Approximation of the distribution of excesses using a generalized probability weighted moment method

Jean Diebolt a, Armelle Guillou b, Imen Rached a

a Université de Marne-la-Vallée, équipe d’analyse et de mathématiques appliquées, bâtiment Copernic, 5, boulevard Descartes, 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 20 June 2004; accepted after revision 18 January 2005

Abstract

The POT (Peaks-Over-Threshold) approach consists of using the generalized Pareto distribution (GPD) to approximate the distribution of excesses over a threshold. In this Note, we consider this approximation using a generalized probability weighted moment (GPWM) method. We study the asymptotic behaviour of our new estimators and also the functional bias of the GPD as an estimate of the distribution function of the excesses. To cite this article: J. Diebolt et al., C. R. Acad. Sci. Paris, Ser. I 340 (2005).

Résumé


Version française abrégée

Le principe de la méthode POT est d’estimer la distribution des excès au-delà d’un seuil \( \mu \) par une loi GPD dépendant de deux paramètres \((\gamma, \sigma)\) après estimation de ces derniers à partir de la loi des excès au-delà de \( \mu \).
Les estimateurs que nous utilisons dans cette note sont basés sur la méthode des moments pondérés généralisés (GPWM). Nous étudions leurs propriétés asymptotiques ainsi que l’erreur, appelée biais fonctionnel, commise en remplaçant la loi des excès au-delà de \( u \) par la GPD avec les paramètres estimés. Notre objectif est double : d’une part, développer cette méthode GPWM dans le cas des excès, car les résultats existant à ce jour dans la littérature concernent uniquement les échantillons issus d’une loi GPD \((\gamma, \sigma)\) et ne sont valables au niveau de la normalité asymptotique que si \( \gamma \in ]-1, 1/2[ \); d’autre part, étendre le domaine de validité à \( \gamma \in ]-1, 3/2[ \) de façon à englober la majorité des applications pratiques.

1. Introduction

A fundamental theorem of extreme value theory states that, if the maximum of a sample \( X_1, \ldots, X_n \) of independent and identically distributed random variables from a distribution function \( F \) properly normalized converges to some limiting distribution function \( H \), then \( H \) depends on a single parameter \( \gamma \) and is one of the three extreme value distributions, namely, Frechet \((\gamma > 0)\), Gumbel \((\gamma = 0)\) and Weibull \((\gamma < 0)\). In such a case, we say that \( F \) is in the maximum domain of attraction of \( H_{\gamma} \), denoted by \( F \in \text{MDA}(H_{\gamma}) \).

The POT (Peaks-Over-Threshold) approach for modelling the tail of a distribution has received considerable attention since it has been shown that the generalized Pareto distribution arises as the limiting distribution of peaks \((\text{or excesses})\) \( X - u \) of a random variable \( X \) over a high threshold \( u \) (Pickands, [3]). In the POT method a generalized Pareto distribution (GPD) is fitted to excesses \( Y_1, \ldots, Y_{N_n} \) \((Y_j = X_{ij} - u > 0, j = 1, \ldots, N_n, \) where \( N_n \) denotes the number of excesses) over a high threshold \( u \). The method is based on the limit law for excess distribution (Balkema and de Haan, [1]; Pickands, [3]).

Let \( F \) be a distribution with a right endpoint \( x_+ \in (0, \infty] \) and excess distribution \( F_u(x) := P(X - u \leq x | X > u) \) for \( 0 < u < x_+ \) and \( 0 < x < x_+ - u \), then

\[
F \in \text{MDA}(H_{\gamma}) \quad \text{iff} \quad \exists (\sigma(u)) > 0: \lim_{u \to x_+} \sup_{0 < x < x_+ - u} \left| F_u(x) - G_{\gamma, \sigma(u)}(x) \right| = 0,
\]

where

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

denotes the distribution function of the GPD\((\gamma, \sigma)\).

The principle of the POT method is to estimate \( F_{u_n} \) by a GPD after estimating the parameters using the excess distribution over a high threshold \( u_n \). These estimators can be expressed using the empirical distribution function of the excesses, defined as

\[
\hat{\mathbb{H}}_{n, u_n}(x) = \frac{1}{N_n} \sum_{j=1}^{N_n} \mathbb{1}_{[Y_j \leq x]}.
\]

The estimators of \( \gamma \) and \( \sigma_n = \sigma(u_n) \) that we use will be based on a generalized probability weighted moment (GPWM) method introduced in Section 2. They will be denoted by \( \hat{\gamma}_{\omega_1, \omega_2, n} \) and \( \hat{\sigma}_{\omega_1, \omega_2, n} \). First, we establish their asymptotic properties. Then, we study the error due to the approximation of \( F_{u_n} \) (unknown) by \( G_{\hat{\gamma}_{\omega_1, \omega_2, n}, \hat{\sigma}_{\omega_1, \omega_2, n}} \). This error, called functional bias, and the limit distribution depend on the functions \( \omega_1 \) and \( \omega_2 \) that we use in the construction of our estimators. The choice of such functions is discussed in Rached [4]. Our aim in this Note is twofold:
Firstly, we introduce the GPWM method and use it in the case of excesses. Note that the results in the literature only concern the PWM method in the case of samples from a GPD (cf. e.g. Hosking and Wallis, [2]) and the asymptotic normality is only valid for $\gamma \in ]-1, 1/2[$.

Secondly, we extend the domain of validity: $\gamma \in ]-1, 3/2]$, which is too restrictive for many applications (as in insurance, ...). All our results are valid for all $\gamma \in ]-1, 3/2[$.

2. Generalized probability weighted moment estimators

Let $\omega$ be a continuous function, null at 0, and which admits a right derivative at 0. The GPWM of a GPD $(\gamma, \sigma)$ with $\gamma < 2$ is defined as

$$v_\omega = E[X \omega(1 - G_{\gamma, \sigma}(X))].$$

Moreover, if we denote by $W$ the primitive of $\omega$ null at 0, then $v_\omega$ can be rewritten as follows:

$$v_\omega = \int_0^\infty W(1 - G_{\gamma, \sigma}(x)) \, dx = \sigma \int_0^1 W(u) u^{-\gamma-1} \, du.$$  \hspace{1cm} (1)

Since $A_{n,u_n}$ is an estimator of the distribution function of the excesses, which can also be approximated by $G_{\gamma, \sigma_n}$, we propose to use the following estimator for $v_\omega$:

$$\hat{v}_{\omega,n} = \int_0^\infty W(1 - A_{n,u_n}(x)) \, dx.$$  \hspace{1cm} (2)

Under suitable assumptions, it is easy to check that there exists a $C^1$-diffeomorphism $T_{(\omega_1, \omega_2)}$ which transforms the GPWMs $(v_{\omega_1}, v_{\omega_2})$ of a GPD $(\gamma, \sigma)$ into $(\gamma, \sigma)$. This is specified in the following proposition.

Proposition 2.1. Let $\omega_1$ and $\omega_2$ be two continuous functions, null at 0 and which admit a right derivative at 0. Consider two GPWMs from a GPD $(\gamma, \sigma)$ defined as in (1). We introduce the following functions $\phi_1, \phi_2$ and $\rho$ for all $\gamma < 2$:

$$\phi_i(\gamma) = \int_0^1 W_i(u) u^{-\gamma-1} \, du \quad \text{for } i \in \{1, 2\} \quad \text{and} \quad \rho(\gamma) = \frac{\phi_1(\gamma)}{\phi_2(\gamma)}.$$  \hspace{1cm} (3)

Then

(i) the functions $\phi_1$ and $\phi_2$ are $C^1$ on $]-\infty, 2[$;

(ii) if, for all $\gamma < 2$, the function $\rho'(\gamma)$ has a constant sign, then there exists a $C^1$-diffeomorphism $T_{(\omega_1, \omega_2)}$ such that:

$$T_{(\omega_1, \omega_2)}(v_{\omega_1}, v_{\omega_2}) = (\gamma, \sigma),$$

where $\gamma$ and $\sigma$ are given respectively by

$$\gamma = \rho^{-1} \left( \frac{v_{\omega_1}}{v_{\omega_2}} \right) \quad \text{and} \quad \sigma = \frac{v_{\omega_2}}{\phi_2 \circ \rho^{-1}(v_{\omega_1}/v_{\omega_2})}.$$  \hspace{1cm} (3)

The Jacobian matrix associated to the diffeomorphism $T_{(\omega_1, \omega_2)}$ is then given by

$$DT_{(\omega_1, \omega_2)}(v_{\omega_1}, v_{\omega_2}) = \frac{1}{\sigma [\phi'_1(\gamma)\phi'_2(\gamma) - \phi'_2(\gamma)\phi'_1(\gamma)]} \begin{pmatrix} \phi_2(\gamma) & -\phi_1(\gamma) \\ -\sigma \phi'_2(\gamma) & \sigma \phi'_1(\gamma) \end{pmatrix}.$$
From (3), it follows that
\[
\left( \hat{\gamma}_{\omega_1, \omega_2, n}, \hat{\sigma}_{\omega_1, \omega_2, n}, \frac{\hat{v}_{\omega_1, n}}{\sigma_n}, \frac{\hat{v}_{\omega_2, n}}{\sigma_n} \right) = T(\omega_1, \omega_2) \left( \hat{\nu}_{\omega_1, n}, \hat{\nu}_{\omega_2, n}, \frac{\nu_{\omega_1, n}}{\sigma_n}, \frac{\nu_{\omega_2, n}}{\sigma_n} \right).
\]

In all the sequel, we denote by \( v_{\omega_1} \) and \( v_{\omega_2} \) the GPWMs of a GPD \((\gamma, 1)\) and by \( A(\omega_1, \omega_2) \) the corresponding Jacobian matrix associated to the \( C^1 \)-diffeomorphism \( T(\omega_1, \omega_2) \) at the point \((v_{\omega_1}^1, v_{\omega_2}^1)\).

We will study the asymptotic behaviour of \( \hat{\gamma}_{\omega_1, \omega_2, n} \) and \( \hat{\sigma}_{\omega_1, \omega_2, n} \) and more specifically the difference \( \overline{\Fun(\sigma_n x)} - \overline{G(\hat{\gamma}_{\omega_1, \omega_2, n}, \hat{\sigma}_{\omega_1, \omega_2, n})} \).

In what follows, we suppose that \( F \) is twice differentiable and that its inverse \( F^{-1} \) exists. Let \( V \) and \( A \) be two functions defined as
\[
V(t) = \overline{F^{-1}(e^{-t})} \quad \text{and} \quad A(t) = \frac{V''(\ln t)}{V'(\ln t)} - \gamma.
\]

We assume the following first and second order conditions:
\[
\lim_{t \to +\infty} A(t) = 0, \quad \text{(4)}
\]
and
\[
A \text{ is of constant sign at } +\infty \text{ and there exists } \rho \leq 0 \text{ such that } |A| \in RV_{\rho}. \quad \text{(5)}
\]

Under these assumptions, it is proved in Worms [5] (Theorem 1.4, page 43) that as \( \sigma_n \to x_+ \)
\[
\overline{F_{\omega_1}(\sigma_n x)} - \overline{G_{\gamma, 1}(x)} = a_n D_{\gamma, \rho}(x) + o(a_n), \quad \text{as } n \to +\infty, \quad \text{(6)}
\]
for all \( x \), when
\[
\sigma_n := \sigma(u_n) = V'(V^{-1}(u_n)), \quad a_n := A(e^{V^{-1}(u_n)})
\]
and
\[
D_{\gamma, \rho}(x) := \begin{cases} C_{0, \rho}(x), & \text{if } \gamma = 0, \\ C_{\gamma, \rho}(\frac{1}{\gamma} \ln(1 + \gamma x)), & \text{if } \gamma \neq 0, \end{cases}
\]
where
\[
C_{\gamma, \rho}(x) := e^{-(1+\gamma)x} I_{\gamma, \rho}(x) \quad \text{and} \quad I_{\gamma, \rho}(x) := \int_{0}^{x} e^{\rho u} \int_{0}^{u} e^{\rho s} \, ds \, du.
\]

First, we establish the asymptotic behaviour of \( (\hat{\nu}_{\omega_1, n}/\sigma_n - v_{\omega_1}^1, \hat{\nu}_{\omega_2, n}/\sigma_n - v_{\omega_2}^1) \).

**Theorem 2.2.** Under assumptions (4) and (5) with \( \gamma \in ]-1, \frac{3}{2}[ \) for all \( C^1 \)-functions \( \omega_1 \) and \( \omega_2 \), null at 0, we have, conditionally on \( N_n = k_n \), for all sequences \( k_n \to \infty \) such that \( \sqrt{k_n} a_n \to \lambda \in \mathbb{R} \)
\[
\sqrt{k_n} \left( \frac{\hat{\nu}_{\omega_1, n}}{\sigma_n} - v_{\omega_1}^1, \frac{\hat{\nu}_{\omega_2, n}}{\sigma_n} - v_{\omega_2}^1 \right) \overset{d}{\to} N(\lambda C, \Gamma),
\]
with
\[
C = \begin{pmatrix} \int_{0}^{\infty} W_1(e^{-u}) e^{\rho u} \int_{0}^{u} e^{\rho s} \, ds \, du & \int_{0}^{\infty} W_2(e^{-u}) e^{\rho u} \int_{0}^{u} e^{\rho s} \, ds \, du \\ \phi_{1}(y+\rho)-\phi_{1}(y) & \phi_{2}(y+\rho)-\phi_{2}(y) \end{pmatrix},
\]
and \( \Gamma \) the variance–covariance matrix of a couple \((Y_{\omega_1}, Y_{\omega_2})\) defined as
\begin{align*}
Y_{\omega_1} &= \int_0^1 t^{-\gamma-1} \omega_1(t) \mathbb{B}(t) \, dt \quad \text{and} \quad Y_{\omega_2} = \int_0^1 t^{-\gamma-1} \omega_2(t) \mathbb{B}(t) \, dt,
\end{align*}

where \( \mathbb{B} \) is a Brownian bridge on \([0, 1]\).

**Proof of Theorem 2.2.** Let \( \alpha_{k_n} \) denote the uniform empirical process based on \( k_n \) random variables. Remark that

\[
\frac{\tilde{V}_{\omega_1,n}}{\sigma_n} - \nu_{\omega_1}^1 = \int_0^\infty \left[ W_1(1 - F_{\alpha_n}(\sigma_n x)) - W_1(1 - G_{\gamma,1}(x)) \right] \, dx - \frac{1}{\sqrt{k_n}} \int_0^\infty \alpha_{k_n} \circ F_{\alpha_n}(\sigma_n x) \omega_1(1 - F_{\alpha_n}(\sigma_n x)) \, dx
\]

\[
+ \frac{1}{\sqrt{k_n}} \int_0^1 (1 - t) \frac{1}{\sqrt{k_n}} (\alpha_{k_n} \circ F_{\alpha_n})^2(\sigma_n x) \omega_1(1 - F_{\alpha_n}(\sigma_n x)) \, dt \, dx
\]

\[
= T_{1,n} - \frac{1}{\sqrt{k_n}} T_{2,n} + \frac{1}{\sqrt{k_n}} T_{3,n}.
\]

Using empirical process arguments, the proof follows from the fact that, for all \( \gamma \in \left(-\frac{1}{2}, 2\right] \), we have (see Rached, [4]):

(i) \( T_{1,n} \to \int_0^\infty W_1(e^{-u}) e^{\nu_{\omega_1}^1} \int_0^u e^{\nu_{\omega_1}^1} \, dt \, du \); (ii) \( T_{2,n} \to \int_0^1 t^{-\gamma-1} \omega_1(t) \mathbb{B}(t) \, dt \); (iii) \( T_{3,n} \to P 0 \).

From this theorem, we can now deduce the asymptotic normality of our estimators \( (\hat{\gamma}_{\omega_1,0,2,n}, \hat{\sigma}_{\omega_1,0,2,n}/\sigma_n) \).

**Corollary 2.3.** Under the same assumptions as in Theorem 2.2, and assuming the existence of the diffeomorphism \( T_{(\omega_1,0,2)} \), we have, conditionally on \( N_n = k_n \), for all sequences \( k_n \to \infty \) such that \( \sqrt{k_n} a_n \to \lambda \in \mathbb{R} \)

\[
\sqrt{k_n} \left( \frac{\tilde{V}_{\omega_1,0,2,n}}{\sigma_{\omega_1,0,2,n}} - \nu_{\omega_1}^1 \right) \to N(\lambda C', \Sigma),
\]

where \( A_{(\omega_1,0,2)} = DT_{(\omega_1,0,2)}(\nu_{\omega_1,0,2}) \), \( \Sigma = A_{(\omega_1,0,2)} \Gamma A_{(\omega_1,0,2)}' \) and \( C' = A_{(\omega_1,0,2)} C \) defined as

\[
\begin{pmatrix}
C_1'(\gamma, \rho) \\
C_2'(\gamma, \rho)
\end{pmatrix} =
\begin{pmatrix}
\phi_1(\gamma) - \phi_1(\gamma + \rho) + \phi_1(\gamma) \phi_2(\gamma + \rho) \\
\rho \phi_1(\gamma) \phi_2(\gamma) - \phi_1(\gamma) \phi_2(\gamma)
\end{pmatrix}
\begin{pmatrix}
\phi_1(\gamma) - \phi_1(\gamma + \rho) + \phi_1(\gamma) \phi_2(\gamma + \rho) \\
\rho \phi_1(\gamma) \phi_2(\gamma) - \phi_1(\gamma) \phi_2(\gamma)
\end{pmatrix}
\begin{pmatrix}
\phi_1(\gamma) - \phi_1(\gamma + \rho) + \phi_1(\gamma) \phi_2(\gamma + \rho) \\
\rho \phi_1(\gamma) \phi_2(\gamma) - \phi_1(\gamma) \phi_2(\gamma)
\end{pmatrix}
\begin{pmatrix}
0 \\
1
\end{pmatrix}
\]

if \( \rho \neq 0 \).

if \( \rho = 0 \).

The proof of Corollary 2.3 is straightforward by a Taylor expansion.

After estimating the parameters of the GPD, the idea is to approximate \( F_{\alpha_n}(\sigma_n x) \) (unknown) by \( \tilde{G}_{\omega_1,0,2,n}(\sigma_{\omega_1,0,2,n}/\sigma_n(x)) \). However, it is clear that this approximation induces a bias, called functional bias, which is studied in the following theorem.

**Theorem 2.4.** Under the same assumptions as in Corollary 2.3, we have, conditionally on \( N_n = k_n \), for all sequences \( k_n \to \infty \) such that \( \sqrt{k_n} a_n \to \lambda \in \mathbb{R} \),
(i) the process
\[ \sqrt{n} \left( \bar{F}_{\text{as}}(\sigma_n x) - \tilde{G}_{\gamma_1,\omega_2,n,\sigma_1,\omega_2,n/\sigma_n}(x) \right) \]
converges in distribution to
\[ \lambda B(x) + Z(x), \]
where
\[ B(x) = D_{\gamma,\rho}(x) + C'_1(\gamma,\rho) \frac{\partial G_{\gamma,1}}{\partial x}(x) - C'_2(\gamma,\rho)x \frac{\partial G_{\gamma,1}}{\partial x}(x), \]
and \( Z(x) \) is a centered Gaussian process defined as
\[ Z(x) = \frac{\partial G_{\gamma,1}}{\partial x}(x) \int_0^1 t^{1(1)}(t)t^{-\gamma-1}B(t)\,dt - \lambda \frac{\partial G_{\gamma,1}}{\partial x}(x) \int_0^1 t^{2(1)}(t)t^{-\gamma-1}B(t)\,dt, \]
with
\[ t^{1(1)}(t) = \frac{\phi_2(\gamma)\omega_1(t) - \phi_1(\gamma)\omega_2(t)}{\phi'_1(\gamma)\phi_2(\gamma) - \phi_1(\gamma)\phi'_2(\gamma)} \]
and
\[ t^{2(1)}(t) = \frac{\phi'_1(\gamma)\omega_2(t) - \phi'_2(\gamma)\omega_1(t)}{\phi'_1(\gamma)\phi_2(\gamma) - \phi_1(\gamma)\phi'_2(\gamma)}. \]
(ii) if \( \rho = 0 \), the functional bias \( B(x) \) in (7) is equal to 0.

The proof of Theorem 2.4 follows from (6) combined with Corollary 2.3 and a Taylor expansion (see Rached, [4]).

Let us give some comments concerning this theorem:

- the error due to the fact that \( \bar{F}_{\text{as}}(\sigma_n x) \) is replaced by \( \tilde{G}_{\gamma_1,\omega_2,n,\sigma_1,\omega_2,n/\sigma_n}(x) \) is of smaller order when \( \rho = 0 \) than in the case \( \rho \neq 0 \);
- the behaviour of the functional bias \( B(x) \) in the case \( \rho = 0 \) is surprising since when estimating the index \( \gamma \), the bias increases when \( \rho \) tends to 0;
- the second order parameter \( \rho \) is zero for many usual distributions in the Gumbel domain of attraction (\( \gamma = 0 \)): e.g. the normal, lognormal, gamma and classical Weibull distributions. In the Frechet domain of attraction (\( \gamma > 0 \)), \( \rho = 0 \) for the loggamma distribution; hence, our result applies to all these distributions;
- this result is closely linked to penultimate approximation established in Worms [6].

References