000 (cf. Figure 2).

4 Conclusion

Extreme value analysis is a powerful tool to examine extreme values and can give valuable insight into how the extreme values of the climate are changing and how humans need to adapt to those. Nonetheless, it also has limits. Since the climate is very complex and chaotic the parameters of the fitted GEVs change with time which can make predictions hard. Furthermore, the model used in the paper is just one way of simulating how the climate could be another model could supply a different result. Not knowing the CO_2 concentration of the future also makes predictions quite troublesome.

In the paper they also investigate if the block length of one year is enough to satisfy equation (1). Furthermore, they also look at how uncertainties behave when the climate model is run for less time than millennia. Due to limited space, we have not included reviews of those in this write-up.

References

Coles, Stuart et al. (2001). *An introduction to statistical modeling of extreme values*. Vol. 208. Springer. Chap. chapter 3.

Fisher, Ronald Aylmer and Leonard Henry Caleb Tippett (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample". In: *Mathematical proceedings of the Cambridge philosophical society*. Vol. 24. 2. Cambridge University Press, pp. 180–190.

Gnedenko, Boris (1943). "Sur la distribution limite du terme maximum d'une serie aleatoire". In: *Annals of mathematics* 44.3, pp. 423–453.

Huang, Whitney K et al. (2016). "Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions". In: *Advances in Statistical Climatology, Meteorology and Oceanography* 2.1, pp. 79–103.

Kharin, Viatcheslav V and Francis W Zwiers (2005). "Estimating extremes in transient climate change simulations". In: *Journal of Climate* 18.8, pp. 1156–1173.