Semi-parametric approach to the Hasofer–Wang and Greenwood statistics in extremes

Cláudia Neves · M. Isabel Fraga Alves

Received: 29 November 2004 / Accepted: 27 February 2005 / Published online: 27 February 2007
© Sociedad de Estadística e Investigación Operativa 2007

Abstract This paper deals with the semi-parametric approach to the problem of statistical choice of extreme domains of attraction. Relying on concepts of regular variation theory, it investigates the asymptotic properties of Hasofer and Wang’s test statistic based on the \( k \) upper extremes taken from a sample of size \( n \), when \( k \) behaves as an intermediate sequence \( k_n \) rather than remaining fixed while the sample size increases. In the process a Greenwood type test statistic is proposed which turns out to be useful in discriminating heavy-tailed distributions. The finite sample behavior of both testing procedures is evaluated in the light of a simulation study. The testing procedures are then applied to three real data sets.

Keywords Max-domain of attraction · Heavy-tailed distribution · Regular variation theory · Hypothesis testing

Mathematics Subject Classification (2000) Primary 62G32 · Secondary 62G10 · 62G20

1 Introduction

The fundamental paper of Gnedenko (1943) establishes the Generalized Extreme Value distribution in the von Mises parametrization \( (G_\gamma, \gamma \in \mathbb{R}) \) as a unified version of all possible non-degenerate weak limits of maxima of sufficiently long sequences

Research partially supported by FCT/POCTI/FEDER.

C. Neves
UIIMA, Department of Mathematics, University of Aveiro, Aveiro, Portugal

M.I. Fraga Alves (✉)
CEAUL, DEIO, Faculdade de Ciências, Universidade de Lisboa, C6, Piso 4, 1749-016 Lisbon, Portugal
e-mail: isabel.alves@fc.ul.pt
of i.i.d. random variables (r.v.’s) $X_1, X_2, \ldots, X_n$. That is, there exist normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that, for all $x$,

$$
\lim_{n \to \infty} P \{ a_n^{-1} (\max(X_1, \ldots, X_n) - b_n) \leq x \} = G_\gamma(x),
$$

with

$$
\begin{cases}
G_\gamma(x) := \exp(-(1 + \gamma x)^{-1/\gamma}), & 1 + \gamma x > 0 \text{ if } \gamma \neq 0; \\
\exp(-\exp(-x)), & x \in \mathbb{R} \text{ if } \gamma = 0.
\end{cases}
$$

For $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$, the $G_\gamma$ d.f. reduces to Weibull, Gumbel and Fréchet distribution functions, respectively. The Fréchet domain of attraction contains distributions with polynomially decaying tails. All d.f.’s belonging to the Weibull domain of attraction are light-tailed with finite right endpoint. The intermediate case $\gamma = 0$ is of particular interest in many applied sciences where extremes are relevant not only because of the simplicity of inference within the Gumbel domain but also for the great variety of distributions possessing an exponential tail whether having finite right endpoint or not. In fact, separating statistical inference procedures according to the most suitable domain of attraction for the sampled distribution has become usual practice. As such, testing procedures for the Gumbel domain against Fréchet or Weibull max-domains have been of great usefulness; among these are, for instance, Castillo et al. (1989), Hasofer and Wang (1992), Fraga Alves and Gomes (1996), Wang et al. (1996) and Marohn (1998a, 1998b).

Take the order statistics (o.s.) $X_{1,n} \leq \cdots \leq X_{n,n}$ as the legacy of the independent random variables $X_1, X_2, \ldots, X_n$ with the same unknown distribution function (d.f.) $F$, after arranging these in nondecreasing order. Following a semi-parametric approach, the only assumption made is that condition (1.1) holds, i.e., the underlying d.f. $F$ is in the domain of attraction of an extreme value distribution $G_\gamma$. In this setup any inference concerning the tail of the underlying distribution $F$ can be based on the $k$ observations above the random threshold $X_{n-k,n}$. The latter contrasts with the alternative possibility of restricting attention to a random number $K_u$ of observations exceeding a given high level $u$. Here the Generalized Pareto distribution would come into play since it stems from the fundamental results of Balkema and de Haan (1974) and Pickands III (1975) as the limiting distribution for the excesses $Y_i = X_i - u | X_i > u, i = 1, \ldots, K_u$ over a sufficiently high threshold $u$.

In the context of statistical choice of extreme domains of attraction Hasofer and Wang (1992) test may be pointed out as one of the most commonly used testing procedure. In particular, Reiss and Thomas (2001) have incorporated it in the “XTREMES” software. Built on the Shapiro–Wilk goodness-of-fit statistic (see Shapiro and Wilk 1965) Hasofer and Wang’s location/scale invariant statistic embodies the reciprocal squared empirical coefficient of variation associated with the sample of the excesses above the random threshold $X_{n-k,n}$. More concretely,

$$
W_n(k) := \frac{1}{k} \frac{\left( k^{-1} \sum_{i=1}^{k} Z_i \right)^2}{k^{-1} \sum_{i=1}^{k} Z_i^2 - \left( k^{-1} \sum_{i=1}^{k} Z_i \right)^2},
$$

where $Z_i := X_{n-i+1,n} - X_{n-k,n}, i = 1, \ldots, k$. The asymptotic statements of the referred authors regarding (1.2) bear on results presented in Weissman (1978) which,