Asymptotics in the symmetrization inequality

Jaap Geluk

Econometric Institute, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Econometric Institute Report EI 2003-44

Abstract

We give a sufficient condition for i.i.d. random variables X_1, X_2 in order to have $P\{X_1 - X_2 > x\} \sim P\{|X_1| > x\}$ as $x \to \infty$. A factorization property for subexponential distributions is used in the proof. In a subsequent paper the results will be applied to model fragility of financial markets.

AMS 1991 subject classification. Primary 60E15; Secondary 62F99.

Key words. symmetrization, heavy tails, convolution, subexponentiality.

1 Introduction and results

Suppose X_1, X_2 are i.i.d. random variables with distribution function (d.f.) F. Very often messy arguments in probability theory can be avoided by using symmetrization. The symmetrization inequalities (see e.g. Feller [6]), state that the tails of F and of the d.f. F_s of the corresponding symmetrized random variable $X_1 - X_2$ are of comparable order.

Theorem 1 (Symmetrization inequalities). *For $x > 0$ we have*

$$P\{|X_1 - X_2| > x\} \le 2P\{|X_1| > \frac{1}{2}x\}. \tag{1}$$

If $a \ge 0$ is chosen so that $P\{X_i \le a\} \ge p$ and also $P\{X_i \ge -a\} \ge p$, then

$$P\{|X_1 - X_2| > x\} \ge pP\{|X_1| > x + a\}. \tag{2}$$

In particular, if 0 is a median for X_i,

$$P\{|X_1 - X_2| > x\} \ge \frac{1}{2}P\{|X_1| > x\}.$$

Looking at these inequalities, it is a natural question to ask how the order of growth of the quantities on the right-hand side of both (1) and (2) compare. This suggests to consider the class of distribution function tails for which $P\{|X_1| > \frac{1}{2}x\}/P\{|X_1| > x\}$ is bounded as $x \to \infty$, i.e. distributions with dominatedly varying tails. Tail behaviour of these random variables is well understood (see e.g. Bingham et al. [2]). However, it is possible to have

$$P\{|X_1 - X_2| > x\} \sim cP\{|X_1| > x\} \text{ as } x \to \infty \tag{3}$$

without dominatedly varying tail behaviour, where c is a constant. In this paper we consider when this happens.

In order to get non-trivial tail behaviour as $x \to \infty$, the random variables need to have unbounded support. First we list some examples:

1. For the standard exponential distribution, the symmetrized random variable has a double exponential distribution, hence (3) is satisfied with $c = 1$.

2. For a Pareto distribution $\overline{F}(x) := P(X > x) = x^{-\alpha}$, $x \ge 1$, (3) holds with $c = 2$.

3. For the normal distribution we have $c = \infty$.

So, in order to obtain a non-trivial result, it seems necessary to restrict the class of distributions to consider.

Here we consider the case of the so called long-tailed distributions $\mathscr{L}$. For a function ϕ we say $\phi \in \mathscr{L}$ to denote $\phi(x + a) \sim \phi(x)$ for $a \in \mathbb{R}$ as $x \to \infty$. A distribution function F is long tailed if it has infinite upper endpoint and $\overline{F} \in \mathscr{L}$, i.e. if $\overline{F}(x + a) \sim \overline{F}(x)$ as $x \to \infty$ (for $a \in \mathbb{R}$). In this case convergence is uniform on compact subsets of $\mathbb{R}$.

The class $\mathscr{L}$ contains the subclass of the subexponential distributions. A distribution function F is subexponential ($F \in \mathscr{S}$) if

$$P(X_1 + X_2 > x) \sim 2P(X_1 > x) \text{ as } x \to \infty. \tag{4}$$

The property of subexponentiality is usually attributed to a.s. non-negative random variables, but we will not assume this. The theory of subexponential distributions is well established by now and its relevance is obvious from applications in various areas of applied probability. For a recent review of applications of subexponentiality in different areas the reader is referred to the books by Asmussen [1] and Embrechts, Klüppelberg and Mikosch [5].

In case X, Y are independent random variables, we write $F * G$ for the d.f. of the convolution $X + Y$, $\overline{F} * \overline{G}(x) = P(X + Y > x)$. Moreover we will write $H = F * G$ for the d.f. of the sum $X + Y$.

It is well-known that the sum of two independent, not identically distributed subexponential random variables X and Y is not necessarily subexponential. See ([9]) for a counterexample. A useful notion is the so called max-sum equivalence, which was introduced in Embrechts and Goldie ([4]).

Definition 1. *F and G are said to be max-sum equivalent, written $F \sim_M G$ if*

$$\overline{F} * \overline{G}(x) \sim \overline{F}(x) + \overline{G}(x), \ \ x \to \infty. \tag{5}$$

Note that $\overline{F}(x) + \overline{G}(x) \sim \overline{F}(x) + \overline{G}(x) + \overline{F}(x)\overline{G}(x)$, so we may rewrite (5) as

$$P(X + Y > x) \sim P(X \vee Y > x) \ (x \to \infty), \tag{6}$$

which explains the terminology. Subexponentiality can be expressed in terms of the relation $\sim_M$ as well: $F \in \mathscr{S}$ if and only if $F \sim_M F$.

The following result shows that for long tailed distributions, asymptotic equivalence in the symmetrization inequality in the sense of (3) with $c = 2$ is connected to membership of the class of subexponential distributions.

In Embrechts and Goldie [4], Thm. 1, it is proved that if $F \in \mathscr{L}, G \in \mathscr{S}$, and $\sup_x \overline{G}(x)/\overline{F}(x) < \infty$, then $F \in \mathscr{S} \Leftrightarrow F * G \in \mathscr{S}$. A key component in the proof of our main result is the following improvement.

Theorem 2. *If $F, G \in \mathscr{L}$ and $\sup_x \overline{G}(x)/\overline{F}(x) < \infty$, then $F \in \mathscr{S} \Leftrightarrow F * G \in \mathscr{S}$.*
In this case we have
$$\overline{H}(x) \sim \overline{F}(x) + \overline{G}(x) \ (x \to \infty). \tag{7}$$

In the following we use the notation F_s for the d.f. of the symmetrized random variable $X_1 - X_2$ and F_a for the distribution function of $|X_1|$. Moreover we denote the d.f. of $-X$ by F^-.

Theorem 3. *Suppose X_1, X_2 are i.i.d. random variables with unbounded support.*

- *Suppose one endpoint is finite and*

$$P(|X_1| > x + a) \sim P(|X_1| > x) \ for \ a \in \mathbb{R} \ as \ x \to \infty. \tag{8}$$

Then (3) holds with $c = 2$.

3

- *Suppose the distribution has support $(-\infty, \infty)$ and $F, F^- \in \mathscr{L}$. Then the following statements are equivalent.*

    1. *$F_a \in \mathscr{S}$*
    2. *$F_s \in \mathscr{S}$*

    Moreover, if 1 or 2 holds, then (3) holds with $c = 2$.

2 Proofs

In proving properties of subexponential distributions, it is often easy to assume that the variables are non-negative. This is not a very restrictive assumption, since subexponentiality is essentially a property of the right tail of the distribution function. This is confirmed by the following two lemmas, which include subexponential distributions since $\mathscr{S} \subset \mathscr{L}$. We will use the notation $a \vee b$ for the maximum $\max(a, b)$.

Lemma 1. *Let X, Y be independent with distribution functions $F, G \in \mathscr{L}$. Then as $x \to \infty$*

$$P(X^+ \vee Y > x) \sim P(X \vee Y > x). \tag{9}$$

and

$$P(X^+ + Y > x) \sim P(X + Y > x). \tag{10}$$

Proof of Lemma 1. The first part is proved by observing that as $x \to \infty$

$$P(X \vee Y > x) \sim \overline{F}(x) + \overline{G}(x) = P(X^+ > x) + P(Y > x) \sim P(X^+ \vee Y > x).$$

In order to prove the second statement note that

$$P(X^+ + Y > x) = P(X + Y > x) + \int_{-\infty}^{0} (\overline{G}(x) - \overline{G}(x - y)) dF(y).$$

Divide the integral on the right-hand side by $\overline{G}(x)$ and let $x \to \infty$. Application of Lebesgue's theorem on dominated convergence shows that the limit is 0. It follows that

$$P(X^+ + Y > x) = P(X + Y > x) + o(\overline{G}(x)), \tag{11}$$

where $o(\overline{G}(x)) \geq 0$. Note that for $x > 0$

$$\overline{G}(x) = P(Y > x) \leq P(X^+ + Y > x),$$

hence (10) follows. $\qquad\square$

The following result is well known (see e.g. Willekens [13], Omey [10]).

Lemma 2. *$F \in \mathscr{S}$ if and only if $F^+ \in \mathscr{S}$, where F^+ is the d.f. of $X^+ = X \vee 0$. Moreover if X and Y are r.v.'s satisfying $P(X > x) \sim P(Y > x)$ $(x \to \infty)$, then $F \in \mathscr{S}$ if and only if $G \in \mathscr{S}$.*

4

Proof of Lemma 2. Twice application of Lemma 1 (take $F = G$) shows that $P(X^+ + Y^+ > x) \sim P(X + Y > x)$ $(x \to \infty)$. The proof of the second statement now follows from the first part and the corresponding property for non-negative random variables with a subexponential distribution function. $\square$

In the lemmas below we assume that X, Y are independent random variables with distribution functions F, G and $H = F * G$ is the distribution function of $X + Y$.

Lemma 3. *Let $F \in \mathscr{S}$ and $G \in \mathscr{S}$. Then the following are equivalent:*

1. $H \in \mathscr{S}$,

2. $F \sim_M G$,

3. $pF + (1 - p)G \in \mathscr{S}$ *for some (all) p satisfying $0 < p < 1$.*

Proof of Lemma 3. Since $F, G \in \mathscr{S} \subset \mathscr{L}$, it follows from Lemmas 1 and 2 that $P(X + Y > x) \sim P(X^+ + Y^+ > x)$. Since $\mathscr{S}$ is closed under asymptotic equivalence (see e.g. Pakes [11] or Teugels [12]), we may assume that X, Y are a.s. non-negative. In this case the lemma is theorem 2 in Embrechts and Goldie [4]. $\square$

Lemma 4.   a *([4], Thm. 3b) $\mathscr{L}$ is closed under convolution.*

   b *([4], Lemma 1) If $H \in \mathscr{S}$ and the d.f. of $X \vee Y$ is in $\mathscr{L}$, then $F \sim_M G$.*

   c *([8], Theorem 2.1) If $F, G \in \mathscr{L}$ and $m \leq \overline{F}(x)/\overline{G}(x) \leq M, x > 0$ for some constants $m, M \in (0, \infty)$, then $F \in \mathscr{S} \Leftrightarrow G \in \mathscr{S}$.*

   d *$F \in \mathscr{S}$ if and only if the d.f. of $aX + b$ is subexponential for some (all) $a > 0$ and $b \in \mathbb{R}$.*

Proof of Lemma 4.   a-c Note that in view of Lemma 2, we may assume that X, Y are a.s. non-negative.

   d This is an easy consequence of the fact that $\mathscr{S} \subset \mathscr{L}$, $\mathscr{S}$ is closed under asymptotic equivalence and (6).

$\square$

Proof of Theorem 2.   a Suppose $F \in \mathscr{S}$. Again we may assume that X, Y are a.s. non-negative. Note that as in the proof of Prop. 1 in Embrechts et al. [3], we find $\overline{H}(x) = O(\overline{F}(x))$ as $x \to \infty$, hence

$$1 \leq \overline{H}(x)/\overline{F}(x) \leq c \text{ for some constant } c > 0. \tag{12}$$

Since $F \in \mathscr{S} \subset \mathscr{L}$, in view of Lemma 4a we have $H \in \mathscr{L}$. Using the inequality (12) and $F, H \in \mathscr{L}, F \in \mathscr{S}$, application Lemma 4c shows that $H \in \mathscr{S}$. Observe that (since $F, G \in \mathscr{L}$) the d.f. of $X \vee Y$ is in $\mathscr{L}$, being asymptotic to $\overline{F}(x) + \overline{G}(x)$. Because $H \in \mathscr{S}$, application of Lemma 4b shows that (7) holds.

b Suppose $H \in \mathscr{S}$. We may assume that X, Y are a.s. non-negative. Note that $P(X \vee Y > x + a) \sim P(X > x + a) + P(Y > x + a) \sim P(X > x) + P(Y > x) \sim P(X \vee Y > x)$. Since $H \in \mathscr{S}$, we can apply Lemma 4b to see that (7) holds. As a consequence we have $\sup_x \overline{H}(x)/\overline{F}(x) < \infty$. Since $\overline{F}(x)/\overline{H}(x) \leq 1$, we may use Lemma 4c to find $F \in \mathscr{S}$.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of Theorem 3.

- In view of the assumption (8) and the fact that the distribution of $X_1 - X_2$ is symmetric around zero, we may assume that the support is $[0, \infty)$. Then $P(X_1 - X_2 > x) = \int_0^\infty P(X_1 > x + y)dF(y)$ and consequently, by (8) and Fatou's Lemma, it follows that
$$\liminf_{x \to \infty} \int_0^\infty \frac{P(X_1 > x + y)}{P(X_1 > x)}dF(y) \geq 1$$
The integral on the left-hand side is bounded above by 1, using monotonicity, which completes the proof.

- *1 $\Rightarrow$ 2* Take X_1, X_2, X_3 i.i.d. with d.f. F. Since $F_a \in \mathscr{L}$ and $F \in \mathscr{L}$ we can apply Theorem 2a (the only if statement) to find that the d.f. of $X_1 + |X_2|$ is subexponential and

$$P(X_1 + |X_2| > x) \sim 2P(X_1 > x) + P(X_1 < -x).$$

Using Theorem 2a again, it follows that $X_1 - X_3 + |X_2|$ has a d.f. in $\mathscr{S}$ and

$$P(X_1 - X_3 + |X_2| > x) \sim 2P(|X_1| > x), \qquad\qquad (13)$$

hence
$$P(X_1 - X_3 + |X_2| - a > x) \sim 2P(|X_1| > x),$$

where a is as in (2). In view of the symmetrization inequality (2) we may apply Theorem 2b to find that $X_1 - X_3$ has d.f. $F_s \in \mathscr{S}$ and

$$P(X_1 - X_3 + |X_2| > x) \sim P(X_1 - X_3 > x) + P(|X_2| > x). \qquad (14)$$

2 $\Rightarrow$ 1 Note that with a the constant in (2) we may apply Theorem 2a to find that $X_1 - X_2 + |X_3| - a$ has a d.f. in $\mathscr{S}$, hence $X_1 - X_2 + |X_3|$ has a subexponential distribution function.

Note that the distributions of $-X_2$ and $X_1 + |X_3|$ are long tailed and with a the same constant as in (2) we have

$$P(X_1 + |X_3| > x) \geq P(X_1 - X_3 > x) \geq \frac{p}{2}P\{|X_2| > x + a\}$$
$$\sim \frac{p}{2}P(|X_2| > x) \geq \frac{p}{2}P(-X_2 > x).$$

6

Application of Theorem 2b shows that $X_1 + |X_3|$ has a subexponential distribution. Another application of Theorem 2b then shows that the d.f of $|X_3|$ is subexponential, hence $F_a \in \mathscr{S}$.

The proof is complete since combination of (13) and (14) gives (3) with $c = 2$.

$\square$

References

[1] S. Asmussen, Ruin probabilities. *World Scientific Publishing Co., Inc., River Edge, NJ* 2000.

[2] N. Bingham, C. Goldie and J. Teugels (1987). Regular Variation. *Cambridge University Press.*

[3] P. Embrechts, C. Goldie, N. Veraverbeke, Subexponentiality and infinite divisibility, Z. Wahrscheinlichkeitstheorie, 49, (1979), 335-347.

[4] P. Embrechts, C. Goldie, On closure and factorization properties of subexponential and related distributions, J. Austr. Math. Soc. 29, (1980), 243-256.

[5] P. Embrechts, C. Klüppelberg, T. Mikosch (1997). Modelling Extremal Events, *Springer Verlag.*

[6] W. Feller (1971). An Introduction to Probability Theory and Its Applications. *Volume II, John Wiley & Sons, Inc.*

[7] C. Goldie, Subexponential distributions and dominated variation tails, J. Appl. Prob. 15, 1978, 440-442.

[8] C. Klüppelberg, Subexponential distributions and integrated tails. J. Appl. Prob. 25, 1988, 132-141.

[9] J.R. Leslie, On the nonclosure under convolution of the subexponential family, J. Appl. Prob., 26, 1989, 58-66.

[10] E. Omey, On the difference between the distribution function of the sum and the maximum of real random variables, Publ. Inst. Math., 71, 2002, 63-77.

[11] A.G. Pakes, On the tails of waiting time distributions, J. Appl. Prob., 12, 1975, 555-564.

[12] J.L. Teugels, The class of subexponential distributions, Ann. Prob., 6, 1975,1000-1011.

[13] E. Willekens, Hogere orde theorie voor subexponentiële verdelingen, Ph.D. Thesis, University of Leuven, 1986.