Maxima of sums and random sums for negatively associated random variables with heavy tails

Dingcheng Wang\textsuperscript{a,}\*, Qihe Tang\textsuperscript{b}

\textsuperscript{a} School of Management and School of Applied Mathematics, University of Electronic Science and Technology of China, Chengdu 610054, Sichuan, PR China

\textsuperscript{b} Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Received 1 December 2003
Available online 26 May 2004

Abstract

This paper obtains some asymptotics for the tail probabilities of the maximum of sums and random sums of negatively associated (NA) random variables with heavy tails, showing that the NA dependence structure does not affect the asymptotic behavior of these tail probabilities.

\copyright 2004 Elsevier B.V. All rights reserved.

Keywords: Asymptotics; Heavy tails; Maximum of sums; Random sums; Negative association; Tail probabilities

1. Introduction

Throughout, let \(X_k; k \geq 1\) be a sequence of random variables (r.v.s) with common distribution function (d.f.) \(F = 1 - \bar{F}\), and let \(S_n\) be its \(n\)th partial sum, \(n \geq 1\). For two positive infinitesimals \(A(\cdot)\) and \(B(\cdot)\), as usual, we write \(A(\cdot) \preceq B(\cdot)\) if \(\lim \sup A(\cdot)/B(\cdot) \leq 1\), \(A(\cdot) \succeq B(\cdot)\) if \(\lim \inf A(\cdot)/B(\cdot) \geq 1\), and \(A \sim B\) if both. All limit relations are for \(x \to \infty\) unless stated otherwise.

A well known notion in extremal value theory, the subexponentiality, describes an important property of the right tail of a distribution. By definition, a d.f. \(F\) concentrated on \([0, \infty)\) is subexponential, denoted by \(F \in \mathcal{S}\), if for the corresponding i.i.d. nonnegative r.v.s \(\{X_k; k \geq 1\}\) and for some \(n \geq 2\) (or equivalently, for any \(n \geq 2\), the relation

\[
\mathbb{P}(S_n > x) \sim n \bar{F}(x)
\]  

(1.1)

\*Corresponding author.

E-mail addresses: wangdc@uestc.edu.cn (D. Wang), q.tang@uva.nl (Q. Tang).
holds. For more details and nice reviews on the subexponentiality, we refer to Embrechts et al. (1997) and the references therein.

The subexponentiality has been extensively investigated by many researchers since it was independently introduced by Chistyakov (1964) and Chover et al. (1973a,b). However, most of the investigations in the literature are restricted to the independence case. One goal of this paper is to prove that relation (1.1) still holds for a certain subclass of the class $\mathcal{S}$ when the corresponding r.v.s are negatively associated. A finite family of random variables $\{X_k; 1 \leq k \leq n\}$ is said to be negatively associated (NA) if for every pair of disjoint subsets $A_1$ and $A_2$ of $\{1, 2, \ldots, n\}$,

$$\text{Cov}\{f_1(X_{k_1}, k_1 \in A_1), f_2(X_{k_2}, k_2 \in A_2)\} \leq 0$$

whenever $f_1$ and $f_2$ are coordinatewise increasing such that the covariance exists. An infinite family is NA if each of its finite subfamilies is negatively associated. This dependence structure was first introduced by Alam and Saxena (1981) and Joag-Dev and Proschan (1983).

Sometimes it is not so reasonable for us to assume that the d.f. $F$ is concentrated on $[0, \infty)$. For example, let $X_k$ be the total underlying net risk due to the $k$th insurance policy, $k \geq 1$. As proposed by Ng et al. (2003) and Kaas and Tang (2003), the $X_k$ can be modelled by

$$X_k = Z_k - (1 + \delta)\mathbb{E}Z_k,$$

where $Z_k$ is the total incurred claims due to the $k$th policy, and $(1 + \delta)\mathbb{E}Z_k$ is the insurance premium paid by the $k$th policy holder with $\delta > 0$ explained as the safety loading. In this way, the d.f. $F$, say, of the r.v. $X_k$ is concentrated on an interval that is bounded from below by a negative number.

Hence, we then consider a more general case where the d.f. $F$ is concentrated on $[s, \infty)$ for some real number $s > -\infty$. Correspondingly, each r.v. in the NA sequence $\{X_k; k \geq 1\}$ is real-valued and is bounded from below by $s$. Let $S_{(n)}$ be the maximum of the first $n$ partial sums, namely,

$$S_{(n)} = \max\{S_k; k = 1, 2, \ldots, n\}.$$

In this latter case we prove that the relation

$$\mathbb{P}(S_{(n)} > x) \sim \mathbb{P}(S_n > x) \sim n\mathbb{P}(x)$$

(1.2)

holds for each $n \geq 1$.

What is more interesting in many application problems is to establish asymptotics for the tail probabilities of random sums or the maximum of sums over a random time horizon. We refer to Ng et al. (2002, 2003), Kaas and Tang (2003), Foss and Zachary (2003), and Ng and Tang (2004) for motivations of this study. Let $\tau$ be a nonnegative and integer-valued r.v. independent of the NA sequence $\{X_k; k \geq 1\}$. Under some mild moment conditions on the common d.f. $F$ and the r.v. $\tau$, we prove as the second main result that the relation

$$\mathbb{P}(S_{(\tau)} > x) \sim \mathbb{P}(S_\tau > x) \sim \mathbb{E}\tau\mathbb{P}(x)$$

(1.3)

holds. As explained by Ng et al. (2003) and Kaas and Tang (2003), the tail probability $\mathbb{P}(S_{(\tau)} > x)$ above can be understood as the finite time ruin probability if we regard the r.v. $\tau$ as the number of claims up to a fixed time horizon.

The remaining part of this paper consists of three sections. Section 2 presents the main results, Section 3 prepares some preliminaries, and Section 4 proves.
2. Main results and their corollary

First of all, we recall two other classes of heavy-tailed distributions, which are crucial for our purpose. We say that a d.f. \( F \) belongs to the class \( \mathcal{D} \) (has dominated variation) if for any \( 0 < l < 1 \) (or equivalently, for some \( 0 < l < 1 \)), it holds that
\[
\limsup_{x \to \infty} \frac{\tilde{F}(lx)}{\tilde{F}(x)} < \infty.
\]
We say that a d.f. \( F \) belongs to the class \( \mathcal{L} \) (is long tailed) if for any \( L > 0 \)
\[
\lim_{x \to \infty} \frac{\tilde{F}(x + L)}{\tilde{F}(x)} = 1.
\]
It is well known that
\[
\mathcal{D} \cap \mathcal{L} \subset \mathcal{S} \subset \mathcal{L}^c;
\]
see Embrechts et al. (1997) for details. We remark that the intersection \( \mathcal{D} \cap \mathcal{L} \) is a useful subclass of the class \( \mathcal{S} \). Specifically, this intersection contains the famous class \( \mathcal{R} \) of d.f.s with regular variation.

Now we are ready to state the main results of this paper.

**Theorem 2.1.** Let \( \{X_k; k \geq 1\} \) be a sequence of nonnegative NA r.v.s with common d.f. \( F \). If \( F \in \mathcal{L} \cap \mathcal{D} \) and \( \mathbb{E} X_1^+ < \infty \) for some \( r > 1 \), then, it holds for each \( n \geq 1 \) that
\[
\mathbb{P}(S_n > x) \sim n \tilde{F}(x).
\]
(2.3)

Throughout the paper, write \( x^+ = \max\{0, x\} \). We can easily obtain a consequence of Theorem 2.1 as follows, which extends the result to the case where the r.v.s are not necessarily nonnegative.

**Corollary 2.1.** Let \( \{X_k; k \geq 1\} \) be a sequence of NA r.v.s with common d.f. \( F \) concentrated on \([s, \infty)\) for some \( s > -\infty \). If \( F \in \mathcal{L} \cap \mathcal{D} \) and \( \mathbb{E}(X_1^+)^{1/r} < \infty \) for some \( r > 1 \). Then, it holds for each \( n \geq 1 \) that
\[
\mathbb{P}(S(n) > x) \sim \mathbb{P}(S_n > x) \sim n \tilde{F}(x).
\]
(2.4)

**Proof.** It follows from Theorem 2.1 that
\[
\mathbb{P}(S(n) > x) \leq \mathbb{P} \left( \sum_{k=1}^{n} X_k^+ > x \right) \leq n \tilde{F}(x).
\]
(2.5)

On the other hand, applying Theorem 2.1 once again, we derive
\[
\mathbb{P}(S_n > x) = \mathbb{P} \left( \sum_{k=1}^{n} (X_k - s) > x - ns \right)
\sim n \mathbb{P}(X_1 - s > x - ns)
\sim n \tilde{F}(x),
\]
(2.6)
where in the last step we have used the condition $F \in \mathcal{L}$. Combining (2.5) and (2.6) yields (2.4). This ends the proof of Corollary 2.1. □

The following result extends Corollary 2.1 to the case of random sums.

**Theorem 2.2.** Let $\{X_k; k \geq 1\}$ be a sequence of NA r.v.s with common d.f. $F$ concentrated on $[s, \infty)$ for some $s > -\infty$ and finite mean $\mu < 0$, and let $\tau$ be a nonnegative and integer-valued r.v. independent of the sequence $\{X_k; k \geq 1\}$. If $F \in \mathcal{L} \cap \mathcal{D}$, $\mathbb{E}(X_1^+) < \infty$ for some $r > 1$, and $\mathbb{E}\tau < \infty$, then it holds that

$$
\mathbb{P}(S(\tau) > x) \sim \mathbb{P}(S > x) \sim \mathbb{E}\tau \mathbb{F}(x).
$$

3. Preliminaries

**Lemma 3.1.** Let $\{X_k; k \geq 1\}$ be a sequence of NA r.v.s with common d.f. $F$ and finite mean $\mu < 0$. If $\mathbb{E}|X_1| < \infty$ for some $r > 1$, then for any $0 < \theta < 1$ such that $(r - 1)/(8\theta) > 1$, there is some constant $C > 0$ independent of $x$ and $n$ such that the inequality

$$
\mathbb{P}\left( S(n) > x, \bigcap_{i=1}^{n} (X_i \leq \theta x) \right) \leq Cx^{\frac{(r-1)}{8\theta} + 1}
$$

(3.1)

holds for all $n \geq 1$ and all $x > 0$.

**Proof.** If we can prove that (3.1) holds for all $n \geq 1$ and all large $x$, say $x \geq x_0$ for some $x_0 > 0$ independent of $n$, then (3.1) holds for all $n \geq 1$ and all $x > 0$.

We apply the approach developed by Tang (2004) with many changes. Write

$$
\tilde{X}_{ki} = X_k \mathbb{I}_{[X_k \leq \theta(x-i\mu)]} + \theta(x - i\mu) \mathbb{I}_{[X_k > \theta(x-i\mu)]} \quad \text{and} \quad a = a(x, i) = \log(x - i\mu),
$$

(3.2)

where $\mathbb{I}_A$ denotes the indicator function of a set $A$. Observe that for each $i \geq 1$, the sequence $\{\tilde{X}_{ki} - \mu, k \geq 1\}$ is still NA. For any fixed $x > 0$, $n \geq 1$, and $h = h(x, i) > 0$ for $1 \leq i \leq n$, we have

$$
\mathbb{P}\left( S(n) > x, \bigcap_{i=1}^{n} (X_i \leq \theta x) \right) \leq \sum_{i=1}^{n} \mathbb{P}\left( \sum_{j=1}^{i} (\tilde{X}_{ji} - \mu) > x - i\mu \right) \leq \sum_{i=1}^{n} (\mathbb{E}\exp\{h(\tilde{X}_{ii} - \mu)\} - 1)^i \exp\{-h(x - i\mu)\}.
$$

(3.3)

Applying an elementary inequality $1 + u \leq e^u$ for all real number $u$, from (3.3) we further obtain

$$
\mathbb{P}\left( S(n) > x, \bigcap_{i=1}^{n} (X_i \leq \theta x) \right) \leq \sum_{i=1}^{n} \exp\{g_i - h(x - i\mu)\},
$$

(3.4)
where

\[
g_i = i \left( \int_{-\infty}^{\theta(x-i\mu)/a'} \frac{\theta(x-i\mu)}{a'} dt + \int_{\theta(x-i\mu)/a'}^{0} \frac{\theta(x-i\mu)}{a'} e^{h(t-\mu)} - 1 \right) F(dt) + ie^{h(t-\mu)} F(\theta(x-i\mu))
\]

\[
\leq i \int_{-\infty}^{\theta(x-i\mu)/a'} \frac{\theta(x-i\mu)}{a'} \left( e^{h(t-\mu)} - 1 \right) F(dt) + i \int_{\theta(x-i\mu)/a'}^{0} \frac{\theta(x-i\mu)}{a'} e^{h(t-\mu)} F(dt) + ie^{h(t-\mu)} F(\theta(x-i\mu))
\]

\[
= g_i^{(1)} + g_i^{(2)},
\]

(3.5)

with \(a = a(x, i) = \log(x - i\mu)\) and \(\gamma > 1\) some fixed constant. First we deal with \(g_i^{(1)}\). Clearly, we can choose some \(A\) large enough such that for all \(x > 0\) and all \(i \geq 1\),

\[
B = \mathbb{E} X_1 \|_{-A < X_1 < 0} - \frac{\mu}{\theta(x-i\mu)/a'} - \mathbb{E} X_1 \|_{-\infty < X_1 < -A} < - \frac{\mu}{64}
\]

(3.6)

Denote \(y(t, i) = y(x, i, t) = \exp\{h(x, i)(t - \mu)\}h(x, i) = e^{h(t-\mu)} h\). Applying another elementary inequality \(e^u - 1 \leq u e^u\) for all real number \(u\), we have, for all \(i \geq 1\),

\[
g_i^{(1)} \leq \left( \int_0^{\theta(x-i\mu)/a'} - \int_{-A}^{0} \right) y(x, t) t F(dt) - i \mu \int_{-\infty}^{\theta(x-i\mu)/a'} y(x, t) F(dt)
\]

\[
\leq i \bar{y} \left( x, \frac{\theta(x-i\mu)}{a'} \right) \mathbb{E} X_1 \|_{0 < X_1 \leq \theta(x-i\mu)/a'} + i \bar{y}(x, -A) \mathbb{E} X_1 \|_{-A < X_1 < 0} - i \mu \gamma \left( x, \frac{\theta(x-i\mu)}{a'} \right)
\]

\[
= i \bar{y} \left( x, \frac{\theta(x-i\mu)}{a'} \right) B + i \left( y(x, -A) - y \left( x, \frac{\theta(x-i\mu)}{a'} \right) \right) \mathbb{E} X_1 \|_{-A < X_1 < 0}.
\]

(3.7)

Now set

\[
h = h(x, i) = \frac{(r+1)a - \gamma r \log a}{\theta(x-i\mu)}
\]

to (3.7). We obtain that, for all \(i \geq 1\) and all large \(x\),

\[
g_i^{(1)} \leq - \frac{\mu}{32} \bar{h}(x, i) - \bar{h}(x, i)
\]

\[
\leq \frac{1}{16}(x - i\mu) \bar{h}(x, i)
\]

\[
\leq \frac{(r+1)a(x, i)}{16\theta}.
\]

(3.8)

Next we turn to \(g_i^{(2)}\). We have, for all \(i \geq 1\) and all large \(x > 0\),

\[
g_i^{(2)} \leq i \left( \bar{F} \left( \frac{\theta(x-i\mu)}{a(x, i)}, \frac{\theta(x-i\mu)}{a'} \right) \right) \exp \{ h(x, i) [\theta(x-i\mu) - \mu] \}
\]

\[
\leq i \mathbb{E} X_1 \| \left( \frac{a(x, i)^{\gamma}}{(x-i\mu)^{\gamma}} + 1 \right) \exp \{ h(x, i) [\theta(x-i\mu) - \mu] \}
\]

\[
\leq C_1.
\]

(3.9)
for some constant $C_1 > 0$. Substituting (3.8) and (3.9) into (3.5) yields that for all $i \geq 1$ and all large $x > 0$,

$$g_i - h(x, i)(x - i\mu) \leq -\frac{(r - 1)a(x, i)}{8\theta}.$$ 

Thus, from (3.4) we conclude that for all $n \geq 1$ and all large $x > 0$,

$$\mathbb{P}\left(S_{(n)} > x; \bigcap_{i=1}^{n} (X_i \leq \theta x)\right) \leq \sum_{i=1}^{n} (x - i\mu)^{-\frac{r-1}{8\theta}} \leq \int_{0}^{x} (x - z\mu)^{-\frac{r-1}{8\theta}}dz.$$ 

Hence, (3.1) holds for some constant $C > 0$. This ends the proof of Lemma 3.1. □

For a d.f. $F$ and any $y > 0$, as done recently by Tang and Tsitsiashvili (2003), we set

$$\tilde{F}_*(y) = \liminf_{x \to \infty} \frac{\tilde{F}(xy)}{\tilde{F}(x)}$$

and then define

$$J^+_F = -\lim_{y \to \infty} \frac{\log \tilde{F}_*(y)}{\log y}. \quad (3.10)$$

We call $J^+_F$ as the upper Matuszewska index of the d.f. $F$. For details we refer to Bingham et al. (1987, Chapter 2.1) and Tang and Tsitsiashvili (2003), among others.

Tang and Tsitsiashvili (2003, Lemma 3.5) proved the following result:

**Lemma 3.2.** For a d.f. $F \in \mathcal{D}$ with its upper Matuszewska index $J^+_F$, it holds for any $v > J^+_F$ that $x^{-v} = o(F(x))$.

**4. Proofs of the main results**

**4.1. Proof of Theorem 2.1**

We formulate the proof of Theorem 2.1 into two lemmas, which provide $\mathbb{P}(S_n > x)$ with the upper and lower bounds, respectively.

**Lemma 4.1.** Under the conditions of Theorem 2.1, it holds for each $n \geq 1$ that

$$\mathbb{P}(S_n > x) \leq n\tilde{F}(x). \quad (4.1)$$
Proof. Clearly, for arbitrarily fixed numbers \( 0 < \theta < 1 \) and \( L > 0 \), we have

\[
P(S_n > x) \leq P\left( \bigcup_{i=1}^n (X_i > x - L) \right) + P\left( S_n > x, \bigcap_{i=1}^n (X_i \leq x - L), \bigcup_{i=1}^n (X_i > \theta x) \right)
\]

\[
+ P\left( S_n > x, \bigcap_{i=1}^n (X_i \leq \theta x) \right)
\]

\[
\leq n \bar{F}(x - L) + \sum_{i=1}^n P(S_n - X_i > L, X_i > \theta x) + P\left( S_n > x, \bigcap_{i=1}^n (X_i \leq \theta x) \right)
\]

\[
= I_1 + I_2 + I_3.
\]  

(4.2)

Now we deal with \( I_3, I_2, \) and \( I_1 \) successively. For \( I_3 \), we choose some \( 0 < \theta < 1 \) such that

\[
\frac{r - 1}{160} - 1 > \frac{1}{\bar{F}}
\]

holds. Applying Lemmas 3.1 and 3.2, for any fixed constant \( a > \mathbb{E}X_1 \) and all large \( x \) we have

\[
I_3 \leq P\left( \sum_{i=1}^n (X_i - a) > x - na, \bigcap_{i=1}^n (X_i - a \leq \theta x) \right)
\]

\[
\leq P\left( \sum_{i=1}^n (X_i - a) > x - na, \bigcap_{i=1}^n (X_i - a \leq \theta (x - na)) \right)
\]

\[
\leq C(x - na) \frac{(r-1)^{160r+1}}{160r+1} = o(\bar{F}(x - na)) = o(\bar{F}(x)).
\]  

(4.3)

For this fixed \( \theta \), we consider \( I_2 \). By the NA property, we have

\[
I_2 \leq \bar{F}(\theta x) \sum_{i=1}^n P(S_n - X_i > L).
\]

Since \( F \in \mathcal{D} \), the factor \( \bar{F}(\theta x) \) has the same order as \( \bar{F}(x) \). Hence, letting \( L \to \infty \) yields that

\[
\lim_{L \to \infty} \limsup_{x \to \infty} \frac{I_2}{\bar{F}(x)} = 0.
\]  

(4.4)

As for \( I_1 \), by the condition \( F \in \mathcal{L} \), it holds for each fixed \( L > 0 \) that

\[
I_1 \sim n \bar{F}(x).
\]  

(4.5)

Hence, substituting (4.3)–(4.5) into (4.2) leads to (4.1). This ends the proof of Lemma 4.1.

\[
\square
\]

Lemma 4.2. Let \( \{X_k; k \geq 1\} \) be a sequence of nonnegative NA r.v.s with common d.f. \( F \). Then, it holds for each \( n \geq 1 \) that

\[
P(S_n > x) \geq n \bar{F}(x).
\]
Proof. By the NA property, we have
\[
\mathbb{P}(S_n > x) \geq \sum_{j=1}^{n} \mathbb{P}(S_n > x, X_j > x) - \sum_{1 \leq i < j \leq n} \mathbb{P}(X_i > x, X_j > x)
\]
\[
= \sum_{j=1}^{n} \mathbb{P}(X_j > x) - \sum_{1 \leq i < j \leq n} \mathbb{P}(X_i > x, X_j > x)
\]
\[
\sim n \bar{F}(x).
\]
This ends the proof of Lemma 4.2. □

We remark that this lemma is of interest on its own right.

4.2. Proof of Theorem 2.2

We only prove the relation
\[
\mathbb{P}(S(t) > x) \sim \bar{F}(x)\mathbb{E}_t \tag{4.6}
\]
since the other relation \(\mathbb{P}(S_t > x) \sim \bar{F}(x)\mathbb{E}_t\) can be proved similarly. Obviously

\[
\mathbb{P}(S(t) > x) = \sum_{n=1}^{\infty} \mathbb{P}(S(n) > x)\mathbb{P}(\tau = n). \tag{4.7}
\]

Recall \(F \in \mathcal{D}\). We choose some \(\theta > 0\) such that
\[
\frac{r - 1}{8\theta} - 1 > \mathbb{P}_F^+.
\]
Then, by virtue of Lemmas 3.1 and 3.2, we conclude that for some constant \(C > 0\),
\[
\mathbb{P}(S(n) > x) \leq n \mathbb{P}(X_1 > \theta x) + \mathbb{P} \left( S(n) > x, \bigcap_{i=1}^{n} (X_i \leq \theta x) \right) \leq Cn \bar{F}(x)
\]
holds for all \(n \geq 1\) and large \(x\). Therefore, applying Corollary 2.1 and the dominated convergence theorem to series (4.7) leads to the desired result (4.6). This ends the proof of Theorem 2.2.

Acknowledgements

We would like to thank Professor Chun Su for his helpful discussions. Wang’s work was supported by the National Science Foundation of China (Project no: 70272001) and the Youth Science and Technology Foundation of UESTC (Project no. JX 03038), Tang’s work was supported by the Dutch Organization for Scientific Research (Project no. NWO 42511013).

References


