Journal of the Korean Statistical Society 45 (2016) 581–594

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Extreme value theory in mixture distributions and a statistical method to control the possible bias

Wonseon Gwak[a], Hyein Goo[a,1], Yang Ho Choi[b], Jae Youn Ahn[a,*]

[a] *Department of Statistics, Ewha Womans University, 11-1 Daehyun-Dong, Seodaemun-Gu, Seoul 120-750, South Korea*

[b] *Department of Actuarial Science, Hanyang University, Ansan, Gyeonggi-do, South Korea*

A R T I C L E I N F O

Article history:
Received 28 July 2015
Accepted 24 April 2016
Available online 14 May 2016

AMS 2000 subject classifications:
primary 62P05
secondary 62G32

Keywords:
Mixture distribution
Extreme behavior
Precipitation
Generalized Extreme distribution

A B S T R A C T

In this paper, extreme behaviors of a mixture distribution are analyzed. We investigate some cases where the mixture distributions are in the proper domain of attraction so that the extreme value of mixture distributions converges to the proper Generalized Extreme Value distribution (GEV). However, in general, there is no guarantee that the distribution of the data is in the proper maximum domain of attraction. Furthermore, since the convergence rate can be slow even with guaranteed asymptotic convergence, GEV estimation method might provide a biased estimation, as shown in Choi et al. (2014). The paper provides a safe method to control the quality of the quantile estimator in extreme values.

1. Introduction

Extreme value theory is a classical field in statistics and focuses on the behaviors of random variables at unusually large levels. Catastrophic disasters such as earthquake and tsunami are usually associated with extreme values. The generalized extreme value (GEV) theory is a typical way of describing extreme values. Specifically, for $X_1, X_2, \cdots, X_n$ being a sequence of i.i.d. non-degenerated random samples with common distribution F and the sample maxima

$$M_n = \max\{X_1, X_2, \ldots, X_n\}.$$

GEV describes the stochastic behavior of M_n as n approaches infinity.

Since extreme values are major concerns in various areas including risk theory and hydrology, GEV is widely studied in the literatures. Often, natural behaviors are explained by mixture distributions. For example, wind speed is known to follow mixture distributions (Akpinar & Akpinar, 2009; Jaramillo & Borja, 2004; Kollu, Rayapudi, Narasimham, & Pakkurthi, 2012; Tian Pau, 2011). Specifically, the distribution of wind speed in Seoul can be explained by two phases: normal circumstances and extreme circumstances caused by extreme meteorological disasters such as typhoon. In line with such an example, two-component mixture extreme value distributions such as Weibull-GEV and GEV-lognormal have been used to model observed wind speed distributions in the literature. While these statistical estimation methods show improvements in accuracy, they lack theoretical reasoning: specifically, though an underlying distribution is a mixture distribution, a distribution of the extreme value does not need to be a mixture distribution.

* Corresponding author.
E-mail address: jaeyahn@ewha.ac.kr (J.Y. Ahn).

[1] Co-first Author.

http://dx.doi.org/10.1016/j.jkss.2016.04.003
1226-3192/

In this paper, we study asymptotic convergence of extreme values (M_n) to a GEV distribution when the underlying distribution function F is a mixture distribution. However, in some cases, especially for some mixture distributions, the convergence rate of extreme values to a GEV distribution can be slow, which can create bias problems (Choi, Gwak, Goo, & Ahn, 2014). Especially, such slow convergence can create huge bias in quantile estimation. Furthermore, since this bias mainly depends on the block size (n), even with an infinite number of duplicated observations of M_n, this bias might not disappear under the GEV estimation method.

To reduce the bias problem in quantile estimation, we can use the nonparametric quantile estimator of the extreme value. Generally, the nonparametric quantile estimator leads to a high variance problem. Furthermore, in extreme value settings, there are few observations; for example, in many cases, observations are the annual maxima. In this paper, by properly combining the nonparametric statistical method and the GEV estimation method, we provide a safe method in quantile estimation of extreme values in order to control both the bias and the variance of the quantile estimators. Simulation studies are provided to support the effectiveness of the suggested method.

This paper is organized as follows. In Section 2, we briefly review the concept and preliminary results of extreme value theory. In Section 3, we discuss the maximum domain of attraction of the mixture distribution. Finally Section 4 concludes the paper with suggestions on the direction of future research.

2. Extreme value theory

This section summarizes classical results in extreme value theory and relevant notations.

2.1. Symbols and notations

Throughout the paper, we use F, with f as a density, to denote the distribution function of the random variable X. Furthermore, let $X_{i,j}$ for any $i, j = \{1, 2, \dots\}$ be independent and identically distributed (i.i.d.) observations from F. We use f_1 and f_2 as density functions of F_1 and F_2, respectively. We also define

$$M_{n,i} = \max\left\{X_{i,1}, \dots, X_{i,n}\right\}.$$

When things are clear, we use M_n to denote $M_{n,i}$. For a given distribution function G, define the essential supremum

$$x_G := \sup\left\{x \in \mathbb{R} \middle| G(x) < 1\right\}$$

and let $\overline{G}$ be a survival function of G. The $(1-p)$-th quantile of the given distribution G is defined as

$$G^{-1}(1-p) := \inf\{x : G(x) \geq 1-p\}$$

and let $Y_1, \dots, Y_m$ be i.i.d observations from G. Define the non-parametric version of G with the observations $Y_1, \dots, Y_m$ as

$$G_m^{-1}(1-p) := Y_{m(i)}, \quad \text{for } 1-p \in \left(\frac{i-1}{m}, \frac{i}{m}\right]$$

where $Y_{m(1)} < \cdots < Y_{m(m)}$ are the order statistics of $Y_1, \dots, Y_m$. Specifically, let $q_{1-p}(M_n)$ be the $(1-p)$-th quantile of M_n and $H_m^{-1}(1-p)$ be the empirical version of $q_{1-p}(M_n)$ with observations $M_{n,1}, \dots, M_{n,m}$. In the paper, n is used for the block size, and m is used for the sample size of extreme values. Finally, we find it to be convenient to define the following assumption.

Assumption 1. For the given distribution functions F_1 and F_2, assume F to be the mixture distribution of F_1 and F_2

$$F(x) := cF_1(x) + (1-c)F_2(x) \quad \text{for some } c \in (0, 1).$$

For convenience, sometimes we call F the c-mixture distribution of F_1 and F_2.

2.2. The generalized extreme value distribution and maximum domain of attraction

The following theorem is the classical result of extreme value theory.

Theorem 1 (*Embrechts, Klüppelberg, & Mikosch, 1997*)**.** *If there exist some norming constants $\{a_n > 0\}$ and $\{b_n\}$ and some nondegenerate distribution function H such that*

$$(M_n - b_n)/a_n \xrightarrow{\text{d}} H \tag{1}$$

then, for $\alpha > 0$, H belongs to one of the following families:

I: *(Gumbel)* $\Lambda(z) = \exp\{-\exp(-z)\}, \ z \in (-\infty, \infty)$

II: *(Fréchet)* $\Phi_\alpha(z) = \begin{cases} 0, & z \leq 0 \\ \exp\left(-z^{-\alpha}\right), & z > 0 \end{cases}$

III: *(Weibull)* $\Psi_\alpha(z) = \begin{cases} \exp\left(-(-z)^{\alpha}\right), & z \leq 0 \\ 1, & z > 0 \end{cases}.$