

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



DISCUSSION
PAPER

2014/26

On the asymptotic distribution of the mean
absolute deviation about the mean

J. SEGERS

On the asymptotic distribution of the mean absolute deviation about the mean

Johan Segers

Université catholique de Louvain
 Institut de Statistique, Biostatistique et Sciences Actuarielles
 Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
 johan.segers@uclouvain.be

June 16, 2014

Abstract

The mean absolute deviation about the mean is an alternative to the standard deviation for measuring dispersion in a sample or in a population. For stationary, ergodic time series with a finite first moment, an asymptotic expansion for the sample mean absolute deviation is proposed. The expansion yields the asymptotic distribution of the sample mean absolute deviation under a wide range of settings, allowing for serial dependence or an infinite second moment.

Key words: central limit theorem; dispersion; ergodicity; regular variation; stable distribution; strong mixing.

1 Introduction

The mean absolute deviation of a sample X_1, \dots, X_n about the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ is given by the statistic

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}_n|. \quad (1.1)$$

If the random variables X_i have common distribution function F with finite mean $\mu = \int_{\mathbb{R}} x \, dF(x)$, then $\hat{\theta}_n$ is an estimator of the mean absolute deviation

$$\theta = \mathbb{E}[|X_1 - \mu|] = \int_{\mathbb{R}} |x - \mu| \, dF(x). \quad (1.2)$$

The (sample) mean absolute deviation is an alternative to the standard deviation for measuring dispersion. Its advantages and drawbacks have been widely discussed in the literature. The standard deviation is motivated mainly from optimality results in the context of independent random

sampling from the normal distribution, an analysis dating back to Fisher and even Laplace (Stigler, 1973). However, the mean absolute deviation may be more appropriate in case of departures from normality or in the presence of outliers (Tukey, 1960; Huber, 1996). It may also offer certain pedagogical advantages. For extensive discussions and comparisons, see Pham-Gia and Hung (2001) and Gorard (2005) and the references therein.

Because of the presence of the absolute value function, finding the asymptotic distribution of the sample mean absolute deviation is surprisingly challenging. In Pollard (1989, Section 2) and van der Vaart (1998, Example 19.25), the exercise is put forward as a showcase for the power of empirical process theory. In a nutshell, their analysis is as follows. Let

$$\tilde{\theta}_n = \frac{1}{n} \sum_{i=1}^n |X_i - \mu|$$

be the version of the sample mean absolute deviation that could be computed if the true mean, μ , were known. Consider the dispersion function

$$D_F(u) = \int_{\mathbb{R}} |x - u| dF(x), \quad u \in \mathbb{R}.$$

Clearly, $\theta = D_F(\mu)$. Under independent random sampling and under the presence of finite second moments, asymptotic uniform equicontinuity of the empirical process $u \mapsto n^{-1/2} \sum_{i=1}^n (|X_i - u| - D_F(u))$ implies that

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n}(\hat{\theta}_n - D_F(\bar{X}_n)) + \sqrt{n}(D_F(\bar{X}_n) - \theta) \\ &= \sqrt{n}(\tilde{\theta}_n - \theta) + \sqrt{n}(D_F(\bar{X}_n) - \theta) + o_p(1), \quad n \rightarrow \infty. \end{aligned} \quad (1.3)$$

If F is continuous at u , then D_F is differentiable at u with derivative $D'_F(u) = 2F(u) - 1$ (Muoz-Perez and Sanchez-Gomez, 1990, Theorem 1). By the delta method, it follows from (1.3) that, if F is continuous at μ ,

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(\tilde{\theta}_n - \theta) + (2F(\mu) - 1) \sqrt{n}(\bar{X}_n - \mu) + o_p(1) \quad (1.4)$$

$$\rightsquigarrow N(0, \sigma_{\hat{\theta}}^2), \quad n \rightarrow \infty, \quad (1.5)$$

the arrow ‘ \rightsquigarrow ’ denoting weak convergence. The asymptotic variance equals

$$\sigma_{\hat{\theta}}^2 = \text{var}(|X_1 - \mu| + (2F(\mu) - 1) X_1).$$

The above proof is elegant and short. However, it rests on a body of advanced empirical process theory. Using direct arguments, Babu and Rao (1992) establish higher-order expansions for $\sqrt{n}(\hat{\theta}_n - \theta)$ under the additional assumption that F is Hölder continuous or even differentiable at μ .

The results described so far are limited to independent random sampling from a distribution F with a finite second moment, and, except for (1.3),

without an atom at its mean μ . In the literature, no results seem to be available on the asymptotic distribution of the sample mean absolute deviation in the case of serial dependence or when the second moment does not exist. Even in the case of independent random sampling from a distribution with finite second moment, the asymptotic distribution of (1.3) in case F has an atom at μ seems not to have been described yet.

The aim of this paper is to derive an asymptotic expansion of $\hat{\theta}_n - \theta$ from first principles and under minimal assumptions (Section 2). The expansion yields the asymptotic distribution of the sample mean absolute deviation under a wide range of settings, including serial dependence, in the infinite-variance case, and without smoothness assumptions (Section 3). Even in the case of independent random sampling from a distribution with finite second moment, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ is found to be non-Gaussian in case F possesses an atom at μ .

2 Asymptotic expansion

Stationarity of the time series $(X_i)_{i \geq 1}$ means that for all positive integers $k \geq 1$ and $h \geq 0$, the distribution of the random vector $(X_{1+h}, \dots, X_{k+h})$ does not depend on h . By the Birkhoff ergodic theorem (Kallenberg, 2002, Theorem 10.6), stationarity and ergodicity imply that, for every Borel measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $\mathbb{E}[|f(X_1)|] < \infty$, we have

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \rightarrow \mathbb{E}[f(X_1)], \quad n \rightarrow \infty, \text{ almost surely.} \quad (2.1)$$

Assume that the stationary distribution F has finite mean, μ , and recall the mean absolute deviation θ in (1.2) and its sample version $\hat{\theta}_n$ in (1.1). Write

$$\hat{\theta}_n - \theta = \frac{1}{n} \sum_{i=1}^n (|X_i - \bar{X}_n| - |X_i - \mu|) + \frac{1}{n} \sum_{i=1}^n (|X_i - \mu| - \theta) \quad (2.2)$$

The second term on the right-hand side is just a sum of centered random variables. It is the first term which poses a challenge.

Lemma 2.1. *Let X_1, X_2, \dots be a stationary, ergodic time series with finite mean $\mu = \mathbb{E}[X_1]$. We have, as $n \rightarrow \infty$, almost surely,*

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (|X_i - \bar{X}_n| - |X_i - \mu|) &= (\bar{X}_n - \mu) (\mathbb{P}[X_1 < \mu] - \mathbb{P}[X_1 > \mu]) \\ &\quad + |\bar{X}_n - \mu| \mathbb{P}[X_1 = \mu] + o(|\bar{X}_n - \mu|). \end{aligned}$$

Proof. Consider the random variables

$$A_n = \min(\bar{X}_n, \mu), \quad B_n = \max(\bar{X}_n, \mu).$$

Let $\text{sign}(z)$ be equal to 1, 0, or -1 according to whether z is larger than, equal to, or smaller than zero, respectively. For $x \in \mathbb{R} \setminus (A_n, B_n)$, that is, for x not between μ and \bar{X}_n , a case-by-case analysis reveals that

$$|x - \bar{X}_n| - |x - \mu| = (\bar{X}_n - \mu) \text{sign}(\mu - x) + |\bar{X}_n - \mu| \mathbb{1}_{\{\mu\}}(x).$$

Consider the partition $\{1, \dots, n\} = \mathcal{K}_n \cup \mathcal{L}_n$, where

$$\begin{aligned} \mathcal{K}_n &= \{i = 1, \dots, n : A_n < X_i < B_n\}, \\ \mathcal{L}_n &= \{1, \dots, n\} \setminus \mathcal{K}_n. \end{aligned}$$

By convention, the sum over the empty set is zero. We find that

$$\begin{aligned} & \sum_{i=1}^n (|X_i - \bar{X}_n| - |X_i - \mu|) \\ &= (\bar{X}_n - \mu) \sum_{i \in \mathcal{L}_n} \text{sign}(\mu - X_i) + |\bar{X}_n - \mu| \sum_{i \in \mathcal{L}_n} \mathbb{1}_{\{\mu\}}(X_i) \\ & \quad + \sum_{i \in \mathcal{K}_n} (|X_i - \bar{X}_n| - |X_i - \mu|) \\ &= (\bar{X}_n - \mu) \sum_{i=1}^n \text{sign}(\mu - X_i) + |\bar{X}_n - \mu| \sum_{i=1}^n \mathbb{1}_{\{\mu\}}(X_i) + R_n \end{aligned}$$

with

$$\begin{aligned} R_n &= \sum_{i \in \mathcal{K}_n} (|X_i - \bar{X}_n| - |X_i - \mu|) \\ & \quad - (\bar{X}_n - \mu) \sum_{i \in \mathcal{K}_n} \text{sign}(\mu - X_i) - |\bar{X}_n - \mu| \sum_{i \in \mathcal{K}_n} \mathbb{1}_{\{\mu\}}(X_i) \end{aligned}$$

By (2.1), we have, as $n \rightarrow \infty$, almost surely,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \text{sign}(\mu - X_i) &\rightarrow \mathbb{E}[\text{sign}(\mu - X_1)] = \mathbb{P}[X_1 < \mu] - \mathbb{P}[X_1 > \mu], \\ \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\mu\}}(X_i) &\rightarrow \mathbb{P}[X_1 = \mu]. \end{aligned}$$

As a consequence, as $n \rightarrow \infty$ and almost surely,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (|X_i - \bar{X}_n| - |X_i - \mu|) &= (\bar{X}_n - \mu) (\mathbb{P}[X_1 < \mu] - \mathbb{P}[X_1 > \mu] + o(1)) \\ & \quad + |\bar{X}_n - \mu| (\mathbb{P}[X_1 = \mu] + o(1)) + \frac{R_n}{n}. \end{aligned}$$

Further, as $||x - \bar{X}_n| - |x - \mu|| \leq |\bar{X}_n - \mu|$ for all $x \in \mathbb{R}$, we obtain

$$|R_n| \leq 3 |\bar{X}_n - \mu| |\mathcal{K}_n|,$$

where $|\mathcal{K}_n|$ is the number of elements of \mathcal{K}_n . The lemma therefore follows if we can prove that

$$\frac{|\mathcal{K}_n|}{n} \rightarrow 0, \quad n \rightarrow \infty, \quad \text{a.s.} \quad (2.3)$$

For $x \in \mathbb{R}$, put

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x), \quad \hat{F}_n(x-) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i < x).$$

We have

$$\frac{|\mathcal{K}_n|}{n} = \hat{F}_n(B_n-) - \hat{F}_n(A_n).$$

By (2.1), \bar{X}_n and thus A_n and B_n converge to μ almost surely. Moreover, we have either $A_n = \mu \leq B_n$ or $A_n \leq \mu = B_n$ (or both, if $\bar{X}_n = \mu$). As a consequence,

$$\frac{|\mathcal{K}_n|}{n} \leq \max\{\hat{F}_n(B_n-) - \hat{F}_n(\mu), \hat{F}_n(\mu-) - \hat{F}_n(A_n)\}.$$

Fix $\delta > 0$. By monotonicity of \hat{F}_n and by ergodicity (2.1), we have, almost surely,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \hat{F}_n(B_n-) &\leq \limsup_{n \rightarrow \infty} \hat{F}_n(\mu + \delta) = F(\mu + \delta), \\ \liminf_{n \rightarrow \infty} \hat{F}_n(A_n) &\geq \liminf_{n \rightarrow \infty} \hat{F}_n(\mu - \delta) = F(\mu - \delta). \end{aligned}$$

Since moreover $\hat{F}_n(\mu) \rightarrow F(\mu)$ and $\hat{F}_n(\mu-) \rightarrow F(\mu-) = \mathbb{P}[X_1 < \mu]$ almost surely, it follows that

$$\begin{aligned} \limsup_{n \rightarrow \infty} (\hat{F}_n(B_n-) - \hat{F}_n(\mu)) &\leq F(\mu + \delta) - F(\mu), \\ \limsup_{n \rightarrow \infty} (\hat{F}_n(\mu-) - \hat{F}_n(A_n)) &\leq F(\mu-) - F(\mu - \delta), \end{aligned}$$

almost surely, and therefore

$$\limsup_{n \rightarrow \infty} \frac{|\mathcal{K}_n|}{n} \leq \max(F(\mu + \delta) - F(\mu), F(\mu-) - F(\mu - \delta))$$

almost surely. Since δ was arbitrary, we obtain (2.3), as required. \square

3 Weak convergence

Combining the expansions in (2.2) and Lemma 2.1, the limit distribution of $\hat{\theta}_n$ follows right away. For the sake of illustration, we consider two settings: strongly mixing time series with finite second moments (Subsection 3.1) and independent random sampling from distributions in the domain of attraction of a stable law with index $\alpha \in (1, 2)$ (Subsection 3.2).

3.1 Strongly mixing time series, finite variance

Let $(X_i)_{i \in \mathbb{Z}}$ be a stationary time series. For $k \in \mathbb{Z}$, consider the σ -fields $\mathcal{F}_k = \sigma(X_i : i \leq k)$ and $\mathcal{G}_k = \sigma(X_i : i \geq k)$. Rosenblatt's mixing coefficients are defined by

$$\alpha_n = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_0, B \in \mathcal{G}_n\}$$

for integer $n \geq 0$. The time series $(X_i)_{i \in \mathbb{Z}}$ is called strongly mixing if $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. Strong mixing implies ergodicity.

Define $\alpha(t) = \alpha_{\lfloor t \rfloor}$, $t \geq 0$, and its inverse function $\alpha^{-1}(u) = \inf\{t \geq 0 : \alpha(t) \leq u\}$ for $u > 0$. Let Q denote the quantile function of the distribution of X_i .

Proposition 3.1. *Let $(X_i)_{i \in \mathbb{Z}}$ be a strongly mixing, stationary time series with finite second moments. If*

$$\int_0^1 \alpha^{-1}(u) \{Q(u)\}^2 du < \infty, \quad (3.1)$$

then, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow Y (\mathbb{P}[X_1 < \mu] - \mathbb{P}[X_1 > \mu]) + |Y| \mathbb{P}[X_1 = \mu] + Z, \quad (3.2)$$

where (Y, Z) is bivariate normal with mean zero and covariance matrix given by

$$\begin{aligned} \text{var}(Y) &= \text{var}(X_0) + 2 \sum_{i=1}^{\infty} \text{cov}(X_i, X_0), \\ \text{var}(Z) &= \text{var}(|X_0 - \mu|) + 2 \sum_{i=1}^{\infty} \text{cov}(|X_i - \mu|, |X_0 - \mu|), \\ \text{cov}(Y, Z) &= \sum_{i \in \mathbb{Z}} \text{cov}(|X_i - \mu|, X_0) = \sum_{i \in \mathbb{Z}} \text{cov}(X_i, |X_0 - \mu|). \end{aligned}$$

all series being absolutely convergent.

Proof. By Theorem 1 in Doukhan et al. (1994) and the remark just after that theorem, we have, as $n \rightarrow \infty$,

$$(Y_n, Z_n) := \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu), \frac{1}{\sqrt{n}} \sum_{i=1}^n (|X_i - \mu| - \theta) \right) \rightsquigarrow (Y, Z), \quad (3.3)$$

with (Y, Z) as in the statement of the proposition. Combining (2.2) and Lemma 2.1 yields the expansion

$$\begin{aligned} & \sqrt{n}(\hat{\theta}_n - \theta) \\ &= Y_n (\mathbb{P}[X_1 < \mu] - \mathbb{P}[X_1 > \mu]) + |Y_n| \mathbb{P}[X_1 = \mu] + o(|Y_n|) + Z_n, \end{aligned} \quad (3.4)$$

as $n \rightarrow \infty$, almost surely. Apply the continuous mapping theorem and Slutsky's lemma to arrive at the result. \square

Condition 3.1 covers many cases and is almost sharp; see the applications on pages 67–68 in Doukhan et al. (1994) as well as their Theorem 2. The case of independent and identically distributed variables is trivially included.

Corollary 3.2. *Let X_1, X_2, \dots be independent and identically distributed random variables with common distribution F . If F has a finite second moment, then weak convergence (3.2) holds, where (Y, Z) is bivariate normal with mean zero and covariance matrix equal to the one of $(X_1 - \mu, |X_1 - \mu|)$.*

Proof. Combine (3.3) and (3.4) with the multivariate central limit theorem, the continuous mapping theorem, and Slutsky's lemma. \square

If $\mathbb{P}[X_1 = \mu] = 0$, i.e., if F is continuous at 0, then the expansion (3.4) is the same as the one in (1.4) and we obtain that $\sqrt{n}(\hat{\theta}_n - \theta)$ is asymptotically normal. In the independence case, the limit distribution coincides with the one in (1.5). However, if $0 < \mathbb{P}[X_1 = \mu] < 1$, then the weak limit in (3.2) is not Gaussian. Moreover, the limit distribution is not centered either, its expectation being $\mathbb{P}[X_1 = \mu] \mathbb{E}[|Y|]$.

3.2 Independent random sampling, infinite variance

One argument in favour of the use of the mean absolute deviation for measuring dispersion is that, unlike the standard deviation, it does not require existence of second moments. Still, in the weak convergence statements (1.5) and (3.2), finite second moments are presupposed. This condition is lifted in the next result. A positive, Borel measurable function L defined on a neighbourhood of infinity is *slowly varying* if $\lim_{x \rightarrow \infty} L(xy)/L(x) = 1$ for all $y > 0$.

Proposition 3.3. *Let X_1, X_2, \dots be independent, identically distributed random variables with common distribution function F . Assume that there*

exist $\alpha \in (1, 2)$, $p \in [0, 1]$, and a slowly varying function L on $(0, \infty)$, such that, for $x > 0$,

$$\left. \begin{aligned} \mathbb{P}[X_1 > x] &= p x^{-\alpha} L(x), \\ \mathbb{P}[X_1 < -x] &= (1-p) x^{-\alpha} L(x). \end{aligned} \right\} \quad (3.5)$$

Then F has a finite mean but an infinite second moment. Defining

$$a_n = \inf\{x > 0 : \mathbb{P}[|X_1| > x] \leq 1/n\},$$

we have, assuming that F is continuous at μ ,

$$\frac{n}{a_n}(\hat{\theta}_n - \theta) \rightsquigarrow G, \quad n \rightarrow \infty,$$

where G is a stable distribution with characteristic function

$$\int_{\mathbb{R}} e^{-isx} dG(x) = \exp\{-\sigma^\alpha |s|^\alpha (1 - i \operatorname{sign}(s) \tan(\alpha\pi/2))\}, \quad s \in \mathbb{R}, \quad (3.6)$$

the scale parameter $\sigma > 0$ being given by

$$\sigma^\alpha = \frac{2^\alpha \{\mathbb{P}[X_1 < \mu]^\alpha p + \mathbb{P}[X_1 > \mu]^\alpha (1-p)\}}{\frac{\Gamma(2-\alpha)}{\alpha-1} |\cos(\alpha\pi/2)|}. \quad (3.7)$$

Proof. The statement about the existence of the moments is well known; see for instance the unnumbered lemma on page 578 in Section XVII.5 in Feller (1971). Since $\mathbb{P}[X_1 = \mu] = 0$, equation (2.2) and Lemma 2.1 imply

$$\hat{\theta}_n - \theta = \frac{1}{n} \sum_{i=1}^n (\xi_i - \theta) + o(|\bar{X}_n - \mu|), \quad n \rightarrow \infty, \text{ a.s.}, \quad (3.8)$$

where, writing $b = \mathbb{P}[X_1 < \mu] - \mathbb{P}[X_1 > \mu]$,

$$\xi_i = |X_i - \mu| + b(X_i - \mu) = |X_i - \mu| (1 + b \operatorname{sign}(X_i - \mu)). \quad (3.9)$$

Note that $\theta = \mathbb{E}[\xi_i]$. Since $-1 < b < 1$, we have $\xi_i \geq 0$. Moreover, by slow variation of L , as $x \rightarrow \infty$,

$$\begin{aligned} \mathbb{P}[\xi_i > x] &= \mathbb{P}[X_i > \mu + x/(1+b)] + \mathbb{P}[X_i < \mu - x/(1-b)] \\ &= ((1+b)^\alpha p + (1-b)^\alpha (1-p) + o(1)) x^{-\alpha} L(x). \end{aligned} \quad (3.10)$$

As $\mathbb{P}[|X_i| > x] = x^{-\alpha} L(x)$, it follows that, up to a multiplicative constant, the tail function of ξ_i is asymptotically equivalent to the one of $|X_i|$. Observe that $1+b = 2\mathbb{P}[X_1 < \mu]$ and $1-b = 2\mathbb{P}[X_1 > \mu]$.

By classical theory on the domains of attraction of non-Gaussian stable distributions, equation (3.5) is equivalent to the weak convergence of

$na_n^{-1}(\bar{X}_n - \mu) = a_n^{-1} \sum_{i=1}^n (X_i - \mu)$ to an α -stable distribution (Gnedenko and Kolmogorov, 1954). The tails of $|X_i|$ and ξ_i being related through (3.10), $a_n^{-1} \sum_{i=1}^n (\xi_i - \theta)$ converges weakly to an α -stable distributions too. The limit distribution can be found for instance from Theorem 1.8.1 in Samorodnitsky and Taqqu (1994) and coincides with G in the statement of the proposition; details are given below. By (3.8), we have

$$\frac{n}{a_n}(\hat{\theta}_n - \theta) = \frac{1}{a_n} \sum_{i=1}^n (\xi_i - \theta) + o_p(1), \quad n \rightarrow \infty.$$

Weak convergence of $na_n^{-1}(\hat{\theta}_n - \theta)$ to G now follows from Slutsky's lemma.

The calculations leading to the expression of the scale parameter σ in equation (3.7) are as follows. Let \tilde{a}_n be such that

$$\lim_{n \rightarrow \infty} n \mathbb{P}[\xi_i > \tilde{a}_n] = \frac{\Gamma(2 - \alpha)}{\alpha - 1} |\cos(\alpha\pi/2)|. \quad (3.11)$$

By Theorem 1.8.1 in Samorodnitsky and Taqqu (1994), we have

$$\frac{1}{\tilde{a}_n} \sum_{i=1}^n (\xi_i - \theta) \rightsquigarrow \tilde{G}, \quad n \rightarrow \infty,$$

where \tilde{G} is an α -stable distribution whose characteristic function has the same form as in (3.6) with σ replaced by $\tilde{\sigma} = 1$. By (3.10) and the definition of a_n , we have

$$\lim_{n \rightarrow \infty} n \mathbb{P}[\xi_i > a_n] = (1 + b)^\alpha p + (1 - b)^\alpha (1 - p). \quad (3.12)$$

The function $x \mapsto \mathbb{P}[\xi_i > x]$ being regularly varying with index $-\alpha$, it follows that a valid choice for \tilde{a}_n in (3.11) is $\tilde{a}_n = \gamma a_n$, where the constant $\gamma > 0$ can be read off by comparing (3.11) and (3.12):

$$\begin{aligned} \gamma^{-\alpha} \{(1 + b)^\alpha p + (1 - b)^\alpha (1 - p)\} &= \lim_{n \rightarrow \infty} n \mathbb{P}[\xi_i > \gamma a_n] \\ &= \frac{\Gamma(2 - \alpha)}{\alpha - 1} |\cos(\alpha\pi/2)|. \end{aligned}$$

If \tilde{Z} denotes a random variable whose distribution function is \tilde{G} , then

$$\frac{1}{a_n} \sum_{i=1}^n (\xi_i - \theta) = \frac{\gamma}{\tilde{a}_n} \sum_{i=1}^n (\xi_i - \theta) \rightsquigarrow \gamma \tilde{Z} = Z, \quad n \rightarrow \infty.$$

By Samorodnitsky and Taqqu (1994, Property 1.2.3), the characteristic function of the law of Z is given by (3.6) with scale parameter $\sigma = \gamma \tilde{\sigma} = \gamma$. \square

The regular variation condition (3.5) covers distributions with power-law tails, such as the Pareto distribution, non-Gaussian stable distributions, the Student t distribution, and the Fréchet distribution.

Remark 3.4. Proposition 3.3 supposes independent random sampling from a heavy-tailed distribution. Extensions to weakly dependent stationary time series are possible too. In Bartkiewicz et al. (2011), for instance, conditions are given which guarantee the weak convergence of the normalized partial sums of a weakly dependent stationary time series to a non-Gaussian stable distribution. These conditions are to be verified for the sequence $(\xi_i)_i$ defined in (3.9). The expansion in (3.8) then allows to obtain the asymptotic distribution of the sample mean absolute deviation.

Remark 3.5. In Proposition 3.3, if F is not continuous at μ , then $na_n^{-1}(\hat{\theta}_n - \theta)$ can still be shown to converge weakly, but the limit law will no longer be stable. As in Corollary 3.2, it will be a non-linear functional of the bivariate stable distribution to which the joint distribution of $(X_1 - \mu, |X_1 - \mu|)$ is attracted (Rvačeva, 1962).

Acknowledgments

The author gratefully acknowledges funding by contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy).

References

- Babu, G. J. and C. R. Rao (1992). Expansions for statistics involving the mean absolute deviations. *Ann. Inst. Statist. Math.* 44(2), 387–403.
- Bartkiewicz, K., A. Jakubowski, T. Mikosch, and O. Wintenberger (2011). Stable limits for sums of dependent infinite variance random variables. *Probab. Theory Related Fields* 150(3-4), 337–372.
- Doukhan, P., P. Massart, and E. Rio (1994). The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.* 30(1), 63–82.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney.
- Gnedenko, B. V. and A. N. Kolmogorov (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley Publishing Company, Inc., Cambridge, Mass. Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob.
- Gorard, S. (2005). Revisiting a 90-year-old debate: the advantages of the mean deviation. *British Journal of Educational Studies* 53(4), 417–430.

- Huber, P. J. (1996). *Robust Statistical Procedures* (Second ed.), Volume 68 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Kallenberg, O. (2002). *Foundations of Modern Probability* (Second ed.). Probability and its Applications (New York). Springer-Verlag, New York.
- Muoz-Perez, J. and A. Sanchez-Gomez (1990). A characterization of the distribution function: the dispersion function. *Statistics & Probability Letters* 10(3), 235–239.
- Pham-Gia, T. and T. L. Hung (2001). The mean and median absolute deviations. *Math. Comput. Modelling* 34(7-8), 921–936.
- Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science* 4(4), pp. 341–354.
- Rvačeva, E. L. (1962). On domains of attraction of multidimensional distributions. *Selected Translations in Math. Statist. Prob. Theory* 2, 183–205.
- Samorodnitsky, G. and M. S. Taqqu (1994). *Stable non-Gaussian Random Processes*. Stochastic Modeling. Chapman & Hall, New York. Stochastic models with infinite variance.
- Stigler, S. M. (1973). Studies in the history of probability and statistics. XXXII. Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika* 60, 439–445.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to probability and statistics*, pp. 448–485. Stanford Univ. Press, Stanford, Calif.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.