

Generalized Probability Weighted Moments in Extreme Value Theory

Pierre Ribereau ¹

Université de Montpellier 2, France

Armelle Guillou

Université de Strasbourg, France

Philippe Naveau

Laboratoire des Sciences du Climat et de l'Environnement, CEA/CNRS, Gif-sur-Yvette, France

Abstract: In 1985 and 1987 Hosking *et al.* estimated with the so-called Probability-Weighted Moments (PWM) method the parameters of the Generalized Extreme Value (GEV) and Generalized Pareto (GPD) distribution, the first being classically fitted to maxima of sequences of independent and identically distributed random variables and the second to excesses over a high threshold. Their approach is still very popular in hydrology and climatology because but its main drawback resides in its limitations when applied to strong heavy-tailed densities. Whenever the shape parameter is larger than 0.5, the asymptotic properties of the PWMs cannot be derived. To broaden the validity domain of the PWM approach, we present an extension of PWM to a larger class of moments, called Generalized PWM (GPWM). This allows us to derive the asymptotic properties of our estimators for larger values of the shape parameter. The performance of our approach is illustrated by studying simulations of small, medium and large GEV and GPD samples. We present also some extensions of our results, such the non-stationary case or re-parametrisation that gives independent estimators.

Keywords: Extreme Value Theory, Generalized Probability Weighted Moments Method, Parameters estimation

1. Introduction

Since the work of Fisher and Tippett in 1928, it is known that the only possible limiting form of a normalized maximum of a random sample (when a non-degenerate limit exists) is captured by the Generalized Extreme Value distribution (GEV)

$$G(x; \sigma, \gamma, \mu) = \begin{cases} \exp\left(-\left\{1 + \gamma \frac{x-\mu}{\sigma}\right\}^{-1/\gamma}\right) & , \text{ if } 1 + \gamma \frac{x-\mu}{\sigma} > 0, \gamma \neq 0, \\ \exp\left(-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right) & , \text{ if } x \in \mathbb{R}, \gamma = 0, \end{cases}$$

with $\mu \in \mathbb{R}$, $\sigma > 0$ and γ are called the location, scale and shape parameters, respectively. The GEV distribution is appropriate when the data consist of a set of maxima. However, there has been some criticism of this approach, because using only maxima leads to loss of the information contained in other large sample values in a given period. This problem

¹Address of correspondence: pribere@math.univ-montp2.fr

is remedied by considering several of the largest order statistics instead of just the largest one: that is, considering all values larger than a given threshold. The differences between these values and a given threshold are called excesses over the threshold. Pickands (1975) showed that the distribution of the excesses can be approached by the Generalized Pareto Distribtuion defined as :

$$F_{\mu,\sigma}(x) = \begin{cases} \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}} & \text{for } \gamma \neq 0 \text{ and } 1 + \gamma x/\sigma > 0, \\ \exp\left(-\frac{x}{\sigma}\right) & \text{for } \gamma = 0, \end{cases}$$

where σ and γ are the scale and shape parameters.

In both case, an important issue is to estimate the parameters of the limiting distribution. The Generalized Probability-Weighted Moments (GPWM) method is a generalization of the Probability-Weighted Moments (PWM) method introduced by Landwehr *et al.* (1979), and developped by Hosking *et al.* (1985, 1987) for the GEV and GPD distribution. The main idea of the PWM method is to match the following moments :

$$\mathbb{E}\left[X^p(F(X))^r(1 - F(X))^s\right], \text{ with } p, r \text{ and } s \text{ real numbers,}$$

with their empirical functionals, similarly to the classical method-of-moments. Hosking and his co-workers asserted that PWMs estimators performed better than a classical Maximum Likelihood estimation (MLE) for small samples. Its conceptual simplicity, its practicability and its good properties for small samples can explain the success of the PWM approach in geosciences and hydrology.

Its main drawback resides in its limitations when applied to strong heavy-tailed densities. Indeed, we can not derive asymptotic normality of the estimators whenever $\gamma > 1/2$, (i.e. finite variance). Here, ours aims are threefold. Firstly, we propose the GPWM estimators for the GEV distribution and establish the asymptotic normality under general conditions ensuring the validity of the method for a large range of values of γ . Secondly, we compare their performances with MLEs and classical PWMs specillay on small and moderate sample sizes. And finally, we propose some extensions in the case of non-stationary data, and re-parametrisation of the GPD distribution that leads to independent estimators.

2. Generalized Probability-Weighted Moments method

Let $X_1, \dots, X_n$ be i.i.d. random variables with distribution function G . The GPWM can be described in the following way

$$\nu_\omega = \mathbb{E}\left(X\omega(G)\right) = \int_{-\infty}^{\infty} x \omega(G(x))dG(x),$$

where ω is a suitable continuous function. By changing variables, this moment can be rewritten as

$$\nu_\omega = \int_0^1 G^{-1}(u)\omega(u)du.$$

Let W be the primitive of ω , null at 0. We propose to estimate ν_ω by

$$\widehat{\nu}_{\omega,n} = \int_0^1 \mathbb{F}_n^{-1}(u)\omega(u)du \tag{1}$$

where $\mathbb{F}_n$ denotes the classical empirical distribution function based on a sample $(X_1, \dots, X_n)$. We are interested in the asymptotic properties of $\widehat{\nu}_{\omega,n}$ for the GEV distribution. To reach this goal, we select a function ω such that

$$\omega(t) = O((1-t)^b) \quad \text{for } t \text{ close to } 1, b \geq 0 \quad (2)$$

and

$$\omega(t) = O(t^{a'}) \quad \text{for } t \text{ close to } 0, a' > 0. \quad (3)$$

These assumptions tie down the functions $G^{-1}(t)$ and $\mathbb{F}_n^{-1}(t)$ at $t = 0$ and $t = 1$. An example of such a function is $\omega(t) = t^a(-\log t)^b, a > a'$. In this case, the GPWM for the GEV distribution can be rewritten as

$$\nu_\omega = \frac{\sigma}{\gamma} \frac{1}{(a+1)^{b-\gamma+1}} \Gamma(b-\gamma+1) - \left(\frac{\sigma}{\gamma} - \mu\right) \frac{1}{(a+1)^{b+1}} \Gamma(b+1).$$

As for the PWMs method, a system of three equations for three different values of a and/or b has to be solved in order to obtain estimators for σ, γ, μ .

Under the conditions (2) and (3), the GPWM ν_ω exists as soon as $\gamma < b+1$. We can also obtain the asymptotic normality as soon as $\gamma < b+1/2$. Details of the result, demonstration and a simulation study can be founded in Diebolt *et al.* (2008). Similar results for the GPD distribution are described in Diebolt *et al.* (2007a). An example of application of the method to extreme quantile estimation is also explained in Diebolt *et al.* (2007b).

3. Extensions

3.1 Extension to non-stationary case

The case when $X_1, \dots, X_n$ are independent but not identically distributed can also be treated using GPWM. Suppose that each X_i follows a $\text{GEV}(\mu_0 + \mu_1 \times i, \sigma, \gamma)$. Using the relation stating that, if $X \sim \text{GEV}(\mu, \sigma, \gamma)$ then $(X - \mu)/\sigma \sim \text{GEV}(0, 1, \gamma)$, one can proposed a specific linear model to treat the problem of non-stationarity. More details can be founded in Ribereau *et al.* (2009).

3.2 Obtaining independent parameters estimation

A major problem of all estimation methods in Extreme value inference is the dependence of the parameters. In the specific case of Maximum Likelihood, a reparametrization of the GPD distribution of the form :

$$\begin{cases} \gamma_{ML}^* = \gamma \\ \xi_{ML}^* = \sigma(1 + \gamma) \end{cases}$$

leads to a diagonal Fisher information matrix and then to independent estimators, making the interpretation of each parameter estimation easier. The case of PWM and GPWM

method is more difficult. For example, for classical PWM, the following reparametrization :

$$\begin{cases} \gamma_{PWM}^* &= \gamma \\ \xi_{PWM}^* &= \sigma \exp\left\{-\frac{1}{2} \log(1 - \gamma) - \frac{11}{28} \log(1 - \gamma + 2\gamma^2) + \right. \\ &\quad \left. \frac{2}{7} \log(2 - \gamma) + \frac{\sqrt{7}}{14} \arctan\left(\frac{-1+4\gamma}{\sqrt{7}}\right)\right\} \end{cases}$$

leads to an asymptotic diagonal variance matrix. In the range $\gamma \in [-0.5, 0.5]$, the second equation can be approximated by $\xi_{PWM}^* = \sigma(\gamma + 1.15)$. Similar reparametrization can be founded for the GPWM approach.

References

- Diebolt J., Guillou A., Naveau P. and Ribereau P. (2008) Improving probability-weighted moment methods for the generalized extreme value distribution, *REVSTAT*, 6, 33–50.
- Diebolt J., Guillou A. and Rached I. (2007a) Approximation of the distribution of excesses through a generalized probability-weighted moments method, *Journal of Statistical Planning and Inference*, 137, 841–857.
- Diebolt J., Guillou A. and Ribereau P. (2007b) Asymptotic normality of extreme quantile estimators based on the peaks-over-threshold approach, *Communication in statistics: theory and method*, 36, 869–886.
- Fisher R.A. and Tippett L.H.C. (1928) Limiting forms of the frequency distribution in the largest particle size and smallest member of a sample, *Proc. Camb. Phil. Soc.*, 24, 180–190.
- Greenwood J.A., Landwehr J.M., Matalas N.C. and Wallis J.R. (1979) Probability-weighted moments: definition and relation to parameters of several distributions expressible in inverse form, *Water resources research*, 15, 1049–1054.
- Hosking J.R.M. and Wallis J.R. (1987) Parameter and quantile estimation for the generalized pareto distribution.
- Hosking J.R.M., Wallis J.R. and Wood E.F. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments.
- Landwehr J., Matalas N. and Wallis J. (1979) Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles, *Water resources research*, 15, 1055–1064.
- Pickands J. (1975) Statistical inference using extreme order statistics, *Annals of statistics*, 3, 119–131.
- Ribereau P., Guillou A. and Naveau P. (2008) Estimating return levels from maxima of non-stationary random sequences, *Non-linear Processes in Geophysics*, Accepted.