THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Subproblems in applied probability

Patrick J. Laub
BE(Software, Hons. I)/BSc(Mathematics), BSc(Mathematics, Hons. I)

A thesis submitted for the degree of Doctor of Philosophy in 2017 at The University of Queensland, School of Mathematics and Physics in association with Aarhus University, Department of Mathematical Sciences

Abstract

Start this section on a new page.

The abstract should outline the main approach and findings of the thesis and must be between 300 and 800 words.

Contents

Chapter 1

Introduction

What are we trying to achieve in this field?

- Typically it is a compromise between: i) creating a realistic model, ii) highly accurate numerical results, iii) simplicity, iv) mathematical elegance, v) generating a "story" / conceptual understanding [& maybe vi) originality]
  - Laplace transform paper: i) no, ii) no, iii) yes, iv) moderately, v) no.
  - Orthogonal polynomial SLN paper: i) no, ii) no, iii) yes, iv) yes, v) somewhat [Dufrense points out the Hermite approximation resembles how one fixes the SLN $\approx$ LN model)
  - Rare maxima: i) no or N/A, ii) a little, iii) yes, iv) yes, v) yes.
  - Weibull sums: i) no, ii) no, iii) yes, iv) yes, v) yes
  - Insurance applications: i) somewhat, ii) somewhat, iii) yes, iv) yes, v) no.
- We rarely ever go and talk to financiers, insurers, etc. to see what exactly they need.

Relevant tradeoffs:

- No point getting highly accurate numerical results for an unrealistic model, e.g. rare-event estimation.
- There is value in having a simple & elegant result, which generates a story, for unrealistic models, e.g. the Black–Scholes formula is not used by financiers to actually price options, but it is a useful semi-justified benchmark/guideline (originally gave investors confidence leading to a boom in options trading), also it generates interesting stories like implied volatility & the volatility smile.

- A supremely complicated model or algorithm will probably not be used outside of academia (unless millions of dollars are on the line). Complexity breeds distrust — intuiting problems with the model ("debugging" the model) is difficult. Also, for most complicated algorithms, there exist very simple alternatives which are just slower (e.g. crude Monte Carlo) or which have bias (the SLN $\approx$ LN approx). One justification for complicated algorithms is they increase coding-time just once (making the library) but reduce run-time for tasks which will be run many times; but Donald Knuth's advice on "premature optimization" would stress that there really needs to be a demand for the optimized algorithm; also, a complicated library usually has a long list of tuning parameters which typically need to be fiddled with in a semi-educated way to achieve success (like CE, or neural networks, or even parallel computing); one good counter-example is KDE code which automatically picks sensible bandwidths (the algorithm just having this one [main] parameter is also nice).

Background:

- Probability background: Distributions. SLN. Dependence. Copulas. Tail properties of copulas.
- Asymptotic analysis. Saddlepoint approximations.
- Orthogonal polynomial expansions.
- Monte Carlo techniques: common random numbers, quasi-Monte Carlo.
- Numerical integration techniques: MCI, Gauss–Hermite quadrature.
- Rare-event simulation: efficiency of estimators, importance sampling, zero-variance estimator.
- Laplace transform inversion techniques.
- Finance background?: Black–Scholes model, option pricing, Value-at-Risk, risk measures.

So three papers:

1. Laplace transform for SLN.
    - Right-tail asymptotics (Leo).
    - Left-tail asymptotics (Gao, then Tankov).

- Laplace transform and derivative approximation.
- $I(\theta)$.
- Estimation: quasi-Monte Carlo, and Gauss–Hermite quadrature.

2. Orthogonal polynomial approximations to SLN density.
   - Gamma reference distribution.
   - Normal reference distribution.
   - Lognormal reference distribution.
3. Maxima of dependent r.v.s.
   -

Part I

“Applied”

Chapter 2

Approximating the Laplace transform of the sum of dependent lognormals

Abstract

Let $(X_1, \ldots, X_n)$ be multivariate normal, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $S_n = \mathrm{e}^{X_1} + \cdots + \mathrm{e}^{X_n}$. The Laplace transform $\mathcal{L}(\theta) = \mathbb{E}\mathrm{e}^{-\theta S_n} \propto \int \exp\{-h_\theta(\boldsymbol{x})\}\mathrm{d}\boldsymbol{x}$ is represented as $\widetilde{\mathcal{L}}(\theta)I(\theta)$, where $\widetilde{\mathcal{L}}(\theta)$ is given in closed form and $I(\theta)$ is the error factor (≈ 1). We obtain $\widetilde{\mathcal{L}}(\theta)$ by replacing $h_\theta(\boldsymbol{x})$ with a second-order Taylor expansion around its minimiser $\boldsymbol{x}^*$. An algorithm for calculating the asymptotic expansion of $\boldsymbol{x}^*$ is presented, and it is shown that $I(\theta) \to 1$ as $\theta \to \infty$. A variety of numerical methods for evaluating $I(\theta)$ is discussed, including Monte Carlo with importance sampling and quasi-Monte Carlo. Numerical examples (including Laplace-transform inversion for the density of S_n) are also given.

2.1 Introduction

The lognormal distribution arises in a wide variety of disciplines such as engineering, economics, insurance and finance, and is often employed in modelling across the sciences [5, 20, 23, 29, 30]. It has a natural multivariate version, namely $(\mathrm{e}^{X_1}, \ldots, \mathrm{e}^{X_n}) \sim \mathsf{LogNormal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ when $(X_1, \ldots, X_n) \sim \mathsf{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In this paper, we consider sums of lognormal random variables, $S_n \stackrel{\text{def}}{=} \mathrm{e}^{X_1} + \cdots + \mathrm{e}^{X_n}$, where the summands exhibit dependence ($\boldsymbol{\Sigma}$ is non-diagonal), using the notation that $S_n \sim \mathsf{SumLogNormal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Such sums have many challenging properties. In particular, there are no closed-form expressions for the density $f(x)$ or Laplace transform $\mathcal{L}(\theta)$ of S_n.

Models using sums of dependent lognormals are widely applicable, though they are particularly important in telecommunications and finance [22, 23]. Indeed, many of the approximations for the Laplace transform of sums of independent lognormals originated from the wireless communications community [16]. This reflects the significance of the SLN distribution within many models, and also that the Laplace transform is of intrinsic interest (engineers frequently work in the Laplace domain). In finance, the value of a portfolio (e.g. a collection of stocks) is SLN-distributed when using the assumptions of the common Black–Scholes framework. Thus the SLN distribution is central to the pricing of certain options (e.g., Asian and basket) [34]. Also, financial risk managers require estimates of $f(x)$ across $x \in (0, \mathbb{E}[S_n])$ to estimate risk measures such as value-at-risk or expected shortfall. Estimation of this kind has long been a legal requirement for many large banks, due to the Basel series of regulations (particularly, Basel II and Basel III), so in this context approximating $\mathcal{L}(\theta)$ is useful as a vehicle for computing the density $f(x)$ or the c.d.f. These issues are carefully explained in [21], [24], and the new Chapter 1 in the recently revised volume of McNeil et al. [33]. Comprehensive surveys of applications and numerical methods for the LN and SLN distributions are in [28, 9, 10].

There exist many approximations to the density of the SLN distribution. Many approximations work from the premise [17] that a sum S_n of lognormals can be accurately approximated by a single lognormal $L \sim \mathrm{LN}(\mu_L, \sigma_L^2)$. We refer to this approach as the *SLN ≈ LN approximation*. Some well-known SLN ≈ LN approximations are the Fenton–Wilkinson [25] and Schwartz–Yeh [35] approaches. These were originally specified for sums of *independent* lognormals, but have since been generalised to the dependent case [3]. A more recent procedure (for the independent case) is the minimax approximation of Beaulieu and Xie [18], calculating the values of μ_L and σ_L which minimise the maximum difference between the densities of S_n and L. However, [18] concludes that the approach is inaccurate in large dimensions or when the X_i have significantly different means or standard deviations. Finally, Beaulieu and Rajwani [17] describe a family of functions which mimic the characteristics of the SLN distribution function (in the independent case) with some success, i.e., high accuracy and closed-form expressions.

Another related avenue of research focuses on the asymptotic behaviour of $f(x)$ in the tails. First, Asmussen and Rojas-Nandayapa [12] characterised the right-tail asymptotics. Next, Gao et al. [26] gave the asymptotic form of the left tail for $n = 2$. Gulisashvili and Tankov [28] then provided the left-tail asymptotics for linear combinations of $n \geq 2$ lognormal variables. Yet these asymptotic forms cannot be used to approximate $f(x)$ with precision; to quote [28, p. 29], "these formulas are not valid for $x \geq 1$ and in practice have very poor accuracy unless x is much smaller than one". Similar numerical experience is reported in Asmussen et al. [10].

The approach taken here is via the Laplace transform. Accurate estimates for the Laplace transform can be numerically inverted to supply accurate density estimates. Asmussen et al. [9, 10] outline a framework to estimate $\mathcal{L}(\theta)$ for $n = 1$ using a modified saddlepoint approximation. In their work, the transform is decomposed into $\mathcal{L}(\theta) = \widetilde{\mathcal{L}}(\theta)I(\theta)$, where $\widetilde{\mathcal{L}}(\theta)$ has an explicit form and an efficient Monte Carlo estimator is given for $I(\theta)$.

This paper generalises the approach of [9, 10] to arbitrary n and dependence. The defining integral for the Laplace transform of S_n is

$$\mathcal{L}(\theta) = \frac{1}{\sqrt{(2\pi)^n|\det(()|\boldsymbol{\Sigma})}} \int_{\mathbb{R}^n} \exp\left\{-\theta \sum_{i=1}^{n} \mathrm{e}^{\mu_i}\mathrm{e}^{x_i} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{x}\right\}\mathrm{d}\boldsymbol{x} \tag{2.1}$$

where $\boldsymbol{D} \overset{\text{def}}{=} \boldsymbol{\Sigma}^{-1}$ (assuming $\boldsymbol{\Sigma}$ to be positive definite so $\boldsymbol{D}$ is well defined). Write the integrand as $\exp\{-h_\theta(\boldsymbol{x})\}$. The idea is then to provide an approximation $\widetilde{\mathcal{L}}(\theta)$ by replacing $h_\theta(\boldsymbol{x})$ by a second-order Taylor expansion around its minimiser $\boldsymbol{x}^*$. Whereas the minimiser x^* has a simple expression in terms of the Lambert W function when $n = 1$, as in [9, 10], the situation is much more complex when $n > 1$. As one of our main results we give a limit result for $\boldsymbol{x}^*$ as $\theta \to \infty$. Further, it is shown that the remainder $I(\theta)$ in the representation $\mathcal{L}(\theta) = \widetilde{\mathcal{L}}(\theta)I(\theta)$ goes to 1, a discussion of efficient Monte Carlo estimators of $I(\theta)$ follows, and numerical results showing the errors of our $\mathcal{L}(\theta)$ and (numerically inverted) $f(x)$ estimators are given.

2.2 Approximating the Laplace transform

Although the definition (2.1) makes sense for all $\theta \in \mathbb{C}$ with $\Re(\theta) > 0$ (we denote this set as $\mathbb{C}_+$), we will restrict the focus to $\theta \in (0, \infty)$. Of particular interest are the terms in the exponent, which in vector form (see Remark 2.2.1 below) are

$$h_\theta(\boldsymbol{x}) \overset{\text{def}}{=} \theta(\mathbf{e}^{\boldsymbol{\mu}})^\top \mathbf{e}^{\boldsymbol{x}} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{x}.$$

An approximation of simple form to $\mathcal{L}(\theta)$—written as $\widetilde{\mathcal{L}}(\theta)$—is available if $h_\theta(\boldsymbol{x})$ is replaced by a second-order Taylor expansion. The expansion is given in the proposition below.

Remark 2.2.1. On vector notation. *All vectors are considered column vectors. Functions applied element-wise to vectors are written in boldface, such as* $\mathbf{e}^{\boldsymbol{x}} \stackrel{\text{def}}{=} (\mathrm{e}^{x_1}, \ldots, \mathrm{e}^{x_n})^\top$ *and* $\mathbf{log}\,\boldsymbol{x} \stackrel{\text{def}}{=} (\log x_1, \ldots, \log x_n)^\top$*. If a vector is to be raised element-wise to a common power, then the power will be boldface, as in* $\boldsymbol{x}^{\boldsymbol{k}} \stackrel{\text{def}}{=} (x_1^k, \ldots, x_n^k)^\top$*. The notation* $\boldsymbol{x} \circ \boldsymbol{y}$ *denotes element-wise multiplication of vectors. The function* $\operatorname{diag}(\cdot)$ *converts vectors to matrices and vice versa, like the* MATLAB *function.*

Proposition 2.2.2. *The second-order Taylor expansion of* $h_\theta(\boldsymbol{x})$ *about its unique minimiser* $\boldsymbol{x}^*$ *is*

$$-\left(\mathbf{1} - \frac{1}{2}\boldsymbol{x}^*\right)^\top \boldsymbol{D}\boldsymbol{x}^* + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^\top (\boldsymbol{\Lambda} + \boldsymbol{D})(\boldsymbol{x} - \boldsymbol{x}^*)$$

where $\boldsymbol{\Lambda} \stackrel{\text{def}}{=} \theta \operatorname{diag}(\mathbf{e}^{\boldsymbol{\mu} + \boldsymbol{x}^*})$.

Proof. As $h_\theta(\boldsymbol{x})$ is strictly convex, a unique minimum exists. Since $\nabla h_\theta(\boldsymbol{x}^*) = \mathbf{0}$, the linear term vanishes in the Taylor expansion, so we have

$$h_\theta(\boldsymbol{x}) \approx h_\theta(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^\top \boldsymbol{H} (\boldsymbol{x} - \boldsymbol{x}^*)$$

where $\boldsymbol{H}$ is defined as the Hessian $\partial^2 h_\theta(\boldsymbol{x})/(\partial x_i\, \partial x_j)$ evaluated at $\boldsymbol{x}^*$. To find the value of $\boldsymbol{H}$, we just take derivatives:

$$\nabla h_\theta(\boldsymbol{x}) = \theta \mathbf{e}^{\boldsymbol{\mu} + \boldsymbol{x}} + \boldsymbol{D}\boldsymbol{x}, \quad \boldsymbol{H} = \boldsymbol{\Lambda} + \boldsymbol{D}.$$

Since $\boldsymbol{\Lambda}$ and $\boldsymbol{D}$ are both positive definite, so is $\boldsymbol{H}$. Also, $\nabla h_\theta(\boldsymbol{x}^*) = \mathbf{0}$ gives

$$-\theta \mathbf{e}^{\boldsymbol{\mu} + \boldsymbol{x}^*} = \boldsymbol{D}\boldsymbol{x}^* \text{ which implies } -\theta (\mathbf{e}^{\boldsymbol{\mu}})^\top \mathbf{e}^{\boldsymbol{x}^*} = \mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}^*. \tag{2.2}$$

Therefore the expansion becomes

$$\begin{aligned} h_\theta(\boldsymbol{x}) &\approx -\mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}^* + \frac{1}{2}(\boldsymbol{x}^*)^\top \boldsymbol{D}\boldsymbol{x}^* + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^\top (\boldsymbol{\Lambda} + \boldsymbol{D})(\boldsymbol{x} - \boldsymbol{x}^*) \\ &= -\left(\mathbf{1} - \frac{1}{2}\boldsymbol{x}^*\right)^\top \boldsymbol{D}\boldsymbol{x}^* + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^\top (\boldsymbol{\Lambda} + \boldsymbol{D})(\boldsymbol{x} - \boldsymbol{x}^*). \end{aligned}$$

□ □

This expansion allows $\mathcal{L}(\theta)$ to be approximated as a constant factor $\exp\{-h_\theta(\boldsymbol{x}^*)\}$ times the integral over a normal density (with inverse covariance $\boldsymbol{\Lambda} + \boldsymbol{D}$), which leads to

$$\mathcal{L}(\theta) \approx \widetilde{\mathcal{L}}(\theta) \stackrel{\text{def}}{=} \frac{1}{\sqrt{|\det(()|\boldsymbol{\Sigma}\boldsymbol{H})}} \exp\left\{\left(\mathbf{1} - \frac{1}{2}\boldsymbol{x}^*\right)^\top \boldsymbol{D}\boldsymbol{x}^*\right\}.$$

We need a suitable error or correction term in order to assess the accuracy of this approximation, so we will decompose the original integral (2.1) into $\mathcal{L}(\theta) = \widetilde{\mathcal{L}}(\theta)I(\theta)$. In the integral of (2.1) change variables such that $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{H}^{-1/2}\boldsymbol{y}$. Then by applying (2.2), multiplying by $\exp\{\mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}^* - \mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}^*\}$, and rearranging, we arrive at

$$\begin{aligned}
\mathcal{L}(\theta) &= \frac{1}{\sqrt{(2\pi)^n |\det(()|\boldsymbol{\Sigma}\boldsymbol{H})}} \int_{\mathbb{R}^n} \exp\Big\{-\theta(\mathrm{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})^\top \mathrm{e}^{\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}} \\
&\qquad - \frac{1}{2}(\boldsymbol{x}^* + \boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y})^\top \boldsymbol{D}(\boldsymbol{x}^* + \boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y})\Big\}\mathrm{d}\boldsymbol{y} \\
&= \widetilde{L}(\theta)I(\theta)
\end{aligned}$$

where

$$I(\theta) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n}} \exp\Big\{(\boldsymbol{x}^*)^\top \boldsymbol{D}\Big(\mathrm{e}^{\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}} - \mathbf{1} - \boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}\Big) - \frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{\Sigma}\boldsymbol{H})^{-1}\boldsymbol{y}\Big\}\mathrm{d}\boldsymbol{y}. \tag{2.3}$$

This equation can be rewritten in ways more convenient for Monte Carlo estimation.

Proposition 2.2.3. *We have that*

$$I(\theta) = \mathbb{E}\Big[g(\boldsymbol{H}^{-\frac{1}{2}}Z)\Big] = \sqrt{|\det(()|\boldsymbol{\Sigma}\boldsymbol{H})}\,\mathbb{E}\Big[v(\boldsymbol{\Sigma}^{\frac{1}{2}}Z)\Big] \tag{2.4}$$

where

$$\begin{aligned}
g(\boldsymbol{u}) &\stackrel{\text{def}}{=} \exp\Big\{(\boldsymbol{x}^*)^\top \boldsymbol{D}(\mathrm{e}^{\boldsymbol{u}} - \mathbf{1} - \boldsymbol{u}) + \frac{1}{2}\boldsymbol{u}^\top \boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{H}^{\frac{1}{2}}\boldsymbol{u}\Big\}, \\
v(\boldsymbol{u}) &\stackrel{\text{def}}{=} \exp\Big\{(\boldsymbol{x}^*)^\top \boldsymbol{D}(\mathrm{e}^{\boldsymbol{u}} - \mathbf{1} - \boldsymbol{u})\Big\},
\end{aligned}$$

and $Z \sim \mathsf{Normal}(\mathbf{0}, \boldsymbol{I})$.

Proof. To show that $I(\theta)$ can be written as the first expectation in (2.4), use $\boldsymbol{H} = \boldsymbol{\Lambda} + \boldsymbol{D}$, then add and subtract a term, to get

$$(\boldsymbol{\Sigma}\boldsymbol{H})^{-1} = [\boldsymbol{\Sigma}(\boldsymbol{D} + \boldsymbol{\Lambda})]^{-1} = (\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{\Lambda})^{-1}\boldsymbol{I} \pm (\boldsymbol{I} + \boldsymbol{\Sigma}\boldsymbol{\Lambda})^{-1}(\boldsymbol{\Sigma}\boldsymbol{\Lambda}) = \boldsymbol{I} - \boldsymbol{H}^{-1}\boldsymbol{\Lambda}$$

and substitute this into the $-\frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{\Sigma}\boldsymbol{H})^{-1}\boldsymbol{y}$ term in (2.3).

To prove $I(\theta)$ equals the second expectation of (2.4), change variables in (2.3) so that $\boldsymbol{y} = (\boldsymbol{\Sigma}\boldsymbol{H})^{1/2}\boldsymbol{z}$, giving

$$I(\theta) = \sqrt{|\det(()|\boldsymbol{\Sigma}\boldsymbol{H})} \int_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n}} \exp\Big\{(\boldsymbol{x}^*)^\top \boldsymbol{D}\Big(\mathrm{e}^{\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{z}} - \mathbf{1} - \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{z}\Big) - \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{I}\boldsymbol{z}\Big\}\mathrm{d}\boldsymbol{z}. \tag{2.5}$$

□ □

Remark 2.2.4. *When $n = 1$, $\mathbf{\Sigma} = \sigma^2$ and $\mu = 0$, (2.5) becomes*

$$I(\theta) = \sqrt{1 + \theta\sigma^2 \mathrm{e}^{x^*}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{x^*}{\sigma^2}(\mathrm{e}^{\sigma z} - 1 - \sigma z) - \frac{1}{2}z^2\right\} \mathrm{d}z.$$

This can be simplified using the Lambert W *function, denoted $\mathcal{W}(\cdot)$, which is defined [19] as the solution to the equation $\mathcal{W}(z)\mathrm{e}^{\mathcal{W}(z)} = z$. With this we have $x^* = -\mathcal{W}(\theta\sigma^2)$. Also, we can manipulate $\sqrt{1 + \theta\sigma^2 \mathrm{e}^{x^*}} = \sqrt{1 - x^*} = \sqrt{1 + \mathcal{W}(\theta\sigma^2)}$, so $I(\theta)$ becomes*

$$I(\theta) = \sqrt{1 + \mathcal{W}(\theta\sigma^2)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\mathcal{W}(\theta\sigma^2)}{\sigma^2}(\mathrm{e}^{\sigma z} - 1 - \sigma z) - \frac{1}{2}z^2\right\} \mathrm{d}z,$$

which coincides with the original result of [9] equation (2.3).

2.3 Asymptotic behaviour of the minimiser $\boldsymbol{x}^*$

We first introduce some notation. For a matrix $\boldsymbol{X}$, we write $\boldsymbol{X}_{i,\bullet}$ and $\boldsymbol{X}_{\bullet,i}$ for the ith row and column. Denote the row sums of $\boldsymbol{D}$ as $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$, that is, $a_i = \boldsymbol{D}_{i,\bullet}\mathbf{1}$. For sets of indices Ω_1 and Ω_2, then $\boldsymbol{X}_{\Omega_1,\Omega_2}$ denotes the submatrix of $\boldsymbol{X}$ containing row/column pairs in $\{(u, v) : u \in \Omega_1, v \in \Omega_2\}$. A shorthand is used for iterated logarithms: $\log_1 \theta \stackrel{\text{def}}{=} \log \theta$ and $\log_n \theta \stackrel{\text{def}}{=} \log \log_{n-1} \theta$ for $n \geq 2$ (note that $\log_k \theta$ is undefined for small or negative θ, but this is no problem as we are considering the case $\theta \to \infty$).

The approach taken to find $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)^\top$ is to set the gradient of $h_\theta(\boldsymbol{x})$ to $\mathbf{0}$, that is, to solve

$$\theta \mathbf{e}^{\boldsymbol{\mu} + \boldsymbol{x}^*} + \boldsymbol{D}\boldsymbol{x}^* = \mathbf{0}. \tag{2.6}$$

We will show that the asymptotics of the x_i^* are of the form

$$x_i^* = \sum_{j=1}^{n} \beta_{i,j} \log_j \theta - \mu_i + c_i + r_i(\theta) \tag{2.7}$$

for some $\boldsymbol{\beta} = (\beta_{i,j}) \in \mathbb{R}^{n \times n}$, $\boldsymbol{c} = (c_1, \ldots, c_n)^\top \in \mathbb{R}^n$ and $\boldsymbol{r}(\theta) = (r_1(\theta), \ldots, r_n(\theta))^\top$ where each $r_i(\theta) = \mathrm{o}(1)$. Before giving the general result, we consider the special case where all $a_i > 0$ since this result and its proof are much simpler.

Proposition 2.3.1. *If all row sums of $\boldsymbol{D}$ are positive then the minimiser $\boldsymbol{x}^*$ takes the form*

$$x_i^* = -\log \theta + \log_2 \theta - \mu_i + \log a_i + r_i(\theta) \tag{2.8}$$

where $r_i(\theta) = \mathcal{O}(\log_2 \theta / \log \theta) = \mathrm{o}(1)$ for $1 \leq i \leq n$, as $\theta \to \infty$.

Proof. Inserting (2.8) in (2.6) we find

$$\theta \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*} + \boldsymbol{D}\boldsymbol{x}^* = (\boldsymbol{a}\log\theta) \circ \mathbf{e}^{\boldsymbol{r}(\theta)} - \boldsymbol{a}\log\theta + \boldsymbol{a}\log_2\theta - \boldsymbol{D}\boldsymbol{\mu} + \boldsymbol{D}\mathbf{log}\,\boldsymbol{a} + \boldsymbol{D}\boldsymbol{r}(\theta) = \mathbf{0}.$$

Looking at these equations we see that we must have

$$\limsup_{\theta} \max_i r_i(\theta) = \liminf_{\theta} \min_i r_i(\theta) = 0,$$

and to remove the $\log_2\theta$ term the main term of $r_i(\theta)$ has to be $-\log_2\theta/\log\theta$. This gives the result of the proposition. □ □

In the general case where some $a_i \leq 0$, the asymptotic form of $\boldsymbol{x}^*$ is different from (2.8) and its derivation is much more intricate.

Theorem 2.3.2. *There exists a partition of $\{1,\dots,n\}$ into $\mathcal{F}_+$ and $\mathcal{F}_-$ such that for $i \in \mathcal{F}_+$,*

$$x_i^* = -\log\theta + \log_{k_i}\theta - \mu_i + c_i + \mathrm{o}(1)$$

for some $1 < k_i \leq n$. All x_i^* in $\mathcal{F}_-$ follow the general form of (2.7). In more detail, there exists a partition of $\mathcal{F}_-$ into $\mathcal{F}_-(1)$ and $\mathcal{F}_- \setminus \mathcal{F}_-(1)$, such that if $i \in \mathcal{F}_-(1)$ then $\beta_{i,1} < -1$, and if $i \in \mathcal{F}_- \setminus \mathcal{F}_-(1)$ then

$$\beta_{i,1} = -1,\ \beta_{i,2} = \cdots = \beta_{i,k_i-1} = 0,\ \beta_{i,k_i} < 0$$

for some $1 < k_i \leq n$. Finally we have, writing subscripts $+$ and $-$ for $\mathcal{F}_+$ and $\mathcal{F}_-$, that $\boldsymbol{x}_- = \boldsymbol{C}\boldsymbol{x}_+ + \mathrm{o}(1)$ where $\boldsymbol{C} = -\boldsymbol{D}_{-,-}^{-1}\boldsymbol{D}_{-,+}$. The sets $\mathcal{F}_+$, $\mathcal{F}_-$, $\mathcal{F}_-(1)$ and the constants $\beta_{i,j}$, c_i, k_i are determined by Algorithm 2.3.3 below.

See Remark 2.3.7 for some further remarks on the role of the signs of the row sums.

Algorithm 2.3.3.

1. *Let $\boldsymbol{\beta}_{\bullet,1}$ be the value of $\boldsymbol{w}$ that minimises $\boldsymbol{w}^\top\boldsymbol{D}\boldsymbol{w}$ over the set $\{\boldsymbol{w} : w_i \leq -1\}$. It will be proved in the appendix that the solution has $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,1} \leq 0$ when $\beta_{i,1} = -1$ and $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,1} = 0$ when $\beta_{i,1} < -1$. Accordingly, we can partition $\{1,\dots,n\}$ into the disjoint sets*

$$\mathcal{F}_+(1) = \emptyset, \quad \mathcal{F}_*(1) = \{i : \boldsymbol{D}_{j,\bullet}\boldsymbol{\beta}_{\bullet,1} < 0\},$$
$$\mathcal{F}_0(1) = \{i : \beta_{i,1} = -1, \boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,1} = 0\}, \quad \mathcal{F}_-(1) = \{i : \beta_{i,1} < -1\}.$$

2. *For $k = 2, \ldots, n$ recursively calculate $\boldsymbol{\beta}_{\bullet,k}$ as the value of $\boldsymbol{w}$ that minimises $\boldsymbol{w}^\top \boldsymbol{D}\boldsymbol{w}$ whilst satisfying*

$$w_i = 0 \text{ for } i \in \mathcal{F}_+(k-1), \quad w_i = 1 \text{ for } i \in \mathcal{F}_*(k-1),$$
$$w_i \leq 0 \text{ for } i \in \mathcal{F}_0(k-1), \quad \boldsymbol{D}_{i,\bullet}\boldsymbol{w} = 0 \text{ for } i \in \mathcal{F}_-(k-1).$$

It will be proved in the appendix that the solution has $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k} \leq 0$ for $i \in \mathcal{F}_0(k-1)$, $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k} = 0$ when $\beta_{i,k} < 0$ for $i \in \mathcal{F}_0(k-1)$, and at least one element of $\mathcal{F}_0(k-1)$ has $\boldsymbol{D}_{i,\bullet}\beta_{\bullet,k} < 0$. This allows us to create a new partition by

$$\begin{aligned}
\mathcal{F}_+(k) &= \mathcal{F}_+(k-1) \cup \mathcal{F}_*(k-1), \\
\mathcal{F}_*(k) &= \{i \in \mathcal{F}_0(k-1) : \beta_{i,k} = 0, \boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k} < 0\}, \\
\mathcal{F}_0(k) &= \{i \in \mathcal{F}_0(k-1) : \beta_{i,k} = 0, \boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k} = 0\}, \\
\mathcal{F}_-(k) &= \mathcal{F}_-(k-1) \cup \{i \in \mathcal{F}_0(k-1) : \beta_{i,k} < 0\}.
\end{aligned}$$

Terminate the loop early if $\mathcal{F}_0(k-1) = \emptyset$.

3. *Say $\mathcal{F}_+ = \mathcal{F}_+(k)$ and $\mathcal{F}_- = \mathcal{F}_-(k)$. For each $i \in \mathcal{F}_+$, let ℓ_i be the index of the first element of $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}$ which is negative, and we have $c_i = \log(-\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,\ell_i})$. Determine the remaining elements (using the same subscript shorthand introduced above) by*

$$\boldsymbol{c}_- = -\boldsymbol{D}_{-,-}^{-1}\boldsymbol{D}_{-,+}(\boldsymbol{c}_+ - \boldsymbol{\mu}_+) + \boldsymbol{\mu}_-. \tag{2.9}$$

Proof of Theorem 2.3.2. We propose a solution of the form (2.7) and show that when the $\beta_{i,j}$ are constructed from Algorithm 2.3.3, the remainder term r_i is $\mathrm{o}(1)$.

The construction allows us to draw the following conclusions for the x_i^*. Let $\mathcal{F}_+$ and $\mathcal{F}_-$ be the sets as defined in Step 3 above. Consider individually the indices which terminated in the $\mathcal{F}_+$ and in the $\mathcal{F}_-$ sets. In the first case, there exists a k_i with $1 < k_i \leq n$ such that

$$\beta_{i,j} = \begin{cases} -1, & j = 1, \\ 1, & j = k_i, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{and} \quad \boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,j} = \begin{cases} 0, & 1 \leq j < k_i - 1, \\ < 0, & j = k_i - 1. \end{cases}$$

Insertion in (2.6) gives

$$\begin{aligned}
0 &= \theta \mathrm{e}^{\mu_i + x_i^*} + \boldsymbol{D}_{i,\bullet}\boldsymbol{x}^* \\
&= -\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k_i-1}\mathrm{e}^{r_i(\theta)} \log_{k_i-1}\theta + \boldsymbol{D}_{i,\bullet}\left(\sum_{j=k_i-1}^{n} \boldsymbol{\beta}_{\bullet,j}\log_j\theta - \boldsymbol{\mu} + \boldsymbol{c} + \boldsymbol{r}(\theta)\right),
\end{aligned}$$

showing that the remainder is $\mathrm{o}(1)$.

In the second case, with $i \in \mathcal{F}_+$,

$$\beta_{i,1} < -1 \quad \text{and} \quad \boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,j} = 0,\ 1 \le j \le n,$$

or there exists $1 < k_i \le n$ such that

$$\beta_{i,j} = \begin{cases} -1, & j = 1, \\ 0, & 2 \le j < k_i, \\ < 0, & j = k_i, \end{cases} \quad \text{and} \quad \boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,j} = 0 \text{ for } 1 \le j \le n.$$

For this case we find $\theta \mathrm{e}^{\mu_i + x_i^*} + \boldsymbol{D}_{i,\bullet}\boldsymbol{x}^* = \mathrm{o}(1) + \boldsymbol{D}_{i,\bullet}\boldsymbol{r}(\theta)$, again showing that the remainder is $\mathrm{o}(1)$. Lastly, to show $\boldsymbol{x}_-$ in terms of $\boldsymbol{x}_+$, consider $\theta \mathbf{e}^{\boldsymbol{\mu}_- + \boldsymbol{x}_-^*} + \boldsymbol{D}_{-,+}\boldsymbol{x}_+ + \boldsymbol{D}_{-,-}\boldsymbol{x}_- = \mathbf{0}$. As $\theta \mathbf{e}^{\boldsymbol{\mu}_- + \boldsymbol{x}_-^*} = \mathrm{o}(1)$, we see that $\boldsymbol{x}_- = -\boldsymbol{D}_{-,-}^{-1}\boldsymbol{D}_{-,+}\boldsymbol{x}_+ + \mathrm{o}(1) = \boldsymbol{C}\boldsymbol{x}_+ + \mathrm{o}(1)$. □ □

In some cases above, we have been able to write the constant c_i as an expression involving $\boldsymbol{D}$ and $\boldsymbol{\mu}$. For example, in Proposition 2.3.1 we have $c_i = \log a_i$, and in Theorem 2.3.2 (2.9) gives the value of c_i for $i \in \mathcal{F}_-$. We can show a similar result in the general case for all $i \in \mathcal{F}_*(1)$, that is, for all i where $x_i^* = -\log\theta + \log_2\theta - \mu_i + c_i + \mathrm{o}(1)$.

Say $\mathcal{F}_* \stackrel{\text{def}}{=} \mathcal{F}_*(1)$ and $\mathcal{F}_\sim \stackrel{\text{def}}{=} \mathcal{F}_*^{\mathsf{c}}$; in the subscripts below, $*$ and $\sim$ refer to these sets. Since $\boldsymbol{D}$ is regular, so is $\boldsymbol{D}_{\sim,\sim}$. Say that $\overline{\boldsymbol{D}} \stackrel{\text{def}}{=} \boldsymbol{D}_{*,*} - \boldsymbol{D}_{*,\sim}\boldsymbol{D}_{\sim,\sim}^{-1}\boldsymbol{D}_{\sim,*}$, and denote the corresponding row sums by $\overline{\boldsymbol{a}} = (\overline{a}_i, i \in \mathcal{F}_*)$.

Corollary 2.3.4. *For all $i \in \mathcal{F}_*$*

$$x_i^* = -\log\theta + \log_2\theta - \mu_i + \log\overline{a}_i + r_i(\theta)$$

where $r_i(\theta) = \mathrm{o}(1)$ and $\overline{a}_i > 0$ as $\theta \to \infty$.

Proof. Let $\boldsymbol{b} = -\boldsymbol{\beta}_{\bullet,1}$. We have

$$b_i = \begin{cases} 1, & i \in \mathcal{F}_*(1) \cup \mathcal{F}_0(1), \\ > 1, & i \in \mathcal{F}_-(1), \end{cases} \qquad \boldsymbol{D}_{i,\bullet}\boldsymbol{b} = \begin{cases} \mathrm{e}^{c_i}, & i \in \mathcal{F}_*(1) = \mathcal{F}_*, \\ 0, & i \in \mathcal{F}_0(1) \cup \mathcal{F}_-(1) = \mathcal{F}_\sim. \end{cases}$$

Split $\boldsymbol{D}$ according to indices in $\mathcal{F}_*$ and $\mathcal{F}_\sim$, then

$$\boldsymbol{D}_{\sim,*}\boldsymbol{b}_* + \boldsymbol{D}_{\sim,\sim}\boldsymbol{b}_\sim = \mathbf{0} \quad \text{and} \quad \boldsymbol{D}_{*,*}\boldsymbol{b}_* + \boldsymbol{D}_{*,\sim}\boldsymbol{b}_\sim = \mathbf{e}^{\boldsymbol{c}_*} > \mathbf{0}.$$

The first equation gives $\boldsymbol{b}_\sim = -\boldsymbol{D}_{\sim,\sim}^{-1}\boldsymbol{D}_{\sim,*}\boldsymbol{b}_*$, and this with the second equation shows that $\overline{\boldsymbol{D}}\boldsymbol{b}_* = \overline{\boldsymbol{D}}\mathbf{1} = \overline{\boldsymbol{a}} = \mathbf{e}^{\boldsymbol{c}_*} > \mathbf{0}$; thus $\overline{\boldsymbol{D}}$ has all row sums positive and $\boldsymbol{c}_* = \mathbf{log}\,(\overline{\boldsymbol{D}}\boldsymbol{b}_*) = \mathbf{log}\,\overline{\boldsymbol{a}}$. □ □

There are some simple forms of $\mathbf{\Sigma}$ which fall into the case where all $a_i > 0$. These include the case where all diagonal elements of $\mathbf{\Sigma}$ are identical, and all non-diagonal elements are identical. Note, by positive definiteness of $\mathbf{\Sigma}$ we must have at least one row-sum of $\boldsymbol{D}$ positive. Also, if $X_1, \ldots, X_n$ is an AR(1) process, the resulting covariance matrix will have all $a_i > 0$. Meanwhile, cases where there exist $a_i \leq 0$ are not difficult to find. For the case $n = 2$ with variances $\sigma_1^2 \leq \sigma_2^2$ and correlation ρ, a simple calculation gives that both row sums are positive when $\rho < \sigma_1/\sigma_2$, and one is negative when $\rho > \sigma_1/\sigma_2$ (see Gao et al. [26] for the expansion of $f(x)$ as $x \downarrow 0$ for these cases). We now list a couple of examples of asymptotic forms of $\boldsymbol{x}^*$ for specific $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ which have some non-positive row sums of $\mathbf{\Sigma}^{-1}$.

Example 2.3.5. *Consider* $\boldsymbol{\mu} = (-10, 0, 10)^\top$ *and*

$$\mathbf{\Sigma} = \begin{pmatrix} 0.5 & 1 & 2 \\ 1 & 3 & 4 \\ 2 & 4 & 10 \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 14 & -2 & -2 \\ -2 & 1 & 0 \\ -2 & 0 & 0.5 \end{pmatrix}.$$

Implementing the algorithm gives that

$$\begin{aligned} x_1^* &= -\log\theta + \log_2\theta + (10 + \log 2) + \mathrm{o}(1), \\ x_2^* &= -2\log\theta + 2\log_2\theta + (20 + 2\log 2) + \mathrm{o}(1), \\ x_3^* &= -4\log\theta + 4\log_2\theta + (40 + 4\log 2) + \mathrm{o}(1), \end{aligned}$$

and

$$(\boldsymbol{\beta} \mid \boldsymbol{c} - \boldsymbol{\mu}) = \left(\begin{array}{ccc|c} -1 & 1 & 0 & 10.69 \\ -2 & 2 & 0 & 21.39 \\ -4 & 4 & 0 & 42.77 \end{array}\right), \quad \boldsymbol{D}(\boldsymbol{\beta} \mid \boldsymbol{c} - \boldsymbol{\mu}) = \left(\begin{array}{ccc|c} -2 & * & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

(where unimportant values of $\boldsymbol{D}(\boldsymbol{\beta} \mid \boldsymbol{c} - \boldsymbol{\mu})$ *are replaced by stars).* □

Example 2.3.6. *Consider* $\boldsymbol{\mu} = (1, 2, 3)^\top$ *and*

$$\mathbf{\Sigma} = \begin{pmatrix} 0.4545 & 0.4545 & 0.4545 \\ 0.4545 & 1.7204 & 1.8470 \\ 0.4545 & 1.8470 & 2.9862 \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 3 & -0.9 & 0.1 \\ -0.9 & 2 & -1.1 \\ 0.1 & -1.1 & 1 \end{pmatrix}.$$

Implementing the algorithm gives that

$$\begin{aligned} x_1^* &= -\log\theta + \log_2\theta - 1 + \log 2.2 + \mathrm{o}(1), \\ x_2^* &= -\log\theta + \log_3\theta - 2 + \log 0.79 + \mathrm{o}(1), \\ x_3^* &= -\log\theta - 0.1\log_2\theta + 1.1\log_3\theta - 3 + c_3 + \mathrm{o}(1), \end{aligned}$$

where $c_3 = 0.9 - 0.1\log 2.2 + 1.1\log 0.79$, *and*

$$(\boldsymbol{\beta} \mid \boldsymbol{c} - \boldsymbol{\mu}) = \left(\begin{array}{ccc|c} -1 & 1 & 0 & -0.2 \\ -1 & 0 & 1 & -2.2 \\ -1 & -0.1 & 1.1 & -2.4 \end{array}\right), \quad \boldsymbol{D}(\boldsymbol{\beta} \mid \boldsymbol{c} - \boldsymbol{\mu}) = \left(\begin{array}{ccc|c} -2.2 & * & * & * \\ 0 & -0.79 & * & * \\ 0 & 0 & 0 & 0 \end{array}\right).$$

□

Remark 2.3.7. The importance of the sign of the row sums of $\boldsymbol{D}$, as illustrated by Proposition 2.3.1, perplexed us for quite some time. However Gulisashvili and Tankov [28] describe an interesting link between the row sums and the *minimum variance portfolio*. They show that the leading asymptotic term of $\mathbb{P}(S_n < x)$ as $x \downarrow 0$ depends upon

$$\overline{\boldsymbol{w}}^\top \boldsymbol{\Sigma}\, \overline{\boldsymbol{w}} = \min_{\boldsymbol{w}\in\Delta} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \text{ where } \Delta \overset{\text{def}}{=} \Big\{ \boldsymbol{w} : \sum_i w_i = 1, w_i \geq 0 \Big\}.$$

The i for which $\overline{w}_i > 0$ indicate which summands in S_n have the 'least variance'. These summands are asymptotically important in the left tail, as they will struggle the most to take very small values. Seen from the viewpoint of modern portfolio theory [32], the solution $\overline{\boldsymbol{w}}$ is viewed as the optimal portfolio weights to create the minimum-variance portfolio. When all $a_i > 0$ then $\overline{w}_i = a_i / \sum_{j=1}^n a_j$ which represents full diversification. However when assets become highly correlated (meaning that some $\boldsymbol{D}$ row sums are non-positive) then there exist $\overline{w}_i = 0$, i.e., some assets are ignored. Thus the asymptotics are qualitatively different when the signs of the row sums change. The exact point where an asset's optimal weight becomes 0 occurs when $a_i = 0$, and this phase change produces a unique and convoluted asymptotic form. As $\mathcal{L}(\theta)$ as $\theta \to \infty$ is related to $\mathbb{P}(S_n < x)$ as $x \downarrow 0$, the behaviour of $\boldsymbol{x}^*$ is explained.

For applications we will need to find $\boldsymbol{x}^*$ for a large number of θ numerically. The results above give a sensible starting point for an iterative solver, such as Newton–Raphson. Another option is based on the following formulation. Let $\boldsymbol{A} \overset{\text{def}}{=} \boldsymbol{D} - \operatorname{diag}(\boldsymbol{D})$ and write the defining equation as

$$\theta \mathrm{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*} + \operatorname{diag}(\boldsymbol{D})\boldsymbol{x}^* = -\boldsymbol{A}\boldsymbol{x}^*.$$

For each row i, all x_i^* are now on the left-hand side. Using properties of the Lambert W function we see that

$$x_i^* = -\mathcal{W}\left(\frac{\theta \mathrm{e}^{\mu_i}}{D_{i,i}} \exp\left\{ -\frac{\boldsymbol{A}_{i,\bullet}\boldsymbol{x}^*}{D_{i,i}} \right\} \right) - \frac{\boldsymbol{A}_{i,\bullet}\boldsymbol{x}^*}{D_{i,i}}.$$

One can use this to perform a component-wise fixed-point iteration as an alternative to the Newton–Raphson scheme.

2.4 Asymptotic behaviour of $I(\theta)$

In order to discuss $I(\theta)$ as $\theta \to \infty$ we will consider it in a form different from Section 2.2. Define $\boldsymbol{\sigma} \overset{\text{def}}{=} \operatorname{diag}(\boldsymbol{H})^{-1/2} \in (0,\infty)^n$ and $\boldsymbol{M} \overset{\text{def}}{=} \operatorname{diag}(\boldsymbol{\sigma})\boldsymbol{H}\operatorname{diag}(\boldsymbol{\sigma}) \in \mathbb{R}^{n\times n}$. In (2.3), substitute $\boldsymbol{H}^{-1/2}\boldsymbol{y} = \boldsymbol{\sigma} \circ \boldsymbol{z}$, so

$$I(\theta) = \int_{\mathbb{R}^n} \frac{\exp\{-\frac{1}{2}\boldsymbol{z}^\top \boldsymbol{M}\boldsymbol{z}\}}{\sqrt{(2\pi)^n |\det(()|\boldsymbol{M}^{-1})}} \exp\left\{-\theta(\mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})^\top \left(\mathbf{e}^{\boldsymbol{\sigma}\circ\boldsymbol{z}} - \mathbf{1} - \boldsymbol{\sigma}\circ\boldsymbol{z} - \frac{1}{2}(\boldsymbol{\sigma}\circ\boldsymbol{z})^{\mathbf{2}}\right)\right\} \mathrm{d}\boldsymbol{z}. \tag{2.10}$$

The limit of this integrand is the density of a multivariate normal distribution, which when integrated is 1. To see this, consider the following. As $\theta \to \infty$ we have $\sigma_i \to 0$ or $\sigma_i \to D_{i,i}^{-1/2} > 0$, so taking $\ell \in (2, \infty)$ means

$$\theta \mathrm{e}^{\mu_i + x_i^*} \sigma_i^\ell = \theta \mathrm{e}^{\mu_i + x_i^*} (\theta \mathrm{e}^{\mu_i + x_i^*} + D_{i,i})^{-\frac{\ell}{2}} = \mathrm{o}(1). \tag{2.11}$$

Consider the second exponent of (2.10). For fixed $\boldsymbol{z}$, $\mathrm{e}^{\sigma_i z_i} - 1 - \sigma_i z_i - \frac{1}{2}\sigma_i^2 z_i^2 = \mathcal{O}(\sigma_i^3)$, and since $\theta \mathrm{e}^{\mu_i + x_i^*} \sigma_i^3 = \mathrm{o}(1)$ by (2.11) we have

$$\theta(\mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})^\top \left(\mathbf{e}^{\boldsymbol{\sigma}\circ\boldsymbol{z}} - \mathbf{1} - \boldsymbol{\sigma}\circ\boldsymbol{z} - \frac{1}{2}(\boldsymbol{\sigma}\circ\boldsymbol{z})^{\mathbf{2}}\right) = \mathrm{o}(1). \tag{2.12}$$

Finally, we consider $\boldsymbol{M}$ as $\theta \to \infty$. Say that $n_+ \stackrel{\text{def}}{=} |\mathcal{F}_+|$ and assume that these are the first n_+ indices. We can then write that $\boldsymbol{M} \to \boldsymbol{M}^* \stackrel{\text{def}}{=} \operatorname{diag}(\boldsymbol{I}_{n_+}, \boldsymbol{F})$ where this $\boldsymbol{F}$ is the bottom-right submatrix of size $(n - n_+) \times (n - n_+)$ of the inverted correlation matrix implied by $\boldsymbol{\Sigma}$. The $\boldsymbol{M}$ matrices are positive definite for all $\theta \in (0, \infty]$; thus the limiting form of the integrand in (2.10) is a non-degenerate multivariate normal density.

Proposition 2.4.1. $\lim_{\theta\to\infty} I(\theta) = 1$.

Proof. We use the dominated convergence theorem. By (2.12) and the paragraph which follows that equation, the exponent of the integrand is bounded by a constant g_1 for $\|\boldsymbol{z}\| < 1$, say, and the exponent is below $-g_2\|\boldsymbol{z}\|$ otherwise ($g_2 > 0$), for $\theta > \theta_0$, say. The latter comes from the positive definiteness of $\boldsymbol{M}^*$, the convergence of $\boldsymbol{M}$ to $\boldsymbol{M}^*$ and the convergence of (2.12). Next, convexity implies that the exponent is bounded by $-g_2\|\boldsymbol{z}\|$ for $\|\boldsymbol{z}\| > 1$. In total we have the bound

$$\exp\left\{g_1 \mathbb{I}_{\{\|\boldsymbol{z}\|\le 1\}} - g_2\|\boldsymbol{z}\|\mathbb{I}_{\{\|\boldsymbol{z}\|>1\}}\right\},$$

which is an integrable function. Thus the conditions for dominated convergence are satisfied and we can safely switch the limit and integral to obtain $I(\theta) \to 1$. □ □

2.5 Estimators of $\mathcal{L}(\theta)$ and $I(\theta)$

The simplest approach is to numerically integrate the original expression in (2.1). This approach is used as a baseline against which the following estimators are compared (the approach can, however, be slow or impossible for large n). The next naïve approach is to estimate the expectation $\mathbb{E}[\mathrm{e}^{-\theta S_n}]$ by crude Monte Carlo (CMC). This would involve simulating random vectors $X_1, \ldots, X_R \overset{\text{i.i.d.}}{\sim} \mathrm{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $X_r = (X_{r,1}, \ldots, X_{r,n})$, and computing

$$\widehat{\mathcal{L}}_{\mathrm{CMC}}(\theta) \overset{\text{def}}{=} \frac{1}{R} \sum_{r=1}^{R} \exp\Big\{ -\theta \sum_{i=1}^{n} X_{r,i} \Big\}.$$

However this estimator is not efficient for large θ, and rare-event simulation techniques are required.

Given the decomposition of $\mathcal{L}(\theta) = \widetilde{\mathcal{L}}(\theta) I(\theta)$, some more accurate estimators can be assessed. Simply using $\widetilde{\mathcal{L}}(\theta)$ gives a biased estimator (which is fast and deterministic) for the transform, however the bias is decreased by estimating $I(\theta)$ with Monte Carlo integration. Proposition 2.2.3 gives two probabilistic representations of $I(\theta)$. We expect the CMC estimator of the first—$\mathbb{E}[g(\boldsymbol{H}^{-1/2} Z)]$—to exhibit infinite variance as $\theta \to \infty$ as this has been proven for $n = 1$ in [9]. Therefore this estimator does not seem promising. The second estimator—$\sqrt{|\det(()|\boldsymbol{\Sigma}\boldsymbol{H})}\, \mathbb{E}[v(\boldsymbol{\Sigma}^{1/2} Z)]$—can be viewed as the first estimator after importance sampling has been applied, so we focus upon this. Taking $Z_1, \ldots, Z_R \overset{\text{i.i.d.}}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$,

$$\widehat{\mathcal{L}}_{\mathrm{IS}}(\theta) \overset{\text{def}}{=} \frac{1}{R} \exp\left\{ \left(\mathbf{1} - \frac{1}{2}\boldsymbol{x}^*\right)^\top \boldsymbol{D}\boldsymbol{x}^* \right\} \sum_{r=1}^{R} \exp\left\{ (\boldsymbol{x}^*)^\top \boldsymbol{D}(\mathbf{e}^{Z_r} - \mathbf{1} - Z_r) \right\}.$$

Many variance-reduction techniques can be applied to increase the efficiency of these estimators. The effect of including control variates into $\widehat{\mathcal{L}}_{\mathrm{IS}}(\theta)$ was considered, using the control variate $(\boldsymbol{x}^*)^\top \boldsymbol{D} Z_r^{\mathbf{2}}$ (note the element-wise square). The variance reduction achieved was small considering the large overhead of computing the variates (and their expectations) so these results have been omitted. Lastly, we considered an estimator based on the Gumbel distribution. Say that $G = (G_1, \ldots, G_n)$ is a vector of i.i.d. standard Gumbel random variables, that is, $\mathbb{P}(G_r < x) = \exp\{-\mathrm{e}^{-x}\}$ for $x \in \mathbb{R}$. Then $\mathcal{L}(\theta)$ can be rewritten as an integral over the density of a vector of standard Gumbel random variables. This estimator was quite accurate, though it had higher relative error and variance than the estimators based on $\widehat{\mathcal{L}}_{\mathrm{IS}}(\theta)$ so it too has been excluded from the results.

The final two variance reduction techniques investigated were *common random numbers* and *quasi-Monte Carlo* applied to $\widehat{\mathcal{L}}_{\mathrm{IS}}(\theta)$; for a detailed explanation of these techniques see [27] or [8]. Both individually achieved significant variance reduction, and together provided the best estimator. Specifically,

$$\widehat{\mathcal{L}}_{\mathrm{Q}}(\theta) \overset{\text{def}}{=} \frac{1}{R} \exp\left\{ \left(\mathbf{1} - \frac{1}{2}\boldsymbol{x}^*\right)^\top \boldsymbol{D}\boldsymbol{x}^* \right\} \sum_{r=1}^{R} \exp\left\{ (\boldsymbol{x}^*)^\top \boldsymbol{D}(\mathbf{e}^{\boldsymbol{q}_r} - \mathbf{1} - \boldsymbol{q}_r) \right\},$$

where $\boldsymbol{q}_r \stackrel{\text{def}}{=} \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Phi}^{-1}(\boldsymbol{u}_r)$, using $\boldsymbol{\Phi}^{-1}(\cdot)$ as the (element-wise) standard normal inverse c.d.f., and where $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots\}$ is the n-dimensional Sobol sequence started at the same point for every θ. Therefore, $\widehat{\mathcal{L}}_{\text{Q}}(\theta)$ is deterministic (for a fixed R and θ), and using this scheme is therefore a kind of numerical quadrature. More sophisticated adaptive quadrature methods could possibly be applied.

2.6 Numerical Results

Relative errors are given for the main estimators of $\mathcal{L}(\theta)$ in the table below. In all estimators the smoothing technique of using common random variables is employed, and all estimators are compared against numerical integration of the relevant integrals to 15 significant digits. See [?] for the software implementation used to create these results.

Table 2.1: Relative error for various approximations of $\mathcal{L}(\theta)$ for $\boldsymbol{\mu} = \boldsymbol{0}$, $\boldsymbol{\Sigma} = [1, 0.5; 0.5, 1]$. The number of Monte Carlo replications R used is 10^6. Note: $*$ indicates that the CMC estimator simply gave an estimate of 0.

θ	100	2,500	5,000	7,500	10,000
$\widetilde{\mathcal{L}}$	-9.89e-3	-1.27e-2	-1.28e-2	-1.27e-2	-1.27e-2
$\widehat{\mathcal{L}}_{\text{CMC}}$	1.29e-2	$*$	$*$	$*$	$*$
$\widehat{\mathcal{L}}_{\text{IS}}$	3.36e-4	2.96e-4	2.57e-4	2.31e-4	2.11e-4
$\widehat{\mathcal{L}}_{\text{Q}}$	-3.19e-6	-5.03e-6	-5.31e-6	-5.56e-6	-5.98e-6

Also, the p.d.f. of S_n can be estimated by numerical inversion of the Laplace transform. As the approximations of $\mathcal{L}(\theta)$ above are valid only for $\theta \in (0, \infty)$, not $\theta \in \mathbb{C}_+$, this restricts the options for Laplace-transform inversion algorithms. The Gaver–Stehfest algorithm [36] and so-called power algorithms [13] can be used. We report on the results of using the Gaver–Stehfest algorithm as implemented by Mallet [31].

Other options for estimating $f(x)$ include numerically integrating the convolution equation (typically this is viable only for small n), the conditional Monte Carlo method (as in Example 4.3 on page 146 of [8]), and kernel density estimation. The following estimators are reported: the conditional Monte Carlo estimator $\widehat{f}_{\text{Cond}}$, $\widetilde{f} \stackrel{\text{def}}{=} \mathcal{L}^{-1}(\widetilde{\mathcal{L}}(\cdot))$, $\widehat{f}_{\text{IS}} \stackrel{\text{def}}{=} \mathcal{L}^{-1}(\widehat{\mathcal{L}}_{\text{IS}}(\cdot))$ and $\widehat{f}_{\text{Q}} \stackrel{\text{def}}{=} \mathcal{L}^{-1}(\widehat{\mathcal{L}}_{\text{Q}}(\cdot))$.

The numerically inverted Laplace transforms are surprisingly accurate. Using common random numbers for the $\mathcal{L}(\theta)$ estimators was necessary, otherwise the inversion algorithms became confused by the non-smooth input. The precision of the inversion algorithms cannot be arbitrarily increased when using standard double-floating-point arithmetic [2],

Table 2.2: Relative errors for estimators of $f(x)$ for $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = [1, 0.5; 0.5, 1]$. The number of Monte Carlo repetitions for each x is $R = 10^4$ for $\widehat{f}_{\text{Cond}}$, $\widehat{f}_{\text{IS}}$ and $\widehat{f}_{\text{Q}}$.

x	0.01	1	1.5	2	3
$\widehat{f}_{\text{Cond}}$	-1.17e-1	2.20e-2	3.72e-3	5.21e-3	-4.60e-3
$\widetilde{f}$	-7.03e-3	2.56e-2	1.79e-2	6.00e-2	3.82e-2
$\widehat{f}_{\text{IS}}$	1.94e-3	1.43e-2	-6.13e-3	4.00e-2	3.68e-3
$\widehat{f}_{\text{Q}}$	2.90e-4	1.11e-2	-9.04e-3	3.70e-2	2.44e-3

so the software suite MATHEMATICA was used. Yet this did not solve the problem of the Gaver–Stehfest algorithm becoming unstable (and very slow) when trying to increase the desired precision. Also, the inversion results became markedly poorer when $f(x)$ exhibited high kurtosis (i.e., when $|\det(()|\boldsymbol{\Sigma})$ became small).

2.7 Closing Remarks

The estimators above give an accurate, relatively simple, and computationally swift method of computing the Laplace transform of the sum of dependent lognormals. We have shown that the approximation's error diminishes to zero ($I(\theta) \to 1$) as $\theta \to \infty$, and that it is still accurate for small values of θ. One can find $\boldsymbol{x}^*$—for each θ examined—using a Newton–Raphson scheme, and Section 2.3 gives an accurate starting value for the iterations.

2.A Remaining steps in the proof of Theorem 2.3.2

First we note that all the minimisations are convex problems and therefore have unique solutions.

For the initial step of the algorithm let $\overline{\boldsymbol{w}}$ be the solution of the minimisation problem and let $\boldsymbol{e}_i$ be the vector with 1 at coordinate i and zero at the other coordinates. Then $g_i(\varepsilon) = (\overline{\boldsymbol{w}} + \varepsilon \boldsymbol{e}_i)^\top \boldsymbol{D}(\overline{\boldsymbol{w}} + \varepsilon \boldsymbol{e}_i)$ is minimised at $\varepsilon = 0$. When $\overline{w}_i < -1$ the vector $\overline{\boldsymbol{w}} + \varepsilon \boldsymbol{e}_i$ is in the search set for all ε small. We therefore have $g_i'(0) = 0$ which gives $\boldsymbol{D}_{i,\bullet}\overline{\boldsymbol{w}} = 0$. When $\overline{w}_i = -1$ the vector $\overline{\boldsymbol{w}} + \varepsilon \boldsymbol{e}_i$ is in the search set only for non-positive values of ε. This implies $g_i'(0) \leq 0$ giving $\boldsymbol{D}_{i,\bullet}\overline{\boldsymbol{w}} \leq 0$.

For the general recursive step we let $\boldsymbol{u} = \boldsymbol{w}_{\mathcal{F}_0(k-1)}$ and express $\boldsymbol{w}_{\mathcal{F}_-(k-1)}$ in terms of $\boldsymbol{u}$ from the equations $\boldsymbol{D}_{i,\bullet}\boldsymbol{w} = 0$, $i \in \mathcal{F}_-(k-1)$. The derivative of $\boldsymbol{w}^\top \boldsymbol{D}\boldsymbol{w}$ with respect to

u_i (i being the index inherited from $\boldsymbol{w}$) is then

$$2\boldsymbol{D}_{i,\bullet}\boldsymbol{w} + 2\frac{\partial \boldsymbol{w}_{\mathcal{F}_-(k-1)}}{\partial u_i}\boldsymbol{D}_{\mathcal{F}_-(k-1)}\boldsymbol{w} = 2\boldsymbol{D}_{i,\bullet}\boldsymbol{w}.$$

As above we find that the derivative of $\boldsymbol{w}^\top \boldsymbol{D}\boldsymbol{w}$ with respect to u_i at the minimising point is zero when $u_i < 0$ and less than or equal to zero when $u_i = 0$.

What is left to prove is that $\mathcal{F}_0(k)$ always has at least one element with $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k+1} < 0$. To this end define $d_1 = -\boldsymbol{\beta}_{\bullet,1}$ and $d_k = d_{k-1} - \boldsymbol{\beta}_{\bullet,k}$ for $k > 1$. From the properties of $\boldsymbol{\beta}$ we find

$$\begin{aligned}
&d_{\mathcal{F}_+(k),k} = 0; \quad d_{\mathcal{F}_*(k),k} = 1 \text{ and } \boldsymbol{D}_{\mathcal{F}_*(k)}d_k > 0;\\
&d_{\mathcal{F}_0(k),k} = 1 \text{ and } \boldsymbol{D}_{\mathcal{F}_0(k)}d_k = 0; \quad \boldsymbol{D}_{\mathcal{F}_-(k)}d_k = 0.
\end{aligned}$$

Assume now that $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k+1} = 0$ for all $i \in \mathcal{F}_0(k)$. We show that this leads to a contradiction. Using the assumption, $\boldsymbol{\beta}_{\bullet,k+1}$ has the properties

$$\begin{aligned}
&\boldsymbol{\beta}_{\mathcal{F}_+(k),k+1} = 0; \quad \boldsymbol{\beta}_{\mathcal{F}_*(k),k+1} = 1;\\
&\boldsymbol{\beta}_{\mathcal{F}_0(k),k+1} \leq 0 \text{ and } \boldsymbol{D}_{\mathcal{F}_0(k)}\boldsymbol{\beta}_{\bullet,k+1} = 0; \quad \boldsymbol{D}_{\mathcal{F}_-(k)}\boldsymbol{\beta}_{\bullet,k+1} = 0.
\end{aligned}$$

Combining the two displays we have

$$\boldsymbol{D}_{\mathcal{F}_0(k)}d_k = \boldsymbol{D}_{\mathcal{F}_0(k)}\boldsymbol{\beta}_{\bullet,k+1}, \quad \boldsymbol{D}_{\mathcal{F}_-(k)}d_k = \boldsymbol{D}_{\mathcal{F}_-(k)}\boldsymbol{\beta}_{\bullet,k+1}.$$

Since d_k and $\boldsymbol{\beta}_{\bullet,k+1}$ are identical on $\mathcal{F}_+(k-1)$ and $\mathcal{F}_*(k-1)$ the equations reduce to

$$\boldsymbol{D}_0\begin{pmatrix} d_{\mathcal{F}_0(k),k} \\ d_{\mathcal{F}_-(k),k}\end{pmatrix} = \boldsymbol{D}_0\begin{pmatrix} \boldsymbol{\beta}_{\mathcal{F}_0(k),k} \\ \boldsymbol{\beta}_{\mathcal{F}_-(k),k}\end{pmatrix}, \text{ where } \boldsymbol{D}_0 = \begin{pmatrix} \boldsymbol{D}_{\mathcal{F}_0(k),\mathcal{F}_0(k)} & \boldsymbol{D}_{\mathcal{F}_0(k),\mathcal{F}_-(k)} \\ \boldsymbol{D}_{\mathcal{F}_-(k),\mathcal{F}_0(k)} & \boldsymbol{D}_{\mathcal{F}_-(k),\mathcal{F}_-(k)}\end{pmatrix}.$$

Since the matrix $\boldsymbol{D}_0$ is positive definite and $d_{\mathcal{F}_0(k),k} \neq \boldsymbol{\beta}_{\mathcal{F}_0(k),k}$, we have reached a contradiction. $\square$

Part II

“Probability”

Chapter 3

Tail asymptotics of light-tailed Weibull-like sums

Abstract

We consider sums of n i.i.d. random variables with tails close to $\exp\{-x^\beta\}$ for some $\beta > 1$. Asymptotics developed by Rootzén (1987) and Balkema, Klüppelberg & Resnick (1993) are discussed from the point of view of tails rather of densities, using a somewhat different angle, and supplemented with bounds, results on a random number N of terms, and simulation algorithms.

3.1 Introduction

Let $X, X_1, \ldots, X_n$ be i.i.d. with common distribution F. A recurrent theme in applied probability is then to determine the order of magnitude of the tail $\mathbb{P}(S_n > x)$ of their sum $S_n = X_1 + \cdots + X_n$.

The results vary according to the heaviness of the tail $\overline{F} = 1 - F$ of F. In the heavy-tailed case, defined as the X for which $\mathbb{E}\mathrm{e}^{sX} = \infty$ for all $s > 0$, there is the subexponential class in which the results take a clean form (see e.g. [**?**] or [7]). In fact, by the very definition of subexponentiality, we have $\mathbb{P}(S_n > x) \sim n\overline{F}(x)$ as $x \to \infty$ where $\overline{F}(x) = \mathbb{P}(X > x)$. The main examples are regularly varying $\overline{F}(x)$, lognormal X, and Weibull tails $\overline{F}(x) = \mathrm{e}^{-cx^\beta}$ where $0 < \beta < 1$.

In the light-tailed case, defined as the X for which $\mathbb{E}\mathrm{e}^{sX} < \infty$ for some $s > 0$, the most standard asymptotic regime is not $x \to \infty$ but rather $x = x_n$ going to ∞ at rate n. For example, let $x_n = nz$ for some z, where typically $z > \mathbb{E}X$ in order to make the problem a rare-event one. Under some regularity conditions, the sharp asymptotics are then given by the saddlepoint approximation $\mathbb{P}(S_n > x) \sim c(z)\mathrm{e}^{-nI(z)}/n^{1/2}$ for suitable $c(z)$ and $I(z)$, cf. [**?**]. This is a large deviations result, describing how likely it is for S_n to be far from the value $n\mathbb{E}X$ predicted by the LLN. However, in many applications the focus is rather on a small or moderate n, i.e. the study of $\mathbb{P}(S_n > x)$ as $x \to \infty$ with n fixed.

The basic light-tailed explicit examples in this setting are the exponential distribution, the gamma distribution, the inverse Gaussian distribution, and the normal distribution. The tail of F is exponential or close-to-exponential for exponential, gamma and inverse Gaussian distributions; this is the borderline between light and heavy tails, and the analysis of tail behaviour is relatively simple in this case (we give a short summary later in Section 3.8). The most standard class of distributions with a lighter tail is formed by the Weibull distributions where $\overline{F}(x) = \mathrm{e}^{-cx^\beta}$ for some $\beta > 1$. For $\beta = 2$, this is close to the normal distribution, where (by its well-known Mill's ratio) $\overline{F}(x) \sim \mathrm{e}^{-x^2/2}/(\sqrt{2\pi}x)$ when $F = \Phi$ is the standard normal law. The earliest study of tail properties of S_n may be that of [**?**] which was later followed up by the mathematically deeper and somewhat general study of Balkema, Klüppelberg, & Resnick [14], henceforth referred to as BKR. The setting of both papers is densities.

Despite filling an obvious place in the theory of tails of sums, it has been our impression that this theory is less known than it should be. This was confirmed by a Google Scholar search which gave only 27 citations of BKR, most of which were even rather peripheral. One reason may be that the title *Densities with Gaussian tails* of BKR is easily misinterpreted, another the heavy analytic flavour of the paper. Also note that the focus of [**?**] is somewhat different and the set of results we are interested in here appears as a by-product at the end of that paper.

The purpose of the present paper is twofold: to present a survey from a somewhat different angle than BKR, in the hope of somewhat remedying this situation; and to supplement the theory with various new results. In the survey part, the aim has been simplicity and intuition more than generality. In particular, we avoid considering convex conjugates and some non-standard central limit theory developed in Section 6 of BKR. These tools are mathematically deep and elegant, but not really indispensable for developing what we see as the main part of the theory. Beyond this expository aspect, our contributions are: to present the main results and their conditions in terms of tails rather than densities; to develop simple upper and lower bounds; to study the case of a random number of terms N, more precisely properties of $\mathbb{P}(S_N > x)$ when N is an independent Poisson r.v.; and to look into simulation aspects.

The precise assumptions on the distribution F in the paper vary somewhat depending on the context and progression of the paper. The range goes from the vanilla Weibull tail $\overline{F}(x) = \mathrm{e}^{-cx^\beta}$ via an added power in the asymptotics, $\overline{F}(x) \sim dx^\alpha \mathrm{e}^{-cx^\beta}$, to the full generality of the BKR set-up. Here cx^β is replaced by a smooth convex function $\psi(x)$ satisfying $\psi'(x) \to \infty$ and the density has the form $\gamma(x)\mathrm{e}^{-\psi(x)}$ for a function γ which is in some sense much less variable than ψ (the precise regularity conditions are given in Section 3.4).

3.2 Heuristics

With heavy tails, the basic intuition on the tail behaviour of S_n is the principle of a single big jump; this states that a large value of S_n is typically caused by one summand being large while the rest take ordinary values. A rigorous formulation of this can be proved in a few lines from the very definition of subexponentiality, see e.g. [7, p. 294]. With light tails, the folklore is that if S_n is large, say $S_n \approx x$, then all X_i are of the same order x/n.

This suggest that the asymptotics of $\mathbb{P}(S_n > x)$ are essentially determined by the form of F locally around x/n. A common type of such local behaviour is that $\overline{F}\Big(x + e(x)y\Big) \sim \overline{F}(x)\mathrm{e}^{-y}$ for some positive function $e(x)$ as $x \to \infty$ with $y \in \mathbb{R}$ fixed; this is abbreviated as $F \in \mathsf{GMDA}(e)$. Equivalently,

$$\Lambda\Big(x + e(x)y\Big) \ \sim \ \Lambda(x) + y \tag{3.1}$$

where $\Lambda(x) = -\log \overline{F}(x)$. Here one can take $e(x) = \mathbb{E}[X - x \mid X > x]$, the so-called *mean excess function*; if F admits a density $f(x)$, an alternative asymptotically equivalent choice is the *inverse hazard rate* $e(x) = 1/\lambda(x)$ where $\lambda(x) = \Lambda'(x) = f(x)/\overline{F}(x)$.

In fact, (3.1) is a necessary and sufficient condition for F to be in $\mathsf{GMDA}(e)$, the maximum domain of attraction of the Gumbel distribution [**?**]. Even if this condition may look special at first sight, it covers the vast majority of well-behaved light-tailed distributions, with some exceptions such as certain discrete distributions like the geometric or Poisson.

From these remarks one may proceed for $n = 2$ from the convolution,

$$
\begin{aligned}
\mathbb{P}(X_1 + X_2 > x) = (f * \overline{F})(x) &= \int_{-\infty}^{\infty} \lambda(z) \exp\Big\{-\Lambda(z) - \Lambda(x-z)\Big\} \, \mathrm{d}z \\
&= \int_{-\infty}^{\infty} \frac{e(x/2)}{e\Big(x/2 + e(x/2)y\Big)} \exp\left\{-\Lambda\Big(\frac{x}{2} + e\Big(\frac{x}{2}\Big)y\Big) - \Lambda\Big(\frac{x}{2} - e\Big(\frac{x}{2}\Big)y\Big)\right\} \mathrm{d}y, \qquad (3.2)
\end{aligned}
$$

where we have substituted $z = x/2 + e(x/2)y$. First note that if $\lambda(x)$ tends to 0 as $x \to \infty$ and is differentiable, we can expand Λ about $y = 0$ as

$$
\Lambda\Big(\frac{x}{2} + e\Big(\frac{x}{2}\Big)y\Big) \sim \Lambda\left(\frac{x}{2}\right) + y + \frac{\lambda'\left(\frac{x}{2}\right)}{2\lambda\left(\frac{x}{2}\right)^2} y^2 .
$$

By defining $\sigma^2(u) = \lambda(u)^2/2\lambda'(u)$ and repeating this argument we get that

$$
\Lambda\Big(\frac{x}{2} \pm e\Big(\frac{x}{2}\Big)y\Big) \sim \Lambda\left(\frac{x}{2}\right) \pm y + \frac{y^2}{4\sigma^2(\frac{x}{2})}. \qquad (3.3)
$$

Also we will use that $e(x)$ is self-neglecting, i.e. $\forall\, t,\ e\Big(x + e(x)t\Big) \sim e(x)$ as $x \to \infty$, as is well-known and easy to prove from (3.1). Combining (3.3) and the self-neglecting property with (3.2) gives us

$$
\begin{aligned}
\mathbb{P}(X_1 + X_2 > x) &\sim \int_{-\infty}^{\infty} 1 \cdot \exp\left\{-2\Lambda\left(\frac{x}{2}\right) - \frac{y^2}{2\sigma^2(\frac{x}{2})}\right\} \mathrm{d}y \\
&= \sqrt{2\pi\sigma^2(x/2)}\, \exp\Big\{-2\Lambda(x/2)\Big\}. \qquad (3.4)
\end{aligned}
$$

In summary, rewriting (3.4) gives

$$
\overline{F^{*2}}(x) \;=\; \mathbb{P}(X_1 + X_2 > x) \;\sim\; \overline{F}(x/2)^2 \sqrt{\pi \frac{\lambda(x/2)^2}{\lambda'(x/2)}}\,. \qquad (3.5)
$$

The key issue in making this precise is to keep better track of the second order term in the Taylor expansion, as discussed later in the paper.

Remark 3.2.1. *The procedure to arrive at* (3.5) *is close to the Laplace method for obtaining integral asymptotics. Classically, the integral in question has the form* $\int_a^b \mathrm{e}^{-\theta h(z)}\, \mathrm{d}z$ *and one proceeds by finding the* z_0 *at which* $h(z)$ *is minimum and performing a second order Taylor expansion around* z_0*. Here, we neglected the* $\lambda(z)$ *in front and took the relevant analogue of* z_0 *as* $x/2$ *which is precisely the minimizer of* $\Lambda(x-z) + \Lambda(z)$. $\diamond$

Remark 3.2.2. *If X_1, X_2 have different distributions F_1, F_2, the above calculations suggest that $X_1 + X_2 > x$ will occur roughly when $X_1 \approx z(x)$, $X_2 \approx x - z(x)$ where $z = z(x)$ is the solution of $\lambda_1(z) = \lambda_2(x - z)$. In fact, this is what is needed to make the first order Taylor terms cancel. For example, if $\overline{F}_1(x) = \mathrm{e}^{-x^{\beta_1}}$, $\overline{F}_2(x) = \mathrm{e}^{-x^{\beta_2}}$ with $\beta_2 < \beta_1$, we get $z(x) \sim cx^\eta$ where $\eta = (\beta_2 - 1)/(\beta_1 - 1) < 1$, $c = (\beta_2/\beta_1)^{1/(\beta_1-1)}$. This type of heuristic is an important guideline when designing importance sampling algorithms, cf. [8, V.1, VI.2].* $\diamond$

3.3 Weibull-like sums

We now make the heuristics of preceding section rigorous for the case of different distributions F_1, F_2 of X_1, X_2 such that the densities f_1, f_2 satisfy

$$f_i(x) \;\sim d_i x^{\alpha_i+\beta-1}\mathrm{e}^{-c_i x^\beta}\,, \quad x \to \infty,\ i = 1, 2 \tag{3.6}$$

for some common $\beta > 1$, where the α_i can take any value in $(-\infty, \infty)$ and c_i, d_i are positive $(i = 1, 2)$.

We start by some analytic preliminaries. Given (3.6), we define

$$\eta = c_1^{1/(\beta-1)} + c_2^{1/(\beta-1)}, \quad \theta_1 = c_2^{1/(\beta-1)}/\eta, \quad \theta_2 = c_1^{1/(\beta-1)}/\eta, \quad \kappa = \frac{\eta^{\beta-1}}{\beta c_1 c_2}\,. \tag{3.7}$$

Note that

$$\overline{F}_i(x) \;\sim\; \frac{d_i}{\beta c_i} x^{\alpha_i}\mathrm{e}^{-c_i x^\beta} \tag{3.8}$$

(hence $c_i = 1$, $d_i = \beta$, $\alpha_i = 0$ corresponds to the traditional Weibull tail e^{-x^β}). Define the excess function of F_i by $e_i(x) = \overline{F}_i(x)/f_i(x)$. Thus $e_i(x)$ is the inverse hazard rate and has asymptotics $x^{1-\beta}/(\beta c_i)$ with limit 0 as $x \to \infty$.

Lemma 3.3.1. *Define $c = c_1\theta_1^\beta + c_2\theta_2^\beta$. Then $c < \min(c_1, c_2)$, $\theta_1 + \theta_2 = 1$, and*

$$e_1(\theta_1 x) \sim e_2(\theta_2 x) \sim \frac{\kappa}{x^{\beta-1}} = \frac{1}{\beta c_1 \theta_1^{\beta-1} x^{\beta-1}} = \frac{1}{\beta c_2 \theta_2^{\beta-1} x^{\beta-1}}. \tag{3.9}$$

Proof. All statements are obvious except $c < \min(c_1, c_2)$. But

$$\begin{aligned} c \;&=\; c_1\theta_1^{\beta-1}\theta_1 + c_2\theta_2^{\beta-1}\theta_2 \;=\; \frac{c_1c_2\theta_1}{\eta^{\beta-1}} + \frac{c_1c_2\theta_2}{\eta^{\beta-1}} \;=\; \frac{c_1c_2}{\eta^{\beta-1}} \\ &<\; \frac{c_1c_2}{\left[c_2^{1/(\beta-1)}\right]^{\beta-1}} \;=\; c_1. \end{aligned} \tag{3.10}$$

Similarly, $c < c_2$. $\square$

Lemma 3.3.2. *$(1+h)^\beta = 1 + h\beta + \dfrac{h^2}{2}\beta(\beta-1)\omega(h)$ where $\omega(h) \to 1$ as $h \to 0$ and $\underline{\omega}_\varepsilon = \inf_{-1+\varepsilon<h<\varepsilon^{-1}} \omega(h) > 0$ for all $\varepsilon > 0$.*

Proof. By standard Taylor expansion results, $\omega(h) = (1+h^*)^{\beta-2}$. where h^* is between 0 and h. The statement on $\underline{\omega}_\varepsilon$ follows from this by considering all four combinations of the cases $h \le 0$ or $h > 0$, $1 < \beta \le 2$ or $\beta \ge 2$ separately. □

The key result is the following. It allows, for example, to determine the asymptotics of the tail or density of F^{*n} in the Weibull-like class by a straightforward induction argument, see Corollary 3.3.5 below.

Theorem 3.3.3. *Under assumption (3.6), $\mathbb{P}(X_1 + X_2 > x) \sim kx^\gamma \mathrm{e}^{-cx^\beta}$ as $x \to \infty$, where $\gamma = \alpha_1 + \alpha_2 + \beta/2$ and $k = d_1 d_2 \theta_1^{\alpha_1}\theta_2^{\alpha_2}\kappa\eta^{1-\beta}(2\pi\sigma^2)^{1/2}/\beta$, with $\theta_1, \theta_2, \kappa, \eta$ as in (3.6), the constant c as in Lemma 3.3.1, and σ^2 determined by*

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad \textit{where } \frac{1}{\sigma_i^2} \;=\; \beta(\beta-1)c_i\theta_i^{\beta-2}\kappa^2 \,.$$

Further the density of $X_1 + X_2$ has asymptotic form $\beta ckx^{\gamma+\beta-1}\mathrm{e}^{-cx^\beta}$.

Remark 3.3.4. If $F_1 = F_2$ and $c_1 = c_2 = 1$, then $\theta_1 = \theta_2 = 1/2$ and $c = 1/2^{\beta-1}$ in accordance with Section 3.2. ◇

Proof. By Lemma 3.3.1, we can choose $0 < a_- < a_+ < 1$ such that $a_+^\beta c_2 > c$, $(1-a_-)c_1 > c$. Then

$$\mathbb{P}\Big(X_1 + X_2 > x, X_1 \notin [a_-x, a_+x]\Big) \;\le \mathbb{P}(X_1 > a_+x) + \mathbb{P}(X_2 > (1-a_-)x)$$

is $\mathrm{o}(x^\gamma \mathrm{e}^{-cx^\beta})$ and so it suffices to show that

$$\mathbb{P}\Big(X_1 + X_2 > x, a_-x < X_1 < a_+x\Big) \;=\; \int_{a_-x}^{a_+x} f_1(z)\overline{F}_2(x-z)\,\mathrm{d}z \tag{3.11}$$

has the claimed asymptotics. The last expression together with $a_- > 0$, $a_+ < 1$ also shows that the asymptotics is a tail property so that w.l.o.g. we may assume that $e_i(\theta_i x) = \kappa/x^{\beta-1}$, implying that (3.9) holds with equality.

Now

$$\begin{aligned}\mathbb{P}\Big(X_1 + X_2 > x, a_- < X_1 < a_+x\Big) \;&=\; \int_{a_-x}^{a_+x} f_1(z)\overline{F}_2(x-z)\,\mathrm{d}z \\ &= \int_{a_-x}^{a_+x} \frac{d_1 d_2}{\beta c_2} z^{\alpha_1+\beta-1}(x-z)^{\alpha_2}\exp\Big\{-c_1 z^\beta - c_2(x-z)^\beta\Big\}\,\mathrm{d}z. \end{aligned} \tag{3.12}$$

Using the substitution $z = \theta_1 x + y\kappa/x^{\beta-1}$, we have $x - z = \theta_2 x - y\kappa/x^{\beta-1}$,

$$c_1 z^\beta + c_2(x-z)^\beta \;=\; c_1\theta_1^\beta x^\beta\Big(1+h_1(x,y)\Big)^\beta + c_2\theta_2^\beta x^\beta\Big(1-h_2(x,y)\Big)^\beta \tag{3.13}$$

where $h_i(x,y) = y\kappa/\theta_i x^\beta$. Taylor expanding $\Big(1 \pm h_i(x,y)\Big)^\beta$ as in Lemma 3.3.2 and using (3.9), the first order term of (3.13) is

$$c_1\theta_1^\beta x^\beta + c_2\theta_2^\beta x^\beta + \beta c_1\theta_1^{\beta-1}\kappa - \beta c_2\theta_2^{\beta-1}\kappa = cx^\beta.$$

Defining $\omega_1(x,y) = \omega\Big(h_1(x,y)\Big)$, $\omega_2(x,y) = \omega\Big(-h_2(x,y)\Big)$, (3.12) becomes

$$\frac{d_1 d_2}{\beta c_2}\int_{y_-(x)}^{y_+(x)} \Big(\theta_1 x + e_1(\theta_1 x)y\Big)^{\alpha_1+\beta-1}\Big(\theta_2 x - e_2(\theta_2 x)y\Big)^{\alpha_2}$$
$$\cdot\ \exp\Big\{-cx^\beta - \frac{y^2}{2\sigma_1^2 x^\beta}\omega_1(x,y) - \frac{y^2}{2\sigma_2^2 x^\beta}\omega_2(x,y)\Big\}\,\frac{\kappa}{x^{\beta-1}}\mathrm{d}y$$

where $y_-(x) = (a_- - \theta_1)x^\beta/\kappa$, $y_+(x) = (a_+ - \theta_1)x/e(\theta_1 x)$. Notice here that $a_- x < z < a_+ x$ ensures the bound

$$h_1(x,y) = \frac{1}{\theta_1 x}(z - \theta_1 x) \geq \frac{a_-}{\theta_1} - 1 \;>\; -1.$$

Similarly, $-h_2(x,y) \geq -a_+/\theta_2 - 1 > 1$. Using Lemmas 3.3.1 and 3.3.2 shows that so that the $\omega_i(x,y)$ are uniformly bounded below, and that $\Big(\theta_i x + e_i(\theta_i x)y\Big)/x$ is bounded in $y_-(x) < y < y_+(x)$ and goes to θ_i as $x \to \infty$. A dominated convergence argument gives therefore that the asymptotics of (3.11) is the same as that of

$$\frac{d_1 d_2\kappa}{\beta c_2}\theta_1^{\alpha_1+\beta-1}\theta_2^{\alpha_2}x^{\alpha_1+\alpha_2}\mathrm{e}^{-cx^\beta}\int_{-\infty}^{\infty}\exp\Big\{-\frac{y^2}{2\sigma^2 x^\beta}\Big\}\,\mathrm{d}y$$
$$= \frac{d_1 d_2 \kappa\eta^{1-\beta}}{\beta}\theta_1^{\alpha_1}\theta_2^{\alpha_2}x^{\alpha_1+\alpha_2}\mathrm{e}^{-cx^\beta}(2\pi\sigma^2 x^\beta)^{1/2} \;=\; kx^\gamma\mathrm{e}^{-cx^\beta}.$$

This proves the assertion on the tail of $X_1 + X_2$, and the proof of the density claim differs only by constants. □

Corollary 3.3.5. *Assume the density f of F satisfies $f(x) \sim dx^{\alpha+\beta-1}\mathrm{e}^{-cx^\beta}$ as $x \to \infty$. Then the tail and the density of an i.i.d. sum satisfy*

$$\overline{F^{*n}}(x) \;=\; \mathbb{P}(S_n > x) \;\sim\; k(n)x^{\alpha(n)}\mathrm{e}^{-c(n)x^\beta}, \tag{3.14}$$
$$f^{*n}(x) \;\sim\; \beta c(n)k(n)x^{\alpha(n)+\beta-1}\mathrm{e}^{-c(n)x^\beta} \tag{3.15}$$

where $c(n) = c/n^{\beta-1}$, $\alpha(n) = n\alpha + (n-1)\beta/2$ and

$$k(n) = \frac{d^n}{\beta c}\Big[\frac{2\pi}{\beta(\beta-1)c}\Big]^{(n-1)/2} n^{\frac{1}{2}(\beta - n(2\alpha+\beta)-1)}. \tag{3.16}$$

Proof. We use induction. The statement is trivial for $n = 1$ so assume it proved for $n-1$. Taking $F_1 = F$, $F_2 = F^{*(n-1)}$ and applying Theorem 3.3.3 implies the result, and provides recurrences for $c(n)$, $\alpha(n)$, and $k(n)$. To be specific, say that the F_i distributions have densities f_i like

$$f_i(x) \sim d_i(n) x^{\alpha_i(n)+\beta-1} \mathrm{e}^{-c_i(n)x^\beta}\,, \quad i = 1, 2.$$

As $F_1 = F$ is fixed, we simply have $c_1(n) = c$, $d_1(n) = d$, $\alpha_1(n) = \alpha$, and for $F_2 = F^{*(n-1)}$ the induction hypothesis gives us

$$c_2(n) = \frac{c}{(n-1)^{\beta-1}}, \qquad d_2(n) = \beta c_2(n-1)k(n-1), \quad \alpha_2(n) = \alpha(n-1).$$

We extend the notation of Theorem 3.3.3 in the obvious way, for example we define $\eta(n) = c_1(n)^{1/(\beta-1)} + c_2(n)^{1/(\beta-1)}$. These simplify to

$$\eta(n) = \frac{nc^{1/(\beta-1)}}{n-1}, \quad \theta_1(n) = \frac{1}{n}, \quad \theta_2(n) = \frac{n-1}{n}, \quad \kappa(n) = \frac{n^{\beta-1}}{\beta c}\,.$$

So $c(n) = c_1(n)\theta_1(n)^\beta + c_2(n)\theta_2(n)^\beta = c/n^\beta + c(n-1)/n^\beta = c/n^{\beta-1}$. Also, we have $\alpha(n) = \alpha_1(n) + \alpha_2(n) + \beta/2 = n\alpha + (n-1)\beta/2$.

The last recursion is less simple. We need the σ constants:

$$\sigma_1^2(n) = \frac{\beta c n^{-\beta}}{\beta-1}, \quad \sigma_2^2(n) = \frac{\beta c(n-1)n^{-\beta}}{\beta-1}, \quad \sigma^2(n) = \frac{\beta c(n-1)n^{-\beta-1}}{\beta-1}.$$

Setting $k(1) = d/(\beta c)$, we get for $n \geq 2$

$$\begin{aligned}
k(n) &= d_1(n)d_2(n)\theta_1(n)^{\alpha_1(n)}\theta_2(n)^{\alpha_2(n)}\kappa(n)\eta(n)^{1-\beta}(2\pi\sigma(n)^2)^{1/2}/\beta \\
&= \left[\frac{2\pi}{\beta(\beta-1)c}\right]^{1/2} d(n-1)^{\alpha(n-1)+\frac{1}{2}(\beta(n-2)+1)} n^{-\alpha n-\frac{1}{2}\beta(n-1)-\frac{1}{2}} k(n-1) \\
&= \frac{d^n}{\beta c}\left[\frac{2\pi}{\beta(\beta-1)c}\right]^{(n-1)/2} \prod_{\ell=2}^{n} (\ell-1)^{\alpha(\ell-1)+\frac{1}{2}(\beta(\ell-2)+1)} \ell^{-\alpha\ell-\frac{1}{2}\beta(\ell-1)-\frac{1}{2}} \\
&= \frac{d^n}{\beta c}\left[\frac{2\pi}{\beta(\beta-1)c}\right]^{(n-1)/2} n^{\frac{1}{2}(\beta-n(2\alpha+\beta)-1)}.
\end{aligned}$$

□

Note that (3.15) is already in Rootzén [?] (see his equations (6.1)–(6.2)). We point out later that the assumptions on the density can be relaxed to $\overline{F}(x) \sim kx^\alpha \mathrm{e}^{-cx^\beta}$ where $k = d/c\beta$.

3.4 Light-tailed sums

We now proceed to the set-up of BKR and first introduce some terminology related to the densities of the form $f(x) \sim \gamma(x)\mathrm{e}^{-\psi(x)}$. The main assumption is that the function ψ is non-negative, convex, C^2, and its first order derivative is denoted λ. Further it is supposed that

$$\lim_{x\to\infty} \lambda(x) = \infty, \tag{3.17}$$

λ' is ultimately positive and $1/\sqrt{\lambda'}$ is self-neglecting, i.e. that for $x \to \infty$

$$\lambda'\Big(x + y/\sqrt{\lambda'(x)}\Big) \;\sim\; \lambda'(x). \tag{3.18}$$

A function γ is called *flat* for ψ if locally uniformly on bounded y-intervals

$$\lim_{x\to\infty} \frac{\gamma\Big(x + y/\sqrt{\lambda'(x)}\Big)}{\gamma(x)} = 1. \tag{3.19}$$

Similar conventions apply to functions denoted ψ_1, ψ_2, etc. For the Weibull case,

$$\psi(x) = ax^\beta, \quad \lambda(x) = a\beta x^{\beta-1}, \quad \gamma(x) = \lambda(x)$$

and so (3.18) and (3.19) are satisfied. Examples beyond Weibull-like distributions are $\psi(x) = x\log x$ and $\psi(x) = \mathrm{e}^{ax}, a > 0$.

Define the class $\mathcal{H}(\gamma,\psi)$ as the class of all distributions F having a density of the form $\gamma(x)\mathrm{e}^{-\psi(x)}$ where ψ is as above and γ a measurable function which is flat for ψ, and let $\overline{\mathcal{H}}(\gamma,\psi)$ be the class of distributions F satisfying $\overline{F}(x) \sim \gamma(x)\mathrm{e}^{-\psi(x)}/\lambda(x)$.

Theorem 3.4.1. (i) *$\mathcal{H}(\gamma,\psi) \subseteq \overline{\mathcal{H}}(\gamma,\psi)$;*
(ii) *Assume $F_1 \in \mathcal{H}(\gamma_1,\psi_1)$, $F_2 \in \mathcal{H}(\gamma_2,\psi_2)$. Then $F_1 * F_2 \in \mathcal{H}(\gamma,\psi)$, where γ,ψ are determined by first solving*

$$q_1 + q_2 = x, \quad \lambda_1(q_1) = \lambda_2(q_2) \tag{3.20}$$

for $q_1 = q_1(x)$, $q_2 = q_2(x)$ and next letting $\psi(x) = \psi_1(q_1) + \psi_2(q_2)$,

$$\gamma(x) \;=\; \sqrt{\frac{2\pi\lambda'(x)}{\lambda_1'(q_1)\lambda_2'(q_2)}}\gamma_1(q_1)\gamma_2(q_2)$$

where $\lambda(x) = \psi'(x) = \lambda_1(q_1) = \lambda_2(q_2)$.
(iii) *Assume $F_1 \in \overline{\mathcal{H}}(\gamma_1,\psi_1)$, $F_2 \in \overline{\mathcal{H}}(\gamma_2,\psi_2)$. Then there exists $H_i \in \mathcal{H}(\gamma_i,\psi_i), H_i \in$ $\mathsf{GMDA}(1/\lambda_i)$ and*

$$\overline{H}_i(x) \sim \overline{F}_i(x), \quad \overline{H_1 * H_2}(x) \sim \overline{F_1 * F_2}(x).$$

Moreover, $F_1 * F_2 \in \overline{\mathcal{H}}(\gamma,\psi)$ with γ,ψ as in (ii) *and $F_1 * F_2 \in \mathsf{GMDA}(1/\lambda)$.*

The proof of Theorem 3.4.1 is in Appendix 3.A. Part (ii) is in BKR, here slightly reformulated, and a number of examples in BKR can be obtained as corollaries of this theorem.

Remark 3.4.2. *Letting* $\tau(y) = \lambda_1^{\leftarrow}(y) + \lambda_2^{\leftarrow}(y)$*, the solution of* (3.20) *can be written*

$$q_1(x) \;=\; \lambda_1^{\leftarrow}\Big(\tau^{\leftarrow}(x)\Big)\,, \quad q_2(x) \;=\; \lambda_2^{\leftarrow}\Big(\tau^{\leftarrow}(x)\Big) \tag{3.21}$$

(here $\cdot^{\leftarrow}$ *means functional inverse).* ◇

3.5 Bounds

There are easy upper- and lower-tail bounds for Weibull sums in terms of the incomplete gamma function $\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1}\mathrm{e}^{-t}\,\mathrm{d}t$ when $\beta > 1$ that in their simplest form just come from thinking about p-norms $\|\boldsymbol{y}\|_p = \Big(|y_1|^p + \cdots + |y_n|^p\Big)^{1/p}$ and the fact that if Y is standard exponential, then $Y^{1/\beta}$ is Weibull with tail e^{-x^β}.

Proposition 3.5.1. *Let* X *have density* $\beta k^{\gamma/\beta}x^{\gamma-1}\mathrm{e}^{-kx^\beta}/\Gamma(\gamma/\beta)$*,* $x > 0$*, where* $k > 0$*,* $\beta \geq 1$*, and* $\gamma > 0$*. Then*

$$\frac{\Gamma(n\gamma/\beta, kx^\beta)}{\Gamma(n\gamma/\beta)} \;\leq\; \mathbb{P}(X_1 + \cdots + X_n > x) \;\leq\; \frac{\Gamma(n\gamma/\beta, kx^\beta/n^{\beta-1})}{\Gamma(n\gamma/\beta)}\,.$$

Proof. An X with the given density has the same distribution as $(Y/k)^{1/\beta}$ where Y is $\mathsf{Gamma}(\alpha, 1)$ with density $y^{\alpha-1}\mathrm{e}^{-y}/\Gamma(\alpha)$, where $\alpha = \gamma/\beta$. Therefore

$$X_1^\beta + \cdots + X_n^\beta \;=\; \|\boldsymbol{X}\|_\beta^\beta \;\stackrel{d}{=}\; \|\boldsymbol{Y}/k\|_1 = Y_1/k + \cdots + Y_n/k\,,$$

where $Y_1, \ldots, Y_n$ are i.i.d. $\mathsf{Gamma}(\alpha, 1)$. From the Jensen and Hölder inequalities we have for $p \geq 1$ and $\boldsymbol{x} \in \mathbb{R}^n$ that

$$\|\boldsymbol{x}\|_p \leq \|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_p n^{1-1/p}.$$

Hence, since further $\|\boldsymbol{Y}\|_1 = Y_1 + \cdots + Y_n$ is $\mathsf{Gamma}(n\alpha, 1)$ with tail $\Gamma(n\alpha, y)/\Gamma(n\alpha)$, one has for any $x > 0$

$$\begin{aligned} \mathbb{P}(X_1 + \cdots + X_n > x) \;&=\; \mathbb{P}\Big(\|\boldsymbol{X}\|_1 > x\Big) \\ &\leq\; \mathbb{P}\Big(\|\boldsymbol{X}\|_\beta^\beta > x^\beta/n^{\beta-1}\Big) = \mathbb{P}\Big(\|\boldsymbol{Y}\|_1 > kx^\beta/n^{\beta-1}\Big)\,, \end{aligned}$$

and similarly for the lower bound. □

The (upper) incomplete gamma function $\Gamma(\alpha, x)$ appearing here is available in most standard software, but note that an even simpler lower bound comes from $\Gamma(\alpha, x) \geq x^{\alpha-1}\mathrm{e}^{-x}$ for $x > 0$ when $\alpha = \gamma/\beta \geq 1$, resp. $\Gamma(\alpha, x) \geq x^{\alpha-1}\mathrm{e}^{-x} \times \big(x/(x+1-\alpha)\big)$ when $\alpha \in (0, 1)$. Moreover, observe that X with the density given in Prop. 3.5.1 has tail probability

$$\overline{F}_X(x) = \mathbb{P}(X > x) = \frac{\Gamma(\gamma/\beta, kx^\beta)}{\Gamma(\gamma/\beta)}\,.$$

Hence, appealing to the fact that $\Gamma(\alpha, x) \sim x^{\alpha-1}\mathrm{e}^{-x}$ as $x \to \infty$, the upper bound in Prop. 3.5.1 is asymptotically

$$\frac{\Gamma(\gamma/\beta)^n}{\Gamma(n\gamma/\beta)}\, n^{n\gamma/\beta-1}\, k^{n-1}\, \left(\frac{x}{n}\right)^{\beta(n-1)} \overline{F}_X(x/n)^n\,.$$

When $\gamma = \beta$ (the ordinary Weibull case), the ratio of this upper bound to the true asymptotic form for $\mathbb{P}(X_1 + \cdots + X_n > x)$ is

$$\frac{n^{(n-1/2)}}{(n-1)!}\left[\frac{(\beta-1)}{2\pi\beta}\right]^{(n-1)/2} k^{(n-1)/2}\, \left(\frac{x}{n}\right)^{\beta(n-1)/2}\,,$$

so the upper bound is out only by a polynomial factor in x, which indicates it is close to the true probability on a logarithmic scale. More precisely, writing $U(x)$ for the upper bound and $P(x)$ for the true probability, it holds trivially that $x^{-1}\log(U(x)) \sim x^{-1}\log(P(x))$ as $x \to \infty$.

It is straightforward to extend Prop. 3.5.1 to the following slightly more general form.

Proposition 3.5.2. *Let $\{X_i\}_{i=1}^n$ be independent random variables with density $\beta k^{\gamma_i/\beta} x^{\gamma_i-1}\mathrm{e}^{-kx^\beta}/\Gamma(\gamma_i/\beta)$ $x > 0$, where $k > 0$, $\beta \geq 1$, and $\gamma_i > 0$, for $i = 1, \ldots, n$. Then with $\gamma_0 = \sum_{i=1}^n \gamma_i$, it holds that*

$$\frac{\Gamma(\gamma_0/\beta, kx^\beta)}{\Gamma(\gamma_0/\beta)} \;\leq\; \mathbb{P}(X_1 + \cdots + X_n > x) \;\leq\; \frac{\Gamma(\gamma_0/\beta, kx^\beta/n^{\beta-1})}{\Gamma(\gamma_0/\beta)}\,.$$

3.6 M.g.f.'s and the exponential family

In this section, we assume that $X \sim F$ has the tail asymptotics $\gamma(x)\mathrm{e}^{-x^\beta}/\lambda(x)$ for some $\beta > 1$ where $\lambda(x) = \beta x^{\beta-1}$. Define

$$\widehat{F}[\theta] \;=\; \mathbb{E}[\mathrm{e}^{\theta X}] \;=\; \int_{-\infty}^{\infty} \mathrm{e}^{\theta z}\, F(\mathrm{d}z)\,, \quad F_\theta(\mathrm{d}z) \;=\; \frac{\mathrm{e}^{\theta z}}{\widehat{F}[\theta]} F(\mathrm{d}z)$$

where expectations with respect to F_θ will be denoted $\mathbb{E}_\theta[\cdot]$. Determining the asymptotics of $\widehat{F}[\theta]$ and characteristics of the exponential family like their moments is easier when taking $\theta = \lambda(x)$. For a general θ, one then just have to substitute $x = \lambda^{\leftarrow}(\theta)$ in the following result.

Proposition 3.6.1. *As $x \to \infty$, it holds that*

$$\widehat{F}\big[\lambda(x)\big] \;\sim\; \sqrt{\frac{2\pi}{\lambda'(x)}}\gamma(x)\mathrm{e}^{(\beta-1)x^\beta}, \tag{3.22}$$

$$\mathbb{E}_{\lambda(x)}X \;\sim\; x. \tag{3.23}$$

Further, we have the following convergence in $\mathbb{P}_{\lambda(x)}$-distribution as $x \to \infty$

$$\sqrt{\lambda'(x)}(X-x) = \sqrt{\beta(\beta-1)x^{\beta-2}}(X-x) \xrightarrow{\mathcal{D}} N\Big(0,1\Big). \tag{3.24}$$

Proof. Suppose for simplicity that X is non-negative. In view of Proposition 3.2 in BKR we can assume w.l.o.g. that $\gamma \in C^\infty$. By Theorem 3.4.1 we have that $\overline{F}(x) \sim \overline{H}(x)$ where H has the density $\gamma(z)\mathrm{e}^{-z^\beta}$ for $z \geq 0$. It follows easily from our proof below that $\mathbb{E}[X^k\mathrm{e}^{\lambda(x)X}] \sim \mathbb{E}[X_*^k\mathrm{e}^{\lambda(x)X_*}]$ for $k \geq 0$ with $X_* \sim H$, so we assume w.l.o.g. that F has the density $f(z) = \gamma(z)\mathrm{e}^{-z^\beta}$ for $z \geq 0$. Lastly, we have that $\gamma(x) = \mathrm{o}(\mathrm{e}^{cx})$ for any $c > 0$, as

$$\lim_{x\to\infty} \frac{\gamma'(x)}{\sqrt{\lambda'(x)}\gamma(x)} \;=\; 0. \tag{3.25}$$

For $g(z) = z^k\mathrm{e}^{\lambda(x)z}$, with $k \geq 0$, it follows by integration by parts that

$$\begin{aligned}
\mathbb{E}[X^k\mathrm{e}^{\lambda(x)X}] &= g(0) + \int_0^\infty g'(z)\overline{F}(z)\mathrm{d}z \\
&= \mathbb{I}\{k=0\} + \int_0^{c_1x} g'(z)\overline{F}(z)\mathrm{d}z + \int_{c_1x}^\infty g'(z)\overline{F}(z)\mathrm{d}z \\
&= \mathcal{O}(\mathrm{e}^{\widetilde{c}_1x^\beta}) + \int_{c_1x}^\infty \Big[kz^{k-1} + \lambda(x)z^k\Big]\mathrm{e}^{\lambda(x)z}\frac{\gamma(z)}{\lambda(z)}\mathrm{e}^{-z^\beta}\mathrm{d}z
\end{aligned} \tag{3.26}$$

for any $0 < c_1 < \widetilde{c}_1 < 1$ sufficiently small.

Consider integrals of the form $\int_{c_1x}^\infty z^k\mathrm{e}^{\lambda(x)z}\gamma(z)\mathrm{e}^{-z^\beta}\mathrm{d}z$ and note that the global maximum of the exponent $\lambda(x)z - z^\beta$ is at $z = x$. We use the substitution, similar to those in Sections 3.2 and 3.3, of $z = x + y/\lambda(x)$ and note that

$$\lambda(x)z - z^\beta \;\sim\; (\beta-1)x^\beta - \frac{y^2\lambda'(x)}{2\lambda(x)^2}.$$

Therefore, for any $D > 0$ we have for $x \to \infty$

$$\begin{aligned}
&\int_{c_1 x}^{\infty} z^k \frac{\gamma(z)}{\lambda(z)} \mathrm{e}^{\lambda(x)z - z^\beta} \mathrm{d}z \;\sim\; \int_{x-D/\lambda(x)}^{x+D/\lambda(x)} z^k \frac{\gamma(z)}{\lambda(z)} \mathrm{e}^{\lambda(x)z - z^\beta} \mathrm{d}z \\
\sim\;& \int_{-D}^{D} \Big(x + \frac{y}{\lambda(x)}\Big)^k \frac{\gamma\Big(x + \frac{y}{\lambda(x)}\Big)}{\lambda\Big(x + \frac{y}{\lambda(x)}\Big)} \exp\Big\{(\beta - 1)x^\beta - \frac{y^2 \lambda'(x)}{2\lambda(x)^2}\Big\} \frac{1}{\lambda(x)} \mathrm{d}y \\
\sim\;& x^k \frac{\gamma(x)}{\lambda(x)^2} \mathrm{e}^{(\beta-1)x^\beta} \int_{-D}^{D} \exp\Big\{-\frac{y^2 \lambda'(x)}{2\lambda(x)^2}\Big\} \mathrm{d}y \;\sim\; \sqrt{\frac{2\pi}{\lambda'(x)}} x^k \frac{\gamma(x)}{\lambda(x)} \mathrm{e}^{(\beta-1)x^\beta}
\end{aligned}$$

where the replacement of the limits $\pm\, D$ by $\pm\,\infty$ follows from $\lambda'(x)/\lambda(x)^2 \to 0$. Combining this integral asymptotic with (3.26) we get

$$\begin{aligned}
\mathbb{E}[X^k \mathrm{e}^{\lambda(x)X}] \;&=\; \mathcal{O}(\mathrm{e}^{\widetilde{c_1} x^\beta}) + k \int_{c_1 x}^{\infty} z^{k-1} \frac{\gamma(z)}{\lambda(z)} \mathrm{e}^{\lambda(x)z - z^\beta} \mathrm{d}z \qquad (3.27)\\
&\qquad + \lambda(x) \int_{c_1 x}^{\infty} z^k \frac{\gamma(z)}{\lambda(z)} \mathrm{e}^{\lambda(x)z - z^\beta} \mathrm{d}z \\
&=\; \mathcal{O}(\mathrm{e}^{\widetilde{c_1} x^\beta}) + \sqrt{\frac{2\pi}{\lambda'(x)}} \gamma(x) \mathrm{e}^{(\beta-1)x^\beta} \Big(x^k + \frac{k}{\lambda(x)} x^{k-1}\Big), \qquad (3.28)
\end{aligned}$$

or to take only the largest term,

$$\mathbb{E}[X^k \mathrm{e}^{\lambda(x)X}] \sim \sqrt{\frac{2\pi}{\lambda'(x)}} \gamma(x) x^k \mathrm{e}^{(\beta-1)x^\beta} \quad \text{as } x \to \infty.$$

From this (3.22)–(3.23) are easy.

Next, we show the asymptotic normality. By the above arguments, we assume for simplicity that F has density $f(z) = \gamma(z)\mathrm{e}^{-z^\beta}$ for all $z > 0$. Similarly, writing instead $z = x + y/\sqrt{\lambda'(x)}$, we have

$$\lambda(x)z - z^\beta \;\sim\; (\beta - 1)x^\beta - \frac{y^2}{2}.$$

For some $D < \min(0, v)$ we obtain

$$\begin{aligned}
&\int_{x+D/\sqrt{\lambda'(x)}}^{x+v/\sqrt{\lambda'(x)}} \gamma(z) \exp\Big\{\lambda(x)z - z^\beta\Big\} \mathrm{d}z \\
\sim\;& \frac{1}{\sqrt{\lambda'(x)}} \int_{D}^{v} \gamma\Big(x + y/\sqrt{\lambda'(x)}\Big) \exp\Big\{(\beta - 1)x^\beta - \frac{y^2}{2}\Big\} \mathrm{d}y \\
\sim\;& \frac{1}{\sqrt{\lambda'(x)}} \gamma(x) \mathrm{e}^{(\beta-1)x^\beta} \int_{D}^{v} \exp\Big\{-\frac{y^2}{2}\Big\} \mathrm{d}y.
\end{aligned}$$

Hence, letting $D \to -\infty$ yields

$$\mathbb{E}[\mathrm{e}^{\lambda(x)X};\ \sqrt{\lambda'(x)}(X-x) \le v] \sim \sqrt{\frac{2\pi}{\lambda'(x)}}\gamma(x)\mathrm{e}^{(\beta-1)x^\beta}\Phi(v).$$

Dividing by (3.22) gives $\mathbb{P}_{\lambda(x)}\Big(\sqrt{\lambda'(x)}(X-x) \le v\Big) \to \Phi(v)$ which is (3.24). □

Remark 3.6.2. *Asymptotic normality for the general case $\overline{F}(x) = \mathrm{e}^{-\psi(x)}$ similar to the result of Proposition 3.6.1 is derived in [15].* ◇

Remark 3.6.3. *The BKR method of proof is modelled after the standard proof of the saddlepoint approximation: exponential change of measure using estimates of the above type. One has*

$$\mathbb{P}(S_n > x) \quad = \quad \widehat{F}[\theta]^n \mathbb{E}_\theta\Big[\mathrm{e}^{-\theta S_n};\ S_n > x\Big] \tag{3.29}$$

and should take θ such that $\mathbb{E}_\theta S_n = x$, i.e. $\theta = \lambda(x/n)$. The approximately normality of $(X_1,\ldots,X_n)$ gives that S_n is approximately normal $\big(x, n/\lambda'(x/n)\big)$. So, one can compute

$$\mathbb{E}_{\lambda(x/n)} \exp\Big\{-a\sqrt{\lambda'(x/n)/n}\, S_n\Big\}$$

for any fixed a but $\theta = \lambda(x/n)$ is of a different order than $\sqrt{\lambda'(x/n)/n}$. Therefore (as for the saddlepoint approximation) a sharper CLT is needed, and this is maybe the most demanding part of the BKR approach. ◇

3.7 Compound Poisson sums

We consider here $S_N = X_1+\cdots+X_N$ where N is Poisson(μ) and independent of $X_1, X_2, \ldots$, where $X_i \sim \mathsf{Weibull}(\beta)$. The asymptotics of $\mathbb{P}(S_N > x)$ are important in many applications, for example actuarial sciences [7], and can be investigated using classical saddlepoint techniques. The relevant asymptotic is the classical Esscher approximation :

$$\mathbb{P}(S_N > x) \sim \frac{\Big(\widehat{F_{S_N}}[\theta] - \mathrm{e}^{-\mu}\Big)\exp\{-\theta x\}}{\theta\sigma_c(\theta)} B_0(\ell)\,, \tag{3.30}$$

where θ is the solution to $\mu\widehat{F}'[\theta] = x$, and $\widehat{F_{S_N}}[\theta] = \exp\{\mu(\widehat{F}[\theta]-1)\}$, $B_0(l) = l\mathrm{e}^{l^2/2}(1-\Phi(l)) \to (2\pi)^{-1/2}$, $\sigma_c^2(\theta) = \mu F''[\theta]$, and $\ell = \theta\sigma_c(\theta)$. See (7.1.10) in [**?**], where also further refinements and variants are given. The issue with implementing (3.30) is that we do not usually have access to $\widehat{F}[\theta]$; note, Mathematica can derive $\widehat{F}[\theta]$ when $\beta = 1.5$, 2, or 3.

For standard $\mathsf{Weibull}(\beta)$ variables, (3.22) simplifies to

$$\widehat{F}[t] \sim \sqrt{\frac{2\pi\beta^{\frac{1}{1-\beta}}}{\beta-1}} t^{\frac{\beta}{2(\beta-1)}} \mathrm{e}^{(\beta-1)(t/\beta)^{\frac{\beta}{\beta-1}}} =: \widetilde{F}[t] \,.$$

Unfortunately $\widehat{F_{S_N}}[t] \not\sim \exp\{\mu(\widetilde{F}[t]-1)\}$, though $\widehat{F_{S_N}}[t] \approx_{\log} \exp\{\mu(\widetilde{F}[t]-1)\}$, where the notation $h_1(x) \approx_{\log} h_2(x)$ means that $\log h_1(x)/\log h_2(x) \to 1$.

One can select the θ which solves $\mu\widetilde{F}'[\theta] = x$, however it seems this must be done numerically. An alternative is the asymptotic forms for $\widehat{F}^{(k)}$ from (3.28). Take

$$\widehat{F}^{(k)}[\theta] = \mathbb{E}[X^k \mathrm{e}^{\theta X}] \sim y^k \widehat{F}[\theta], \quad \text{for } k \in \mathbb{N} \tag{3.31}$$

where we've written $\theta = \lambda(y)$ as in Section 3.6. Thus if we set θ as the solution to $\mu y \widetilde{F}[\lambda(y)] = x$ then we get

$$y = 2^{-1/\beta} \left[\frac{(\beta+2)}{(\beta-1)\beta} \mathcal{W}\left(\frac{(\beta-1)\beta}{(\beta+2)} \left(\frac{2^{\frac{1}{\beta}+\frac{1}{2}} x}{c_1} \right)^{\frac{2\beta}{\beta+2}} \right) \right]^{1/\beta} \tag{3.32}$$

where $\mathcal{W}$ is the Lambert W function and $c_1 = \mu\sqrt{2\pi}\beta/\sqrt{(\beta-1)\beta}$.

With this choice of θ, we can say $\widehat{F}^{(k)}[\theta] \sim xy^{k-1}$, so $\sigma_c^2(\theta) \sim \mu xy$ and $\ell \sim \lambda(y)\sqrt{\mu xy}$, and substituting this into (3.30) gives us

$$\mathbb{P}(S_N > x) \approx_{\log} \frac{\mathrm{e}^{-\mu}\Big(\exp\{\mu x/y\} - 1\Big)\exp\{-\theta x\}}{\lambda(y)\sqrt{\mu xy}} B_0(\ell) \,. \tag{3.33}$$

Preliminary numerical work indicates that (3.33) is not particular accurate in the whole range of relevant parameters. The problem derives from the fact we only have log-asymptotics for $\widehat{F_{S_N}}[\theta]$; finding more accurate asymptotics is left for future work.

A further interesting extension could be the asymptotic form of $\mathbb{P}\Big(Z(t) > x\Big)$ where Z is a Lévy process where the Lévy measure has tail $\gamma(x)\mathrm{e}^{-\psi(x)}$.

3.8 The exponential class of distributions

For $F \in \mathsf{GMDA}(e)$ in the previous sections we have discussed the case that $e(x) = 1/\lambda(x)$ with

$$\lim_{x\to\infty} e(x) = 0.$$

If $\lim_{x\to\infty} e(x) = \infty$, then F is long-tailed in the sense that $\overline{F}(x-y) \sim \overline{F}(x)$ for any fixed y. Convolutions of distributions with long-tailed are well-understood. The intermediate case is that

$$\lim_{x\to\infty} e(x) = 1/\gamma, \quad \gamma > 0.$$

For such F we have

$$\overline{F}(x+s) \sim \mathrm{e}^{-\gamma s}\overline{F}(x), \quad x \to \infty$$

for any $s \in \mathbb{R}$, which is also denoted as $F \in \mathcal{L}(\gamma)$. Note in passing that any distribution $F \in \mathsf{GMDA}(e)$ with upper endpoint infinity satisfies (see e.g. [?, Prop. 1.4])

$$\overline{F}(x) \sim \overline{H}(x) = C \exp\Big(-\int_0^x \frac{1}{u(t)}\mathrm{d}t\Big), \quad x \to \infty \tag{3.34}$$

for some $C > 0$, where u is absolutely continuous with respect to Lebesgue measure, with density u' satisfying $\lim_{x\to\infty} u'(x) = 0$. Such H is commonly referred to as a von Mises distribution.

It is well-known ([?], [?]) that the class of distributions $\mathcal{L}(\gamma)$ is closed under convolution. In the particular case that the X_i have tails

$$\overline{F}_i(x) = \ell_i(x)x^{\gamma_i-1}\mathrm{e}^{-kx^\beta}, \quad 1 \le i \le n, \tag{3.35}$$

where ℓ_i's are positive slowly varying functions and $\beta = 1, \gamma_i > 0, i \le n, k > 0$ we have in view of Theorem 2.1 in [?] (see also Theorem 6.4 ii) in [4])

$$\mathbb{P}\left(S_n > x\right) \sim \frac{k^{n-1}}{\Gamma(\gamma_0)} x^{\gamma_0-1} \prod_{i=1}^n \ell_i(x)\mathrm{e}^{-kx^\beta}. \tag{3.36}$$

where $\gamma_0 = \sum_{i=1}^n \gamma_i$. If (3.36) holds with $\beta > 1$, then for non-negative X_i's using the β-norm argument we have as in Section 3.5

$$\mathbb{P}\left(S_n > x\right) \le \mathbb{P}\left(X_1^\beta + \cdots + X_n^\beta > x^\beta/n^{\beta-1}\right) \tag{3.37}$$

for any $x > 0$. Since $\mathbb{P}(X_1^\beta > x) \sim \ell_i(x^{1/\beta})x^{(\gamma_i-1)/\beta}\mathrm{e}^{-kx}$, then by (3.36) and Theorem 3.4.1

$$\ln \mathbb{P}\left(S_n > x\right) \sim \ln \mathbb{P}\left(X_1^\beta + \cdots + X_n^\beta > x^\beta/n^{\beta-1}\right) \sim kn(x/n)^\beta$$

and thus the upper bound in (3.37) is logarithmic asymptotically exact.

3.9 Applications to Monte Carlo simulation

In this section, we write $h_1(x) \approx_{\log} h_2(x)$ if $\log h_1(x)/\log h_2(x) \to 1$ and $\le_{\log}$ if the lim sup of the ratio of log's is at most 1, and we take the summands to have a density like $\gamma(x)\mathrm{e}^{-x^\beta}$ as $x \to \infty$.

Algorithms for tails $\mathbb{P}(S_n > x)$ with large x are one of the traditional objects of study of the rare-event simulation literature. An *estimator* is a r.v. $Z(x)$ with $\mathbb{E}Z(x) = \mathbb{P}(S_n > x)$ and its efficiency is judged by ratios of the form $r_p(x) = \mathbb{E}Z(x)^2/\mathbb{P}(S_n > x)^p$. The estimator will improve upon crude Monte Carlo simulation if $r_1(x) \to 0$ as $x \to \infty$. It is said to have bounded relative error if $r_2(x)$ stays bounded as $x \to \infty$ and to exhibit logarithmic efficiency if $r_{2-\varepsilon}(x) \to 0$ for all $\varepsilon > 0$ which in turn will hold if $\mathbb{E}Z(x)^2 \approx_{\log} \mathbb{P}(S_n > x)^2$. These two concepts are usually considered in some sense optimal. For a survey, see Chapters V–VI in [8].

The conventional light-tailed rare-event folklore says that a particular kind of importance sampling, *exponential tilting*, is often close to optimal. Here instead of $\mathbb{I}(S_n > x)$ one returns

$$Z_\theta(x) \;=\; \mathbb{I}\{S_n > x\} \times L_\theta \quad \text{where} \;\; L_\theta = \widehat{F}[\theta]^n \exp\{-\theta S_n\}$$

where $X_1, \ldots, X_n$ are i.i.d. with density $f_\theta(y) = \mathrm{e}^{\theta y} f(y)/\widehat{F}(\theta)$ rather than the given density $f(x)$, and θ is chosen such that $\mathbb{E}_\theta X = x/n$, that is, $\theta = \lambda(x/n)$. The standard efficiency results do, however, require both $n \to \infty$ and $x \to \infty$ such that $nx \sim z$ for some $z > \mathbb{E}X$ and therefore do not deal with a fixed n, the object of this paper. It is believed that the scheme is still often close to optimal in this setting, but very few rigorous results in this direction has been formulated. We give one such in Proposition 3.9.2 below.

One problem that arises is how to simulate from f_θ. Proposition 3.6.1 tells us that f_θ is asymptotically normal with mean x/n and variance $1/\lambda'(x/n)$ when $\theta = \lambda(x/n)$. So we simulate using acceptance–rejection with a moment-matched gamma distribution as proposal, and our acceptance ratio will increase to 1 as $x \to \infty$. To be specific, we take a $\mathsf{Gamma}(a, b)$ proposal, which has a density $f_{a,b}(y) \propto y^{a-1}\mathrm{e}^{-by}$, where $a = x^2\lambda'(x/n)/n^2$, and $b = x\lambda'(x/n)/n$. The reason we do not directly use a the limiting normal distribution as a proposal is that the tail of the normal distribution is too light when $\beta \in (1, 2)$.

Remark 3.9.1. *The acceptance ratio can be improved for small x by locally searching for the optimal proposal, that is, the distribution with parameters*

$$(\mu^*, \sigma^*) = \underset{\mu,\sigma>0}{\arg\min} \; \max_{y \geq 0} \; \frac{f_{\lambda(x/n)}(y)}{f_{\mathrm{Prop}}(y; \mu, \sigma^2)} \,.$$

The asymptotic $(\mu, \sigma) = (x/n, 1/\sqrt{\lambda'(x/n)})$ can be used as the initial search point. In experiments, it seems that the asymptotic variance is close to optimal, whereas some efficiency can be gained by adjusting the mean parameter. $\diamond$

Proposition 3.9.2. *The estimator $Z_\theta(x)$ exhibits logarithmic efficiency.*

Proof. We first note that

$$\mathbb{E}_\theta[Z_\theta(x)^2] \;=\; \mathbb{E}_\theta[L_\theta^2;\, S_n > x] \;=\; \mathbb{E}[L_\theta;\, S_n > x] \;\leq\; \mathrm{e}^{-\theta x}\widehat{F}[\theta]^n \mathbb{P}(S_n > x)\,.$$

By Corollary 3.3.5 and (3.22),

$$\overline{F}^{*n}(x) \;\approx_{\log}\; \exp\{n(x/n)^\beta\}\,, \quad \widehat{F}[\lambda(x/n)]^n \;\approx_{\log}\; \exp\{n(\beta-1)(x/n)^\beta\}\,.$$

From $\theta = \lambda(x/n) = \beta(x/n)^{\beta-1}$ we then get

$$\begin{aligned}\frac{\mathbb{V}\mathrm{ar}_\theta(Z_\theta(x))}{\mathbb{P}(S_n > x)} \;&\le\; \frac{\mathbb{E}_\theta[Z_\theta(x)^2]}{\mathbb{P}(S_n > x)} \\ &\le_{\log}\; \exp\Big\{-\theta x + n(\beta-1)(x/n)^\beta + n(x/n)^\beta\Big\} \\ &=\; \exp\Big\{-\beta(x/n)^{\beta-1}x + n\beta(x/n)^\beta\Big\} \;=\; 1,\end{aligned}$$

completing the proof. □

Some estimators based on conditional Monte Carlo ideas are discussed in [6] and efficiency properties derived in some special cases. The algorithms do improve upon crude Monte Carlo, though logarithmic efficiency is not obtained. The advantage is, however, that they are much easier implemented than the above exponential tilting scheme. The next two propositions extend results of [6] to more general tails.

Proposition 3.9.3. *Consider the conditional Monte Carlo estimator $Z_{\mathrm{Cd}}(x) = \overline{F}(x - S_{n-1})$ of $\mathbb{P}(S_n > x)$. Then $\limsup r_p(x) < \infty$ whenever $p < p_n$ where $p_n = n^{\beta-1}c_n$ with c_n given by* (3.38) *below. Here $p_n > 1$.*

Proof. We have $\mathbb{E}Z_{\mathrm{Cd}}(x)^2 = \int \overline{F}(x-y)^2 f^{*(n-1)}(y)\,\mathrm{d}y$ where the asymptotics of the integral is covered by Theorem 3.3.3. In the setting there, $c_1 = 2$, $c_2 = 1/(n-1)^{\beta-1}$ which gives $\theta_1 = 1/(1+\mu)$, $\theta_2 = \mu/(1+\mu)$ where $\mu = 2^{1/(\beta-1)}(n-1)$. The result gives that $\mathbb{E}Z_{\mathrm{Cd}}(x)^2 \approx_{\log} \mathrm{e}^{-c_n x^\beta}$ where

$$c_n \;= c_1\theta_1^\beta + c_2\theta_2^\beta \;=\; \frac{2 + 2^{\beta/(\beta-1)}(n-1)}{\Big(1 + 2^{1/(\beta-1)}(n-1)\Big)^\beta}\,. \tag{3.38}$$

Since $\mathbb{P}(S_n > x) \approx_{\log} \mathrm{e}^{-x^\beta/n^{\beta-1}}$, this implies the first assertion of the proposition. To see that $p_n > 1$, note that for $a > 1$

$$n^{\beta-1}\frac{a^{\beta-1} + a^\beta(n-1)}{\Big(1 + a(n-1)\Big)^\beta} \;=\; \left[\frac{na}{1 + a(n-1)}\right]^{\beta-1} \;>\; \left[\frac{na}{na}\right]^{\beta-1} \;=\; 1$$

and take $a = 2^{1/(\beta-1)}$. □

We finally consider the so-called Asmussen–Kroese estimator

$$Z_{\mathrm{AK}}(x) \;=\; n\,\overline{F}\Big(M_{n-1} \vee (x - S_{n-1})\Big)\,. \tag{3.39}$$

where $M_{n-1} = \max(X_1, \dots, X_{n-1})$. It was initially developed in [11] with heavy tails in mind, but it was found empirically in [6] that it also provides some variance reduction for light tails, in fact more than $Z_{\mathrm{Cd}}(x)$. We have:

Proposition 3.9.4. *Consider the estimator $Z_{\mathrm{AK}}(x)$ of $\mathbb{P}(S_n > x)$ with $n = 2$. Then $\limsup r_p(x) < \infty$ whenever $p < 3/2$.*

Proof. When $n = 2$, we have $M_{n-1} = S_{n-1} = X_1$ and so the analysis splits into an $X_1 > x/2$ and an $X_1 \le 2$ part. The first is

$$\begin{aligned}
\mathbb{E}\Big[Z_{\mathrm{AK}}(x)^2;\, X_1 > x/2\Big] \;&=\; 4\int_{x/2}^{\infty} \overline{F}(y)^2 f(y)\,\mathrm{d}y \\
&\approx_{\log}\; \int_{x/2}^{\infty} \mathrm{e}^{-2y^\beta}\mathrm{e}^{-y^\beta}\,\mathrm{d}y \;\approx_{\log}\; \mathrm{e}^{-3x^\beta/2^\beta}.
\end{aligned}$$

The second part is

$$\begin{aligned}
\mathbb{E}\Big[Z_{\mathrm{AK}}(x)^2;\, X_1 \le x/2\Big] \;&=\; 4\int_{-\infty}^{x/2} \overline{F}(x-y)^2 f(y)\,\mathrm{d}y \\
&=\; 4\int_{x/2}^{\infty} \overline{F}(y)^2 f(x-y)\,\mathrm{d}y \;=\; 4I_1 + 4I_2
\end{aligned}$$

where I_1 is the integral over $[x/2, ax)$ and I_2 is the one over $[ax, \infty)$. Here we take $a = (3/2)^{1/\beta}/2$; since $\beta > 1$, we have $a < 3/4 < 1$. Let further $b = a - 1/2$. Then

$$\begin{aligned}
I_2 \;&=\; \int_{ax}^{\infty} \overline{F}(y)^2\mathcal{O}(1)\mathrm{d}y \;\approx_{\log}\; \int_{ax}^{\infty} \mathrm{e}^{-2x^\beta}\mathcal{O}(1)\mathrm{d}y \\
&\approx_{\log}\; \mathrm{e}^{-2a^\beta x^\beta} \;=\; \mathrm{e}^{-3x^\beta/2^\beta}\,, \\
I_1 \;&\approx_{\log}\; \int_{x/2}^{ax} \exp\Big\{-2y^\beta - (x-y)^\beta\Big\} \\
&=\; \int_0^{bx} \exp\Big\{-2(x/2+z)^\beta - (x/2-z)^\beta\Big\}\,\mathrm{d}z.
\end{aligned}$$

By convexity of $v \mapsto v^\beta$, we have

$$(u+v)^\beta \;=\; u^\beta(1+v/u)^\beta \;\ge u^\beta(1+\beta v/u) \;=\; u^\beta + \beta v u^{\beta-1}$$

for $u > 0$ and $-u < v < \infty$. Taking $u = x/2$ gives

$$I_2 \;\le_{\log}\; \int_0^{bx} \exp\Big\{-3x^\beta/2^\beta - \beta z(x/2)^{\beta-1}\Big\}\,\mathrm{d}z \;=\; \mathrm{e}^{-3x^\beta/2^\beta}\mathrm{o}(1)\,,$$

completing the proof. □

3.A Proof of Theorem 3.4.1

For the proof of Theorem 3.4.1, we first note that, as shown in BKR, that as $x \to \infty$

$$\frac{\lambda'(x)}{\lambda(x)^2} \;\to\; 0. \tag{3.40}$$

$$\frac{\gamma'(x)}{\sqrt{\lambda'(x)}\gamma(x)} \;\to\; 0. \tag{3.41}$$

In view of Proposition 3.2 in BKR, (3.41) need not hold for γ itself but does for a tail equivalent version, with which γ can be replaced w.l.o.g. This implies

$$\lambda \text{ is flat for } \psi. \tag{3.42}$$

Indeed, given y it holds for some x^* between 0 and $x + y/\sqrt{\lambda'(x)}$ that

$$\lambda\Big(x + y/\sqrt{\lambda'(x)}\Big) \;=\; \lambda(x) + \frac{\lambda'(x^*)}{\sqrt{\lambda'(x)}}y \;=\; \lambda(x) + \mathcal{O}\Big(\sqrt{\lambda'(x)}\Big) \;= \lambda(x)\Big(1 + \mathrm{o}(1)\Big)$$

where the $\mathcal{O}(\cdot)$ estimate follows from a known uniformity property of self-neglecting functions and the $\mathrm{o}(\cdot)$ estimate by (3.40). Using further (3.40) we have that $e = 1/\lambda$ is self-neglecting.

Proof of Theorem 3.4.1 (i). Write $\overline{H}(x) = \gamma(x)\mathrm{e}^{-\psi(x)}/\lambda(x)$. Then

$$\begin{aligned}\overline{H}'(x) \;&=\; \left[\gamma(x) + \frac{\gamma'(x)}{\psi'(x)} - \frac{\gamma(x)\psi''(x)}{\psi'(x)^2}\right]\mathrm{e}^{-\psi(x)} \\ &=\; \gamma(x)\left[1 + \frac{\gamma'(x)}{\gamma(x)\psi'(x)} - \frac{\psi''(x)}{\psi'(x)^2}\right]\mathrm{e}^{-\psi(x)}.\end{aligned}$$

Here the last term in $[\cdot]$ goes to 0 according to (3.40). This together with (3.41) also gives

$$\frac{\gamma'(x)}{\gamma(x)\psi'(x)} \;=\; \frac{\gamma'(x)\psi''^{-1/2}}{\gamma(x)} \cdot \frac{\psi''^{1/2}}{\psi'(x)} \;=\; \mathrm{o}(1)\cdot\mathrm{o}(1) \;=\; \mathrm{o}(1).$$

Thus $\overline{H}'(x) \sim f(x)$ which implies $\overline{H}(x) \sim \overline{F}(x)$. $\square$

We also have this an alternative proof for part (i).

Proof of Theorem 3.4.1 (i). Using integrations by parts yields

$$\begin{aligned}\int_x^\infty f(y)\,\mathrm{d}y &= \int_0^\infty \frac{\gamma(x+y)}{\psi'(x+y)} \cdot \psi'(x+y)\mathrm{e}^{-\psi(x+y)}\,\mathrm{d}y \\ &= \frac{\gamma(x)}{\psi'(x)}\mathrm{e}^{-\psi(x)} - \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{\gamma(x+y)}{\psi'(x+y)}\right] \cdot \mathrm{e}^{-\psi(x+y)}\,\mathrm{d}y.\end{aligned}$$

But by the same estimates as in Proof 1, the first part of the integrand is $\mathrm{o}\big(\gamma(x)\big)$ so that the whole integral is $\mathrm{o}\big(\overline{F}(x)\big)$. □

The following lemma is just a reformulation of part (ii) of the theorem, proved in BKR.

Lemma 3.A.1. *For any two pairs (γ_1, ψ_1), (γ_2, ψ_2) satisfying the assumptions of Section 3.1, it holds that*

$$\int_{-\infty}^\infty \gamma_1(z)\mathrm{e}^{-\psi_1(z)} \cdot \gamma_2(x-z)\mathrm{e}^{-\psi_2(x-z)}\,\mathrm{d}z \tag{3.43}$$

has the asymptotics given by Theorem 3.4.1(ii).

Proof of Theorem 3.4.1 (ii). This is a reformulation of Theorem 1.1 in BKR. Since by (3.20) $q_1' + q_2' = 1$ we have the claimed relation between λ and λ_1, λ_2, namely

$$\lambda(x) = \lambda_1\big(q_1(x)\big)q_1'(x) + \lambda_2\big(q_2(x)\big)q_2'(x) = \lambda_1(q_1) = \lambda_2(q_2) \tag{3.44}$$

establishing the proof. □

Proof of Theorem 3.4.1 (iii). We have that $\mathrm{e}^{-\psi_i(x)}, i = 1, 2$ is a von-Mises function (see (3.34)) and thus $\mathrm{e}^{-\psi_i(x)} \in \mathsf{GMDA}(e_i), i = 1, 2$ with $e_i = 1/\lambda_i$. Since further e_i's are self-neglecting and by (3.40) $r_i(x) = \sqrt{\lambda_i(x)}/\lambda_i(x) \to 0$ as $x \to \infty$ we have that

$$\lim_{x\to\infty} \frac{\gamma_i(x + e_i(x)y)}{\gamma_i(x)} = \lim_{x\to\infty} \frac{\gamma_i(x + yr_i(x)/\sqrt{\lambda_i(x)})}{\gamma_i(x)} = 1$$

uniformly on bounded y-intervals. Hence $F_i \in \mathsf{GMDA}(e_i)$. In view of Proposition 3.2 in BKR we can find smooth γ_i^*'s such that $\overline{H}_i(x) = \gamma_i^*(x)\mathrm{e}^{-\psi_i(x)}/\lambda_i(x)$ is asymptotically equivalent to $\overline{F}_i(x)$ as $x \to \infty$. Since also $H_i \in \mathsf{GMDA}(e_i)$ and $\lim_{x\to\infty} \lambda_i(x) = \infty$, then for any $c > 0$ we have

$$\lim_{x\to\infty} \frac{\overline{H}_i(x+c)}{\overline{H}_i(x)} = 0, \quad i = 1, 2.$$

Consequently, Corollary 1 in [?] yields $\overline{H_1 * H_2}(x) \sim \overline{F_1 * F_2}(x)$ and thus the claim follows from ii).
By the above, we can find the asymptotics of $\overline{F_1 * F_2}(x)$ assuming that F_i's possess a density, so alternatively we have

$$\overline{F_1 * F_2}(x) \;=\; \int_{-\infty}^{\infty} \gamma_1(z)\mathrm{e}^{-\psi_1(x)} \cdot \frac{\gamma_2(x-z)}{\lambda_2(x-z)}\mathrm{e}^{-\psi_2(x-z)}\,\mathrm{d}z \qquad (3.45)$$

But by (3.42), γ_2/λ_2 is flat for ψ_2, so using Lemma 3.20 with γ_2 replaced by γ_2/λ_2 gives that this integral asymptotically equals $\gamma(x)\mathrm{e}^{-\psi(x)}/\gamma_2\big(q_2(x)\big)$. But in view of (3.44) this is the same as $\gamma(x)\mathrm{e}^{-\psi(x)}/\lambda(x)$. This completes the proof. $\square$

Conclusion

Testing [1]

Bibliography

[1] Joseph Abate and Ward Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on computing*, 7(1):36–43, 1995.

[2] Joseph Abate and Ward Whitt. A unified framework for numerically inverting Laplace transforms. *INFORMS Journal on Computing*, 18(4):408–421, 2006.

[3] Adnan A Abu-Dayya and Norman C Beaulieu. Outage probabilities in the presence of correlated lognormal interferers. *IEEE Transactions on Vehicular Technology*, 43(1):164–173, 1994.

[4] André Adler and Anthony G Pakes. On relative stability and weighted laws of large numbers. *Extremes*, 20(1):1–31, 2017.

[5] John Aitchison and James AC Brown. *The Lognormal Distribution with Special Reference to its Uses in Economics*. Cambridge University Press, 1957.

[6] Søren Asmussen. Conditional monte carlo for sums, with applications to insurance and finance. *Annals of Actuarial Science*, 2017. Submitted, available from `thiele.au.dk/publications`.

[7] Søren Asmussen and Hansjörg Albrecher. *Ruin Probabilities*. World Scientific, 2 edition, 2010. Advanced series on statistical science & applied probability; v. 14.

[8] Søren Asmussen and Peter W Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability series*. Springer, 2007.

[9] Søren Asmussen, Jens Ledet Jensen, and Leonardo Rojas-Nandayapa. On the Laplace transform of the Lognormal distribution. *Methodology and Computing in Applied Probability*, 16:1–18, 2014.

[10] Søren Asmussen, Jens Ledet Jensen, and Leonardo Rojas-Nandayapa. Exponential family techniques in the lognormal left tail. *Scandinavian Journal of Statistics*, 2015. Pending revision.

[11] Søren Asmussen and Dirk P Kroese. Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability*, 38(2):545–558, 2006.

[12] Søren Asmussen and Leonardo Rojas-Nandayapa. Asymptotics of sums of lognormal random variables with Gaussian copula. *Statistics & Probability Letters*, 78(16):2709–2714, 2008.

[13] Efstathios Avdis and Ward Whitt. Power algorithms for inverting Laplace transforms. *INFORMS Journal on Computing*, 19(3):341–355, 2007.

[14] August A Balkema, Claudia Klüppelberg, and Sidney I Resnick. Densities with gaussian tails. *Proceedings of the London Mathematical Society*, 3(3):568–588, 1993.

[15] August A Balkema, Claudia Klüppelberg, and Sidney I Resnick. Domains of attraction for exponential families. *Stochastic processes and their applications*, 107(1):83–103, 2003.

[16] Norman C Beaulieu, Adnan A Abu-Dayya, and Peter J McLane. Estimating the distribution of a sum of independent lognormal random variables. *IEEE Transactions on Communications*, 43(12):2869–2873, 1995.

[17] Norman C Beaulieu and Faruq Rajwani. Highly accurate simple closed-form approximations to lognormal sum distributions and densities. *IEEE Communications Letters*, 8(12):709–711, 2004.

[18] Norman C Beaulieu and Qiong Xie. An optimal lognormal approximation to lognormal sum distributions. *IEEE Transactions on Vehicular Technology*, 53(2):479–489, 2004.

[19] Robert M Corless, Gaston H Gonnet, David EG Hare, David J Jeffrey, and Donald E Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5(1):329–359, 1996.

[20] Edwin L Crow and Kunio Shimizu. *Lognormal Distributions: Theory and Applications*. Dekker New York, 1988.

[21] Klaus Duellmann. *Regulatory Capital*, volume IV, pages 1525–1538. Wiley, 2010.

[22] Daniel Dufresne. The log-normal approximation in financial and other computations. *Advances in Applied Probability*, 36(3):747–773, 2004.

[23] Daniel Dufresne. Sums of lognormals. Technical report, Centre for Actuarial Sciences, University of Melbourne, 2009.

[24] Paul Embrechts, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. An academic response to Basel 3.5. *Risks*, 2(1):25–48, 2014.

[25] Lawrence Fenton. The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems*, 8(1):57–67, 1960.

[26] Xin Gao, Hong Xu, and Dong Ye. Asymptotic behavior of tail density for sum of correlated lognormal variables. *International Journal of Mathematics and Mathematical Sciences*, 2009. Volume 2009, Article ID 630857, 28 pages.

[27] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*, volume 53 of *Stochastic Modelling and Applied Probability series*. Springer, 2003.

[28] Archil Gulisashvili and Peter Tankov. Tail behavior of sums and differences of log-normal random variables. *Bernoulli*, 2015. To appear, accessed online on 26th August 2015 at `http://www.e-publications.org/ims/submission/BEJ/user/submissionFile/17119?confirm=ef609013`.

[29] Norman L Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous Univariate Distributions*, volume 1. New York: Wiley, 2 edition, 1994.

[30] Eckhard Limpert, Werner A Stahel, and Markus Abbt. Log-normal distributions across the sciences: Keys and clues. *BioScience*, 51(5):341–352, 2001.

[31] Arnaud Mallet. Numerical inversion of Laplace transform. 2000. Mathematica package, accessed online on 26th August 2015 at `http://library.wolfram.com/infocenter/MathSource/2691`.

[32] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[33] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2nd edition, 2015.

[34] Moshe Arye Milevsky and Steven E Posner. Asian options, the sum of lognormals, and the reciprocal gamma distribution. *Journal of Financial and Quantitative Analysis*, 33(03):409–422, 1998.

[35] Stuart C Schwartz and Yu-Shuan Yeh. On the distribution function and moments of power sums with log-normal components. *Bell System Technical Journal*, 61(7):1441–1462, 1982.

[36] Harald Stehfest. Algorithm 368: Numerical inversion of Laplace transforms [d5]. *Communications of the ACM*, 13(1):47–49, 1970.