THIELE CENTRE

FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Distinguishing log-concavity from heavy tails

Søren Asmussen and Jaakko Lehtomaa

16

Distinguishing log-concavity from heavy tails

Søren Asmussen and Jaakko Lehtomaa

Department of Mathematics, Aarhus University,
{asmus@math.au.dk, jaakkolehtomaa@gmail.com}

Abstract

Well-behaved densities are typically log-convex with heavy tails and log-concave with light ones. We discuss a benchmark for distinguishing between the two cases, based on the observation that large values of a sum $X_1 + X_2$ occur as result of a single big jump with heavy tails whereas X_1, X_2 are of equal order of magnitude in the light-tailed case. The method is based on the ratio $|X_1 - X_2|/(X_1 + X_2)$, for which sharp asymptotic result are presented as well as a visual tool for distinguishing between the two cases. The study supplements modern non-parametric density estimation methods where log-concavity plays a main role, as well as heavy-tailed diagnostics such as the mean excess plot.

Keywords: Heavy-tailed; log-concave; Mean excess function; Principle of a single big jump Please add some keywords

1 Introduction

General interest towards non-parametric thinking has increased over the last few years. One example is density estimation under shape constraints instead of requiring the membership of a parametric family. Here a particular robust alternative to parametric tests is provided by searching for the best fitting log-concave density. Another example is the mean excess plot which aims at distinguishing light and heavy tails.

Throughout the paper, we consider i.i.d. random variables $X, X_1, X_2, \ldots > 0$ with common distribution F having density f and tail $\overline{F}(x) = \mathbb{P}(X > x)$. Then X is (right) *heavy-tailed* if $\mathbb{E}\,e^{sX} = \infty$ for all $s > 0$ and *light-tailed* otherwise. The density f is *log-concave*, if $f(x) = e^{\phi(x)}$, where ϕ is a concave function. If ϕ is convex, then f is *log-convex*. This paper aims to illustrate that light-tailed asymptotic behaviour is associated with log-concave densities. Likewise, log-convexity seems to be connected to heavy-tailed behaviour. One can use the connection to assess potential heavy-tailedness by searching for patterns that are typically present among distributions with log-concave or log-convex densities.

Log-concavity is a widely studied topic in its own right [2, 23]. There also exists a substantial literature of its connections to probability theory and statistics [16, 18]. Several papers concentrate on the statistical estimation of density functions

assuming log-concavity [11, 24]. This is due to the fact that log-concavity provides desirable statistical properties for estimators. For instance, maximum likelihood estimation becomes applicable and the estimate is unique. The topic is discussed in detail in the beginning of [6]. Unfortunately, much less emphasis seems to be put on verification of the log-concavity property itself. Specifically, it seems to be relatively little studied if it is feasible that the sample be generated by a log-concave distribution. See, however, [17, 13].

A distribution with a log-concave density f is necessarily light-tailed. In contrast, f is log-convex in the tail in the standard examples of heavy tails such as regular variation, the lognormal distribution and Weibull case $\overline{F}(x) = \mathrm{e}^{-x^{\alpha}}$ with $\alpha < 1$. An important class of heavy-tailed distributions are the subexponential ones defined by $\mathbb{P}(X_1 + X_2 > d) \sim 2\overline{F}(d)$. The intuition underlying this definition is the principle of a single big jump: $X_1 + X_2$ is large if one of X_1, X_2 is large whereas the other remains typical. This motivates that then

$$R = \frac{|X_1 - X_2|}{X_1 + X_2} \tag{1.1}$$

is close to 1. In contrast, the folklore is that X_1, X_2 contribute equally to $X_1 + X_2$ with light tails. We are not aware of general rigorous formulations of this principle, but it is easily verified in explicit examples like a gamma or normal F, see further below, and for a large number of summands rather than just 2 it is supported by conditioned limit theorems, see e.g. [4, VI.5]. However, it was recently shown in [18] that these properties of R hold in greater generality and that asymptotic properties of the corresponding conditioned random variable

$$Y_d = R \,\big|\, X_1 + X_2 > d. \tag{1.2}$$

provide a sharp borderline between log-convexity and log-concavity.

In this paper we provide a wider perspective in terms of both sharper and more general limit results and of the usefulness for visual statistical data exploration. To this end, we propose a feature based nonparametric test. It can be used as a visual aid in identification of log-concavity or heavy-tailed behaviour. It complements earlier ways to detect signs of heavy-tailedness such as the mean excess plot [15]. Further tests based on probabilistic features have been previously utilised in e.g. [10, 12, 14].

2 Background

A property holds *eventually*, if there exists a number y_0 so that the property holds in the set $[y_0, \infty)$. Standard asymptotic notation is used for limiting statements. These and basic properties of regularly varying functions with parameter α, denoted $\mathrm{RV}(\alpha)$, can be recalled from e.g. [8].

We note that principle of a single big jump relates to the fact that joint distributions of independent random variables concentrate probability mass to different regions. For example, a distribution with tail function $\overline{F}(x) = \mathrm{e}^{-x^{\alpha}}$ satisfies

$$\overline{F}(x) = \mathrm{o}(\overline{F}(x/2)^2)$$

2

for $\alpha > 1$ and

$$\overline{F}(x/2)^2 = \mathrm{o}(\overline{F}(x))$$

for $\alpha \in (0,1)$, as $x \to \infty$. We refer to [1, 3, 9, 22] for related work in this direction. It is shown in Lemma 1.2 of [18] that log-concavity or log-concavity of the density is closely related to the occurrence of the principle of a single big jump. A further observation in this direction is the following lemma. It states that contour lines of joint densities of independent variables behave differently for log-concave and log-convex densities, and thereby leads naturally to different concentration of probability mass of joint densities (recall that a contour line corresponding to a value $p \in \mathbb{R}$ of joint density $f \colon \mathbb{R}^2 \to \mathbb{R}$ is the set of points in the plane defined as $\{(x,y) \in \mathbb{R}^2 : f(x,y) = p\}$.

Lemma 2.1. *Suppose X_1 and X_2 are i.i.d. unbounded non-negative random variables. Assume further that they have a common twice differentiable density function f of the form*

$$f(x) = \mathrm{e}^{-h(x)},$$

where h is strictly increasing function.

If f is log-concave (log-convex) then for any fixed $p \in (0, \mathrm{e}^{-h(0)})$ there exists a convex (concave) function ψ_p defining a contour line of f_{X_1, X_2} corresponding to p such that $f_{X_1, X_2}(x, \psi_p(x)) = p$ for all $x \in [0, h^{-1}(-\log p - h(0))]$.

Lemma 2.1 implies that log-convex and log-concave densities cause maximal points of joint densities to accumulate into different regions in the plane. Log-convex densities tend to put probability mass near the axis, while log-concave densities have a tendency to concentrate mass near the graph of the identity function. The exponential density is the limiting case where all contour lines are straight lines. More generally, for $f_\alpha(x) = C_\alpha e^{-x^\alpha}$, where $C_a > 0$ is an integration constant, the contour lines are circles for $\alpha = 2$, straight lines for $\alpha = 1$ and parabolas, for $\alpha = 1/2$.

3 Theoretical Results

The emphasis of the paper is in the mathematical formulation of the connection between log-convexity and the principle of a single big jump. However, some additional theoretical results are provided concerning convergence rates of the conditional ratio defined in (3.1). These rates, or estimates for the rates, are obtained in some standard distribution classes. Their proofs are mainly based on sharp asymptotics of subexponential distributions obtained in [5, 7, 20]. Recall that some main classes of such distribution are RV(α), meaning regularly varying ones where $\overline{F}(x) = L(x)/x^\alpha$ with $\alpha > 0$ and $L(\cdot)$ slowly varying, Weibull tails with $\overline{F}(x) = \mathrm{e}^{-x^\alpha}$ for some $\alpha \in (0,1)$,and lognormal tails which are close to the case $\gamma = 2$ of $\overline{F}(x) = \mathrm{e}^{-\log^\gamma x}$ for $x \geq 1$ and some $\gamma > 1$; we refer in the following to this class as *lognormal type* tails.

3.1 Convergence Properties

Define the function $g\colon (0, \infty) \to [0, 1]$ by

$$g_X(d) = g(d) = \mathbb{E}\left[\frac{|X_1 - X_2|}{X_1 + X_2} \,\middle|\, X_1 + X_2 > d\right] \tag{3.1}$$

It can be viewed as a generalisation of the function f_{Z_d} considered in [18]. and has the same interpretation as in the case with densities: if both X_1 and X_2 contribute equally to the sum $X_1 + X_2$, then g should eventually obtain values close to 0; similarly, if only one of the variables tends to be the same magnitude as the whole sum, then g is close to 1 for large d. Note also that g is scale independent in the sense that $g_{aX}(d) = g_X(d/a)$ for all $a > 0$. Due to this property, two or more samples can be standardised to have, say, equal means in order to obtain graphs in the same scale.

In Proposition 3.1 sharp asymptotic forms of g are exhibited in some classes of distributions.

Proposition 3.1. *The following convergence rates hold for g defined in* (3.1).

1. *Let X be $RV(\alpha)$ with $\alpha > 1$ and eventually decreasing density f. Then*

$$g(d) = 1 - \frac{c}{d} + o(1/d) \qquad \text{where } c = \frac{2\alpha\, \mathbb{E}[X]}{\alpha + 1}.$$

2. *Let X be Weibull distributed . Then*

$$g(d) = 1 - o(d^{\alpha - 1}). \tag{3.2}$$

3. *Let X be of lognormal type. Then*

$$g(d) = 1 - o(\log^{\gamma - 1} d/d). \tag{3.3}$$

Remark 3.2. In the case of Weibull and lognormal distributions, the implication is that $g(d)$ converges to 1 at a larger rate than their associated hazard rates tend to zero. In addition, inspection of the proof shows

$$\liminf_{d \to \infty} d\,|g(d) - 1| > 0.$$

This implies that the actual convergence rate can not be substantially larger than in the regularly varying case, where the leading term is explicitly identified.

The light-tailed case appears to be more difficult to study than the heavy-tailed case. Difficulty arises mainly from the lack of good asymptotical approximations for probabilities of the form $\mathbb{P}(X_1 + X_2 > d)$ when $\mathbb{P}(X_1 > d)$ decays much faster than e^{-d}. Interestingly, the full asymptotic form of g can be recovered in the special case of the normal distribution if we allow X to obtain negative values.

Proposition 3.3. *Suppose that X is normally distributed with $\mathbb{E}[X] = 0$ and $\mathbb{V}\mathrm{ar}[X] = 1/\sqrt{2}$. Then*

$$g(d) = \frac{c}{d} + o(1/d), \qquad \text{where } c = \mathbb{E}\left[|X_1 - X_2|\right]. \tag{3.4}$$

The following theorem can be used to assess if a sample is coming from a source with log-concave density. It can be seen as a natural contiuation as well as a generalisation to [18].

Theorem 3.4. *Assume the density f is twice differentiable and eventually log-concave. Then*

$$\limsup_{d\to\infty} g(d) \leq \tfrac{1}{2}. \tag{3.5}$$

Similarly, if f is eventually log-convex, then

$$\liminf_{d\to\infty} g(d) \geq \tfrac{1}{2}. \tag{3.6}$$

4 Statistical Application: Visual Test

Suppose $(X_1, Y_2), (X_2, Y_2), \ldots$ is a sequence of i.i.d. vectors whose components are also i.i.d. One can formulate the empirical counterpart of (3.1) by setting

$$\hat{g}(d,n) = \frac{\displaystyle\sum_{k=1}^{n} R_k \mathbf{1}(X_k + Y_k > d)}{\displaystyle\sum_{k=1}^{n} \mathbf{1}(X_k + Y_k > d)}, \tag{4.1}$$

where

$$R_k = \frac{|X_k - Y_k|}{X_k + Y_k}$$

and $\mathbf{1}(A)$ is the indicator function of the event A.

Remark 4.1. Equation (4.1) requires as input a two-dimensional sequence of random variables. One can form such a sequence from a real valued i.i.d. source $Z_1, Z_2, \ldots, Z_N$ using any pairing of the Z_i. Obvious examples are to take $X_k = Z_{2k-1}$, $Y_k = Z_{2k}$, to take the set $\{(X_k, Y_k\}$ as all pairings of the Z_k or as a randomly sampled subset of these $N(N-1)/2$ pairings. If the data is truly i.i.d, this should not have any effect to the outcome.

4.1 Examples and Applications

The function $\hat{g}(d,n)$ can be used to determine if the data support the density being log-concave or light-tailed behaviour. According to Theorem 3.4, the graph should then stay below $1/2$.

The test method is visual. A similar idea has been used at least in the classical mean excess plot, where one visually assesses if the tail excess in the sample points is increasing in the level, as is the case for heavy tails.
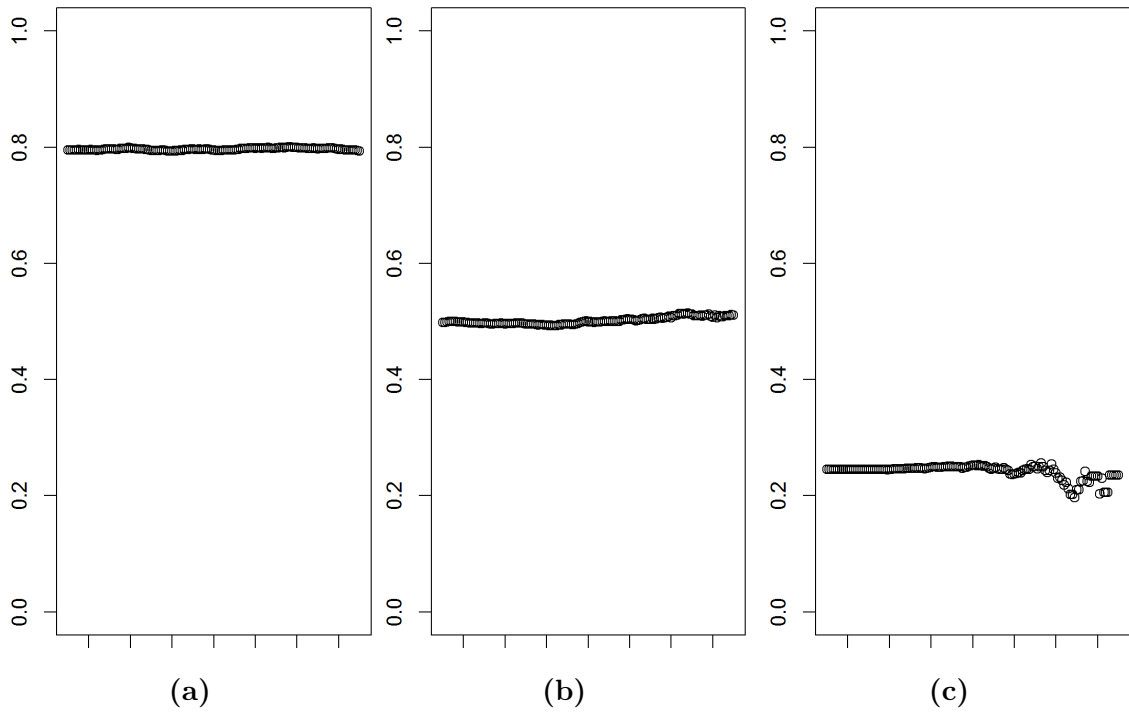
Figure 1: Graphs of $\hat{g}(d, n)$ for $n = 10000$ for Gamma distributed random variables with shapes $0.2, 1$ and 5 in figures (a), (b) and (c), respectively. All variables are standardised to have mean 3.

Figure 2: Graphs of $\hat{g}(d, n)$ for $n = 10000$ for Weibull distributed random variables with shapes $0.2, 1$ and 5 in figures (a), (b) and (c), respectively. All variables are standardised to have mean 3.
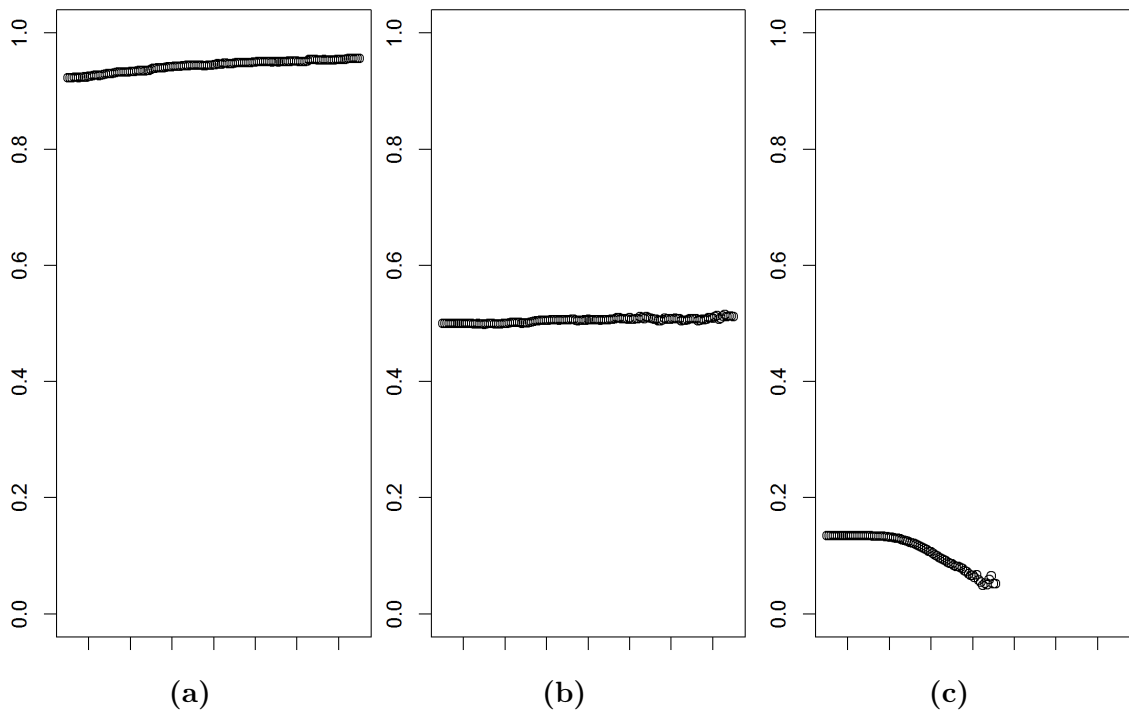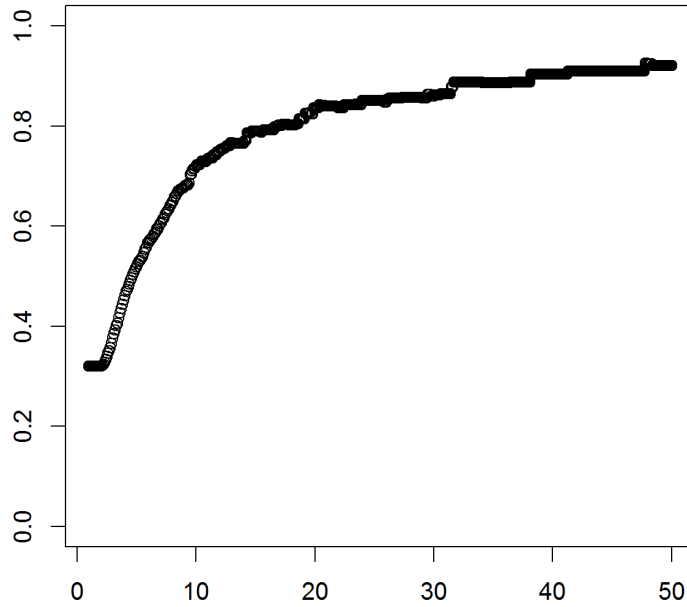
Figure 3: Graph of $\hat{g}(d, n)$ from an classical set of Danish fire insurance data that can be obtained for instance from data set 'danish' in the R package [21]. The sample is traditionally used to illustrate how heavy-tailed data behaves. A similar set of data was previously used in [19]. The graph supports the usual finding that the data set is heavy-tailed.
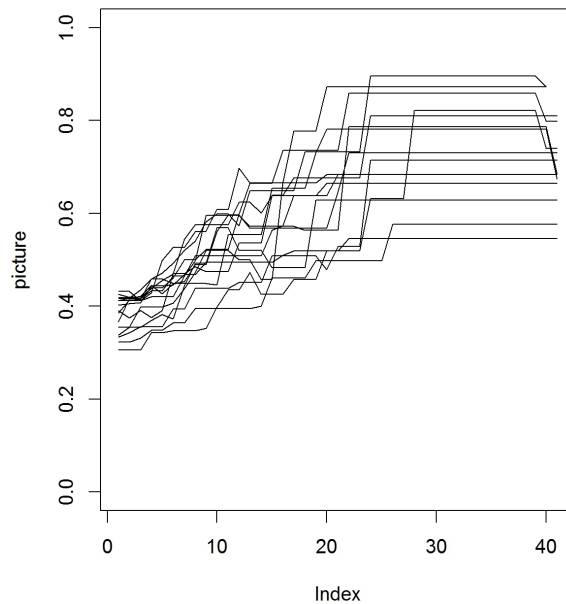

Figure 4: The graphs of multiple versions of $\hat{g}(d, n)$ based on a dataset obtained from Hansjörg Albrecher (private communication) and related to occurrences of floods in a particular area. The sample size is $n = 39$. Bivariate vectors $(X_1, Y_1), \ldots, (X_{19}, Y_{19})$ were sampled several times randomly without replacement from the original data. The overall appearance of the paths points to the data being heavy- rather than light-tailed.

7

5 Proofs

Proof of Lemma 2.1. Suppose h is concave and $p \in (0, 1)$. The contour line corresponding to value p is formed as the set of points (x, y) that satisfy $f_{X_1, X_2}(x, y) = p$, or equivalently

$$h(x) + h(y) = -\log p. \tag{5.1}$$

For any such pair (x, y) one can solve (5.1) for y to obtain

$$y = h^{-1}(-\log p - h(x)). \tag{5.2}$$

Firstly, h^{-1} is convex as the inverse of an increasing concave function. Secondly, the composition of an increasing convex function and a convex function remains convex. Thus, as a function of x, Expression (5.2) defines a convex function when $x \in [0, h^{-1}(-\log p - h(0))]$. So, one can define $\psi_p(x) = h^{-1}(-\log p - h(x))$.

If h is convex the proof is analogous. $\qquad\square$

The following technical lemma is needed in the proof of Proposition 3.1. It applies to Pareto, Weibull and lognormal type distributions. Indeed, condition (5.3) follows from Proposition 1.2. (ii) of [7] and further needed assumptions are easily verified apart from strong subexponentiality, which is known to hold in the mentioned examples.

Lemma 5.1. *Suppose X_1 and X_2 are non-negative i.i.d. variables with a common density f, where the hazard rate $r(d) = f(d)/\overline{F}(d)$ is eventually decreasing with $r(d) = \mathrm{o}(1)$. Assume further that*

$$\mathbb{P}(X_1 + X_2 > d) - 2\,\mathbb{P}(X_1 > d) \sim 2\,\mathbb{E}[X]f(d). \tag{5.3}$$

Then

$$\frac{2\,\mathbb{P}(X_1 > d) + 2f(d)\,\mathbb{E}[X]}{\mathbb{P}(X_1 + X_2 > d)} = 1 + \mathrm{o}(r(d)). \tag{5.4}$$

If in addition $\overline{F}(d/2)^2 = \mathrm{o}(\overline{F}(d))$, then

$$\frac{\mathbb{P}(X_1 \le d/2, X_2 \le d, X_1 + X_2 > d)}{\mathbb{P}(X_1 > d)} = \mathbb{E}[X]r(d) + \mathrm{o}(r(d)). \tag{5.5}$$

Proof. Equality (5.3) implies subexponentiality of X_1. Writing

$$\frac{2\,\mathbb{P}(X_1 > d) + 2\,\mathbb{E}[X]f(d)}{\mathbb{P}(X_1 + X_2 > d)}$$
$$= 1 + \frac{-(\mathbb{P}(X_1 + X_2 > d) - 2\,\mathbb{P}(X_1 > d)) + 2\,\mathbb{E}[X]f(d)}{\mathbb{P}(X_1 + X_2 > d)}$$

and observing that the nominator on the right hand side is of order

$$\mathbb{E}[X]r(d)\mathrm{o}(1)2\overline{F}(d)$$

proves (5.4) since $2\,\mathbb{P}(X_1 > d)/\mathbb{P}(X_1 + X_2 > d) \to 1$ by subexponentiality.

Equality (5.3) implies

$$\frac{\mathbb{P}(X_1 + X_2 > d)}{2\,\mathbb{P}(X_1 > d)} = 1 + \frac{2\,\mathbb{E}[X]f(d)(1 + \mathrm{o}(1))}{2\,\mathbb{P}(X_1 > d)}$$

$$= 1 + \mathbb{E}[X]r(d) + \mathrm{o}(r(d)). \qquad (5.6)$$

On the other hand, writing

$$\{X_1 + X_2 > d, X_2 > X_1\} = \{X_1 \le d/2, X_2 > d\}$$
$$\cup \{X_1 \le d/2, X_2 \le d, X_1 + X_2 > d\}$$
$$\cup \{X_1 > d/2, X_2 > d/2\}$$
$$= A_1 \cup A_2 \cup A_3$$

gives

$$\frac{\mathbb{P}(X_1 + X_2 > d)}{2\,\mathbb{P}(X_1 > d)} = \frac{2\,\mathbb{P}(A_1) + 2\,\mathbb{P}(A_2) + \mathbb{P}(A_3)}{2\,\mathbb{P}(X_1 > d)} = 1 + \frac{\mathbb{P}(A_2)}{\mathbb{P}(X_1 > d)} + \mathrm{o}(r(d)).$$

Since we know from (5.6) that the first order error term must be $\mathbb{E}[X]r(d)$, Equation (5.5) holds. $\qquad\square$

Proof of Proposition 3.1. Suppose X is regularly varying with index α. In light of Lemma 5.1 we only need to establish

$$\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_1 + X_2 > d\right] = \overline{F}(d)\left(1 - \frac{c}{d} + \frac{\alpha}{d}\,\mathbb{E}\,X + \mathrm{o}(1/d)\right). \quad (5.7)$$

The contribution to (5.7) from

$$\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_2 > Ad, X_1 + X_2 > d\right]$$

is of order $O\big(\overline{F}(d/2)\overline{F}(Ad)\big) = O\big(d^{-2\alpha+\epsilon}\big)$ for any $A > 0$ and $\epsilon > 0$. So, it can be neglected. We are left with estimating

$$\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_2 \le Ad, X_1 + X_2 > d\right]$$

$$= \mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_2 \le Ad, X_1 + X_2 > d\right]$$

$$= \int_0^{Ad} \mathbb{E}\left[\frac{X_1 - y}{X_1 + y}; X_1 + y > d\right] f(y)\,\mathrm{d}y.$$

We will bound this quantity from above and below, assuming $A < 1/2$.

Firstly,

$$\int_0^{Ad} \mathbb{E}\left[\frac{X_1 - y}{X_1 + y}; X_1 + y > d\right] f(y)\,\mathrm{d}y$$

$$\le \int_0^{Ad} \mathbb{E}\left[\left(1 - \frac{2(1-A)y}{X_1}\right); X_1 + y > d\right] f(y)\,\mathrm{d}y.$$

9

Now given $X > x$, $X - x$ is approximately distributed as xE for large x where $\mathbb{P}(E > z) = 1/(1 + z)^\alpha$. Hence dominated convergence gives

$$\mathbb{E}\left[\frac{1}{X} \,\Big|\, X > z\right] \sim \mathbb{E}\left[\frac{1}{z(1 + E)}\right], \qquad z \to \infty.$$

We get

$$\int_0^{Ad} \mathbb{E}\left[\left(1 - \frac{2(1 - A)y}{X_1}\right); X_1 + y > d\right] f(y) \, \mathrm{d}y$$

$$\leq \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{(d - y)}(1 + o(1)) \,\mathbb{E}\,\frac{1}{1 + E}\right) \overline{F}(d - y) f(y) \, \mathrm{d}y$$

$$= \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{(d - y)}\frac{\alpha}{\alpha + 1}\right) \overline{F}(d - y) f(y) \, \mathrm{d}y + \eta_1(d)$$

$$\leq \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{d}\frac{\alpha}{\alpha + 1}\right) \left(\overline{F}(d) + y f(d)\right) f(y) \, \mathrm{d}y + \eta_1(d) + \eta_2(d)$$

$$\leq \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{d}\frac{\alpha}{\alpha + 1}\right) \left(\overline{F}(d) + \frac{y\alpha}{d}\overline{F}(d)(1 + A)\right) f(y) \, \mathrm{d}y + \eta_1(d) + \eta_2(d)$$

$$\leq \overline{F}(d)\left(1 - \frac{2(1 - A)\,\mathbb{E}\,X}{d}\frac{\alpha}{\alpha + 1} + \frac{\alpha\,\mathbb{E}\,X(1 + A)}{d}\right) + \eta_1(d) + \eta_2(d).$$

Here the error terms $\eta_1(d)$ and $\eta_2(d)$ are or order $o(\overline{F}(d)/d)$. The latter error comes from Taylor expansion of function $\overline{F}(d - y)$ around point $y = 0$. The fact that f is assumed eventually decreasing guarantees that $f(x) \sim \alpha x^{-\alpha - 1} L(x)$, when $\overline{F}(x) = x^{-\alpha} L(x)$.

Secondly, for the lower bound, we have that

$$\int_0^{Ad} \mathbb{E}\left[\frac{X_1 - y}{X_1 + y}; X_1 + y > d\right] f(y) \, \mathrm{d}y \geq \int_0^{Ad} \mathbb{E}\left[\left(1 - \frac{2y}{X_1}\right); X_1 + y > d\right] f(y) \, \mathrm{d}y.$$

As before, we get

$$\int_0^{Ad} \mathbb{E}\left[\left(1 - \frac{2y}{X_1}\right); X_1 + y > d\right] f(y) \, \mathrm{d}y$$

$$\geq \int_0^{Ad} \left(1 - \frac{2y}{(d - y)}(1 + o(1)) \,\mathbb{E}\,\frac{1}{1 + E}\right) \overline{F}(d - y) f(y) \, \mathrm{d}y$$

$$= \int_0^{Ad} \left(1 - \frac{2y}{(d - y)}\frac{\alpha}{\alpha + 1}\right) \overline{F}(d - y) f(y) \, \mathrm{d}y + \eta_1(d)$$

$$\geq \int_0^{Ad} \left(1 - \frac{2y}{d}\frac{\alpha}{\alpha + 1}\right) \left(\overline{F}(d) + y f(d)\right) f(y) \, \mathrm{d}y + \eta_1(d) + \eta_2(d)$$

$$\geq \int_0^{Ad} \left(1 - \frac{2y}{d}\frac{\alpha}{\alpha + 1}\right) \left(\overline{F}(d) + \frac{y\alpha}{d}\overline{F}(d)(1 - A)\right) f(y) \, \mathrm{d}y + \eta_1(d) + \eta_2(d)$$

$$\geq \overline{F}(d)\left(1 - \frac{2\,\mathbb{E}\,X}{d}\frac{\alpha}{\alpha + 1} + \frac{\alpha\,\mathbb{E}\,X(1 - A)}{d}\right) + \eta_1(d) + \eta_2(d)$$

for error terms η_1 and η_2 of order $o(\overline{F}(d)/d)$.

Repeating the argument with arbitrarily small $A > 0$ and combining the upper and lower estimates allows one to deduce

$$d \left| g(d) - \left(1 - \frac{c}{d}\right) \right| \to 0,$$

as $d \to \infty$, which proves the claim.

Suppose then that X is Weibull distributed. Now assumptions of Lemma 5.1 are satisfied with $r(d) = \alpha d^{\alpha-1}$. Since $\overline{F}(d/2)^2 = \mathrm{O}(e^{-cx^\alpha})$ for some $c > 1$ depending on α, we only need to find the order of

$$\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_2 \le d/2, X_1 + X_2 > d\right]. \tag{5.8}$$

In fact, proceeding similarly as in the regularly varying case, it can be seen that (5.8) equals

$$\int_0^{d/2} \mathbb{E}\left[1 - \frac{2y}{X_1 + y} \,\Big|\, X_1 + y > d\right] \overline{F}(d - y) f(y) \, \mathrm{d}y. \tag{5.9}$$

It is known that $(X_1 - z)/e(z)|X_1 > z$, where $e(z) = 1/(\alpha z^{\alpha-1})$, converges in distribution to a standard exponential variable, as $z \to \infty$. Because $e(z/2)/z = \mathrm{o}(1)$, it holds for $y \in [0, d/2]$ that

$$\mathbb{E}\left[1 - \frac{2y}{X_1 + y} \,\Big|\, X_1 + y > d\right]$$

$$= 1 - \frac{2y}{d} \mathbb{E}\left[\frac{1}{\frac{e(d-y)}{d} \frac{X_1 - (d-y)}{e(d-y)} + 1} \,\Big|\, X_1 + y > d\right]$$

$$= 1 - \frac{2y}{d}(1 + \mathrm{o}(1)),$$

(the interchange of expectation and convergence is justified by dominated convergence). In addition, the same error term can be used for any y.

So, (5.9) can be written as

$$\int_0^{d/2} \left(1 - \frac{2y}{d}(1 + \mathrm{o}(1))\right) \left[\overline{F}(d) + \int_{d-y}^d f(s) \, \mathrm{d}s\right] f(y) \, \mathrm{d}y$$

$$= \overline{F}(d) \int_0^{d/2} \left(1 - \frac{2y}{d}\right) f(y) \, \mathrm{d}y$$

$$+ \overline{F}(d) \frac{\int_0^{d/2} \left[\int_{d-y}^d \left(1 - \frac{2y}{d}(1 + \mathrm{o}(1))\right) f(s) \, \mathrm{d}s\right] f(y) \, \mathrm{d}y}{\overline{F}(d)} + \mathrm{o}(\overline{F}(d)/d).$$

Now, using the definition of A_2 from Lemma 5.1 with Equality (5.5) we get

$$\frac{\int_0^{d/2} \left[\int_{d-y}^d f(s) \, \mathrm{d}s\right] f(y) \, \mathrm{d}y}{\overline{F}(d)} = \frac{\mathbb{P}(A_2)}{\mathbb{P}(X_1 > d)} = \mathbb{E}[X] r(d) + \mathrm{o}(r(d))$$

and

$$\frac{\int_0^{d/2} (2y/d) \left[\int_{d-y}^d f(s) \, \mathrm{d}s\right] f(y) \, \mathrm{d}y}{\overline{F}(d)} = 2\,\mathbb{E}[X_1/d \,|\, A_2] \frac{\mathbb{P}(A_2)}{\mathbb{P}(X_1 > d)} = \mathrm{o}(r(d)),$$

11

since

$$\mathbb{E}[X_1/d \mid A_2] = o(1) \tag{5.10}$$

Equation (5.10) follows from the fact that conditionally to A_2, all probability mass concentrates near small values of X_1/d.

Gathering estimates and using Equation (5.4) of Lemma 5.1 yields

$$
\begin{aligned}
g(d) &= (1 + o(r(d))) \frac{2\,\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_2 \le d/2, X_1 + X_2 > d\right]}{2\,\mathbb{P}(X_1 > d) + 2f(d)\,\mathbb{E}[X]} \\
&= (1 + o(r(d))) \frac{2\overline{F}(d)\left[1 - 2\,\mathbb{E}[X]/d + o(1/d) + \mathbb{E}[X]r(d) - o(r(d)) + o(e^{-(c-1)d^\alpha})\right]}{2\overline{F}(d)(1 + \mathbb{E}[X]r(d))} \\
&= 1 + o(r(d)) = 1 + o(d^{\alpha-1}).
\end{aligned}
$$

This shows (3.2), and (3.3) can be obtained using similar calculations with $e(z) = z/\log^{\gamma-1} z$. $\qquad\square$

Proof of Proposition 3.3. Note first that $X_1 + X_2$ and $X_1 - X_2$ are independent in the normal case. Denote $Z = X_1 + X_2$ so that $Z \sim \mathcal{N}(0, 2)$. Let $e(d) = \overline{F}(d)/f(d)$ be the mean excess function of Z (inverse hazard rate). It is then standard that $e(d)$ is of order $1/d$ and that $(Z - d)e(d)|Z > d$ converges in distribution to a standard exponential. Writing

$$g(d) = \frac{\mathbb{E}[|X_1 - X_2|]}{d}\,\mathbb{E}\left[\frac{1}{\frac{e(d)}{d}\frac{Z-d}{e(d)} + 1}\,\middle|\,Z > d\right], \tag{5.11}$$

it follows in the same way as in the proof or Proposition 3.1 that the r.h.s. of (5.11) is $(c/d)(1 + o(1))$. This proves the claim. $\qquad\square$

Proof of Theorem 3.4. Suppose f is log-concave and twice differentiable. Since

$$g(d) = \frac{\displaystyle\int_d^\infty \int_0^z |1 - 2y/z|\, f(z-y)f(y)\,\mathrm{d}y\,\mathrm{d}z}{\displaystyle\int_d^\infty \int_0^z f(z-y)f(y)\,\mathrm{d}y\,\mathrm{d}z},$$

it suffices to show that for a fixed z it holds that

$$\int_0^1 |1 - 2s| f_{Z_z}(s)\mathrm{d}s \le \tfrac{1}{2},$$

where

$$f_{Z_z}(s) = \frac{f(zs)f(z(1-s))}{\displaystyle\int_0^1 f(zx)f(z(1-x))\,\mathrm{d}x}, \qquad s \in [0,1]. \tag{5.12}$$

In fact, by symmetry, one only needs to show

$$\int_0^{1/2} (1 - 2s)(f_{Z_z}(s) - 1)\mathrm{d}s \le 0. \tag{5.13}$$

It is known from the proof of Proposition 2.1 of [18] that f_{Z_z} is increasing in $[0, 1/2]$. Since f_{Z_z} is non-negative and integrates to one over interval $[0, 1]$, there exists a number $a \in (0, 1/2)$ such that $f_{Z_z}(s) \le 1$ when $s \le a$ and $f_{Z_z}(s) > 1$ when $s > a$. So,

$$\int_0^{1/2} (1 - 2s)(f_{Z_z}(s) - 1) \, \mathrm{d}s$$

$$= \int_0^a (1 - 2s)(f_{Z_z}(s) - 1) \, \mathrm{d}s + \int_a^{1/2} (1 - 2s)(f_{Z_z}(s) - 1) \, \mathrm{d}s$$

$$\le (1 - 2a) \left[\int_0^a (f_{Z_z}(s) - 1) \, \mathrm{d}s + \int_0^a (f_{Z_z}(s) - 1) \, \mathrm{d}s \right] = 0,$$

which proves (5.13). Generally, if f is log-concave and twice differentiable in the set $[x_0, \infty)$, then f_{Z_z} is increasing in the set $[x_0/z, 1/2]$. The difference to the presented calculation vanishes in the limit $d \to \infty$ and thus (3.5) holds.

If f is eventually log-convex the proof is analogous and (3.6) holds. $\qquad \square$

References

[1] ALBRECHER, H., ROBERT, C. Y., AND TEUGELS, J. L. Joint asymptotic distributions of smallest and largest insurance claims. *Risks 2*, 3 (2014), 289–314.

[2] AN, M. Y. Logconcavity versus logconvexity: a complete characterization. *J. Econom. Theory 80*, 2 (1998), 350–369.

[3] ARMENDÁRIZ, I., AND LOULAKIS, M. Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic Process. Appl. 121*, 5 (2011), 1138–1147.

[4] ASMUSSEN, S., AND GLYNN, P. W. *Stochastic simulation: algorithms and analysis*, vol. 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

[5] ASMUSSEN, S., AND KORTSCHAK, D. Error rates and improved algorithms for rare event simulation with heavy Weibull tails. *Methodol. Comput. Appl. Probab. 17*, 2 (2015), 441–461.

[6] BALABDAOUI, F., RUFIBACH, K., AND WELLNER, J. A. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist. 37*, 3 (2009), 1299–1331.

[7] BALTRŪNAS, A., AND OMEY, E. The rate of convergence for subexponential distributions and densities. *Liet. Mat. Rink. 42*, 1 (2002), 1–18.

[8] BINGHAM, N. H., GOLDIE, C. M., AND TEUGELS, J. L. *Regular variation*, vol. 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[9] BOBKOV, S. G., AND CHISTYAKOV, G. P. On concentration functions of random variables. *J. Theoret. Probab. 28*, 3 (2015), 976–988.

[10] CROVELLA, M. E., AND TAQQU, M. S. Estimating the heavy tail index from scaling properties. *Methodol. Comput. Appl. Probab. 1*, 1 (1999), 55–79.

[11] CULE, M., SAMWORTH, R., AND STEWART, M. Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol. 72*, 5 (2010), 545–607.

[12] DEL CASTILLO, J., DAOUDI, J., AND LOCKHART, R. Methods to distinguish between polynomial and exponential tails. *Scand. J. Stat. 41*, 2 (2014), 382–393.

[13] DÜMBGEN, L., AND RUFIBACH, K. Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Bernoulli 15*, 1 (2009), 40–68.

[14] GEL, Y. R., MIAO, W., AND GASTWIRTH, J. L. Robust directed tests of normality against heavy-tailed alternatives. *Comput. Statist. Data Anal. 51*, 5 (2007), 2734–2746.

[15] GHOSH, S., AND RESNICK, S. A discussion on mean excess plots. *Stochastic Process. Appl. 120*, 8 (2010), 1492–1517.

[16] GUPTA, R. C., AND BALAKRISHNAN, N. Log-concavity and monotonicity of hazard and reversed hazard functions of univariate and multivariate skew-normal distributions. *Metrika 75*, 2 (2012), 181–191.

[17] HAZELTON, M. L. Assessing log-concavity of multivariate densities. *Statist. Probab. Lett. 81*, 1 (2011), 121–125.

[18] LEHTOMAA, J. Limiting behaviour of constrained sums of two variables and the principle of a single big jump. *Statist. Probab. Lett. 107* (2015), 157–163.

[19] MCNEIL, A. Estimating the tails of loss severity distributions using extreme value theory. *Astin Bull. 27*, 01 (1997), 117–137.

[20] OMEY, E., AND WILLEKENS, E. Second order behaviour of the tail of a subordinated probability distribution. *Stochastic Process. Appl. 21*, 2 (1986), 339–353.

[21] PFAFF, B., AND MCNEIL, A. *evir: Extreme Values in R*, 2012. R package version 1.7-3, `https://CRAN.R-project.org/package=evir`.

[22] PRUSS, A. R. Comparisons between tail probabilities of sums of independent symmetric random variables. *Ann. Inst. H. Poincaré Probab. Statist. 33*, 5 (1997), 651–671.

[23] SAUMARD, A., AND WELLNER, J. A. Log-concavity and strong log-concavity: a review. *Stat. Surv. 8* (2014), 45–114.

[24] WALTHER, G. Inference and modeling with log-concave distributions. *Statist. Sci. 24*, 3 (2009), 319–327.