

Article

# Distinguishing Log-Concavity from Heavy Tails

Søren Asmussen \* and Jaakko Lehtomaa

Department of Mathematics, Aarhus University, Ny Munkegade 118, DK-8000 Aarhus C, Denmark; jaakkolehtomaa@gmail.com

\* Correspondence: asmus@math.au.dk; Tel.: +45-2066-3866

Academic Editor: Qihe Tang

Received: 14 November 2016; Accepted: 17 January 2017; Published: 7 February 2017

**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 results 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

**MSC:** 60E05; 60G50; 62G20

## 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 particularly robust alternative to parametric tests is provided by searching for the best fitting log-concave density. Another example is the mean excess plot that aims at distinguishing light and heavy tails.

Throughout the paper, we consider i.i.d. random variables  $X, X_1, X_2, \dots > 0$  with common distribution  $F$  having density  $f$  and tail  $\bar{F}(x) = \mathbb{P}(X > x)$ . 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  $\phi$  is a concave function. If  $\phi$  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 [1,2]. There also exists substantial literature regarding its connections to probability theory and statistics [3,4]. Several papers concentrate on the statistical estimation of density functions assuming log-concavity [5,6]. 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 [7]. 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, for example [8,9].

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  $\bar{F}(x) = e^{-x^\alpha}$  with  $\alpha < 1$ . An important class of heavy-tailed distributions are the

subexponential ones defined by  $\mathbb{P}(X_1 + X_2 > d) \sim 2\bar{F}(d)$ . 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 then motivates

$$R = \frac{|X_1 - X_2|}{X_1 + X_2} \tag{1}$$

being 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., ([10] (VI.5))). However, it was recently shown in [4] that these properties of  $R$  hold in greater generality and that asymptotic properties of the corresponding conditioned random variable

$$Y_d = R \mid X_1 + X_2 > d \tag{2}$$

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 [11]. Further tests based on probabilistic features have been previously utilised in e.g., [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  $\alpha$ , denoted  $RV(\alpha)$ , can be recalled from e.g., [15].

We note that the 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  $\bar{F}(x) = e^{-x^\alpha}$  satisfies

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

for  $\alpha > 1$  and

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

for  $\alpha \in (0, 1)$ , as  $x \rightarrow \infty$ . We refer to [16–19] for related work in this direction. It is shown in Lemma 1.2 of [4] that log-concavity or log-convexity 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 concentrations of probability mass of joint densities (recall that a contour line corresponding to a value  $p \in \mathbb{R}$  of joint density  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  is the set of points in the plane defined as  $\{(x, y) \in \mathbb{R}^2 : f(x, y) = p\}$ ).

**Lemma 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

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

where  $h$  is a strictly increasing function.

If  $f$  is log-concave (log-convex), then, for any fixed  $p \in (0, e^{-h(0)})$ , there exists a convex (concave) function  $\psi_p$  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 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  $C_\alpha > 0$  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 on 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 Equation (3). 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 [20–22]. Recall that some main classes of such distributions are  $RV(\alpha)$ , meaning regularly varying ones, where  $\bar{F}(x) = L(x)/x^\alpha$  with  $\alpha > 0$  and  $L(\cdot)$  are slowly varying, Weibull tails with  $\bar{F}(x) = e^{-x^\alpha}$  for some  $\alpha \in (0, 1)$ , and lognormal tails which are close to the case  $\gamma = 2$  of  $\bar{F}(x) = e^{-\log^\gamma x}$  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  $g: (0, \infty) \rightarrow [0, 1]$  by

$$g_X(d) = g(d) = \mathbb{E} \left[ \frac{|X_1 - X_2|}{X_1 + X_2} \mid X_1 + X_2 > d \right]. \tag{3}$$

It can be viewed as a generalisation of the function  $f_{Z_d}$  considered in [4] 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 on the same scale.

In Proposition 1, sharp asymptotic forms of  $g$  are exhibited in some classes of distributions.

**Proposition 1.** The following convergence rates hold for  $g$  defined in Equation (3).

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

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

2. Let  $X$  be Weibull distributed with  $\alpha < 1$ . Then,

$$g(d) = 1 - o(d^{\alpha-1}). \tag{4}$$

3. Let  $X$  be of lognormal type. Then,

$$g(d) = 1 - o(\log^{\gamma-1} x/x). \tag{5}$$

**Remark 1.** 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

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

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 2.** Suppose that  $X$  is normally distributed with  $\mathbb{E}[X] = 0$  and  $\text{Var}[X] = 1/\sqrt{2}$ . Then,

$$g(d) = \frac{c}{d} + o(1/d), \quad \text{where } c = \mathbb{E}[|X_1 - X_2|]. \tag{6}$$

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 continuation as well as a generalisation to [4].

**Theorem 1.** Assume the density  $f$  is twice differentiable and eventually log-concave. Then,

$$\limsup_{d \rightarrow \infty} g(d) \leq \frac{1}{2}. \tag{7}$$

Similarly, if  $f$  is eventually log-convex, then

$$\liminf_{d \rightarrow \infty} g(d) \geq \frac{1}{2}. \tag{8}$$

#### 4. Statistical Application: Visual Test

Suppose  $(X_1, Y_2), (X_2, Y_2), \dots$  is a sequence of i.i.d. vectors whose components are also i.i.d. One can formulate the empirical counterpart of Quantity (3) by setting

$$\hat{g}(d, n) = \frac{\sum_{k=1}^n R_k \mathbf{1}(X_k + Y_k > d)}{\sum_{k=1}^n \mathbf{1}(X_k + Y_k > d)}, \tag{9}$$

where

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

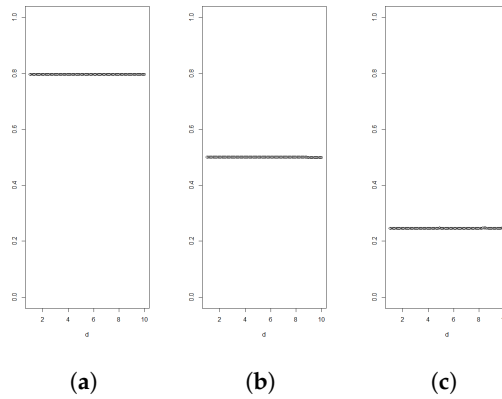
and  $\mathbf{1}(A)$  is the indicator function of the event  $A$ .

**Remark 2.** Equation (9) requires as input a two-dimensional sequence of random variables. One can form such a sequence from a real valued i.i.d. source  $Z_1, Z_2, \dots, Z_N$  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 on the outcome.

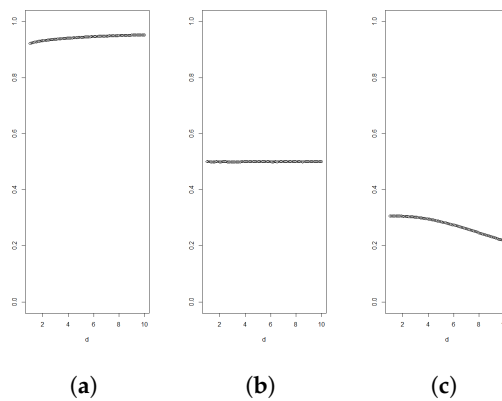
##### 4.1. Examples and Applications

A graph of  $\hat{g}(d, n)$  as function of  $d$  can be used to determine if the data support the density being log-concave or light-tailed behaviour. According to Theorem 1, the graph should then stay below  $1/2$ . Figures 1–4 illustrate such graphs using experimental data.

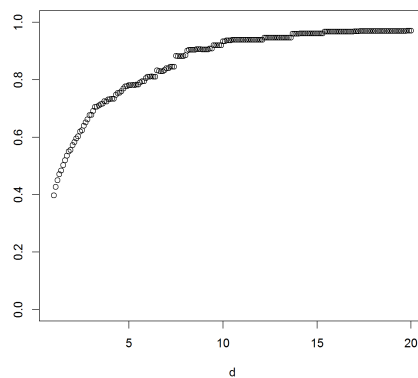
The test method is visual. A similar idea has been used at least in the classical mean excess plot, where one visually assesses if the tail excess in the sample points is increasing at the level, as is the case for heavy tails.



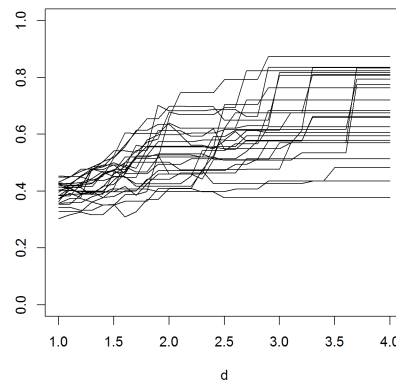
**Figure 1.** Graphs of  $\hat{g}(d, n)$  for  $n = 10,000$  for Gamma distributed random variables with shapes 0.2, 1 and 5 in figures (a), (b) and (c), respectively. All variables are standardised to have mean 3.



**Figure 2.** Graphs of  $\hat{g}(d, n)$  for  $n = 10,000$  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 a classical set of Danish fire insurance data that can be obtained for instance from data set ‘danish’ in the R package [23]. The data is scaled to have mean 1. The sample is traditionally used to illustrate how heavy-tailed data behaves. A similar set of data was previously used in [24]. 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 data is scaled to have mean 1. The sample size is  $n = 39$ . Bivariate vectors  $(X_1, Y_1), \dots, (X_{19}, Y_{19})$  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.

#### 4.2. Finer Diagnostics

The idea of plotting  $\hat{g}(d, n)$  as a function of  $d$  was introduced as a graphical test for distinguishing between heavy-tailed and log-concave light tailed distributions. It seems reasonable to ask if the plot can be used for finer diagnostics, in particular to further specify the tail behaviour of  $F$  when  $F$  was found to be heavy-tailed. Such an idea would be based on the rate of convergence of  $\hat{g}(d, n)$  to 1.

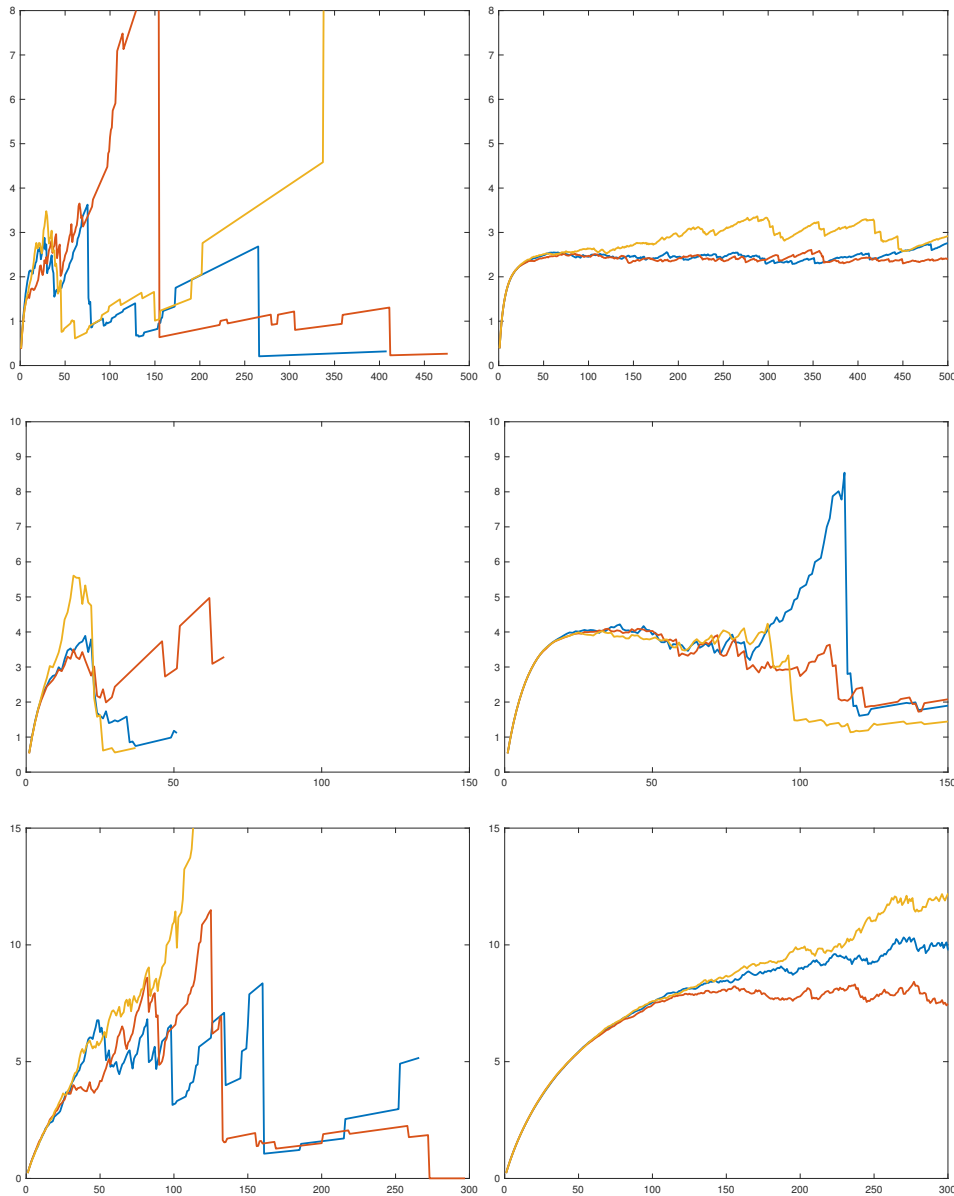
To gain some preliminary insight, we simulated  $R = 5000 = 5 \cdot 10^3$  and  $R = 5 \cdot 10^6$  i.i.d pairs of r.v.'s  $X_1, X_2$  from an  $F$  which was either  $RV(1.5)$ ,  $\text{lognormal}(0,1)$  or  $\text{Weibull}(0.4)$ . The results are in Figure 5 as plots of  $d(1 - \hat{g}(d, n))$ , with three runs in each subfigure.

A first conclusion is that a sample size of  $R = 5000$  is grossly insufficient for drawing conclusions about the way in which  $\hat{g}(d, n)$  approaches 1—random fluctuations take over long before a systematic trend is apparent. The sample size  $R = 5 \cdot 10^6$  is presumably unrealistic in most cases, but even for this, the picture is only clear in the  $RV$  case. Here,  $d(1 - \hat{g}(d, n))$  seems to have a limit  $c$ , as it should be, and the plot is in good agreement with the value  $2.4 = 2\alpha\mathbb{E}[X]/(\alpha + 1)$  of  $c$  predicted by Proposition 1.

Whether a limit exists in the  $\text{lognormal}$  or  $\text{Weibull}$  case is less clear. The results of Proposition 1 are less definite here, but, actually, a heuristic argument suggests that the limit  $c$  should exist and be  $2\mathbb{E}[X]$ . To this end, let  $X_{(1)} < X_{(2)}$  be the order statistics. According to subexponential theory (see, in particular, [25]),  $X_{(1)}, X_{(2)}$  are asymptotically independent given  $X_1 + X_2 > d$ , with  $X_{(1)}$  having asymptotic distribution  $F$  and  $X_{(2)}$  being of the form  $d + e(d)E$  with  $e(d)$ , and  $E$  as in the proof of Proposition 1. For large  $d$ , this gives

$$\frac{|X_1 - X_2|}{X_1 + X_2} \sim \frac{X_{(2)} - X_{(1)}}{X_{(2)}} \sim \frac{1 - X_{(1)}/X_{(2)}}{1 + X_{(1)}/X_{(2)}} \sim 1 - 2X_{(1)}/X_{(2)}. \tag{10}$$

In the  $\text{lognormal}$  or  $\text{Weibull}$  case, one has  $e(d) = o(d)$  and so  $d + e(d)E \sim d$ . Taking expectations gives the conjecture.



**Figure 5.**  $d(1 - \hat{g}(d, n))$ . Pareto in the first row, lognormal in the second, and Weibull in the last.  $R = 5000$  (left),  $R = 5 \times 10^6$  (right).

For the Weibull,  $2\mathbb{E}[X] = 6.65$ , and the  $R = 5 \cdot 10^6$  part of Figure 5 is rather inconclusive concerning the conjecture. We did one more run with  $R = 5 \cdot 10^7$  over a range of parameters (the variance  $\sigma^2$  of  $\log X$  for the lognormal and  $\beta$  for the Weibull). All  $X_1, X_2$  were normalized to have mean 1 so that the conjecture would assert convergence to  $c = 2$ . This is not seen in the results in Figure 6. Large values of  $\sigma^2$  and small values of  $\beta$  could appear to give convergence, but not to 2.

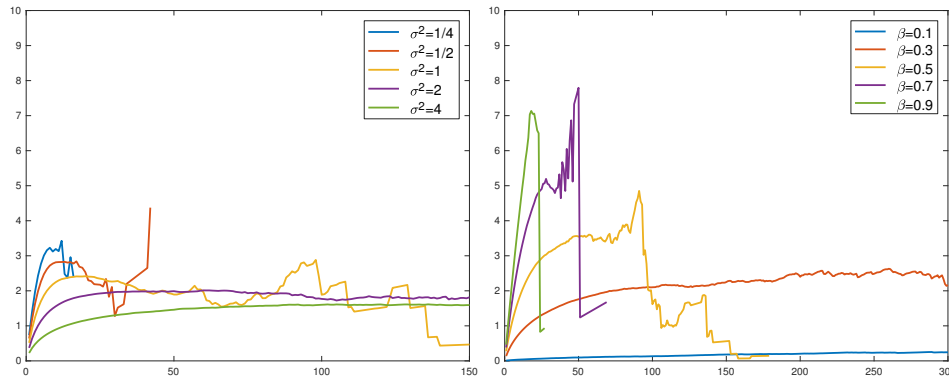


Figure 6.  $R = 5 \times 10^7$ . Lognormal (left), Weibull (right).

It should be noted that the heuristics give the correct result in the RV case. Namely, here we can take  $e(d) = d$  and  $\mathbb{P}(E > x) = 1/(1+x)^\alpha$ . This easily gives  $\mathbb{E}1/(1+E) = \alpha/(\alpha+1)$  so that Quantity (10) is approximately  $1 - 2\alpha\mathbb{E}[X]/d(\alpha+1)$ , as rigorously verified in Proposition 1.

The overall conclusion is that the finer diagnostic value of the method is quite limited, and restricted to RV and sample sizes which may be unrealistically large in many contexts.

5. Proofs

**Proof of Lemma 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

$$h(x) + h(y) = -\log p. \tag{11}$$

For any such pair  $(x, y)$  one can solve Equation (11) for  $y$  to obtain

$$y = h^{-1}(-\log p - h(x)). \tag{12}$$

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 (12) defines a convex function when  $x \in [0, h^{-1}(-\log p - h(0))]$ . Thus, one can define  $\psi_p(x) = h^{-1}(-\log p - h(x))$ .

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

The following technical lemma is needed in the proof of Proposition 1. It applies to Pareto, Weibull and lognormal type distributions. Indeed, condition (13) follows from Proposition 1.2; (ii) of [21] and further needed assumptions are easily verified apart from strong subexponentiality, which is known to hold in the mentioned examples.

**Lemma 2.** Suppose  $X_1$  and  $X_2$  are non-negative i.i.d. variables with a common density  $f$ , where the hazard rate  $r(d) = f(d)/\bar{F}(d)$  is eventually decreasing with  $r(d) = o(1)$ . Assume further that

$$\mathbb{P}(X_1 + X_2 > d) - 2\mathbb{P}(X_1 > d) \sim 2\mathbb{E}[X]f(d). \tag{13}$$

Then,

$$\frac{2\mathbb{P}(X_1 > d) + 2f(d)\mathbb{E}[X]}{\mathbb{P}(X_1 + X_2 > d)} = 1 + o(r(d)). \tag{14}$$

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

$$\frac{\mathbb{P}(X_1 \leq d/2, X_2 \leq d, X_1 + X_2 > d)}{\mathbb{P}(X_1 > d)} = \mathbb{E}[X]r(d) + o(r(d)). \tag{15}$$



**Proof.** Equality (13) 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  $\mathbb{E}[X]r(d)o(1)2\bar{F}(d)$  proves Equation (14) since  $2\mathbb{P}(X_1 > d)/\mathbb{P}(X_1 + X_2 > d) \rightarrow 1$  by subexponentiality.

Equality (13) implies

$$\begin{aligned} \frac{\mathbb{P}(X_1 + X_2 > d)}{2\mathbb{P}(X_1 > d)} &= 1 + \frac{2\mathbb{E}[X]f(d)(1 + o(1))}{2\mathbb{P}(X_1 > d)} \\ &= 1 + \mathbb{E}[X]r(d) + o(r(d)). \end{aligned} \tag{16}$$

On the other hand, writing

$$\begin{aligned} \{X_1 + X_2 > d, X_2 > X_1\} &= \{X_1 \leq d/2, X_2 > d\} \\ &\cup \{X_1 \leq d/2, X_2 \leq d, X_1 + X_2 > d\} \\ &\cup \{X_1 > d/2, X_2 > d/2\} \\ &= A_1 \cup A_2 \cup A_3 \end{aligned}$$

gives

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

Since we know from Equation (16) that the leading term tending to zero must be  $\mathbb{E}[X]r(d)$ , Equation (15) holds.  $\square$

**Proof of Proposition 1.** Suppose  $X$  is regularly varying with index  $\alpha$ . In light of Lemma 2, we only need to establish

$$\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_1 + X_2 > d\right] = \bar{F}(d)\left(1 - \frac{c}{d} + \frac{\alpha}{d}\mathbb{E}X + o(1/d)\right). \tag{17}$$

The contribution to the l.h.s of Equation (17) 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  $O(\bar{F}(d/2)\bar{F}(Ad)) = O(d^{-2\alpha+\epsilon})$  for any  $A > 0$  and  $\epsilon > 0$ . Thus, it can be neglected. We are left with estimating

$$\begin{aligned} &\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_2 \leq Ad, X_1 + X_2 > d\right] \\ &= \mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_2 \leq Ad, X_1 + X_2 > d\right] \\ &= \int_0^{Ad} \mathbb{E}\left[\frac{X_1 - y}{X_1 + y}; X_1 + y > d\right]f(y) dy. \end{aligned}$$

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

Firstly,

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

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

$$\mathbb{E}\left[\frac{1}{X} \mid X > z\right] \sim \mathbb{E}\left[\frac{1}{z(1 + E)}\right], z \rightarrow \infty.$$

We get

$$\begin{aligned} & \int_0^{Ad} \mathbb{E}\left[\left(1 - \frac{2(1 - A)y}{X_1}\right); X_1 + y > d\right] f(y) dy \\ & \leq \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{(d - y)}(1 + o(1))\mathbb{E}\frac{1}{1 + E}\right) \bar{F}(d - y) f(y) dy \\ & = \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{(d - y)} \frac{\alpha}{\alpha + 1}\right) \bar{F}(d - y) f(y) dy + \eta_1(d) \\ & \leq \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{d} \frac{\alpha}{\alpha + 1}\right) (\bar{F}(d) + yf(d)) f(y) dy + \eta_1(d) + \eta_2(d) \\ & \leq \int_0^{Ad} \left(1 - \frac{2(1 - A)y}{d} \frac{\alpha}{\alpha + 1}\right) (\bar{F}(d) + \frac{y\alpha}{d} \bar{F}(d)(1 + A)) f(y) dy + \eta_1(d) + \eta_2(d) \\ & \leq \bar{F}(d) \left(1 - \frac{2(1 - A)\mathbb{E}X}{d} \frac{\alpha}{\alpha + 1} + \frac{\alpha\mathbb{E}X(1 + A)}{d}\right) + \eta_1(d) + \eta_2(d). \end{aligned}$$

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

Secondly, for the lower bound, we have that

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

As before, we get

$$\begin{aligned} & \int_0^{Ad} \mathbb{E}\left[\left(1 - \frac{2y}{X_1}\right); X_1 + y > d\right] f(y) dy \\ & \geq \int_0^{Ad} \left(1 - \frac{2y}{(d - y)}(1 + o(1))\mathbb{E}\frac{1}{1 + E}\right) \bar{F}(d - y) f(y) dy \\ & = \int_0^{Ad} \left(1 - \frac{2y}{(d - y)} \frac{\alpha}{\alpha + 1}\right) \bar{F}(d - y) f(y) dy + \eta_1(d) \\ & \geq \int_0^{Ad} \left(1 - \frac{2y}{d} \frac{\alpha}{\alpha + 1}\right) (\bar{F}(d) + yf(d)) f(y) dy + \eta_1(d) + \eta_2(d) \\ & \geq \int_0^{Ad} \left(1 - \frac{2y}{d} \frac{\alpha}{\alpha + 1}\right) (\bar{F}(d) + \frac{y\alpha}{d} \bar{F}(d)(1 - A)) f(y) dy + \eta_1(d) + \eta_2(d) \\ & \geq \bar{F}(d) \left(1 - \frac{2\mathbb{E}X}{d} \frac{\alpha}{\alpha + 1} + \frac{\alpha\mathbb{E}X(1 - A)}{d}\right) + \eta_1(d) + \eta_2(d) \end{aligned}$$

for error terms  $\eta_1$  and  $\eta_2$  of order  $o(\bar{F}(d)/d)$ .

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

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

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

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

$$\mathbb{E}\left[\frac{X_1 - X_2}{X_1 + X_2}; X_1 > X_2, X_2 \leq d/2, X_1 + X_2 > d\right]. \tag{18}$$

In fact, proceeding similarly as in the regularly varying case, it can be seen that Quantity (18) equals

$$\int_0^{d/2} \mathbb{E}\left[1 - \frac{2y}{X_1 + y} \mid X_1 + y > d\right] \bar{F}(d - y) f(y) dy. \tag{19}$$

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

$$\begin{aligned} & \mathbb{E}\left[1 - \frac{2y}{X_1 + y} \mid X_1 + y > d\right] \\ &= 1 - \frac{2y}{d} \mathbb{E}\left[\frac{1}{\frac{e(d-y)}{d} \frac{X_1 - (d-y)}{e(d-y)} + 1} \mid X_1 + y > d\right] \\ &= 1 - \frac{2y}{d} (1 + o(1)) \end{aligned}$$

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

Thus, Quantity (19) can be written as

$$\begin{aligned} & \int_0^{d/2} \left(1 - \frac{2y}{d} (1 + o(1))\right) \left[\bar{F}(d) + \int_{d-y}^d f(s) ds\right] f(y) dy \\ &= \bar{F}(d) \int_0^{d/2} \left(1 - \frac{2y}{d}\right) f(y) dy \\ &+ \frac{\bar{F}(d) \int_0^{d/2} \left[\int_{d-y}^d \left(1 - \frac{2y}{d} (1 + o(1))\right) f(s) ds\right] f(y) dy}{\bar{F}(d)} + o(\bar{F}(d)/d). \end{aligned}$$

Now, using the definition of  $A_2$  from Lemma 2 with Equality (15), we get

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

and

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

since

$$\mathbb{E}[X_1/d \mid A_2] = o(1). \tag{20}$$

Equation (20) 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 (14) of Lemma 2 yields

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

This shows Equation (4), and Equation (5) can be obtained using similar calculations with  $e(z) = z/\log^{\gamma-1} z$ .  $\square$

**Proof of Proposition 2.** 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  $e(d) = \bar{F}(d)/f(d)$  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

$$g(d) = \frac{\mathbb{E}[|X_1 - X_2|]}{d} \mathbb{E}\left[\frac{1}{\frac{e(d)}{d} \frac{Z-d}{e(d)} + 1} \mid Z > d\right], \tag{21}$$

it follows in the same way as in the proof of Proposition 1 that the r.h.s. of Equation (21) is  $(c/d)(1 + o(1))$ . This proves the claim.  $\square$

**Proof of Theorem 1.** Suppose  $f$  is log-concave and twice differentiable. Since

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

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

$$\int_0^1 |1 - 2s| f_{Z_z}(s) ds \leq \frac{1}{2},$$

where

$$f_{Z_z}(s) = \frac{f(zs)f(z(1-s))}{\int_0^1 f(zx)f(z(1-x)) dx}, \quad s \in [0, 1]. \tag{22}$$

In fact, by symmetry, one only needs to show

$$\int_0^{1/2} (1 - 2s)(f_{Z_z}(s) - 1) ds \leq 0. \tag{23}$$

It is known from the proof of Proposition 2.1 of [4] 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  $f_{Z_z}(s) \leq 1$  when  $s \leq a$  and  $f_{Z_z}(s) > 1$  when  $s > a$ . Therefore,

$$\begin{aligned} &\int_0^{1/2} (1 - 2s)(f_{Z_z}(s) - 1) ds \\ &= \int_0^a (1 - 2s)(f_{Z_z}(s) - 1) ds + \int_a^{1/2} (1 - 2s)(f_{Z_z}(s) - 1) ds \\ &\leq (1 - 2a) \left[ \int_0^a (f_{Z_z}(s) - 1) ds + \int_0^a (f_{Z_z}(s) - 1) ds \right] = 0, \end{aligned}$$

which proves Inequality (23). 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 from the presented calculation vanishes in the limit  $d \rightarrow \infty$ , and thus Inequality (7) holds.

If  $f$  is eventually log-convex, the proof is analogous and Inequality (8) holds.  $\square$

**Author Contributions:** The authors contributed equally to this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- An, M.Y. Logconcavity versus logconvexity: A complete characterization. *J. Econom. Theory* **1998**, *80*, 350–369.
- Saumard, A.; Wellner, J.A. Log-concavity and strong log-concavity: A review. *Stat. Surv.* **2014**, *8*, 45–114.
- Gupta, R.C.; Balakrishnan, N. Log-concavity and monotonicity of hazard and reversed hazard functions of univariate and multivariate skew-normal distributions. *Metrika* **2012**, *75*, 181–191.
- Lehtomaa, J. Limiting behaviour of constrained sums of two variables and the principle of a single big jump. *Stat. Probab. Lett.* **2015**, *107*, 157–163.
- Cule, M.; Samworth, R.; Stewart, M. Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2010**, *72*, 545–607.
- Walther, G. Inference and modeling with log-concave distributions. *Stat. Sci.* **2009**, *24*, 319–327.
- Balabdaoui, F.; Rufibach, K.; Wellner, J.A. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Stat.* **2009**, *37*, 1299–1331.
- Hazelton, M.L. Assessing log-concavity of multivariate densities. *Stat. Probab. Lett.* **2011**, *81*, 121–125.
- Dümbgen, L.; Rufibach, K. Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **2009**, *15*, 40–68.
- Asmussen, S.; Glynn, P.W. Stochastic simulation: Algorithms and analysis. In *Stochastic Modelling and Applied Probability*; Springer: New York, NY, USA, 2007; Volume 57.
- Ghosh, S.; Resnick, S. A discussion on mean excess plots. *Stoch. Process. Appl.* **2010**, *120*, 1492–1517.
- Crovella, M.E.; Taqqu, M.S. Estimating the heavy tail index from scaling properties. *Methodol. Comput. Appl. Probab.* **1999**, *1*, 55–79.
- Del Castillo, J.; Daoudi, J.; Lockhart, R. Methods to distinguish between polynomial and exponential tails. *Scand. J. Stat.* **2014**, *41*, 382–393.
- Gel, Y.R.; Miao, W.; Gastwirth, J.L. Robust directed tests of normality against heavy-tailed alternatives. *Comput. Stat. Data Anal.* **2007**, *51*, 2734–2746.
- Bingham, N.H.; Goldie, C.M.; Teugels, J.L. Regular variation. In *Encyclopedia of Mathematics and Its Applications*; Cambridge University Press: Cambridge, UK, 1989; Volume 27.
- Albrecher, H.; Robert, C.Y.; Teugels, J.L. Joint asymptotic distributions of smallest and largest insurance claims. *Risks* **2014**, *2*, 289–314.
- Armendáriz, I.; Loulakis, M. Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stoch. Process. Appl.* **2011**, *121*, 1138–1147.
- Bobkov, S.G.; Chistyakov, G.P. On concentration functions of random variables. *J. Theor. Probab.* **2015**, *28*, 976–988.
- Pruss, A.R. Comparisons between tail probabilities of sums of independent symmetric random variables. *Ann. Inst. Henri Poincaré Probab. Stat.* **1997**, *33*, 651–671.
- Asmussen, S.; Kortschak, D. Error rates and improved algorithms for rare event simulation with heavy Weibull tails. *Methodol. Comput. Appl. Probab.* **2015**, *17*, 441–461.
- Baltrūnas, A.; Omey, E. The rate of convergence for subexponential distributions and densities. *Liet. Math. J.* **2002**, *42*, 1–14.
- Omey, E.; Willekens, E. Second order behaviour of the tail of a subordinated probability distribution. *Stoch. Process. Appl.* **1986**, *21*, 339–353.
- Pfaff, B.; McNeil, A. evir: Extreme Values in R. 2012. R Package Version 1.7-3. Available online: <https://CRAN.R-project.org/package=evir> (accessed on 14 November 2016).

24. McNeil, A. Estimating the tails of loss severity distributions using extreme value theory. *Astin Bull.* **1997**, *27*, 117–137.
25. Albrecher, H.; Asmussen, S. Ruin probabilities. In *Advanced Series on Statistical Science & Applied Probability*; World Scientific Publishing: Hackensack, NJ, USA, 2010; Volume 14.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).