Available online at www.sciencedirect.com
SCIENCE DIRECT
C. R. Acad. Sci. Paris, Ser. I 338 (2004) 629–634
COMPTES RENDUS MATHEMATIQUE

Statistics/Probability Theory

# A new look at probability-weighted moments estimators

Jean Diebolt [a], Armelle Guillou [b], Imen Rached [a]

[a] *Université de Marne-la-Vallée, équipe d'analyse et de mathématiques appliquées, 5, boulevard Descartes, batiment Copernic, Champs-sur-Marne, 77454 Marne-la-Vallée cedex 2, France*

[b] *Université Paris VI, laboratoire de statistique théorique et appliquée, boîte 158, 175, rue du Chevaleret, 75013 Paris, France*

Received 30 October 2003; accepted after revision 19 February 2004

Presented by Paul Deheuvels

**Abstract**

In this article, we propose to study, in more generality, the probability-weighted moments method used by Hosking and Wallis (1987) in the case of generalized Pareto distributions which depend on two parameters $\gamma$ and $\sigma$. The objective is to extend the domain of validity: $\gamma < 1/2$ required in order to obtain the asymptotic properties of their estimators. By simulations, we show the efficiency of our technique. ***To cite this article: J. Diebolt et al., C. R. Acad. Sci. Paris, Ser. I 338 (2004).***


**Résumé**

**Un nouvel aperçu sur les estimateurs des moments pondérés.** Dans cet article, nous proposons d'étudier, dans un cadre plus général, la méthode des moments pondérés utilisée par Hosking et Wallis (1987) dans le cas de distributions de Pareto généralisées dépendant de deux paramètres $\gamma$ et $\sigma$. L'objectif est d'élargir le domaine d'applications : $\gamma < 1/2$ indispensable pour obtenir les propriétés asymptotiques de leurs estimateurs. Nous montrons l'efficacité de notre technique par le biais de simulations. ***Pour citer cet article : J. Diebolt et al., C. R. Acad. Sci. Paris, Ser. I 338 (2004).***


## Version française abrégée

Les lois de Pareto généralisées ont été introduites par Pickands [6] pour modéliser les excès au-delà d'un seuil. Elles dépendent de deux paramètres $\gamma$ et $\sigma$. Leur évaluation n'est généralement pas un problème facile. Hosking et Wallis [5] ont proposé d'utiliser une version simplifiée des estimateurs des moments pondérés (PWM), où seuls les deux premiers moments sont utilisés, et par simulations ils ont montré l'efficacité de leur technique par rapport à des approches plus standards, comme la méthode du maximum de vraisemblance ou la méthode des moments. Le problème essentiel de leur approche est le domaine de validité : $\gamma < 1/2$, indispensable afin d'obtenir les propriétés asymptotiques de leurs estimateurs. Le but de cet article est donc d'utiliser la méthode PWM dans toute sa généralité de façon à élargir le domaine d'applications et de permettre ainsi un plus large éventail d'utilisations de cette technique. Nous illustrons par simulations l'efficacité de cette nouvelle approche.

*E-mail address:* guillou@ccr.jussieu.fr (A. Guillou).

1631-073X/$ – see front matter 
doi:10.1016/j.crma.2004.02.011

## 1. Introduction

The distribution of the largest values of certain natural phenomena (e.g., waves, earthquakes, floods, ...) is of interest in many practical applications. This interest has given rise to a rapid development of extreme value theory in recent years. The traditional approach to the analysis of extreme values in a given population is based on the family of generalized extreme value (GEV) distributions (see Fisher and Tippett [3]).

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 the loss of information contained in other large-sample values in a given period. This problem 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 exceedances over the threshold. These exceedances are typically assumed to have a generalized Pareto distribution $\mathrm{GPD}(\gamma,\sigma)$ whose distribution function is defined by

$$G_{\gamma,\sigma}(x) = \begin{cases} 1-\left(1+\dfrac{\gamma}{\sigma}x\right)^{-1/\gamma} & \text{if } \gamma \neq 0,\ \sigma > 0, \\ 1-\exp\left(-\dfrac{x}{\sigma}\right) & \text{if } \gamma = 0,\ \sigma > 0, \end{cases} \tag{1}$$

where $x \in [0,\infty[$ if $\gamma \geqslant 0$ and $x \in [0,-\sigma/\gamma[$ if $\gamma < 0$, $\sigma$ and $\gamma$ being the scale and shape parameters.

Among all the ad-hoc methods used in parameter estimation, the method of moments has attracted a lot of interest. In full generality, it consists in equating model-moments based on $G_{\gamma,\sigma}$ to the corresponding empirical moments based on the data. Their general properties are notoriously unreliable on account of the poor sampling properties of second- and higher-order sample moments. Another currently favored method consists in using the maximum likelihood (ML) approach, whose justification is based on large-sample theory, and therefore there has been little assessment of the performance of this method when applied to small or moderate samples. Since neither method is completely satisfactory, we propose to use, in this paper, the probability-weighted moments method in its full generality. First we explain the 'classical' probability-weighted moments (PWM) method, which was introduced by Greenwood et al. [4] and used by Hosking and Wallis [5]. Note that all these methods (moments, ML, PWM) were also used in order to estimate the parameters of the GEV distribution and that recently, Coles and Dixon [1] introduced a penalized likelihood method which improves on the small-sample properties of the classical ML method.

The probability-weighted moments of a positive random variable $X$ with distribution function $F(x) = \mathbb{P}(X \leqslant x)$ are the quantities

$$M_{p,r,s} = \mathbb{E}\big(X^p\big(F(X)\big)^r\big(1-F(X)\big)^s\big),$$

where $p$, $r$ and $s$ are positive real numbers.

Greenwood et al. [4] exhibited several distributions (the Gumbel, logistic and Weibull distributions, among others) for which the relationship between the parameters of the distribution and the PWMs $M_{1,r,s}$ is simpler than the relationship between the parameters and the conventional moments $M_{p,0,0}$. When $r$ and $s$ are integers, $F^r(1-F)^s$ may be expressed as a linear combination of either powers of $F$ or powers of $(1-F)$, so it is natural to summarize a distribution either by the moments $M_{1,r,0}$ $(r = 0,1,\ldots)$ or by $M_{1,0,s}$ $(s = 0,1,\ldots)$.

Hosking and Wallis [5] proposed to base estimation on the first two probability-weighted moments:

$$\mu_s = \mathbb{E}\big[X\big(1-G_{\gamma,\sigma}(X)\big)^s\big] = \frac{\sigma}{(s+1)(s+1-\gamma)} \quad \text{with } s = 0 \text{ and } s = 1.$$

The parameters $(\gamma,\sigma)$ can be recovered by

$$\gamma = 2 - \frac{\mu_0}{\mu_0 - 2\mu_1} \quad \text{and} \quad \sigma = \frac{2\,\mu_0\mu_1}{\mu_0 - 2\mu_1}.$$

The PWM estimators $\hat{\gamma}_n$ and $\hat{\sigma}_n$ are obtained by replacing $\mu_0$ and $\mu_1$ by estimators based on an observed sample of size $n$. For instance, Hosking and Wallis [5] proposed to use $\hat{\mu}_s := \frac{1}{n}\sum_{j=1}^{n}(1-p_{j,n})^s X_{j,n}$, where $X_{1,n} \leqslant \cdots \leqslant X_{n,n}$ denotes the ordered sample and $p_{j,n} = (j-0.35)/n$. For $\gamma < 1/2$, the probability-weighted moments estimators $\hat{\gamma}_n$ and $\hat{\sigma}_n$ are asymptotically normally distributed with a variance-covariance matrix given by

$$\frac{1}{(1-2\gamma)(3-2\gamma)}\begin{pmatrix} (1-\gamma)(2-\gamma)^2(1-\gamma+2\gamma^2) & \sigma(\gamma-2)(2-6\gamma+7\gamma^2-2\gamma^3) \\ \sigma(\gamma-2)(2-6\gamma+7\gamma^2-2\gamma^3) & \sigma^2(7-18\gamma+11\gamma^2-2\gamma^3) \end{pmatrix}.$$

The PWM estimators have several advantages over existing methods of estimation. They are fast and straightforward to compute and always yield feasible values for the estimated parameters. The biases of the estimators are small and they decrease rapidly as the sample size increases. The standard deviations of the PWM estimators are comparable with those of the ML estimators for moderate sample sizes and are often less than those of the ML estimators for small samples.

The main problem with this method is the range of validity: $\gamma < 1/2$, in order to derive the asymptotic properties of the PWM estimators. This condition is restrictive for many applications (as in insurance, ...), where at least the domain $\gamma \leqslant 1$ should be covered. In order to solve this problem, we propose to use the PWM method in its full generality, i.e. the approach used should not be reduced to the two first moments only. With this aim, instead of defining the estimators $(\hat{\gamma}_n, \hat{\sigma}_n)$ of $(\gamma, \sigma)$ on the basis of $(\hat{\mu}_0, \hat{\mu}_1)$, we propose to define new estimators $(\hat{\gamma}_{s_1,s_2,n}, \hat{\sigma}_{s_1,s_2,n})$ based on $(\hat{\mu}_{s_1,n}, \hat{\mu}_{s_2,n})$ defined as

$$\hat{\mu}_{s_1,n} := \frac{1}{s_1+1}\int_0^\infty \big(1-\mathbb{F}_n(x)\big)^{s_1+1}\,\mathrm{d}x, \tag{2}$$

where $\mathbb{F}_n$ is the classical empirical distribution function based on $X_1, \ldots, X_n$, a sample from a GPD$(\gamma, \sigma)$, and $s_1$ and $s_2$ are real numbers such that $1 \leqslant s_1 < s_2$.

In Section 2, we establish the main asymptotic properties of $(\hat{\mu}_{s_1,n}, \hat{\mu}_{s_2,n})$ from which we deduce those of $(\hat{\gamma}_{s_1,s_2,n}, \hat{\sigma}_{s_1,s_2,n})$. Then, in Section 3, some simulations are proposed in order to give an indication about the choice of the parameters $(s_1, s_2)$. The couple $(s_1, s_2) = (1, 1.5)$ seems to give, in all our simulated examples, estimators with small variances. Finally, we study the efficiency of $\hat{\gamma}_{1,1.5,n}$ in a MSE-sense.

## 2. Main results

Let $1 \leqslant s_1 < s_2$. We denote by $T_{(s_1,s_2)}$ the $C^1$-diffeomorphism which transforms $(\mu_{s_1}, \mu_{s_2})$ into $(\gamma, \sigma)$:

$$T_{(s_1,s_2)} : ]0,\infty[^2 \cap \big\{(x,y):\ (s_1+1)x > (s_2+1)y\big\} \longrightarrow ]-\infty, s_1+1[ \times ]0,\infty[,$$

$$\gamma = \frac{(s_1+1)^2\mu_{s_1} - (s_2+1)^2\mu_{s_2}}{(s_1+1)\mu_{s_1} - (s_2+1)\mu_{s_2}} \quad \text{and} \quad \sigma = \frac{(s_2+1)(s_1+1)(s_2-s_1)\mu_{s_1}\mu_{s_2}}{(s_1+1)\mu_{s_1} - (s_2+1)\mu_{s_2}}.$$

Now, we need to define an estimator for the PWM $\mu_{s_1}$. For practical reasons, we propose to use the estimator $\hat{\mu}_{s_1,n}$ defined in (2) and we establish in our Theorem 2.1 the asymptotic normality of $(\hat{\mu}_{s_1,n}, \hat{\mu}_{s_2,n})$.

**Theorem 2.1.** *Let $\mu_{s_1}$ and $\mu_{s_2}$, $1 \leqslant s_1 < s_2$, be the PWM of a random variable $X$ from a* GPD$(\gamma, \sigma)$. *Denoting their estimates by $\hat{\mu}_{s_1,n}$ and $\hat{\mu}_{s_2,n}$ respectively, then for all $\gamma < s_1 + 1/2$ :*

$$\sqrt{n}(\hat{\mu}_{s_1,n} - \mu_{s_1}, \hat{\mu}_{s_2,n} - \mu_{s_2}) \longrightarrow^d \left(\sigma\int_0^1 t^{s_1-\gamma-1}B(t)\,\mathrm{d}t, \sigma\int_0^1 t^{s_2-\gamma-1}B(t)\,\mathrm{d}t\right),$$

*where $B$ is a Brownian bridge and the variance-covariance matrix of the limiting distribution is given by*

$$\Gamma_{s_1,s_2}=\begin{pmatrix}\frac{\sigma^2}{(2s_1+1-2\gamma)(s_1+1-\gamma)^2} & \frac{\sigma^2}{(s_1+1-\gamma)(s_2+1-\gamma)(s_2+s_1+1-2\gamma)}\\ \frac{\sigma^2}{(s_1+1-\gamma)(s_2+1-\gamma)(s_2+s_1+1-2\gamma)} & \frac{\sigma^2}{(2s_2+1-2\gamma)(s_2+1-\gamma)^2}\end{pmatrix}.$$

**Proof.** A Taylor expansion of order 2 with remainder gives

$$\sqrt{n}(\hat\mu_{s_1,n}-\mu_{s_1}) =^d -\int_0^\infty \alpha_n\big(F(x)\big)\big(1-F(x)\big)^{s_1}\,\mathrm{d}x + \frac{s_1}{\sqrt{n}}\int_0^\infty\int_0^1 (1-t)\big(\alpha_n\big(F(x)\big)\big)^2\left(1-F(x)-\frac{t}{\sqrt{n}}\alpha_n\big(F(x)\big)\right)^{s_1-1}\mathrm{d}t\,\mathrm{d}x =: T_{1,n}+T_{2,n},$$

where $\alpha_n(\cdot)$ denotes the classical empirical process based on uniform random variables. Then, we prove that, for all $\gamma < s_1 + 1/2$,

$$T_{1,n} \longrightarrow^d -\sigma\int_0^1 t^{s_1-\gamma-1}B(t)\,\mathrm{d}t,$$

$$T_{2,n} \longrightarrow^P 0.$$

To study $T_{1,n}$, we decompose the integral into three parts: the integral from 0 to $\frac{1}{n}$, from $\frac{1}{n}$ to $1-\frac{1}{n}$ and from $1-\frac{1}{n}$ to 1. The first and the last integral can be treated using Jaeschke's theorem (see, e.g., Shorack and Wellner [7], p. 600). For the second one, we use Einmahl and Mason's result [2]. For the convergence of $T_{2,n}$ to 0 in probability, we first remark that, after changing variables,

$$0 \leqslant T_{2,n} \leqslant \frac{s_1\sigma}{\sqrt{n}}\int_0^1 \big|\alpha_n(1-u)\big|\int_0^1\big|\alpha_n(1-u)\big|\left(u-\frac{t}{\sqrt{n}}\alpha_n(1-u)\right)^{s_1-1}u^{-\gamma-1}\,\mathrm{d}t\,\mathrm{d}u$$
$$\leqslant -\sigma\int_0^1 \alpha_n(1-u)\left(\left(u-\frac{1}{\sqrt{n}}\alpha_n(1-u)\right)^{s_1}-u^{s_1}\right)u^{-\gamma-1}\,\mathrm{d}u.$$

Let $U_{1,n}\leqslant\cdots\leqslant U_{n,n}$ be the uniform random variables on which the process $\alpha_n(\cdot)$ is based. We decompose the latter integral into two parts: from 0 to $U_{1,n}$ and from $U_{1,n}$ to 1. The convergence to 0 in probability of $T_{2,n}$ then follows by application of Jaeschke's theorem and use of Kiefer's theorem and Robbins and Siegmund's theorem (see, e.g., Shorack and Wellner [7], pp. 407–408). Theorem 2.1 can then easily be deduced.

From this theorem, we can now establish our main result which is the asymptotic normality of our estimators $(\hat\gamma_{s_1,s_2,n},\hat\sigma_{s_1,s_2,n})$, for all $\gamma < s_1+1/2$.

**Corollary 2.2.** *Let $\hat\gamma_{s_1,s_2,n}$ and $\hat\sigma_{s_1,s_2,n}$ be the estimators of the* GPD$(\gamma,\sigma)$ *deduced from the estimators $\hat\mu_{s_1,n}$ and $\hat\mu_{s_2,n}$ using the $C^1$-diffeomorphism $T_{(s_1,s_2)}$. Then, for all $1\leqslant s_1<s_2$ and all $\gamma<s_1+1/2$,*

$$\sqrt{n}(\hat\gamma_{s_1,s_2,n}-\gamma,\hat\sigma_{s_1,s_2,n}-\sigma)\longrightarrow^d \mathcal{N}(0,\Sigma_{s_1,s_2}),$$

*where*

$$\Sigma_{s_1,s_2}=A_{s_1,s_2}\Gamma_{s_1,s_2}A^t_{s_1,s_2},$$

*with $A_{s_1,s_2}=DT_{(s_1,s_2)}(\mu_{s_1},\mu_{s_2})$.*

**Proof.** From Theorem 2.1, we derive that

$$(\hat{\gamma}_{s_1,s_2,n}, \hat{\sigma}_{s_1,s_2,n}) \stackrel{d}{=} T_{(s_1,s_2)}\left((\mu_{s_1}, \mu_{s_2}) + \frac{1}{\sqrt{n}}\big(\xi_1 + \xi_1^{(n)}, \xi_2 + \xi_2^{(n)}\big)\right),$$

where $\xi = (\xi_1, \xi_2)$ follows a $\mathcal{N}(0, \Gamma_{s_1,s_2})$ distribution and $\xi^{(n)} = (\xi_1^{(n)}, \xi_2^{(n)})$ converges to $(0,0)$ in distribution as $n \to \infty$. A Taylor expansion of $T_{(s_1,s_2)}$ gives our Corollary 2.2.

## 3. Simulations

Our aim now is to give some indications about the choice of the parameters $(s_1, s_2)$. If we study the expression of $\Sigma_{s_1,s_2}$ as a function of $(s_1, s_2)$, we find that there is no value of $(s_1, s_2)$ which minimizes the asymptotic variance of $\hat{\gamma}_{s_1,s_2,n}$ and $\hat{\sigma}_{s_1,s_2,n}$ simultaneously in the domain $\{(s_1, s_2){:}\ 1 \leqslant s_1 < s_2,\ \gamma < s_1 + 1/2\}$. However, a graphical study of these functions for different values of $\gamma$ $(-1/2, 0, 1/2, 3/4, 1)$ shows that they have almost the same behaviour. Considering the optimal values for $(s_1, s_2)$ in these cases, the choice of $(s_1, s_2) = (1, 1.5)$ seems to be a good compromise for practical use. For example, for $\gamma = 0$, the optimal value is $(s_1, s_2) = (1, 1.1)$, and for $\gamma = 1$, it is $(1, 2)$. Note that this means that the asymptotic properties of our estimators are valid for all $\gamma < 3/2$, which is enough for many applications. As an example, we give in Fig. 1 the results for $(\gamma = 1, \sigma = 1)$.

A computer simulation experiment was run using the values $(s_1, s_2) = (1, 1.5)$ for the GPD. Simulations were performed for sample sizes $n = 25, 50, 100, 200$ and $500$ with the shape parameter taking the values $\gamma = -0.4, 0, 0.4, 1$. The scale parameter $\sigma$ was set to 1 throughout. Since the method is equivariant under scale changes of the data, setting $\sigma = 1$ involves no loss of generality. For each combination of values of $n$ and $\gamma$, 50 000 random samples were generated from the GPD. Our simulation results are summarized in Table 1, which gives the root mean squared error (RMSE) of $\hat{\gamma}_{s_1,s_2,n}$ (similar results can be obtained for $\hat{\sigma}_{s_1,s_2,n}$). We can see that the generalized PWM method seems to give small RMSE in all the cases and these RMSE decrease quickly as the sample size increases.


## Acknowledgement

The authors are very grateful to the referees for their comments.


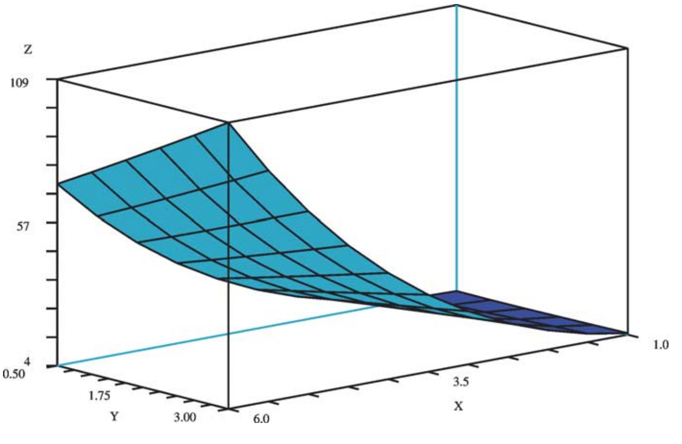


Fig. 1. Graph of the asymptotic variance of $\hat{\gamma}_{s_1,s_2,n}$ for $(\gamma = 1, \sigma = 1)$ where $X = s_1$, $Y = s_2 - s_1$.

Table 1
RMSE of $\hat{\gamma}_{s_1,s_2,n}$ for different values of $n$ and $\gamma$ $(\sigma = 1)$

| $n$ | $\gamma$ | | | |
|---|---|---|---|---|
| | $-0.4$ | 0 | 0.4 | 1 |
| 25 | 0.56 | 0.46 | 0.39 | 0.40 |
| 50 | 0.36 | 0.30 | 0.26 | 0.30 |
| 100 | 0.24 | 0.20 | 0.18 | 0.23 |
| 200 | 0.17 | 0.15 | 0.12 | 0.17 |
| 500 | 0.10 | 0.088 | 0.078 | 0.11 |

## References

[1] S.G. Coles, M. Dixon, Likelihood-based inference for extreme value models, Extremes 2 (1999) 5–23.
[2] U. Einmahl, D.M. Mason, Approximation to permutation and exchangeable processes, J. Theor. Probab. 5 (1992) 101–126.
[3] R.A. Fisher, L.H.C. Tippett, Limiting forms of the frequency distribution in the largest particle size and smallest member of a sample, Proc. Cambridge Philos. Soc. 24 (1928) 180–190.
[4] J.A. Greenwood, J.M. Landwehr, N.C. Matalas, J.R. Wallis, Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form, Water Resources Res. 15 (1979) 1049–1054.
[5] J.R.M. Hosking, J.R. Wallis, Parameter and quantile estimation for the generalized Pareto distribution, Technometrics 29 (1987) 339–349.
[6] J. Pickands III, Statistical inference using extreme order statistics, Ann. Statist. 3 (1975) 119–131.
[7] G.R. Shorack, J.A. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.