A sufficient condition for observation noise not to affect the extreme value distribution *

Yoichi Nishiyama† and Takaaki Shimura

The Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

nisiyama@ism.ac.jp shimura@ism.ac.jp

December 8, 2008

Running title: Extreme value distribution of data with noise

Abstract

We present a sufficient condition under which the rescaled sample maxima of the data $\{X_i\}$ has the same asymptotic distribution as that of the data $\{X_i + e_i\}$ with observation noise e_i. The condition is described in terms of the Orlicz norm.

Keywords: Extreme value distribution, sample maxima, observation noise, Orlicz norm.

*This version is the same as *Research Memorandum* **1081**, Inst. Statist. Math. The Japanese version will be published in *Proc. Inst. Statist. Math.* (2009©Inst. Statist. Math.), **57** No. 2, which will be available at `http://www.ism.ac.jp/editsec/toukei.html`

†Corresponding author.

Let $\{X_i\}$ be any sequence of real valued random variables. Suppose that there exist some sequences a_n and b_n of constants and a non-degenerate distribution G such that the weak convergence

$$\frac{M_n - b_n}{a_n} \to^d G$$

holds for the sample maxima $M_n = \max_{1\leq i\leq n} X_i$. In this paper, we consider the case $a_n \to \infty$ only. In the case where $\{X_i\}$ is an independently, identically distributed (i.i.d.) sequence, this assumption is equivalent to that the distribution is long tailed, that is, for every real number k the asymptotic relation $P(X > x + k) \sim P(X > x)$ holds as $x \to \infty$; this is satisfied if the distribution is subexponential. (See Shimura (2008).)

In this paper, we consider the situation where the perturbed data $\tilde{X}_i = X_i + e_i$ with noise e_i is observed. Here $\{e_i\}$ is any sequence of real valued random variables which may depend on $\{X_i\}$. Our aim is to give a sufficient condition under which the convergence

$$\frac{\tilde{M}_n - b_n}{a_n} \to^d G \tag{1}$$

holds for the same constants a_n, b_n and distribution G as above, where $\tilde{M}_n = \max_{1\leq i\leq n} \tilde{X}_i$.

In order to state our main result, let us recall the definition of the Orlicz norm $||Z||_\psi$. Let ψ be a convex, non-decreasing, non-zero function such that $\psi(0) = 0$ and that

$$\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty \tag{2}$$

for a constant c. For a real valued random variable Z, we define

$$||Z||_\psi = \inf\{C > 0 : \ E\psi(|Z|/C) \leq 1\}.$$

This defines a norm, called the Orlicz norm corresponding to ψ. When $\psi(z) = z^p$, the corresponding Orlicz norm is the L_p norm. Another important case is $\psi_p(z) = e^{z^p} - 1$ for $p \geq 1$. In this case $\psi_p^{-1}(z) = (\log(1 + z))^{1/p}$, and if

$$P(|Z| > x) \leq Ke^{-Cx^p}, \quad \forall x \geq 0$$

then $||Z||_{\psi_p} \leq ((1+K)/C)^{1/p}$. (See e.g. Lemma 2.2.1 of van der Vaart and Wellner (1996).) In general, it is easy to see that $E|Z| \leq \psi^{-1}(1)||Z||_\psi$ by the definition and Jensen's inequality. See Chapter 2.2 of van der Vaart and Wellner (1996) for more imformation on the Orlicz norms.

Now we give our main theorem.

Theorem 1 *Assume $\psi^{-1}(n)/a_n \to 0$. If $\sup_i ||e_i||_\psi < \infty$ then (1) holds.*

Proof. We denote by $i_0 = i_0(\omega)$ the index i which achieves the maximum of $X_i(\omega)$ for $1 \le i \le n$. Since

$$\max_{1\le i\le n} X_i - \max_{1\le i\le n} |e_i| \le X_{i_0} + e_{i_0} \le \max_{1\le i\le n}\{X_i + e_i\} \le \max_{1\le i\le n} X_i + \max_{1\le i\le n} |e_i|,$$

we have $|M_n - \tilde{M}_n| \le \max_{1\le i\le n} |e_i|$. So it is enough to show that $\max_{1\le i\le n} |e_i|/a_n \to^p 0$. For every $\varepsilon > 0$, it holds that

$$\begin{aligned} P\left(\frac{\max_{1\le i\le n}|e_i|}{a_n} > \varepsilon\right) &\le \frac{E\max_{1\le i\le n}|e_i|}{a_n\varepsilon} \\ &\le \frac{\psi^{-1}(1)||\max_{1\le i\le n}|e_i|||_\psi}{a_n\varepsilon}. \end{aligned}$$

Here, by Lemma 2.2.2 of van der Vaart and Wellner (1996), there exists a constant K depending only on ψ such that the right hand side is bounded by

$$\frac{\psi^{-1}(1)K\psi^{-1}(n)\max_{1\le i\le n}||e_i||_\psi}{a_n\varepsilon},$$

which tends to zero as $n \to \infty$ by assumption. The proof is completed. □

Although one may think that the above theorem might be already well known since the proof is so simple, it is appearently new. Indeed, the key inequality, namely Lemma 2.2.2 of van der Vaart and Wellner (1996), seems to have newly presented in thier book. The assumption (2) on which their proof is based is not imposed in the usual contexts of Orlicz norms. Our approach using Orlicz norms for the extreme vlaue theory seems novel, and we believe the importance of our simple and new result.

Let us give some examples.

Proposition 2 *Assume $a_n \sim n^{1/\alpha}$ for some $\alpha > 0$. If $\sup_i E|e_i|^p < \infty$ for some $p > (\alpha \vee 1)$, then (1) holds.*

Proof. Set $\psi(z) = z^p$ in Theorem 1. □

When $\{X_i\}$ is an i.i.d. sequence, typical examples of Proposition 2 are Cauchy distribution ($\alpha = 1$) and Pareto distribution ($\alpha > 0$ is general), and in these cases the extreme value distribution G is the Fréchet distribution. The tail of the latter is light when α is larger, and in that case p has to be chosen to be larger. (So the tail of $\{e_i\}$ has to be light.) For a distribution in the maximum domain of attracton of the Fréchet distribution, the constant a_n is given by $n^{1/\alpha}L(n)$ where $L(\cdot)$ is a slowly varying function (see e.g. page 131 of Embrechts *et al.* (1997)). If we choose $p > (\alpha \vee 1)$ corresponding to this α, then $\sup_i E|e_i|^p < \infty$ implies (1).

The following proposition is concerned with cases of lighter tailed distributions.

Proposition 3 *Assume* $(\log n)^{1/p}/a_n \to 0$ *for some* $p \geq 1$. *If there exist* $K, C > 0$ *such that for all* i

$$P(|e_i| > x) \leq Ke^{-Cx^p}, \quad \forall x \geq 0,$$

then (1) holds.

Proof. Set $\psi(z) = \psi_p(z) = e^{z^p} - 1$ in Theorem 1, and note that $||e_i||_{\psi_p} \leq ((1+K)/C)^{1/p}$. □

The examples of Proposition 3 in the case where $\{X_i\}$ is an i.i.d. sequence include (some cases of) Weibull-like distributions, Benktander-type-I distributions and and Benktander-type-II distributions, and in these cases the extreme value distribution G is the Gumbel distribution. (See pages 153–157 of Embrechts *et al.* (1997).) We refer to Theorem 3.3.26 of Embrechts *et al.* (1997) for the characterization of the maximum domain of attraction of the Gumbel distribution and the corresponding constants a_n and b_n.

In summary, for given a_n if we choose a function ψ such that $\psi^{-1}(n)/a_n \to 0$, then it is sufficient for (1) that the Orlicz norms of the noises with respect to ψ are bounded. Our result may be applied also to the cases where $\{X_i\}$ is not i.i.d. For example, Theorem 4.4.6 of Embrechts *et al.* (1997) gives a general theory of extreme value distribution for stationary sequences, and our result can treat such sequences with noises.

References

[1] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events.* Springer-Verlag, Berlin Heidelberg.

[2] Shimura, T. (2008). Discretization of distributions in the maximum domain of attraction. *In preparation.*

[3] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.