A note on max-sum equivalence

Jinzhu Li\textsuperscript{a,b}, Qihe Tang\textsuperscript{b,∗}

\textsuperscript{a} School of Mathematical Science and LPMC, Nankai University, Tianjin 300071, PR China
\textsuperscript{b} Department of Statistics and Actuarial Science, The University of Iowa, 241 Schaeffer Hall, Iowa City, IA 52242, USA

**Article Info**

Article history:
Received 1 April 2010
Received in revised form 15 July 2010
Accepted 21 July 2010
Available online 5 August 2010

**MSC:**
primary 60E05
secondary 62E20

**Keywords:**
Max-sum equivalence
Subexponentiality

**Abstract**

For finitely many independent real-valued random variables, if their maximum follows a subexponential distribution, then the tail probabilities of their sum and maximum are asymptotically equivalent.

© 2010 Elsevier B.V. All rights reserved.

doi:10.1016/j.spl.2010.07.015

1. Main result

Throughout this note, all limit relations hold as \( x \to \infty \) unless stated otherwise. The relation \( a(x) \sim b(x) \) means that the quotient of the two sides converges to 1. For a (cumulative) distribution function \( F \), write \( F_n = 1 - F \) as its decumulative distribution function.

Let \( X_1, \ldots, X_n \) be \( n \) independent real-valued random variables with distributions \( F_1, \ldots, F_n \), respectively. Denote by \( G_n \) the distribution of \( \max\{X_1, \ldots, X_n\} \) and by \( H_n \) the distribution of \( X_1 + \cdots + X_n \) (i.e. \( H_n = F_1 * \cdots * F_n \)). These random variables are said to be max-sum equivalent if \( H_n(x) \sim G_n(x) \). This is connected to the well-known principle of a single big jump in extreme value theory. See Embrechts and Goldie (1980), Cline (1986, 1987), Leslie (1989), and Geluk (2009) for interesting discussions on or related to max-sum equivalence.

The subexponential class, which is one of the most important classes of heavy-tailed distributions, naturally appears when studying max-sum equivalence. By definition, a distribution \( F \) on \([0, \infty)\) is subexponential, written as \( F \in \mathcal{S} \), if the relation

\[
F_n(x) \sim nF(x)
\]

holds for some (or, equivalently, for all) \( n = 2, 3, \ldots \), where \( F_n^{*} \) is the \( n \)-fold convolution of \( F \). More generally, a distribution \( F \) on \((-\infty, \infty)\) is still said to be subexponential if \( F(x)1_{(x>0)} \) is. In the latter case, relation (1) still holds.

Clearly, if the random variables \( X_1, \ldots, X_n \) are independent and identically distributed (i.i.d.) with common subexponential distribution, then they are max-sum equivalent. The goal of this note is to point out a sufficient condition for max-sum equivalence for the non-identical case.

∗ Corresponding author. Tel.: +1 319 335 0730; fax: +1 319 335 3017.
E-mail addresses: lijinzhunk@gmail.com (J. Li), qtang@stat.uiowa.edu (Q. Tang).

0167-7152/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2010.07.015
Theorem 1. Let $X_1, \ldots, X_n$ be $n \geq 2$ independent real-valued random variables as given above. If $G_n \in \mathcal{S}$ then $X_1, \ldots, X_n$ are max-sum equivalent, that is,

$$
\overline{H_n}(x) \sim \overline{G_n}(x) \sim \sum_{i=1}^{n} \overline{F_i}(x).
$$

The second relation in (2) automatically holds for distributions $F_i$ with ultimate right tails. In addition, it is known that the class $\mathcal{S}$ is closed under tail equivalence; see Theorem 3 of Teugels (1975) and Lemma A3.15 of Embrechts et al. (1997). Thus, the condition $G_n \in \mathcal{S}$ is easily verifiable in concrete cases.

2. Remarks

Other important classes of heavy-tailed distributions include the class $\mathcal{L}$ of long-tailed distributions, characterized by the relation $\overline{F}(x-y) \sim \overline{F}(x)$ for all real $y$, and the class $\mathcal{D}$ of distributions with dominatedly varying tails, characterized by the relation $\overline{F}(xy) = 0(\overline{F}(x))$ for all $0 < y < 1$. It is well known that

$$
\mathcal{L} \cap \mathcal{D} \subset \mathcal{S} \subset \mathcal{L}.
$$

See page 50 of Embrechts et al. (1997) for these inclusions for the case with $F$ supported on $[0, \infty)$. It is easy to see that they are still valid for the general case with $F$ supported on $(-\infty, \infty)$. Note that for the general case, relation (1) only does not imply $F \in \mathcal{S}$; see, e.g., page 19 of Borovkov and Borovkov (2008) for a simple counterexample. However, this implication holds if restricted to the scope of $F \in \mathcal{L}$.

Lemma 1 of Embrechts and Goldie (1980) shows that if $G_2 \in \mathcal{L}$ and $H_2 \in \mathcal{S}$ then $\overline{H_2}(x) \sim \overline{F_1}(x) + \overline{F_2}(x)$. This result does not cover, and is not covered by, our Theorem 1.

Theorem 2.1 of Cai and Tang (2004) shows that if $F_1 \in \mathcal{L} \cap \mathcal{D}$ and $F_2 \in \mathcal{L} \cap \mathcal{D}$ then $\overline{H_2}(x) \sim \overline{F_1}(x) + \overline{F_2}(x)$. This corresponds to a special case of our Theorem 1.

Another similar result is Theorem 3 of Geluk and de Vries (2006), showing that relation (2) holds if $F_i \ast F_j \in \mathcal{S}$ for all $i, j = 1, \ldots, n$. Actually, their condition implies $G_n \in \mathcal{S}$ and, hence, their result also corresponds to a special case of our Theorem 1. To see this, first notice that their condition implies $F_i \in \mathcal{S}$ for all $i = 1, \ldots, n$ because of the closure of $\mathcal{S}$ under the convolution root (see Theorem 2 of Embrechts et al. 1979). Write $K = n^{-1} \sum_{i=1}^{n} F_i$. Clearly, $\overline{G_n}(x) \sim n\overline{K}(x)$. Thus, $G_n \in \mathcal{S}$ if and only if $K \in \mathcal{S}$. The latter can be verified as follows. First, $K \in \mathcal{L}$ is immediate. Second, we have

$$
\overline{K^{2+}}(x) = \frac{1}{n^2} \left( \sum_{i=1}^{n} \overline{F_i^{2+}}(x) + 2 \sum_{1 \leq j < k \leq n} \overline{F_j F_k}(x) \right)
$$

$$
\sim \frac{2}{n^2} \left( \sum_{i=1}^{n} \overline{F_i}(x) + \sum_{1 \leq j < k \leq n} (\overline{F_j}(x) + \overline{F_k}(x)) \right)
$$

$$
= 2 \overline{\overline{K}}(x),
$$

where in the second step we used Theorem 2 of Embrechts and Goldie (1980).

3. Proof of Theorem 1

The following lemma will be used in the proof of Theorem 1:

Lemma 1. Let $Y_1, \ldots, Y_n$ be $n \geq 2$ i.i.d. real-valued random variables with common distribution $G \in \mathcal{S}$. Then

$$
\lim_{c \to \infty} \lim_{x \to \infty} \frac{\Pr \left( \sum_{i=1}^{n} Y_i > x, Y_1 > c, Y_2 > c \right)}{\bar{G}(x)} = 0.
$$

Proof. For every $x \geq 0$ and $c \geq 0$, write

$$
\Pr \left( \sum_{i=1}^{n} Y_i > x, Y_1 > c, Y_2 > c \right) = \Pr \left( \sum_{i=1}^{n} Y_i > x \right) - 2 \Pr \left( \sum_{i=1}^{n} Y_i > x, Y_1 \leq c \right)
$$

$$
+ \Pr \left( \sum_{i=1}^{n} Y_i > x, Y_1 \leq c, Y_2 \leq c \right) = I_1(x) - 2I_2(x, c) + I_3(x, c).
$$

By the definition of subexponentiality, $I_1(x) \sim n\overline{G}(x)$. Furthermore,
where in the last step we used $G \in \mathcal{L}$ and the dominated convergence theorem. Similarly,

$$I_3(x, c) = \int_0^c \int_0^c \Pr \left( \sum_{i=1}^{n-2} Y_i > x - y_1 - y_2 \right) G(dy_1) G(dy_2)$$

$$\sim (n-2)G(c)^2 \bar{G}(x),$$

which is understood as $I_3(x, c) = o(\bar{G}(x))$ in the case $n = 2$. Plugging these estimates into (3) yields the desired result. $\square$

**Proof of Theorem 1.** We only need to prove the first relation in (2). For every $x \geq 0$ and $0 \leq c \leq x/n$,

$$\bar{H}_n(x) = \Pr \left( \sum_{i=1}^{n} X_i > x, \bigcup_{j=1}^{n} (X_j > c) \right).$$

According to whether or not there is exactly only one $(X_j > c)$ occurring in the union, we split the probability on the right-hand side into two parts as

$$\bar{H}_n(x) = J_1(x, c) + J_2(x, c).$$

First we deal with $J_1(x, c)$. On the one hand, by $G_n \in \mathcal{L}$,

$$J_1(x, c) = \sum_{j=1}^{n} \Pr \left( \sum_{i=1}^{n} X_i > x, X_j > c, \bigcap_{k=1, k \neq j}^{n} (X_k \leq c) \right)$$

$$\leq \sum_{j=1}^{n} \bar{F}_j(x - (n-1)c)$$

$$\sim \frac{G_n(x)}{\bar{G}_n(x)}.$$ 

On the other hand, for arbitrarily fixed $d > 0$,

$$J_1(x, c) \geq \sum_{j=1}^{n} \Pr \left( \sum_{i=1}^{n} X_i > x, X_j > c, \bigcap_{k=1, k \neq j}^{n} (-d < X_k \leq c) \right)$$

$$\geq \sum_{j=1}^{n} \Pr \left( X_j > x + (n-1)d, \bigcap_{k=1, k \neq j}^{n} (-d < X_k \leq c) \right)$$

$$\geq \prod_{k=1}^{n} \Pr(-d < X_k \leq c) \sum_{j=1}^{n} \bar{F}_j(x + (n-1)d)$$

$$\sim \prod_{k=1}^{n} \Pr(-d < X_k \leq c) \bar{G}_n(x),$$

where in the last step we used $G_n \in \mathcal{L}$ again. By the arbitrariness of $d$, it follows that

$$\prod_{k=1}^{n} F_k(c) \leq \liminf_{x \to \infty} \frac{j_1(x, c)}{G_n(x)} \leq \limsup_{x \to \infty} \frac{j_1(x, c)}{G_n(x)} \leq 1. \quad (5)$$

Next we turn to $J_2(x, c)$. Introduce $Y = \max\{X_1, \ldots, X_n\}$, which is distributed by $G_n \in \mathcal{L}$, and let $Y_1, \ldots, Y_n$ be i.i.d. copies of $Y$. Clearly,

$$J_2(x, c) = \Pr \left( \sum_{i=1}^{n} X_i > x, \bigcup_{1 \leq j < k \leq n} (X_j > c, X_k > c) \right)$$

$$\leq \sum_{1 \leq j < k \leq n} \Pr \left( \sum_{i=1}^{n} Y_i > x, Y_j > c, Y_k > c \right).$$
Thus, by Lemma 1,

$$\limsup_{c \to \infty} \limsup_{x \to \infty} \frac{J_2(x, c)}{G_n(x)} = 0.$$  \hspace{1cm} (6)

A simple combination of (4)–(6) gives the desired result. \hfill \Box

Acknowledgement

The authors are grateful to the referee for his/her useful comments.

References


