THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Computational methods for sums of random variables

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.

Declaration by author

(All candidates to reproduce this section in their thesis verbatim)

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the Copyright Act 1968.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Start this section on a new page.

List your publications arising from your candidature using the standard citation format for your discipline. Divide your publications into sections as appropriate in your discipline e.g. peer-reviewed papers, book chapters, conference abstracts. If you have no publications from your candidature then state "No publications".

Publications that are detailed in the section below (Publications included in this thesis) should also be listed in this section.

Publications included in this thesis

If you choose to incorporate your publications into your thesis as per UQ policy (PPL 4.60.07 Alternative Thesis Format Options) use this section to detail the publication, where it appears in the thesis and your contribution to the authorship if you are not a sole author. Describe the contributions of all authors to the publication as per the requirements in the UQ Authorship Policy (PPL 4.20.04 Authorship).

For example;
Publication citation âĂŞ incorporated as Chapter 3.

Contributor	Statement of contribution
Author XXXX (Candidate)	Designed experiments (60%) Wrote the paper (70%)
Author YYYY	Designed experiments (40%) Wrote and edited paper (30%)
Author ZZZZ	Statistical analysis of data in tables 2 and 3

If you have no publications included in this thesis then state "No publications included".

Contributions by others to the thesis

List the significant and substantial inputs made by others to the research, work and writing represented and/or reported in the thesis. These could include significant contributions to: the conception and design of the project; non-routine technical work; analysis and interpretation of research data; drafting significant parts of the work or critically revising it so as to contribute to the interpretation. If no one contributed significantly then state "No contributions by others."

Statement of parts of the thesis submitted to qualify for the award of another degree

If part of the thesis work was used for another award, you must list the relevant parts of the thesis, the degree name, year and institution, and the outcome of the submission of material. If no parts of the thesis have been submitted in this way then state "None".

For example:

Chapters two and three submitted for MPhil, The University of Queensland, 2006, thesis withdrawn from assessment and project converted to PhD.
Chapter Five, page 88 paragraph 2 to page 89 paragraph 4 submitted for BSc Honours, Queensland University of Technology, 2005, degree awarded 19 December 2005.

Acknowledgements

Start this section on a new page.

Free text section for you to record your acknowledgment and gratitude for the more general academic input and support; financial support from grants and scholarships; and the non-academic support you have received during the course of your candidature.

PJL was supported by an Australian Government Research Training Program Scholarship and an Australian Research Council Centre of Excellence for Mathematical & Statistical Frontiers Scholarship.

... We are grateful to Jakob Schach Møller for helpful discussions on orthonormal expansions...

Keywords

Maximum 10 words; user lower case throughout, separating words/phrases with commas.

For example: word, word word, word, word, word word

Australian and New Zealand Standard Research Classifications (ANZSRC)

Provide data that links your thesis to the disciplines and discipline clusters in the Federal GovernmentâĂŹs Excellence in Research for Australia (ERA) initiative.

Please allocate the thesis a maximum of 3 Australian and New Zealand Standard Research Classifications (ANZSRC) codes at the 6 digit level and include the descriptor and a percent weighting for each code. Total percent must add to 100. `http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/6BB427AB9696C225CA2574180004463E?opendocument`

Example:
ANZSRC code: 060101, Analytical Biochemistry, 60%
ANZSRC code: 060104, Cell Metabolism, 20%
ANZSRC code: 060199, Biochemistry and Cell Biology not elsewhere classified, 20%

Fields of Research Classification

Allows for categorisation of the thesis according to the field of research.

Please allocate the thesis a maximum of 3 Fields of Research (FoR) Codes at the 4 digit level and include the descriptor and a percent weighting for each code. Total percent must add to 100.

`http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/6BB427AB9696C225CA257418000446`
`opendocument`

Example:
FoR code: 0601, Biochemistry and Cell Biology, 80%
FoR code: 0699, Other Biological Sciences, 20%

Contents

List of Abbreviations used in the thesis

TODO: Add list of notation here.

Chapter 1

Introduction

> "In theory there is no difference between theory and practice. In practice there is." Yogi Berra

1.1 Motivations

1.2 Foundational background

- Probability background: Distributions. SLN.
- Asymptotic analysis. Saddlepoint approximations.
- Monte Carlo techniques: common random numbers, quasi-Monte Carlo.
- Numerical integration techniques: MCI, Gauss–Hermite quadrature.

1.3 Contributions

Pushout paper

We introduce a novel unbiased estimator for the density of a sum of random variables. Our estimator possesses several advantages over the conditional Monte Carlo approach. Specifically, it applies to the case of dependent random variables, allows for transformations of random variables, is computationally faster to run, and is simpler to implement. We provide several numerical examples that illustrate these advantages.

SLN orthogonal paper

Approximations for an unknown density g in terms of a reference density f_ν and its associated orthonormal polynomials are discussed. The main application is the approximation of the density f of a sum S of lognormals which may have different variances or be dependent. In this setting, g may be f itself or a transformed density, in particular that of $\log S$ or an exponentially tilted density. Choices of reference densities f_ν that are considered include normal, gamma and lognormal densities. For the lognormal case, the orthonormal polynomials are found in closed form and it is shown that they are not dense in $\mathcal{L}^2(f_\nu)$, a result that is closely related to the lognormal distribution not being determined by its moments. This therefore warns against the most obvious choice of taking f_ν as lognormal. Numerical examples are presented and comparisons are made to an established approach, the Fenton–Wilkinson method, and a recent approach, the log skew normal approximation. Also, the extensions to density estimation for statistical data sets and non-Gaussian copulas are outlined.

SLP orthogonal paper

Two numerical methods are proposed to evaluate numerically the survival function of a compound distribution and the stop-loss premium associated to a non-proportional global reinsurance treaty. The first method relies on a representation of the probability density function in terms of Laguerre polynomial and the gamma density, the second is a numerical inversion of the Laplace transform. Numerical comparison are conducted at the end of the paper.

SLN Laplace transform paper

Let $(X_1, \dots, X_n)$ 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.

Weibull sums paper

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.

Maxima paper

We consider the general problem of estimating probabilities which arise as a union of dependent events. We propose a flexible series of estimators for such probabilities, and describe variance reduction schemes applied to the proposed estimators. We derive efficiency results of the estimators in rare-event settings, in particular those associated with extremes. Finally, we examine the performance of our estimators in a numerical example.

Chapter 2

Monte Carlo Estimation of the Density of the Sum of Dependent Random Variables

2.1 Introduction

Sums of random variables are fundamental to modeling stochastic phenomena. In finance, risk managers need to predict the distribution of a portfolio's future value which is the sum of multiple assets; similarly, the distribution of the sum of an individual asset's returns over time is needed for valuation of some exotic (e.g. Asian) options [103, 124]. In insurance, the probability of ruin (i.e. bankruptcy) is determined by the distribution of aggregate losses (sums of individual claims of random size) [88, 13]. Lastly, wireless system engineers model total interference in a wireless communications network as the sum of all interfering signals (often lognormally distributed) [62].

In this article, we consider estimating the probability density function (pdf) of sums of random variables (rvs). A major motivation for obtaining accurate pdf estimates of a rv is to produce confidence intervals for quantiles. For example, the US Nuclear Regulatory Commission specifies regulations in terms of the "95/95" rule, i.e. the upper 95% confidence interval for a 95% quantile. The most common approach [12] is to first estimate the cumulative distribution function (cdf) via

$$\widehat{F}_X(x) = \frac{1}{R}\sum_{r=1}^{R} \mathbb{I}_{\{X^{[r]} \le x\}} \quad \text{for } X^{[1]}, \ldots, X^{[R]} \overset{\text{i.i.d.}}{\sim} F_X\,,$$

and then the quantile $\widehat{q}_\alpha = \widehat{F}_X^{-1}(\alpha)$. In the obvious notation, we then have the convergence

in distribution:

$$\sqrt{R}(\widehat{q}_\alpha - q_\alpha) \xrightarrow{\mathcal{D}} \mathsf{N}\left(0, \frac{\alpha(1-\alpha)}{f_X(q_\alpha)^2}\right) \quad \text{as } R \to \infty,$$

where the limiting variance depends on the unknown density $f_X(q_\alpha)$. Thus, any confidence intervals for $\widehat{q}_\alpha$ require estimation of the density $f_X(q_\alpha)$, which is a highly nontrivial problem.

In general, the pdf of a sum of rvs is only available via an n-dimensional convolution. The convolution cannot be computed analytically or numerically (via quadrature), except in the special cases of normal or exponential rvs. For this reason, one has to resort to density estimation methods such as kernel density estimation [34], Conditional Monte Carlo [12], or a modification of the Asmussen–Kroese estimator [20].

The purpose of this work is to present a novel Monte Carlo estimator of the pdf of the sum of (in)dependent rvs. There are three main advantages of the proposed estimator. First, we show that the estimator often enjoys smaller variance than its competitors. Second, the estimator only requires evaluation of the joint pdf up to an (typically unknown) normalizing constant, a situation similar to the application of Markov chain Monte Carlo. As a result of this, the estimator is useful in estimating posterior marginal densities in Bayesian inference (Section 2.4.2). Finally, when the rvs have a copula dependence, the proposed estimator is simpler to implement than its Conditional Monte Carlo counterpart (Section 2.3.1). Note that the source code used in this paper is available online [125].

Throughout the paper, we use lowercase boldface letters like $\boldsymbol{c}$, $\boldsymbol{x}$, $\boldsymbol{y}$ for non-random vectors and uppercase boldface letters like $\boldsymbol{X}$ for random vectors, and $\mathbf{1}$ for the vector of 1's. If $\boldsymbol{X}$ is of length n, we write: $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$. The inner-product is denoted $\boldsymbol{x} \cdot \boldsymbol{y}$. For a differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$, we write

$$\nabla f(\boldsymbol{z}) = \left(\partial f(\boldsymbol{x})/\partial x_1, \ldots, \partial f(\boldsymbol{x})/\partial x_n\right)^\top \Big|_{\boldsymbol{x}=\boldsymbol{z}},$$

and use $\nabla_i f(\boldsymbol{z})$ to denote the i'th component of $\nabla f(\boldsymbol{z})$.

2.2 Proposed Push–Out Estimator

Our method is derived from the "Push-Out" method [122, 90] in sensitivity analysis of Discrete Event Systems, where a judiciously chosen change of variable allows differentiation of an otherwise non- smooth function. We thus tackle the pdf estimation problem by viewing it as a special type of sensitivity analysis. We note that a similar insight was used in Example 5.7 of [15], but that estimator is strictly restricted to iid sums of positive random variables, with the further requirement that the pdf $f_X(0) > 0$. This positivity

condition is quite restrictive — it excludes cases such as Pareto rvs or Weibull rvs with shape parameter greater than one. As we shall see, none of these restrictions apply to our estimator.

Assumption 2.2.1. The random vector $\boldsymbol{X}$ has a density $f_{\boldsymbol{X}}$ (each X_i is supported either on the entire real line or a half-real line), the gradient $\nabla f_{\boldsymbol{X}}$ is a continuous function on the support of $\boldsymbol{X}$, and we have the integrability condition $\mathbb{E}\,|\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X})| < \infty$ (here $\boldsymbol{X} \sim F_{\boldsymbol{X}}$). $\diamond$

The proposed estimator is based on the following simple formulas, proved in the appendix.

Proposition 2.2.2. *For the rv $S = \sum_{i=1}^n X_i = \mathbf{1} \cdot \boldsymbol{X}$ where $\boldsymbol{X}$ satisfies Assumption 2.2.1,*

$$f_S(s) = \quad \frac{1}{s}\,\mathbb{E}\left\{\mathbb{I}_{\{\mathbf{1}\cdot\boldsymbol{X}\le s\}}[\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X}) + n]\right\} \tag{2.1}$$

for any $s \neq 0$. $\diamond$

Corollary 2.2.3. *For the rv $S = \sum_{i=1}^n c_i X_i = \boldsymbol{c} \cdot \boldsymbol{X}$ where $\boldsymbol{X}$ satisfies Assumption 2.2.1,*

$$f_S(s) = \quad \frac{1}{s}\,\mathbb{E}\left\{\mathbb{I}_{\{\boldsymbol{c}\cdot\boldsymbol{X}\le s\}}[\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X}) + n]\right\} \tag{2.2}$$

for any $s \neq 0$, and where each $c_i \neq 0$. $\diamond$

More generally, for rvs of the form $S = \sum_{i=1}^n h_i(X_i)$, where each h_i is invertible on its support, and the transformed random variables $h(X_i)$ each obey Assumption 2.2.1, we have that

$$f(s) = \quad \tfrac{1}{s}\,\mathbb{E}\left\{\mathbb{I}_{\{S\le s\}}[\tfrac{\boldsymbol{h}(\boldsymbol{X})}{\boldsymbol{h}'(\boldsymbol{X})} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X}) + \mathbf{1} \cdot \nabla \tfrac{\boldsymbol{h}(\boldsymbol{X})}{\boldsymbol{h}'(\boldsymbol{X})}]\right\}. \tag{2.3}$$

Since the more general case (2.3) is, after some rearrangement, equivalent to the simpler one (2.1), we henceforth only consider sums of the form $S = \sum_{i=1}^n X_i$.

It is straightforward to show that (2.1), (2.2), (2.3) still hold if the indicators $\mathbb{I}_{\{\cdot\}}$ are replaced by $-(1 - \mathbb{I}_{\{\cdot\}})$. This suggests the pair of (unbiased) estimators ($\boldsymbol{X} \sim F_{\boldsymbol{X}}$):

$$\begin{aligned}
\widehat{f}_1(s) &= \quad \frac{1}{s}\mathbb{I}_{\{\mathbf{1}\cdot\boldsymbol{X}\le s\}}\Big[\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}\Big(\boldsymbol{X}\Big) + n\Big] \\
\widehat{f}_2(s) &= -\frac{1}{s}\mathbb{I}_{\{\mathbf{1}\cdot\boldsymbol{X}> s\}}\Big[\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}\Big(\boldsymbol{X}\Big) + n\Big]
\end{aligned}$$

A weighted combination of the latter two yields our proposed "push-out" estimator:

$$\widehat{f}(s; p) = p\widehat{f}_1(s) + (1 - p)\widehat{f}_2(s), \tag{2.4}$$

where $p \in \mathbb{R}$ is chosen to (approximately) minimize the overall variance, as follows.

Let Σ be the sample covariance matrix of $(\widehat{f}_1, \widehat{f}_2)$ and note that

$$\widehat{\mathbb{V}\mathrm{ar}}(p) = (\Sigma_{11} + \Sigma_{22} - 2\Sigma_{12})\, p^2 + 2(\Sigma_{12} - \Sigma_{22})\, p + \Sigma_{22}$$

is an unbiased estimator of the variance of (2.4). Then, we chose p to be

$$p^\star = \arg\min_p \widehat{\mathbb{V}\mathrm{ar}}(p) = (\Sigma_{22} - \Sigma_{12})/(\Sigma_{11} + \Sigma_{22} - 2\Sigma_{12}). \tag{2.5}$$

In order to ensure the unbiasedness of (2.4), we may, for example, estimate $p^\star$ from a pilot (independent) sample, as explained in Section 2.4.

We note that a downside of our push-out estimator is that the integrability condition in Assumption 2.2.1 is difficult to verify in many practical settings. As a matter of fact, in our simulations in Section 2.4 we implicitly assume without verifying the stronger condition $\mathbb{E}\,|\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X})|^4 < \infty$, which not only ensures the finite variance of the push-out estimator (2.4), but also the reliability of its sample (empirical) variance estimator (the variance of the sample variance has to be finite).

2.3 Competitor Methods

In the following Sections 2.3.1 and 2.3.2 we describe our main competitors — the Conditional Monte Carlo and Asmussen–Kroese estimators [12]. We then use these methods as benchmarks to illustrate the performance of the proposed estimator in various settings.

2.3.1 Conditional Monte Carlo estimator

The Conditional Monte Carlo estimator [12] takes the form

$$\widehat{f}_{\mathrm{Cond}}(s) = \frac{1}{n} \sum_{i=1}^{n} f_{X_i|\boldsymbol{X}_{-i}}(s - S_{-i}), \quad \boldsymbol{X} \sim F_{\boldsymbol{X}},$$

where the notation $\boldsymbol{X}_{-i}$ denotes the vector $\boldsymbol{X}$ with the i-th component removed and $S_{-i} = \mathbf{1} \cdot \boldsymbol{X}_{-i}$. This is particularly simple for the independent case, as $f_{X_i|\boldsymbol{X}_{-i}} = f_{X_i}$.

We now examine the dependent case where $\boldsymbol{X}$'s dependence structure is given by an Archimedean copula with generator ψ; i.e., the cdf yields

$$\mathbb{P}(X_1 \le F_{X_1}^{-1}(u_1), \ldots, X_n \le F_{X_n}^{-1}(u_n)) = \varphi\Big(\textstyle\sum_{i=1}^n \psi(u_i)\Big), \quad \boldsymbol{u} \in [0,1]^n,$$

where $\varphi \equiv \psi^{-1}$ is the functional inverse of ψ. The conditional densities of $\boldsymbol{X}$ can be calculated from the formula ($\varphi^{(n)}$ denotes n-th derivative)

$$f_{X_i|\boldsymbol{X}_{-i}}(x_i|\boldsymbol{x}_{-i}) = f_{X_i}(x_i)\psi^{(1)}(F_{X_i}(x_i))\frac{\varphi^{(n)}(\sum_{j=1}^{n}\psi(F_{X_j}(x_j)))}{\varphi^{(n-1)}(\sum_{j\neq i}\psi(F_{X_j}(x_j)))}\,. \tag{2.6}$$

Some Archimedean copulas, such as the Clayton and Gumbel–Hougaard copulas, have what is called a Marshall–Olkin representation. An Archimedean copula is in the Marshall–Olkin representation class if $\varphi(s) = \mathbb{E}[\mathrm{e}^{-sZ}]$ for some positive rv Z with cdf F_Z. Then an $\boldsymbol{X}$ with this dependence structure can be simulated via

$$\boldsymbol{X} = \Big(F_{X_1}^{-1}\Big(\varphi\Big(\frac{E_1}{Z}\Big)\Big), \ldots, F_{X_n}^{-1}\Big(\varphi\Big(\frac{E_n}{Z}\Big)\Big)\Big), \quad E_i \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(1),\ Z \sim F_Z\,. \tag{2.7}$$

For this case, Asmussen [12, Proposition 8.3] conditions upon the Z as well as $\boldsymbol{X}_{-i}$ to obtain what we call the *extended Conditional Monte Carlo estimator*

$$\widehat{f}_{\mathrm{ExtCond}}(s) = \frac{1}{n}\sum_{i=1}^{n} f_{X_i|\boldsymbol{X}_{-i},Z}(s - S_{-i}), \tag{2.8}$$

where $f_{X_i|\boldsymbol{X}_{-i},Z}(x_i) = -z\psi'(F_i(x_i))f_{X_i}(x_i)\,\mathrm{e}^{-z\psi(F_i(x_i))}$ and $\boldsymbol{X}$ is given by (2.7).

We will use this estimator as a benchmark in our comparisons later on.

2.3.2 Asmussen–Kroese estimator

The Asmussen–Kroese estimator [20] (typically for tail probabilities) is defined as

$$\widehat{F}_{\mathrm{AK}}(s) = 1 - \sum_{i=1}^{n} \overline{F}_{X_i|\boldsymbol{X}_{-i}}(\max\{M_{-i}, s - S_{-i}\})$$

where: $M_{-i} = \max\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$ and $\overline{F}_{X_i|\boldsymbol{X}_{-i}}(x) = 1 - F_{X_i|\boldsymbol{X}_{-i}}(x)$. Each $\overline{F}_{X_i|\boldsymbol{X}_{-i}}(\max\{M_{-i}, s - S_{-i}\}) = \overline{F}_{X_i|\boldsymbol{X}_{-i}}(s - S_{-i})$, whenever $M_{-i} + S_{-i} < s$. Thus, we can take the derivative of this piecewise estimator to obtain

$$\widehat{f}_{\mathrm{AK}}(s) = \sum_{i=1}^{n} f_{X_i|\boldsymbol{X}_{-i}}(s - S_{-i})\mathbb{I}_{\{M_{-i}+S_{-i}\leq s\}},$$

which can be viewed as alternative conditional estimator. When it is applicable, we use the "extended" form of this estimator where $f_{X_i|\boldsymbol{X}_{-i}}$ is replaced with $f_{X_i|\boldsymbol{X}_{-i},Z}$ as in Section 2.3.1. Notice that the term $1/n$ in (2.8) does not appear here.

2.4 Numerical Experiments and Extensions to Marginal Distributions

In Section 2.4.1, for various distributions of $\boldsymbol{X}$ we compare: i) our proposed method, ii) the conditional MC estimator, iii) the Asmussen–Kroese (AK) estimator, and iv) the default kernel-density estimator (KDE) in MATHEMATICA. Following this, Section 2.4.2 extends the estimator to the case of marginal density estimation in the context of Bayesian statistics.

2.4.1 Copula examples

When the Marshall–Olkin representation (2.7) is available, we simulate $\boldsymbol{X}$ using this form and give results for the extended version (2.8) of the conditional MC estimator. If this representation is unavailable, we use the standard version (2.6) of the conditional MC estimator where $\boldsymbol{X}$ is sampled using MATHEMATICA's built-in routines (to replicate this in another language one could simulate $\boldsymbol{X}$ using the "conditional distribution method" [109, p. 41]). KDE is provided by MATHEMATICA's KernelMixtureDistribution function with default bandwidth; to keep the support positive, we reflect this estimator about the origin. We report on all distributions and copulas as they are parametrized in MATHEMATICA.

We conduct 4 experiments, each one depicted on Figures 1 to 4 below. Each experiment uses $R = 10^5$ iid replicates of $\boldsymbol{X}$ which are common to all estimators (our estimator uses the first 5% of these to obtain the $p^\star$ coefficient, as in (2.5), and the remaining samples for pdf estimation as in (2.4)).

Our primary measure of performance is (square root of) the *work-normalized relative variance*:

$$\mathrm{WNRV}(\widehat{f}(x)) = (\mathrm{CPU\_Time}) \times \mathbb{V}\mathrm{ar}(\widehat{f}(x))/(R[\widehat{f}(x)]^2)$$

For each experiment we also display a subplot of the estimated density function, as well as the estimated standard deviation.

Figures 1–4 show that our proposed estimator consistently has the smallest WNRV. In Figures 3 and 4, it also has the smallest standard deviation. In Figure 1, the simple case for the sum of iid gamma rvs, the standard deviation of our estimator is similar to the Conditional MC method. In Figure 2, the sum of non-identical lognormal variables under a Frank copula, the AK estimator has the smallest standard deviation (it was designed for such subexponential distributions). In general, the Conditional MC method does not perform well in the case of heavy-tailed summands, as in Figures 2 and 3.

Figure 2.1: Sum of $n = 40$ iid $\mathsf{Gamma}(3, 2)$ random variables.

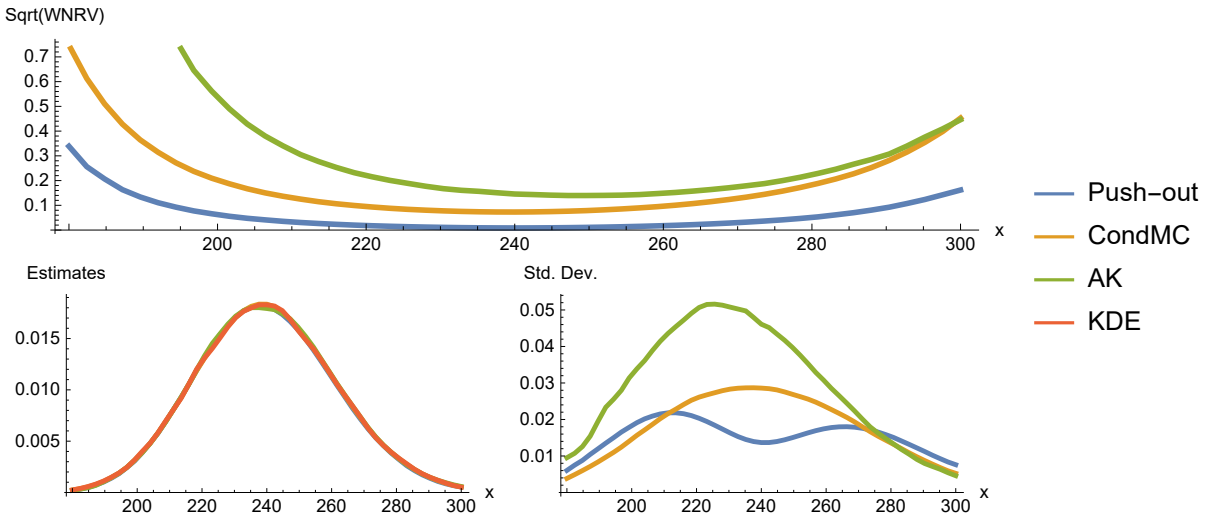


Figure 2.2: Sum of $n = 10$ random variables where $X_i \sim \mathsf{Lognormal}(i - 10, \sqrt{i})$ with a $\mathsf{Frank}(1/1000)$ copula. The choice of marginals mimic the challenging (and somewhat pathological) example considered in [14].

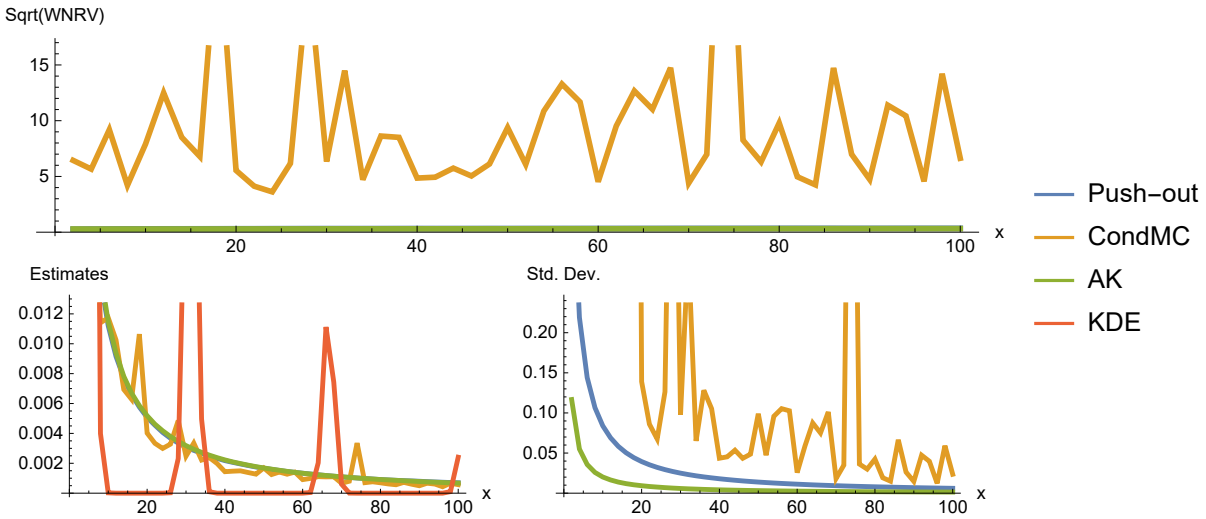


Figure 2.3: Sum of $n = 10$ Weibull$(0.3, 1)$ random variables with a Clayton$(1/5)$ copula.

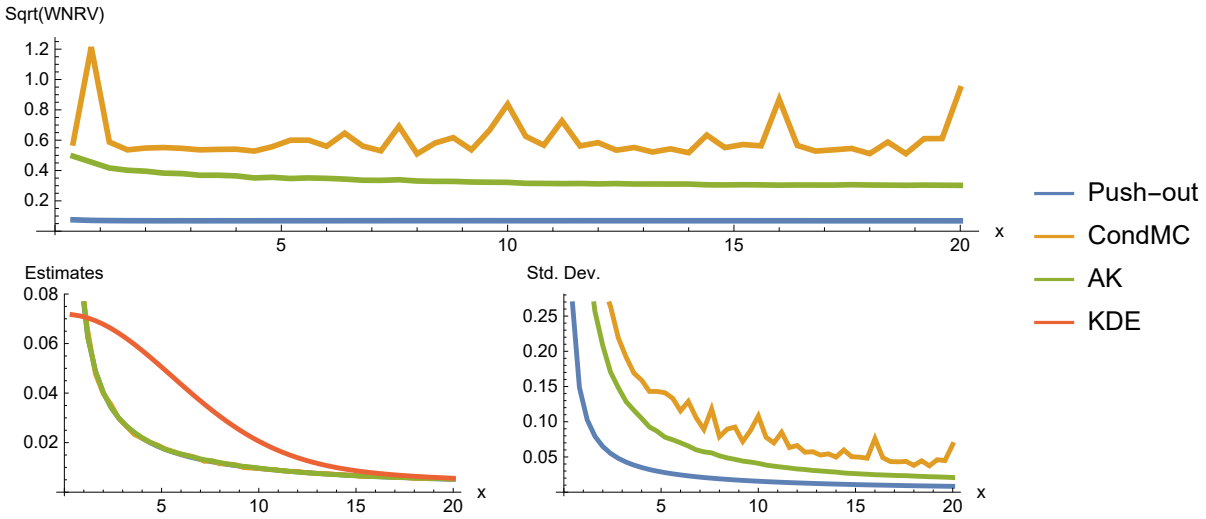
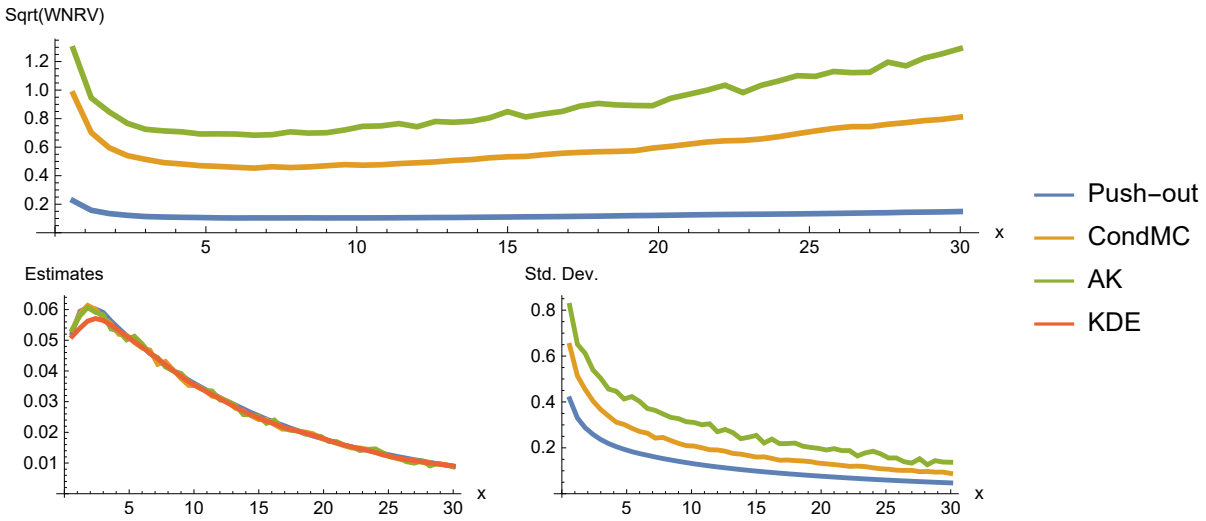


Figure 2.4: Sum of $n = 15$ Exp(1) random variables with a GumbelHougaard(5) copula.



It is important to note that due to the $1/s$ term, the proposed push-out estimator can have large variance for very small s, even when $F(s)$ or $1 - F(s)$ is not close to zero. This problem can be resolved with a simple linear shift, as follows. If one element, say X_1, is supported on $\mathbb{R}$, then $f_S(s) = f_{\widetilde{S}}(s-a)$ for $a \in \mathbb{R}$, where $\widetilde{S} = (X_1 + a) + X_2 + \cdots + X_n$. We

can then use the original estimator (with shifted values of s and X_1) to obtain estimates of the density of S near or at zero.

2.4.2 Estimating Marginal Distributions with Bayesian Applications

One extension of the estimator is in the estimation of marginal densities.

Proposition 2.4.3. *For an $\boldsymbol{X}$ which satisfies Assumption 2.2.1, the marginal densities are given by*

$$f_{X_i}(s) = \frac{1}{s}\,\mathbb{E}\left\{\mathbb{I}_{\{X_i \le s\}}\Big(X_i \nabla_i \log f_{\boldsymbol{X}}(\boldsymbol{X}) + 1\Big)\right\} \tag{2.9}$$

for $i = 1, \ldots, n$, and $s \neq 0$. ◇

We use the weighted estimator of the form (2.4) which is based on (2.9). A nice feature of the corresponding estimator is that, due to the presence of the $\nabla \log f_{\boldsymbol{X}}(\boldsymbol{x})$ term, the normalizing constant of f need not be known.

As an example, we use Markov Chain Monte Carlo to obtain samples from the posterior density of a Bayesian model, and use these to estimate the posterior marginal pdfs with our push-out estimator.

We consider the well-known "Pima Indians" dataset (standardized), which records a binary response variable (the incidence of diabetes) for 532 women, along with seven possible predictors. We specify a Logistic Regression model with predictors: *Number of Pregnancies*, *Plasma Glucose Concentration*, *Body Mass Index*, *Diabetes Pedigree Function*, and *Age* (see [63] for justification). The prior is $\boldsymbol{\beta} \sim \mathsf{N}(\mathbf{0}, \mathrm{I})$, as in [63].

To obtain samples from the posterior density, we implement an isotropic Random Walk sampler, using a radially symmetric Gaussian density with $\sigma^2 = 7.5 \times 10^{-3}$ (trace plots indicate this choice mixes well for the model). We note that by introducing auxiliary random variables, it is possible to simulate from this Bayesian posterior using a complicated Gibbs sampler [81, Equation 8], which requires simulation of costly (non-standard) *Kolmogorov-Smirnov*-distributed random variables, and truncated normal and truncated logistic random variables. In contrast, the Random Walk sampler is simpler to implement.

We ran the Random Walk sampler for 10^3 steps for burn-in, then used the next 5×10^4 samples (without any thinning) to obtain a KDE, as well as density estimates using our push-out estimator. As a benchmark, we compare the accuracy with a KDE constructed using every 50-th sample from an MCMC chain of length $50 \times 5 \times 10^6$.

The result of this comparison is depicted on Figure 2.5 below.

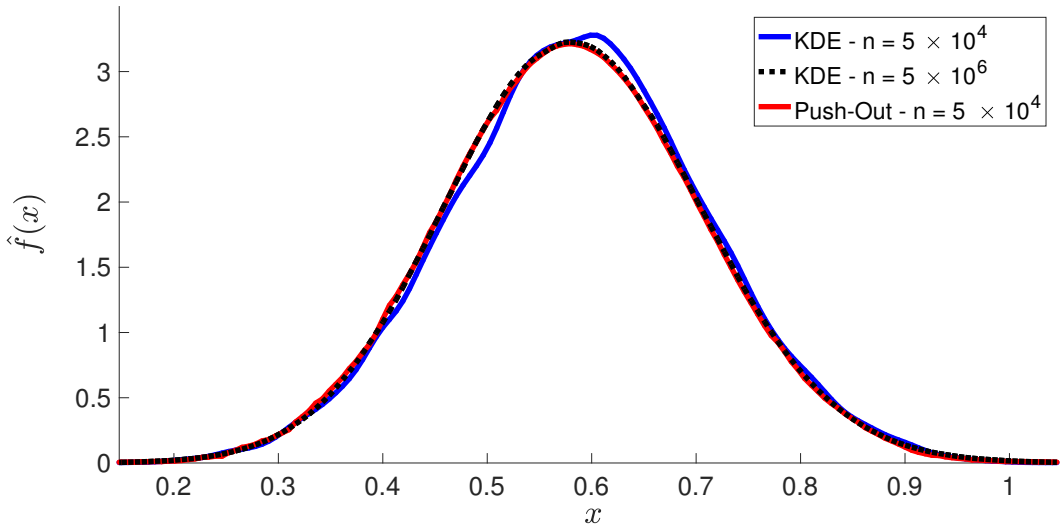


Figure 2.5: Density estimation of posterior marginal corresponding to the coefficient parameter of the *Body Mass Index* predictor variable.

As expected, using the same set of samples, the push-out estimator yields a more accurate estimate than KDE. The reason for the lower accuracy of KDE in this context is well-known — a mean square error convergence of $\mathcal{O}(n^{-4/5})$, instead of the canonical Monte Carlo rate of $\mathcal{O}(1/n)$, due to the presence of non-negligible bias in the KDE estimator (see [34], for example).

2.4.4 Discussion

In our experiments, the proposed push-out estimator is usually as accurate, or more accurate, than the Conditional MC and AK estimators. However, when computing time is taken into account, then our push-out estimator truly excels. For example, in the test with the Clayton copula on Figure 2.3, our proposed estimator took 3.03 minutes to compute, whereas the extended Conditional MC estimator took 9.16 minutes, and the Asmussen–Kroese estimator took 9.81 minutes (for source code see [125]). One reason for this lower computing cost is that evaluation of the gradient of the joint log-pdf, $\nabla \log f_{\boldsymbol{X}}$, is often much faster than evaluation of the conditional pdfs $f_{X_i|\boldsymbol{X}_{-i}}$.

A practical difficulty with the extended Conditional MC estimator (2.8) is the simulation of the auxiliary variable Z. For example, simulation is difficult and time-consuming for the $\mathsf{GumbelHougaard}(\theta)$ copula with generator $\psi(u) = [-\log(u)]^\theta$, because in that case $Z \sim \mathsf{Stable}(1, \theta^{-1}, 1, 0, \cos(\frac{\pi}{2\theta})^\theta)$ is a complicated Lévy α-stable distribution. (On the other

hand, simulation is simple for the Clayton(θ) copula with generator $\psi(u) = \theta(u^{-1/\theta} - 1)$, because $Z \sim$ Gamma$(\theta, 1/\theta)$.) It is also worth noting that many Archimedean copulas, e.g. Frank copula, do not have a Marshall–Olkin representation.

We also remark that one must be cautious with the reported estimate of estimator variance when using Conditional Monte Carlo. In heavy-tailed settings, the estimator has large variance, as illustrated using a sum of 30 iid Weibull rvs on Figure 2.6.

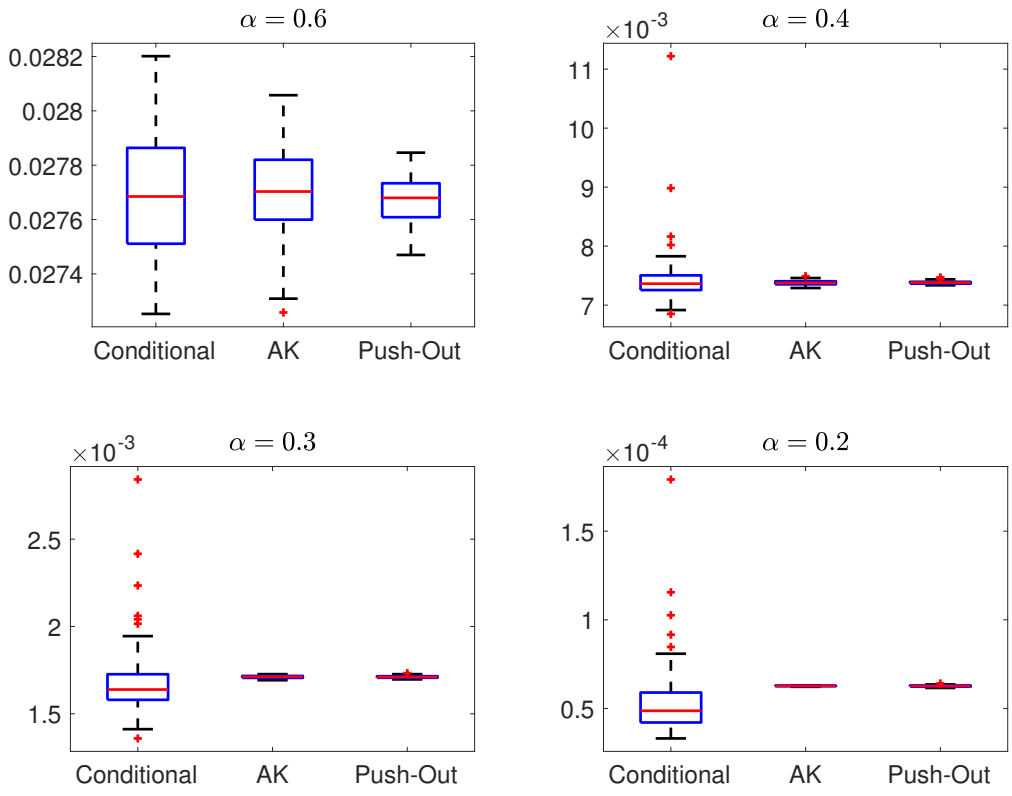


Figure 2.6: Results from 100 runs, each with $R = 10^5$ samples, of Conditional Monte Carlo, Asmussen Kroese, and our proposed method, for 30 iid Weibull$(\alpha, 1)$ rv's for varying shape parameter α. We estimate the pdf at $s = \mathbb{E}[S] = 30\,\Gamma(1 + (1/\alpha))$ in each case.

2.5 Conclusion

We have introduced a novel "push-out" Monte Carlo estimator for the unbiased estimation of the pdf of sums of random variables, and given several examples of its implementation.

The numerical experiments suggest that in terms of WNRV the proposed push-out estimator is preferable to the existing competitors (Conditional Monte Carlo, Asmussen-Kroese, and kernel density estimators). The main reason for this is that typically the evaluation of the gradient of the joint pdf $f_{\boldsymbol{X}}$ is faster to compute than the conditional pdfs of $f_{\boldsymbol{X}}$. On the other hand, a shortcoming of our proposed estimator is that verifying the theoretical finiteness of its variance is difficult. In particular, in our numerical experiments we have implicitly assumed the moment condition $\mathbb{E}\left|\boldsymbol{X}\cdot\nabla\log f_{\boldsymbol{X}}(\boldsymbol{X})\right|^4<\infty$.

2.A Proofs

Proposition 2.2.2. Define the cdf $F_S(s)=\int_{\mathbf{1}\cdot\boldsymbol{x}\le s} f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$, so that the pdf is $f_S(s)=\frac{\mathrm{d}}{\mathrm{d}s}F_S(s)$. The change of variables $\boldsymbol{x}=s\boldsymbol{y}$ yields:

$$F_S(s)=\int_{\mathcal{R}_s} f_{\boldsymbol{X}}(s\boldsymbol{y})|s|^n\mathrm{d}\boldsymbol{y}\quad s\neq 0,$$

where the notation $\int_{\mathcal{R}_s}$ means $\int_{\mathbf{1}\cdot\boldsymbol{y}\le 1}$ if $s>0$, else $\int_{\mathbf{1}\cdot\boldsymbol{y}>1}$ for $s<0$.

Let $\varphi(s):=\int_{\mathcal{R}_s}\frac{\mathrm{d}}{\mathrm{d}s}\left(f_{\boldsymbol{X}}(s\boldsymbol{y})|s|^n\right)\mathrm{d}\boldsymbol{y}$. We will use the fact that $\varphi(s)=f_S(s)$ almost everywhere (i.e. except possibly on sets of zero Lebesgue measure) on $s\notin(-\varepsilon,\varepsilon)$ for an arbitrarily small $\varepsilon>0$.

In order to justify the identity $\varphi(s)=f_S(s)$ (almost everywhere) in the case of $s>\varepsilon$ (similar arguments apply for $s<\varepsilon$), we use the Fubini-Tonelli theorem for exchanging the order of integration. This exchange holds under the integrability condition

$$\int_\varepsilon^s\int_{\mathbf{1}\cdot\boldsymbol{y}\le 1}\left|\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{\boldsymbol{X}}(t\boldsymbol{y})t^n\right)\right|\mathrm{d}\boldsymbol{y}\mathrm{d}t<\infty \tag{2.10}$$

and the existence of a continuous $\nabla f_{\boldsymbol{X}}$, both of which follow from Assumption 2.2.1 (verified at the end of this proof). Using the Fubini-Tonelli theorem [121] we then write:

$$\begin{aligned}\int_\varepsilon^s\varphi(t)\mathrm{d}t&=\int_\varepsilon^s\int_{\mathbf{1}\cdot\boldsymbol{y}\le 1}\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{\boldsymbol{X}}(t\boldsymbol{y})t^n\right)\mathrm{d}\boldsymbol{y}\mathrm{d}t\\&=\int_{\mathbf{1}\cdot\boldsymbol{y}\le 1}\int_\varepsilon^s\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{\boldsymbol{X}}(t\boldsymbol{y})t^n\right)\mathrm{d}t\mathrm{d}\boldsymbol{y}\\&=\int_{\mathbf{1}\cdot\boldsymbol{y}\le 1}(f_{\boldsymbol{X}}(s\boldsymbol{y})s^n-f_{\boldsymbol{X}}(\varepsilon\boldsymbol{y})\varepsilon^n)\mathrm{d}\boldsymbol{y}=F_S(s)-F_S(\varepsilon)\end{aligned}$$

Hence, by the fundamental theorem of Calculus, φ equals the derivative of F_S up to a set of measure zero. In other words, $\varphi(s)=f_S(s), s>\varepsilon$ almost everywhere.

To proceed, we write $\mathrm{sign}(x) = x/|x| = \frac{\mathrm{d}}{\mathrm{d}x}|x|$

$$\begin{aligned} f_S(s) = \varphi(s) &= \int_{\mathcal{R}_s} |s|^n \boldsymbol{y} \cdot \nabla f_{\boldsymbol{X}}(s\boldsymbol{y}) + n|s|^{n-1}\mathrm{sign}(s) f_{\boldsymbol{X}}(s\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\ &= \textstyle\int_{\mathcal{R}_s} \left[\boldsymbol{y} \cdot \nabla \log f_{\boldsymbol{X}}(s\boldsymbol{y}) + \frac{n\,\mathrm{sign}(s)}{|s|}\right] |s|^n f_{\boldsymbol{X}}(s\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \end{aligned}$$

so after a change of variables $\boldsymbol{y} = \boldsymbol{x}/s$ and using $\mathrm{sign}(x)/|x| = 1/x$, we obtain

$$f_S(s) = \int_{\mathbf{1}\cdot\boldsymbol{x}\le s} \left[\frac{\boldsymbol{x}}{s} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{x}) + \frac{n}{s}\right] f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \frac{1}{s}\,\mathbb{E}\left\{\mathbb{I}_{\{\mathbf{1}\cdot\boldsymbol{X}\le s\}}[\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X}) + n]\right\}.$$

To verify (2.10), note that after using the change of variable above, it can be upper bounded by

$$\textstyle\int_\varepsilon^s \frac{1}{t}\,\mathbb{E}\left\{\mathbb{I}_{\{\mathbf{1}\cdot\boldsymbol{X}\le t\}}|\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X}) + n|\right\}\mathrm{d}t \le (\mathbb{E}\left|\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X})\right| + n)\int_\varepsilon^s \frac{1}{t}\mathrm{d}t < \infty,$$

which is bounded by assumption. □

Corollary 2.2.3. Consider $\widetilde{\boldsymbol{X}} := \boldsymbol{c}\boldsymbol{X}$, with density $f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = f_{\boldsymbol{X}}(\boldsymbol{c}^{-1}\widetilde{\boldsymbol{x}})/|\prod_{i=1}^n c_i|$. As $\boldsymbol{X}$ satisfies Assumption 2.2.1 then so does $\widetilde{\boldsymbol{X}}$, so we apply Proposition 2.2.2 to $\widetilde{\boldsymbol{X}}$ to get

$$f_S(s) = \frac{1}{s}\,\mathbb{E}_{f_{\widetilde{\boldsymbol{X}}}}\left\{\mathbb{I}_{\{\mathbf{1}\cdot\widetilde{\boldsymbol{X}}\le s\}}[\widetilde{\boldsymbol{X}} \cdot \nabla \log f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{X}}) + n]\right\} = \frac{1}{s}\,\mathbb{E}_{f_{\boldsymbol{X}}}\left\{\mathbb{I}_{\{\boldsymbol{c}\cdot\boldsymbol{X}\le s\}}[\boldsymbol{X} \cdot \nabla \log f_{\boldsymbol{X}}(\boldsymbol{X}) + n]\right\},$$

where the second equality uses $\nabla \log f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = \boldsymbol{c}^{-1}\nabla \log f_{\boldsymbol{X}}(\boldsymbol{c}^{-1}\widetilde{\boldsymbol{x}})$. □

Proposition 2.4.3. Note that the marginal density of X_i can be written as

$$f_{X_i}(s) = \int_{x_{-i}\in\mathbb{R}^{n-1}} \frac{\mathrm{d}}{\mathrm{d}s}\int_{x_i\le s} f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}x_i\mathrm{d}\boldsymbol{x}_{-i}.$$

Applying the same push-out technique (now using the change of variables $x_1 = sy_1$) as in Proposition 2.2.2 for the inner integral, and then rearranging yields

$$f_{X_i}(s) = \frac{1}{s}\int_{\mathbb{R}^n} \mathbb{I}_{\{x_i\le s\}}(x_i\nabla_i \log f_{\boldsymbol{X}}(\boldsymbol{x}) + 1) f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \frac{1}{s}\,\mathbb{E}\left\{\mathbb{I}_{\{X_i\le s\}}[\Big(X_i\nabla_i \log f_{\boldsymbol{X}}(\boldsymbol{X}) + 1\Big)]\right\}.$$

□

Chapter 3

Orthonormal polynomial expansions and lognormal sum densities

3.1 Introduction

The lognormal distribution arises in a wide variety of disciplines such as engineering, economics, insurance, finance, and across the sciences [8, 48, 87, 99, 53]. Therefore, it is natural that sums S of n lognormals come up in a number of contexts. A basic example in finance is the Black–Scholes model, which assumes that security prices are lognormals, and hence the value of a (long-only) portfolio with n securities has the form of a SLN (sum of lognormals). In insurance, individual claim sizes are often taken as independent lognormals, so the total claim amount in a certain period is again of the SLN form [133]. A further example occurs in telecommunications, where the inverse of the signal-to-noise ratio (a measure of performance in wireless systems) can be modeled as a sum of i.i.d. lognormals [69].

The SLN distribution function is, however, not available in explicit form, and evaluating it numerically or approximating it is considered to be a challenging problem with a long history. The classical approach is to use an approximation with another lognormal distribution. This goes back at least to [61] and it is nowadays known as the *Fenton–Wilkinson method* as according to [102] this approximation was already used by Wilkinson in 1934. However, it can be rather inaccurate when the number of summands is rather small, or when the dispersion parameter is too high. Also tail approximations have been extensively discussed, with the right tail being a classical example in subexponential theory [59], and the study of the left tail being more recent [19, 70].

This paper discusses a different method where one approximates the SLN probability den-

sity function (p.d.f.) f using polynomials $\{Q_k\}_{k\in\mathbb{N}_0}$ which are orthonormal w.r.t. some reference measure ν. In the general formulation, one is interested in approximating a target density g using the density f_ν of ν as reference. One then finds a series representation of g/f_ν of the form $\sum_{k=0}^{\infty} a_k Q_k$, and then the approximation of g is

$$\widehat{g}(x) \;=\; f_\nu(x) \sum_{k=0}^{K} a_k Q_k(x), \tag{3.1}$$

for some suitable K. The most obvious connection to the lognormal sum problem is $g = f$, but for some choices of ν we must take a different target g. In one case we set g as the density of $\log S$ and transform back to get the approximation $\widehat{f}(x) = \widehat{g}(\log x)/x$. In another case we set g as the exponentially tilted SLN p.d.f. The choice of ν is a crucial step, and three candidates for ν are investigated: the normal, the gamma, and the lognormal distributions.

The form of the Q_k is classical for the normal distribution where it is the Hermite polynomials and for the gamma where it is the generalised Laguerre polynomials, but for the lognormal distributions it does not appear to be in the literature and we give here the functional expression (Theorem 3.2.6). The Fenton–Wilkinson method may be seen as the $K = 2$ case of f_ν being lognormal (with $g = f$), and this choice of f_ν may be the most obvious one. However, we show that in the lognormal case the orthonormal polynomials are not dense in $\mathcal{L}^2(f_\nu)$. This result is closely related to the lognormal distribution not being determined by its moments [80, 30] and indicates that a lognormal f_ν is potentially dangerous. For this reason, the rest of the paper concentrates on taking the reference distribution as normal (using the logarithmic transformation) or gamma (using exponential tilting).

After discussing the details of the orthonormal polynomials expansions in Sections 3.2 and 3.3, we proceed in Section 3.4 to show a number of numerical examples. The polynomial expansions are compared to existing methods as Fenton–Wilkinson and a more recent approximation in terms of log skew normal distributions [78], as well as to exact values obtained by numerical quadrature in cases where this is possible or by Monte Carlo density estimation. Section 3.4 also outlines an extension to statistical data sets and non-Gaussian copulas. Appendix A contains a technical proof and Appendix B some new material on the SLN Laplace transform.

3.2 Orthogonal polynomial representation of probability density functions

Let X be a random variable which has a density f with respect to some measure $\lambda \geq 0$ (typically Lebesgue measure on an interval or counting measure on a subset of $\mathbb{Z}$). If f is unknown but the distribution of X is expected to be close to some probability measure ν with p.d.f. f_ν, one may use f_ν as a first approximation to f and next try to improve by invoking suitable correction terms.

In the setting of this paper X is the sum of lognormal r.v.s and the correction terms are obtained by expansions in terms of orthonormal polynomials. Before going into the details of the lognormal example, let us consider the general case.

Assuming all moments of ν to be finite, the standard Gram–Schmidt orthogonalization technique shows the existence of a set of polynomials $\{Q_k\}_{k\in\mathbb{N}_0}$ which are orthonormal in $\mathcal{L}^2(\nu)$ equipped with the usual inner product $\langle g, h\rangle = \int gh \,\mathrm{d}\nu$ and the corresponding norm $\|g\|^2 = \langle g, g\rangle$. That is, the Q_k satisfy

$$\left\langle Q_i, Q_j \right\rangle = \int Q_i(x)Q_j(x)\,\mathrm{d}\nu(x) = \delta_{ij}\,, \quad i,j \in \mathbb{N}_0, \tag{3.2}$$

where δ_{ij} denotes the Kronecker symbol. If there exists an $\alpha > 0$ such that

$$\int \mathrm{e}^{\alpha|x|}\,\mathrm{d}\nu(x) < \infty\,, \tag{3.3}$$

the set $\{Q_k\}_{k\in\mathbb{N}_0}$ is complete in $\mathcal{L}^2(\nu)$, cf. Chapter 7 of the book by Szökefalvi-Nagy [132]. The implication is that if f/f_ν is in $\mathcal{L}^2(\nu)$, that is, if

$$\int \frac{f(x)^2}{f_\nu(x)^2}\,\mathrm{d}\nu(x) \;=\; \int \frac{f(x)^2}{f_\nu(x)}\,\mathrm{d}\lambda(x) \;<\; \infty\,, \tag{3.4}$$

we may expand f/f_ν as $\sum_{k=0}^{\infty} a_k Q_k$ where

$$a_k \;= \langle f/f_\nu, Q_k\rangle \;=\; \int fQ_k \,\mathrm{d}\lambda \;=\; \mathbb{E}\left[Q_k(X)\right]\,. \tag{3.5}$$

This suggests that we use the form of (3.1) as an approximation of f in situations where the p.d.f. of X is unknown but the moments are accessible.

Remark 3.2.1. If the first m moments of X and ν coincide, one has $a_k = 0$ for $k = 1, \ldots, m$. When choosing ν, a possible guideline is therefore to match as many moments as possible. ◇

Due to the Parseval relationship $\sum_{k=0}^{\infty} a_k^2 = \|f/f_\nu\|^2$, the coefficients of the polynomial expansion, $\{a_k\}_{k\in\mathbb{N}_0}$, tend toward 0 as $k \to \infty$. The accuracy of the approximation (3.1), for a given order of truncation K, depends upon how swiftly the coefficients decay; note that the $\mathcal{L}^2(\nu)$ loss of the approximation of f/f_ν is $\sum_{K+1}^{\infty} a_k^2$. Note also that the orthogonal polynomials can be specified recursively (see Thm. 3.2.1 of [131]) which allows a reduction of the computing time required for the coefficients' evaluation and makes it feasible to consider rather large K.

3.2.2 Normal reference distribution

A common choice as a reference distribution is the normal $\mathcal{N}(\mu, \sigma^2)$. The associated orthonormal polynomial are given by

$$Q_k(x) = \frac{1}{k! 2^{k/2}} H_k\left(\frac{x-\mu}{\sigma\sqrt{2}}\right), \tag{3.6}$$

where $\{H_k\}_{k\in\mathbb{N}_0}$ are the (physicists') Hermite polynomials, defined in [131] for instance. If f is continuous, a sufficient (and close to necessary) condition for $f/f_\nu \in \mathcal{L}^2(\nu)$ is

$$f(x) \;=\; \mathcal{O}\Big(\mathrm{e}^{-ax^2}\Big) \quad \text{as } x \to \pm\infty \quad \text{with } a > \Big(4\sigma^2\Big)^{-1}. \tag{3.7}$$

Indeed, we can write the integral in (3.4) as $I_1 + I_2 + I_3$, the integrals over $(-\infty, -A)$, $[-A, A]$, resp. (A, ∞). Note that $I_2 < \infty$ follows since the integrand is finite by continuity, whereas the finiteness of I_1, I_3 is ensured by the integrands being $\mathcal{O}(\mathrm{e}^{-bx^2})$ where $b = 2a - 1/2\sigma^2 > 0$. Similar arguments apply to conditions (3.10) and (3.13) below.

Remark 3.2.3. The expansion formed by a standard normal baseline distribution and Hermite polynomials is known in the literature as Gram–Charlier expansion of type A, and the application to a standardised sum is the Edgeworth expansion, cf. [47], [26]. $\diamond$

3.2.4 Gamma reference distribution

If X has support $(0, \infty)$, it is natural to look for a ν with the same property. An obvious candidate is the gamma distribution, denoted $\mathsf{Gamma}(r, m)$ where r is the shape parameter and m the scale parameter. The p.d.f. is

$$f_\nu(x) = \frac{x^{r-1}\mathrm{e}^{-x/m}}{m^r \Gamma(r)}, \;\; x \in \mathbb{R}_+ . \tag{3.8}$$

The associated polynomials are given by

$$Q_n(x) = (-1)^n \left[\frac{\Gamma(n+r)}{\Gamma(n+1)\Gamma(r)}\right]^{-1/2} L_n^{r-1}(x/m), \;\; n \in \mathbb{N}_0, \tag{3.9}$$

where $\{L_n^{r-1}\}_{n\in\mathbb{N}_0}$ denote the *generalised Laguerre polynomials*, see [131]; in Mathematica these are accessible via the `LaguerreL` function. Similarly to (3.7), one has the following sufficient condition for $f/f_\nu \in \mathcal{L}^2(\nu)$:

$$\begin{aligned} f(x) &= \mathcal{O}\Big(\mathrm{e}^{-\delta x}\Big) \quad \text{as } x\to\infty \quad \text{with } \delta > 1/2m\,, \text{ and} \\ f(x) &= \mathcal{O}\Big(x^{\beta}\Big) \quad \text{as } x\to 0 \quad \text{with } \beta > r/2-1\,. \end{aligned} \tag{3.10}$$

3.2.5 Lognormal reference distribution

The lognormal distribution $\mathrm{LN}(\mu,\sigma^2)$ is the distribution of e^Y where $Y\sim\mathsf{Normal}(\mu,\sigma^2)$. It has support on $\mathbb{R}_+$. The polynomials orthogonal to the $\mathrm{LN}(\mu,\sigma^2)$ are given in the following proposition, to be proved in the Appendix:

Theorem 3.2.6. *The polynomials orthonormal with respect to the lognormal distribution are given by*

$$Q_k(x) = \frac{\mathrm{e}^{-\frac{k^2\sigma^2}{2}}}{\sqrt{\left[\mathrm{e}^{-\sigma^2},\mathrm{e}^{-\sigma^2}\right]_k}} \sum_{i=0}^{k} (-1)^{k+i} \mathrm{e}^{-i\mu-\frac{i^2\sigma^2}{2}} e_{k-i}\left(1,\ldots,\mathrm{e}^{(k-1)\sigma^2}\right) x^i, \tag{3.11}$$

for $k\in\mathbb{N}_0$ where

$$e_i\left(X_1,\ldots,X_k\right) = \begin{cases} 1 & \text{for } i=0, \\ \sum_{1\le j_1<\ldots<j_i\le k} X_{j_1}\ldots X_{j_i}, & \text{for } 1\le i\le k, \\ 0, & \text{for } i>k, \end{cases} \tag{3.12}$$

are the elementary symmetric polynomials and $[x,q]_n = \prod_{i=0}^{n-1}\left(1-xq^i\right)$ is the q-Pochhammer symbol.

Remark 3.2.7. The result of Theorem 3.2.6 does not appear to be in the literature; the closest reference seems to be a 1923 paper by Wigert [135] who considers the distribution with p.d.f. $\ell\mathrm{e}^{-\ell^2\ln^2(x)}/\sqrt{\pi}$, for $x>0$, introduced by Stieljes [130, pp. 507–508] (later called the Stieltjes–Wigert distribution). These polynomials are also mentioned in [42, pp. 172–175]. ◇

The equivalent of condition (3.7) for $f/f_\nu\in\mathcal{L}^2(\nu)$ now becomes

$$f(x) = \mathcal{O}\Big(\mathrm{e}^{-b\log^2 x}\Big) \quad \text{for } x\to 0 \text{ and } \infty \quad \text{with } b > \left(4\sigma^2\right)^{-1}, \tag{3.13}$$

which is rather mild. However, a key difficulty in taking the reference distribution as lognormal is the following result related to the fact that the lognormal and the Stieltjes–Wigert distributions are not characterised by their moments, see [80, 30, 41, 43]. Hence,

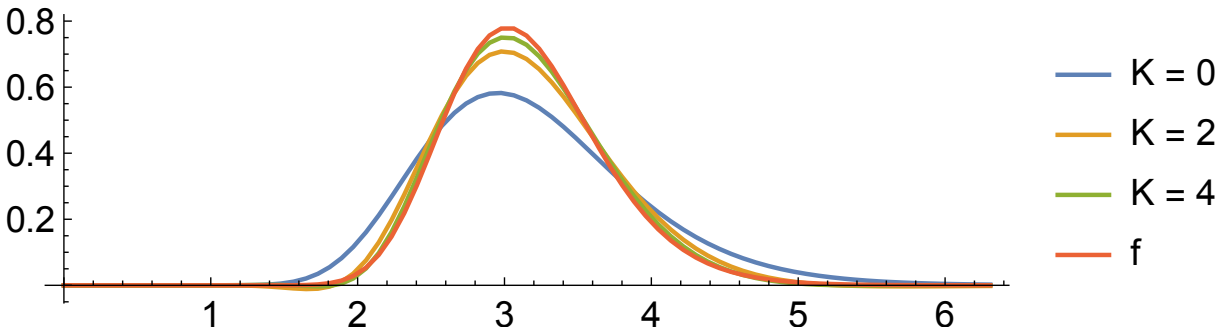


Figure 3.1: Examples of orthogonal polynomial approximations using a $\mathsf{Normal}(1.13, 0.23^2)$ reference converging to the target f with increasing K.

the orthogonal polynomials associated with the lognormal p.d.f. and the Stieltjes–Wigert p.d.f. are also the orthogonal polynomials for some other distribution.

Proposition 3.2.8. *The set of orthonormal polynomials in Theorem 3.2.6 is incomplete in* $\mathcal{L}^2(\nu)$. *That is,* $\operatorname{span}\{Q_k\}_{k\in\mathbb{N}_0}$ *is a proper subset of* $\mathcal{L}^2(\nu)$.

Proof. Let Y be a r.v. whose distribution is the given lognormal ν and X a r.v. with a distribution different from Y but with the same moments. According to [30, pp. 201–202] such an X can be chosen such that f_X/f_ν is bounded and hence in $\mathcal{L}^2(\nu)$. The projection of f_X/f_ν onto $\operatorname{span}\{Q_k\}$ is then

$$\begin{aligned}\sum_{k=0}^{\infty} \langle f_X/f_\nu, Q_k\rangle\, Q_k \;&=\; \sum_{k=0}^{\infty} \mathbb{E}\left[Q_k(X)\right] Q_k \;=\; \sum_{k=0}^{\infty} \mathbb{E}\left[Q_k(Y)Q_0(Y)\right] Q_k \\ &=\; Q_0 \;=\; 1 \neq f_X/f_\nu,\end{aligned}$$

where the first step used (3.5) and the second that the moments are the same (and that $Q_0(\cdot) = 1$). This implies $f_X/f_\nu \in \mathcal{L}^2(\nu) \setminus \operatorname{span}\{Q_k\}$ and the assertion. □

3.2.9 Convergence of the estimators w.r.t. K

Orthogonal polynomial approximations generally become more accurate as the order of the approximation K increases. Figure 3.1 shows a specific orthogonal polynomial approximation, $\widehat{f}_{\mathcal{N}}$ (which involves a logarithmic transformation and is described in Section 3.3.4), converging to the true SLN density f for increasing K. In this example, we take the SLN distribution with $\mu = (0, 0, 0)^\top$, $\Sigma_{ii} = 0.1$, and $\rho = -0.1$.

Proposition 3.2.8 implies that orthogonal polynomial approximations with a lognormal reference distribution cannot be relied upon to converge to the desired target density but

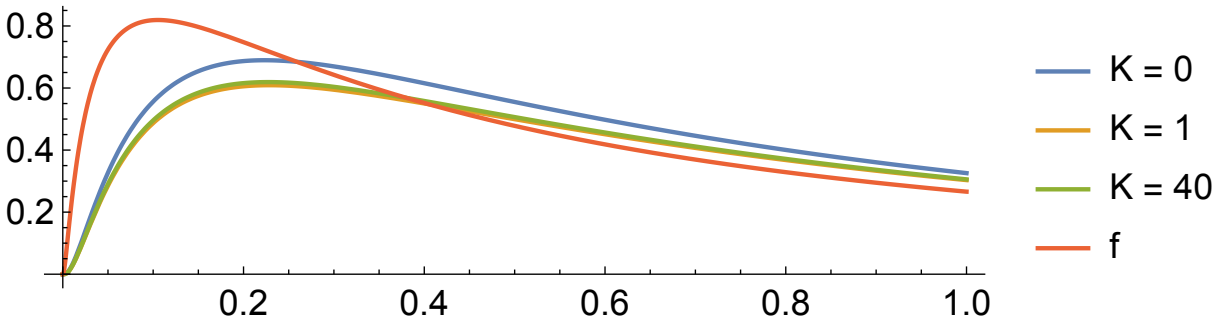


Figure 3.2: Example of orthogonal polynomial approximations of f using a $\mathrm{LN}(0, 1.22^2)$ reference not converging to the $\mathrm{LN}(0, 1.50^2)$ target.

may have a different limit (the orthogonal projection described there). The next plot, Figure 3.2, illustrates this phenomenon. The approximation appears to converge, but not to the target density. Our theoretical discussion suggests that this incorrect limit density has the same moments as the target lognormal distribution, and this was verified numerically for the first few moments.

Lastly, it must be noted that we cannot in practice take K arbitrarily large, due to numerical errors incurred in calculating the $\{a_k\}$ coefficients. Obviously this can be overcome by using infinite precision operations, however this swiftly becomes prohibitively slow. Software tools like `Mathematica` allow for arbitrarily large but finite precision, which gives on the flexibility to choose a desired accuracy/speed trade-off. We use this technology and select $K \leq 40$.

3.3 Application to lognormal sums

We now turn to our main case of interest where $X = S$ is a lognormal sum. Specifically,

$$S = \mathrm{e}^{X_1} + \ldots + \mathrm{e}^{X_n}\,, \quad n \geq 2\,, \tag{3.14}$$

where the vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is governed by a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ is the mean vector and $\boldsymbol{\Sigma} = \left(\sigma_{ij}\right)$ the covariance matrix. We write this distribution as $\mathrm{SLN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and hereafter denote its p.d.f. as f. We are interested in computing the p.d.f. when the summands exhibit dependency ($\boldsymbol{\Sigma}$ is non-diagonal). This is an ambitious goal given that the p.d.f. of the sum of two i.i.d lognormally distributed random variables is already unknown. The validity of the polynomial approximations rely on the $\mathcal{L}^2$ integrability condition (3.4), which is difficult to check because the p.d.f. of S is not available. We will need asymptotic results describing the left and the right tail of the distribution of S, which we collect in the following subsection.

3.3.1 Tail asymptotics of lognormal sums

The tail asymptotics of $f(x)$ are given in the following lemma, which simply collects the results from Corollary 2 of [70] and Theorem 1 of [22].

Lemma 3.3.2. *We have*

$$f(x) = \mathcal{O}(\exp\{-c_1 \ln(x)^2\}) \text{ as } x \to 0 \text{ and} \tag{3.15}$$
$$f(x) = \mathcal{O}(\exp\{-c_2 \ln(x)^2\}) \text{ as } x \to \infty \tag{3.16}$$

where

$$c_1 = \left[2 \min_{\boldsymbol{w} \in \Delta} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}\right]^{-1} \text{ and } \quad c_2 = \left[2 \max_{i=1,\ldots,n} \sigma_{ii}\right]^{-1},$$

with the notation that $\Delta = \{\boldsymbol{w} \,|\, w_i \in \mathbb{R}_+, \sum_{i=1}^n w_i = 1\}$.

We are also interested in the asymptotic behaviour of $Z = \ln(S)$ later in the paper. Writing the p.d.f. of Z as f_Z we have $f_Z(z) = \mathrm{e}^z f(\mathrm{e}^z)$. Together with L'Hôpital's rule this gives the following results (extending [64]):

Corollary 3.3.3. *We have*

$$f_Z(z) = \mathcal{O}(\exp\{-c_1 z^2\}) \text{ as } z \to -\infty \text{ and} \tag{3.17}$$
$$f_Z(z) = \mathcal{O}(\exp\{-c_2 z^2\}) \text{ as } z \to +\infty \tag{3.18}$$

where the constants are as in Lemma 3.3.2.

3.3.4 Lognormal sums via a normal reference distribution

Consider transforming S to $Z = \ln(S)$ and expanding this density with orthogonal polynomials using a normal distribution as reference. That is, our approximation to f using a $\mathsf{Normal}(\mu, \sigma^2)$ reference is

$$\widehat{f}_{\mathcal{N}}(x) = \frac{1}{x} \widehat{f}_Z(\ln x) \quad \text{where} \quad \widehat{f}_Z(z) = \varphi_{\mu,\sigma^2}(z) \sum_{i=1}^K a_i \, Q_i(z)\,,$$

with the normal p.d.f. $\varphi_{\mu,\sigma^2} = f_\nu$. The following result tells us when the integrability condition $f_Z/f_\nu \in \mathcal{L}^2(\nu)$ is satisfied. It follows immediately by combining (3.7) and Corollary 3.3.3

Proposition 3.3.5. *Consider* $Z = \ln(S)$ *where* S *is* $\mathcal{SLN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distributed. Let* ν *be the probability measure associated to the normal distribution* $\mathcal{N}(\mu, \sigma^2)$. *We have* $f_Z/f_\nu \in \mathcal{L}^2(\nu)$ *if*

$$2\sigma^2 > (2c_2)^{-1} = \max_{i=1,\ldots,n} \Sigma_{ii}\,. \tag{3.19}$$

Computing the $\{\widehat{a}_k\}_{k\in\mathbb{N}_0}$ coefficients can be done using Crude Monte Carlo (CMC), as in

$$\widehat{a}_k = \frac{1}{R}\sum_{r=1}^{R} Q_n(\ln S_r)\,, \quad S_1,\ldots,S_R \overset{\text{i.i.d.}}{\sim} \mathrm{SLN}(\boldsymbol{\mu},\boldsymbol{\Sigma})\,,$$

for $k = 0,\ldots,K$. We can use the same S_1, …, S_R for all $\widehat{a}_k$ together with a smoothing technique called *common random numbers* [15, 65]. Note that a non-trivial amount of computational time is typically spent just constructing the Hermite polynomials. Incorporating the Hermite polynomial's recurrence relation in our calculations achieved a roughly 40× speed-up compared with using Mathematica's `HermiteH`.

3.3.6 Lognormal sums via a gamma reference distribution

When ν is $\mathsf{Gamma}(r,m)$, it makes little sense to expand f in terms of $\{Q_k\}_{k\in\mathbb{N}_0}$ and f_ν as the integrability condition (3.10) fails, $f/f_\nu \notin \mathcal{L}^2(\nu)$. The workaround consists in using orthogonal polynomials to expand the *exponentially tilted* distribution, denoted $\mathrm{SLN}_\theta(\boldsymbol{\mu},\boldsymbol{\Sigma})$. This distribution's p.d.f. is

$$f_\theta(x) = \frac{\mathrm{e}^{-\theta x} f(x)}{\mathcal{L}(\theta)}\,, \quad \theta \geq 0, \tag{3.20}$$

where $\mathcal{L}(\theta) = \mathbb{E}[\mathrm{e}^{-\theta S}]$ is the Laplace transform of S. Asmussen et al. [19] investigated the use of $f_\theta(x)$ in approximating the left tail of S, and developed asymptotic forms and Monte Carlo estimators of this density.

Remark 3.3.7. The use of gamma distribution and Laguerre polynomials links our approach to a well established technique called the *Laguerre method*. The expansion is an orthogonal projection onto the basis of Laguerre functions constructed by multiplying Laguerre polynomials and the square root of the exponential distribution with parameter 1. The method is described in [1]. Note also that the damping procedure employed when integrability problems arise is quite similar to considering the exponentially tilted distribution instead of the real one. The use of the gamma distribution as reference is applied to actuarial science in [68, 67]. ◇

Using (3.10), we immediately obtain the following result which sheds light on how to tune the parameters of the reference gamma distribution so the integrability condition $f_\theta/f_\nu \in \mathcal{L}^2(\nu)$ is satisfied.

Proposition 3.3.8. *Consider the r.v. S_θ with the exponentially-tilted $\mathcal{SLN}_\theta(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution. Let ν be the probability measure associated with the* $\mathsf{Gamma}(r,m)$ *distribution. We have $f_\theta/f_\nu \in \mathcal{L}^2(\nu)$ if $m > 1/2\theta$.*

Hereafter we assume that the parameters r and m of $f_\nu \sim \mathsf{Gamma}(r, m)$ are chosen to satisfy Proposition 3.3.8's conditions.

Our approximation—based upon rearranging (3.20)—is of the form

$$\widehat{f}(x) = \mathrm{e}^{\theta x}\mathcal{L}(\theta)\widehat{f}_\theta(x) = \mathrm{e}^{\theta x}\mathcal{L}(\theta)\sum_{k=0}^{K} a_k Q_k(x) f_\nu(x)\,. \tag{3.21}$$

The coefficients $a_k = \mathbb{E}[Q_k(S_\theta)]$ can be estimated in (at least) three different ways: (i) using CMC, (ii) using MC while importance sampling from the original $\mathrm{SLN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, or (iii) by directly computing the moments $\mathbb{E}[S_\theta^k]$. The first method is nontrivial, as simulating from f_θ likely requires using acceptance-rejection (as in [19]). Options (ii) and (iii) use

$$a_k = \mathbb{E}[Q_k(S_\theta)] =: q_{k0} + q_{k1}\,\mathbb{E}[S_\theta] + \cdots + q_{kk}\,\mathbb{E}[S_\theta^k] \tag{3.22}$$

where $\{q_{ki}\}$ are the coefficients in Q_k, and

$$\mathbb{E}[S_\theta^i] = \frac{\mathbb{E}[S^i \mathrm{e}^{-\theta S}]}{\mathcal{L}(\theta)} =: \frac{\mathcal{L}_i(\theta)}{\mathcal{L}(\theta)}\,.$$

The $\mathcal{L}_i(\theta)$ notation was selected to highlight the link between $\mathbb{E}[S_n^i \mathrm{e}^{-\theta S_n}]$ and the ith derivative of $\mathcal{L}(\theta)$.

All three methods require access to the Laplace transform, and method (iii) requires $\mathcal{L}_i(\theta)$, however none of $\mathcal{L}(\theta)$ or $\mathcal{L}_i(\theta)$ are available in closed form. Our approach to circumvent these problems is presented in the Appendix.

3.4 Numerical illustrations

We take several approximations $\widehat{f}$ and compare them against the benchmark of numerical integration. One form of f particularly useful for numerical integration, in terms of the $\mathrm{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ density $f_{\mathcal{LN}}$, is as a surface integral, $f(s) = n^{-\frac{1}{2}} \int_{\Delta_n^s} f_{\mathcal{LN}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$, where $\Delta_n^s = \{\boldsymbol{x} \in \mathbb{R}_+^n : ||\boldsymbol{x}||_1 = s\}$. Mathematica integrates this within a reasonable time for $n = 2$ to 4 using `NIntegrate` and `ParametricRegion`). For $n > 4$ we qualitatively assess the performance of the estimators by plotting them.

The error measure used is the $\mathcal{L}^2([0, \mathbb{E}[S]])$ norm of $\widehat{f} - f$. We focus on this region as it is the hardest to approximate (indeed, Lemma 3.3.2 shows that just a single lognormal is a theoretically justified approximation of the SLN right tail) and due to its special relevance in applications, see for example the introduction of [19] and the references therein.

3.4.1 The estimators

We will compare the following approximations:

- the Fenton-Wilkinson approximation $\widehat{f}_{\mathrm{FW}}$, cf. [61], consists in approximating the distribution of S by a single lognormal with the same first and second moment;
- the log skew normal approximation $\widehat{f}_{\mathrm{Sk}}$, cf. [78][1], is a refinement of Fenton–Wilkinson by using a log skew normal as approximation and fitting the left tail in addition to the first and second moment;
- the conditional Monte Carlo approximation $\widehat{f}_{\mathrm{Cond}}$, cf. Example 4.3 on p. 146 of [15], uses the representation $f(x) = \mathbb{E}\big[\mathbb{P}(S \in \mathrm{d}x \,|\, Y)\big]$ for some suitable Y (here chosen as one of the normal r.v.s X_i occurring in (3.14)) and simulates the conditional expectation;
- $\widehat{f}_{\mathcal{N}}$ is the approximation described in Section 3.3.4 using a logarithmic transformation and the Hermite polynomials with a normal reference distribution;
- $\widehat{f}_{\Gamma}$ is the approximation described in Section 3.3.6 using exponential tilting and the generalised Laguerre polynomials with a gamma reference distribution.

These approximations are all estimators of functions (i.e., not pointwise estimators, such as in [91]) and they do not take excessive computational effort to construct. The first two, $\widehat{f}_{\mathrm{FW}}$ and $\widehat{f}_{\mathrm{Sk}}$, only need $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and do not have any Monte Carlo element. Similarly, the estimator $\widehat{f}_{\Gamma}$ when utilising the Gauss–Hermite quadrature described in (3.29) in the Appendix does not use Monte Carlo. For the remaining approximations we utilise the *common random numbers* technique, meaning that the same $R = 10^5$ i.i.d. $\mathrm{SLN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ samples $\boldsymbol{S} = (S_1, \ldots, S_R)^\top$ are given to each algorithm. Lastly, all the estimators except $\widehat{f}_{\Gamma}$ satisfy $\int \widehat{f}(x)\,\mathrm{d}x = 1$. One problem with the orthogonal polynomial estimators is that they can take negative values; this can easily be fixed, but we do not make that adjustment here.

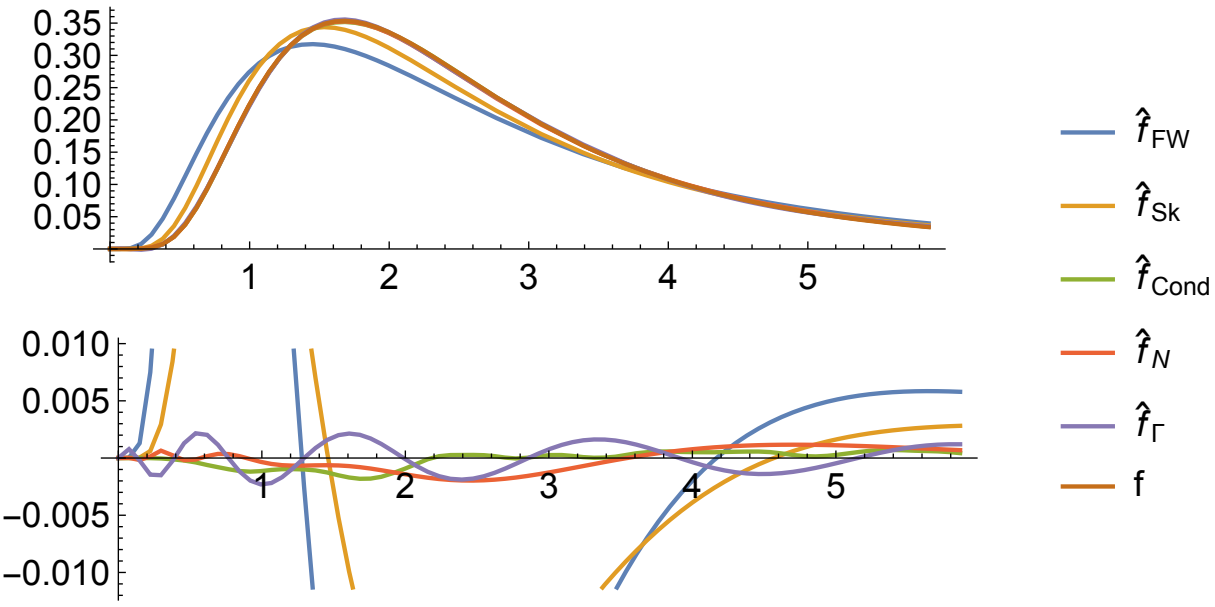
For $\widehat{f}_{\mathcal{N}}$, we take $\mu = \mathbb{E}[Z]$ and $\sigma^2 = \mathbb{V}\mathrm{ar}[Z]$, calculated using numerical integration. The $\widehat{f}_{\Gamma}$ case is more difficult. Equation (3.21) shows that we must impose $\theta m < 1$ to ensure that $\widehat{f}_{\Gamma}(x) \to 0$ as $x \to \infty$. Exploring different parameter selections showed that fixing $\theta = 1$ worked reasonably well. Moment matching f_θ to f_ν leads to the selection of m and r. The moments of f_θ, $\widehat{\mathbb{E}\, f_\theta} = \widehat{\mathcal{L}}_1(\theta)/\widehat{\mathcal{L}}_0(\theta)$ and $\widehat{\mathbb{V}\mathrm{ar}\, f_\theta} = \widehat{\mathcal{L}}_2(\theta)/\widehat{\mathcal{L}}_0(\theta) - \widehat{\mathbb{E}\, f_\theta}^{\,2}$ can be approximated using the Gauss–Hermite quadrature of (3.29); for this we use $H = 64$, 32, 16 for $n = 2$, 3, 4 respectively (and CMC for $n > 4$).

[1]Note that in [78], the formula for $\varepsilon_{\mathrm{opt}}$ contains an typographic error.

With these regimes, parameter selection for the reference distributions is automatic, and the only choice the user must make is in selecting K. In these tests we examined various K from 1 to 40, and show the best approximations found. The source code for these tests is available online at [16], and we invite readers to experiment with the effect of modifying K, θ and the parameters of the reference distributions.

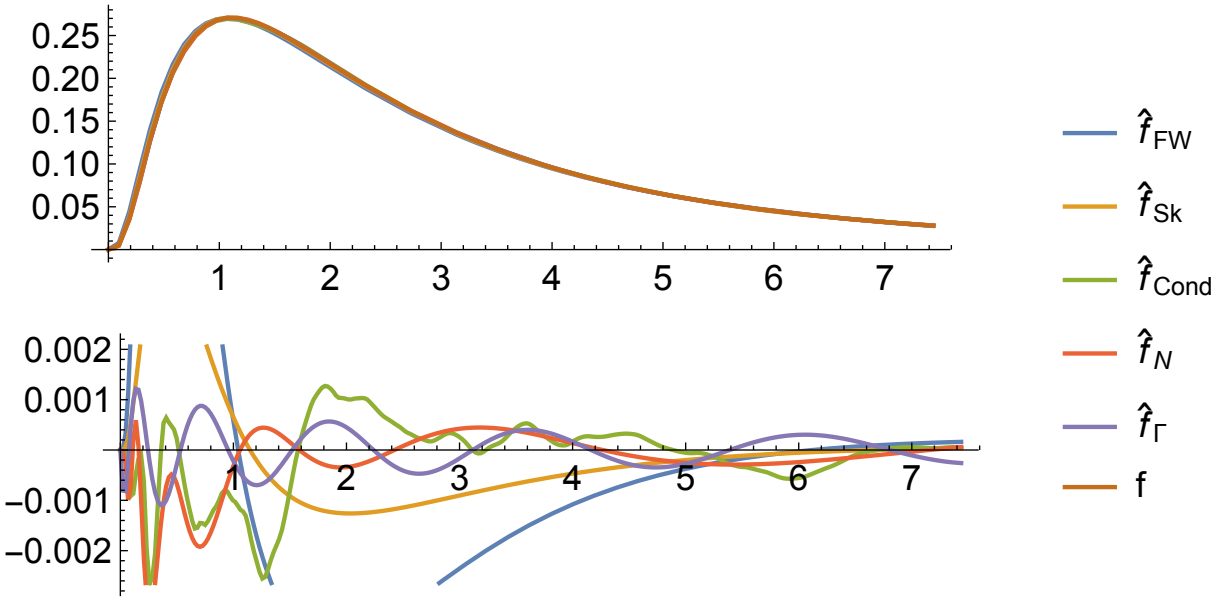
3.4.2 Results

For each test case with $n \leq 4$ we plot the $\widehat{f}(x)$ and $f(x)$ together and then $\widehat{f}(x) - f(x)$ over $x \in (0, 2\,\mathbb{E}[S])$. A table then shows the $\mathcal{L}^2$ errors over $(0, \mathbb{E}[S])$. A version of the paper with figures in colour is in the online supplement [16].



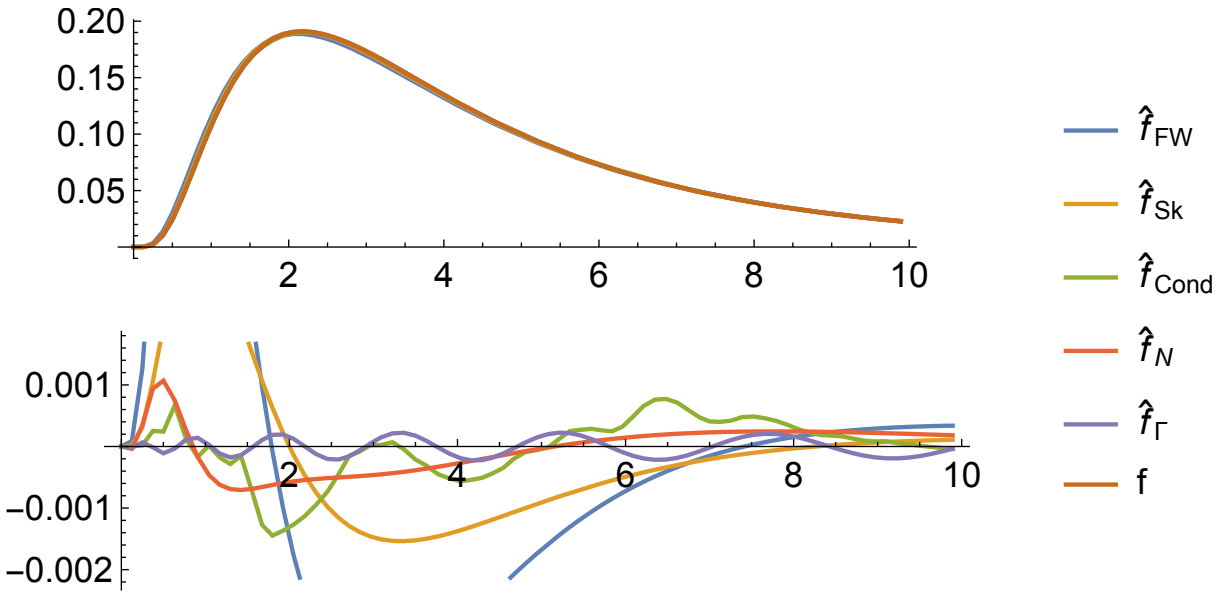
	$\widehat{f}_{\mathrm{FW}}$	$\widehat{f}_{\mathrm{Sk}}$	$\widehat{f}_{\mathrm{Cond}}$	$\widehat{f}_{\mathcal{N}}$	$\widehat{f}_{\Gamma}$
$\mathcal{L}^2$	8.01×10^{-2}	4.00×10^{-2}	1.56×10^{-3}	1.94×10^{-3}	2.28×10^{-3}

Test 1: $\boldsymbol{\mu} = (0, 0)$, $\mathrm{diag}(\boldsymbol{\Sigma}) = (0.5, 1)$, $\rho = -0.2$. Reference distributions used are $\mathsf{Normal}(0.88, 0.71^2)$ and $\mathsf{Gamma}(2.43, 0.51)$ with $K = 32$, 16 resp.



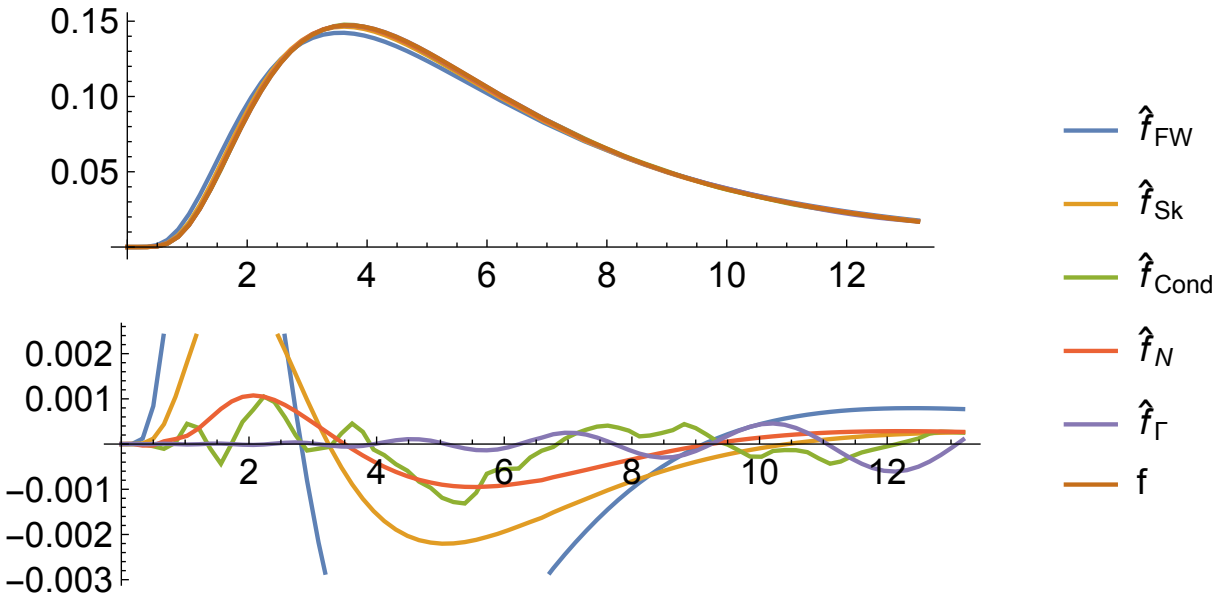
	$\widehat{f}_{\mathrm{FW}}$	$\widehat{f}_{\mathrm{Sk}}$	$\widehat{f}_{\mathrm{Cond}}$	$\widehat{f}_{\mathcal{N}}$	$\widehat{f}_{\Gamma}$
$\mathcal{L}^2$	1.02×10^{-2}	3.49×10^{-3}	1.78×10^{-3}	7.86×10^{-4}	7.24×10^{-4}

Test 2: $\boldsymbol{\mu} = (-0.5, 0.5)$, $\mathrm{diag}(\boldsymbol{\Sigma}) = (1, 1)$, $\rho = 0.5$. Reference distributions used are $\mathsf{Normal}(0.91, 0.90^2)$ and $\mathsf{Gamma}(2.35, 0.51)$ with $K = 32$, 16 resp.



	$\widehat{f}_{\mathrm{FW}}$	$\widehat{f}_{\mathrm{Sk}}$	$\widehat{f}_{\mathrm{Cond}}$	$\widehat{f}_{\mathcal{N}}$	$\widehat{f}_{\Gamma}$
$\mathcal{L}^2$	9.48×10^{-3}	3.71×10^{-3}	1.60×10^{-3}	1.18×10^{-3}	3.53×10^{-4}

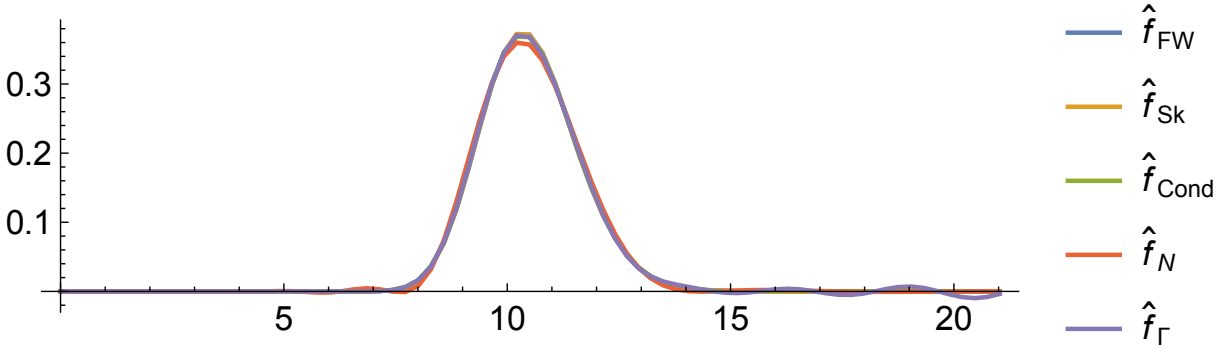
Test 3: $n = 3$, $\mu_i = 0$, $\Sigma_{ii} = 1$, $\rho = 0.25$. Reference distributions used are $\mathsf{Normal}(1.32, 0.74^2)$ and $\mathsf{Gamma}(3, 0.57)$ with $K = 7$, 25 resp.



	$\widehat{f}_{\mathrm{FW}}$	$\widehat{f}_{\mathrm{Sk}}$	$\widehat{f}_{\mathrm{Cond}}$	$\widehat{f}_{\mathcal{N}}$	$\widehat{f}_{\Gamma}$
$\mathcal{L}^2$	1.82×10^{-2}	6.60×10^{-3}	1.90×10^{-3}	1.80×10^{-3}	1.77×10^{-4}

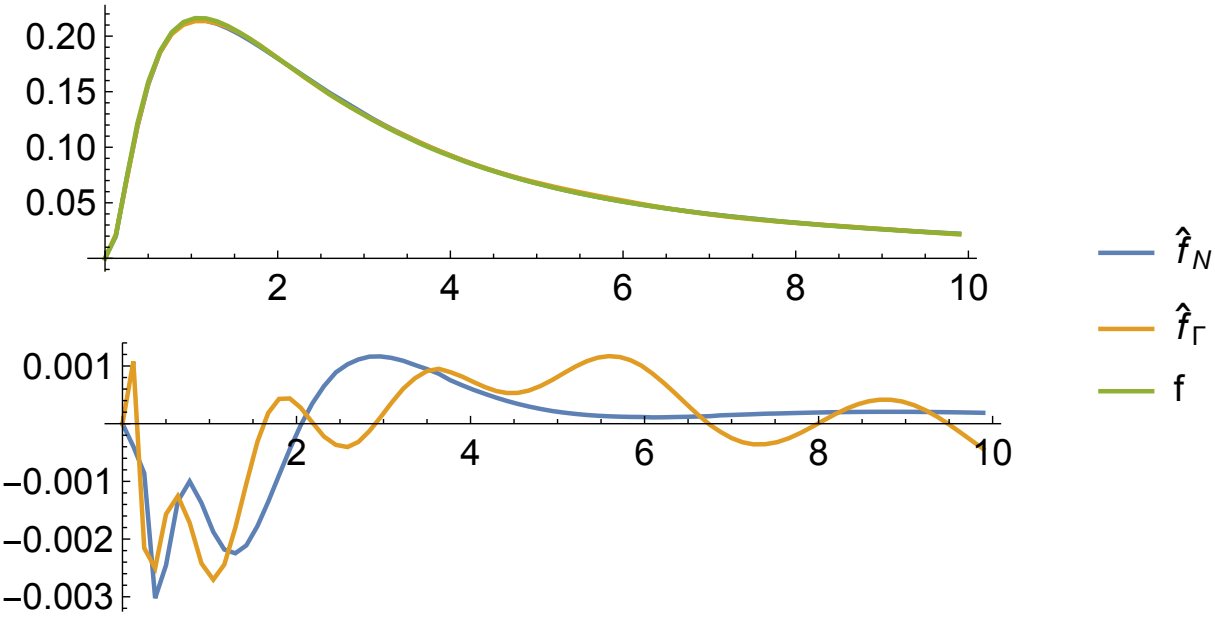
Test 4: $n = 4$, $\mu_i = 0$, $\Sigma_{ii} = 1$, $\rho = 0.1$. Reference distributions used are $\mathsf{Normal}(1.32, 0.74^2)$ and $\mathsf{Gamma}(3.37, 0.51)$ with $K = 18$, 18 resp.

The following test case shows the density approximations for a large n.



Test 5: Sum of 10 i.i.d. $\mathrm{LN}(0, 0.1)$ r.v.s. Reference distributions used are $\mathsf{Normal}(2.35, 0.23^2)$ and $\mathsf{Gamma}(12.61, 0.25)$ with $K = 18$, 35 resp.

Finally, we fit $\widehat{f}_{\mathcal{N}}$ and $\widehat{f}_{\Gamma}$ to simulated data (10^5 replications) for the sum of lognormals with a non-Gaussian dependence structure. Specifically, we take the sum of $n = 3$



Test 6: Sum of 3 LN$(0,1)$ r.v.s with $C_{10}^{\mathrm{Cl}}(\cdot)$ copula (i.e., $\tau = \frac{5}{6}$). Reference distributions used are Normal$(1.46, 0.71^2)$ and Gamma$(8.78, 0.25)$ with $K = 40$. The $\mathcal{L}^2$ errors of $\widehat{f}_{\mathcal{N}}$ and $\widehat{f}_{\Gamma}$ are 2.45×10^{-3} and 2.04×10^{-3} respectively.

standard lognormal r.v.s with a *Clayton copula*, defined by its distribution function

$$C_\theta^{\mathrm{Cl}}(u_1, \ldots, u_n) = \left(1 - n + \sum_{i=1}^{n} u_i^{-\theta}\right)^{-1/\theta}, \qquad \text{for } \theta > 0\,.$$

The Kendall's tau correlation of the C_θ^{Cl} copula is $\tau = \theta/(\theta+2)$ [103].

Our overall conclusion of the numerical examples is that no single method can be considered as universally superior. Of the methods in the literature, the log skew normal approximations is generally better than Fenton-Wilkinson, which is unsurprising given it is an extension introducing one more parameter. The estimators, $\widehat{f}_{\mathcal{N}}$ and $\widehat{f}_{\Gamma}$, based on orthogonal polynomial approximation techniques, are very flexible. They also display at least as good and sometimes better p.d.f. estimates over the interval $(0, \mathbb{E}[S])$ and their periodic error indicates that they would supply even more accurate c.d.f. estimates. One should note, however, that their performance relies on the tuning of parameters and that somewhat greater effort is involved in their computation (though this is mitigated through the availability of the software in [16]).

An interesting feature of $\widehat{f}_{\mathcal{N}}$ and $\widehat{f}_{\Gamma}$ is that the Clayton copula example indicates some robustness to the dependence structure used. In view of the current interest in financial applications of non-Gaussian dependence this seems a promising line for future research.

Appendix A: Proof of Proposition 3.2.6

Proof. The polynomials which are orthogonal with respect to the lognormal distribution will be derived using the general formula

$$Q_n(x) = \frac{1}{\sqrt{D_{n-1,n-1}D_{n,n}}} \begin{vmatrix} s_0 & s_1 & \cdots & s_n \\ s_1 & s_2 & \cdots & s_{n+1} \\ \vdots & \vdots & & \vdots \\ s_{n-1} & s_n & \cdots & s_{2n-1} \\ 1 & x & \cdots & x^n \end{vmatrix}, \; n \geq 1, \tag{3.23}$$

where $\{s_n\}_{n\in\mathbb{N}_0}$ denotes the moment sequence of the lognormal distribution and $D_{n,n} = \left|\left\{s_{k,l}\right\}_{0\leq k,l\leq n}\right|$ is a Hankel determinant. Formula (3.23) is obtained through orthogonalization of the sequence of linearly independent functions $1, x, x^2, \ldots$, see [131, pp. 26–27]. The moments of the lognormal distribution are given by $s_n = p^n q^{n^2}$, where $p = \mathrm{e}^{\mu}$ and $q = \mathrm{e}^{\frac{\sigma^2}{2}}$. Consider

$$D_{n,n} = \begin{vmatrix} 1 & pq & \cdots & p^n q^{n^2} \\ pq & p^2q^4 & & p^{n+1}q^{(n+1)^2} \\ \vdots & \vdots & & \vdots \\ p^{n-1}q^{(n-1)^2} & p^n q^{n^2} & \cdots & p^{2n-1}q^{(2n-1)^2} \\ p^n q^{n^2} & p^{n+1}q^{(n+1)^2} & \cdots & p^{2n}q^{(2n)^2} \end{vmatrix}, \; n \geq 1, \tag{3.24}$$

and denote by R_k the kth row and by C_ℓ the ℓth column. We apply the elementary operations $R_{k+1} \to p^{-k}q^{-k^2}R_{k+1}$, and $C_{\ell+1} \to p^{-\ell}q^{-\ell^2}C_{\ell+1}$ for $k, \ell = 0, \ldots, n$ to get a Vandermonde type determinant. Thus we have

$$D_{n,n} = \mathrm{e}^{n(n+1)\mu}\mathrm{e}^{\frac{n(n+1)(2n+1)}{3}\sigma^2} \prod_{k=0}^{n-1} \left[\mathrm{e}^{-\sigma^2}; \mathrm{e}^{-\sigma^2}\right]_k \tag{3.25}$$

We expand the determinant in (3.23) with respect to the last row to get

$$Q_n(x) = \frac{1}{\sqrt{D_{n-1,n-1}D_{n,n}}} \sum_{k=0}^{n} (-1)^{n+k} x^k D_{n-1,n}^{-k}, \tag{3.26}$$

where $D_{n-1,n}^{-k}$ is $D_{n,n}$ with the last row and the $(k+1)$th column deleted. We perform on $D_{n-1,n}^{-k}$ the following operations: $R_{j+1} \to p^{-j}q^{-j^2}R_{j+1}$, for $j = 0, \ldots, n-1$, $C_{j+1} \to$

$p^{-j}q^{-j^2}C_{j+1}$, for $j = 0, \ldots, k-1$, and finally $C_j \to p^{-j}q^{-j^2}C_j$, for $j = k+1, \ldots, n$. We obtain

$$D_{n-1,n}^{-k} = p^{n^2-k}q^{\frac{2n^3+n}{3}-k^2} \begin{vmatrix} 1 & \alpha_0 & \cdots & \alpha_0^{k-1} & \alpha_0^{k+1} & \cdots & \alpha_0^n \\ 1 & \alpha_1 & \cdots & \alpha_1^{k-1} & \alpha_1^{k+1} & \cdots & \alpha_1^{n+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \alpha_{n-1} & \cdots & \cdots & \cdots & \cdots & \alpha_{n-1}^n \end{vmatrix},$$

where $\alpha_k = q^{2k}$, for $k = 0, \ldots, n-1$. Expanding the polynomial $B(x) = \prod_{i=0}^{n-1}(x - \alpha_i)$, we get

$$B(x) = x^n + \beta_{n-1}x^{n-1} + \ldots + \beta_0,$$

where $\beta_k = (-1)^{n-k}e_{n-k}(\alpha_0, \ldots, \alpha_{n-1})$, and $e_k(X_1, \ldots, X_n)$ denotes the elementary symmetric polynomial, defined previously in (3.12). We apply the elementary operation $C_n \to C_n + \sum_{j=0}^{k-1} a_j C_{j+1} + \sum_{j=k+1}^{n-1} a_j C_j$, followed by $n-k$ cyclic permutations to get

$$D_{n-1,n}^{-k} = p^{n-k}q^{n^2-k^2}e_{n-k}\left(1, \ldots, q^{2(n-1)}\right) D_{n-1,n-1}. \tag{3.27}$$

Inserting (3.25) and (3.27) into (3.26) leads to (3.11). □

Appendix B: Computing the coefficients of the expansion $\{a_k\}_{k\in\mathbb{N}_0}$ in the gamma case

We extend here the techniques developed in [91] to construct an approximation for $\mathcal{L}_i(\theta)$. We note that $\mathcal{L}_i(\theta) \propto \int_{\mathbb{R}^n} \exp\{-h_{\theta,i}(\boldsymbol{x})\}\, \mathrm{d}\boldsymbol{x}$ where

$$h_{\theta,i}(\boldsymbol{x}) = -i\ln(\mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}}) + \theta \mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x}\,, \quad i \in \mathbb{N}_0\,.$$

This uses the notation $\mathbf{e}^{\boldsymbol{x}} = (\mathrm{e}^{x_1}, \ldots, \mathrm{e}^{x_n})^\top$. Next, define $\boldsymbol{x}^*$ as the minimiser of $h_{\theta,i}$ (calculated numerically), and consider a second order Taylor expansion of $h_{\theta,i}$ about $\boldsymbol{x}^*$. Denote $\widetilde{\mathcal{L}}_i(\theta)$ as the approximation of $\mathcal{L}_i(\theta)$ in which $h_{\theta,i}$ is replaced by this Taylor expansion. Simplifying yields

$$\widetilde{\mathcal{L}}_i(\theta) = \frac{\exp\{-h_{\theta,i}(\boldsymbol{x}^*)\}}{\sqrt{|\boldsymbol{\Sigma}\boldsymbol{H}|}} \tag{3.28}$$

where $\boldsymbol{H}$, the Hessian of $h_{\theta,i}$ evaluated at $\boldsymbol{x}^*$, is

$$\boldsymbol{H} = i\frac{\mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*}(\mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})^\top}{(\mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})^2} + \boldsymbol{\Sigma}^{-1} - \mathrm{diag}(\boldsymbol{\Sigma}^{-1}\boldsymbol{x}^*)\,.$$

As $\theta \to \infty$ we have $\widetilde{\mathcal{L}}_i(\theta) \to \mathcal{L}_i(\theta)$. We can rewrite $\mathcal{L}_i(\theta) = \widetilde{\mathcal{L}}_i(\theta) I_i(\theta)$ and estimate $I_i(\theta)$ as in [91].

Proposition 3..3. *The moments of the exponentially-tilted distribution* $\mathrm{SLN}_\theta(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, denoted* $\mathcal{L}_i(\theta)$*, can be written as* $\mathcal{L}_i(\theta) = \widetilde{\mathcal{L}}_i(\theta) I_i(\theta)$*, where* $\widetilde{\mathcal{L}}_i(\theta)$ *is in* (3.28)*, and*

$$I_i(\theta) = \sqrt{|\boldsymbol{\Sigma}\boldsymbol{H}|}\, v(\mathbf{0})^{-1}\ \mathbb{E}[v(\boldsymbol{\Sigma}^{\frac{1}{2}} Z)]$$

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

$$v(\boldsymbol{z}) = \exp\{i \ln(\mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*+\boldsymbol{z}}) - \theta \mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*+\boldsymbol{z}} - (\boldsymbol{x}^*)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{z}\}\,.$$

Proof. We begin by substituting $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{H}^{-\frac{1}{2}} \boldsymbol{y}$ into $\mathcal{L}_i(\theta)$, then multiply by $\exp\{\pm \text{ some constants}\}$:

$$\begin{aligned}
\mathcal{L}_i(\theta) &= \int_{\mathbb{R}^n} \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{|\boldsymbol{\Sigma}|}} \exp\{i \log(\mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}}) - \theta \mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}} - \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}\}\, \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{|\boldsymbol{\Sigma}\boldsymbol{H}|}} \exp\{i \log(\mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*+\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}}) - \theta \mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*+\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}} \\
&\qquad - \frac{1}{2} (\boldsymbol{x}^* + \boldsymbol{H}^{-\frac{1}{2}} \boldsymbol{y})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}^* + \boldsymbol{H}^{-\frac{1}{2}} \boldsymbol{y})\}\, \mathrm{d}\boldsymbol{y} \\
&= \widetilde{\mathcal{L}}_i(\theta) \exp\{-i \log(\mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*}) + \theta \mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*}\} \\
&\qquad \times \int_{\mathbb{R}^n} (2\pi)^{-\frac{n}{2}} \exp\{i \log(\mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*+\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}}) - \theta \mathbf{1}^\top \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*+\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}} \\
&\qquad\qquad - (\boldsymbol{x}^*)^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{H}^{-\frac{1}{2}} \boldsymbol{y} - \frac{1}{2} \boldsymbol{y}^\top (\boldsymbol{\Sigma}\boldsymbol{H})^{-1} \boldsymbol{y}\}\, \mathrm{d}\boldsymbol{y}\,.
\end{aligned}$$

That is, $\mathcal{L}_i(\theta) = \widetilde{\mathcal{L}}_i(\theta) I_i(\theta)$. In $I_i(\theta)$, take the change of variable $\boldsymbol{y} = (\boldsymbol{\Sigma}\boldsymbol{H})^{\frac{1}{2}} \boldsymbol{z}$, and the result follows. □

Remark 3..4. The form of $I_i(\theta)$ naturally suggests evaluation using *Gauss–Hermite* quadrature:

$$\widehat{\mathcal{L}}_i(\theta) = \frac{\exp\{-h_{\theta,i}(\boldsymbol{x}^*)\}}{v(\mathbf{0})\, \pi^{n/2}} \sum_{i_1=1}^{H} \cdots \sum_{i_n=1}^{H} v(\boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{z}) \prod_{j=1}^{n} w_{i_j} \tag{3.29}$$

where $\boldsymbol{z} = (z_{i_1}, \ldots, z_{i_n})^\top$, the set of weights and nodes $\{(w_i, z_i) : 1 \le i \le H\}$ is specified by the Gauss–Hermite quadrature algorithm, and $H \ge 1$ is the order of the approximation. This approximation is accurate, especially so when the i in $\mathcal{L}_i$ becomes large. Even for $\mathcal{L}$ $(= \mathcal{L}_0)$ this method appears to outperform the quasi-Monte Carlo scheme outlined in [91]. ◇

Thus, with $\widehat{\mathcal{L}}_i(\theta)$ given in (3.29), we can now estimate the coefficients. The three methods correspond to

1. $\widehat{a}_k = R^{-1} \sum_{r=1}^{R} Q_k(S_r)$, for $S_1, \ldots, S_R \overset{\text{i.i.d.}}{\sim} f_\theta(x)$,
2. $\widehat{a}_k = \sum_{j=0}^{k} q_{kj} \widehat{\mathbb{E}[S_\theta^j]} = q_{k0} + (R\,\widehat{\mathcal{L}}(\theta))^{-1} \sum_{j=1}^{k} q_{kj} \sum_{r=1}^{R} S_r^j \mathrm{e}^{-\theta S_r}$, from (3.22), where $S_1, \ldots, S_R \overset{\text{i.i.d.}}{\sim} f(x)$,
3. $\widehat{a}_k = q_{k0} + \widehat{\mathcal{L}}(\theta)^{-1} \sum_{j=1}^{k} q_{kj}\, \widehat{\mathcal{L}}_j(\theta)$.

In the numerical illustrations, we switched between using methods (2) and (3) for large and small n respectively. Algorithms for efficient simulation from f_θ is work in progress.

Chapter 4

Two numerical methods to evaluate stop-loss premiums

4.1 Introduction

Consider the random variable (r.v.)

$$S_N = \sum_{k=1}^{N} U_k,$$

where N is a counting r.v. and $\{U_k\}_{k\in\mathbb{N}_+}$ is a sequence of r.v.s which are independent and identically distributed (i.i.d.), non-negative, and independent of N. We denote the probability density function (p.d.f.) of S_N as f_{S_N}, and its survival function (s.f.) as

$$\overline{F}_{S_N}(x) = \mathbb{P}(S_N > x), \quad \text{for } x \geq 0.$$

This paper concerns approximations of f_{S_N} and $\overline{F}_{S_N}$, though we begin with a discussion of how S_N is used in actuarial science.

Frequently S_N models the aggregated losses of a non-life insurance portfolio over a given period of time—here N represents the number of claims and U_k the claim sizes—yet other applications also exist. Actuaries and risk managers typically want to quantify the risk of large losses by a single comprehensible number, a risk measure.

One popular risk measure is the Value-at-Risk (VaR). In actuarial contexts, the VaR at level $\alpha \in (0, 1)$ is defined such that the probability of (aggregated) losses exceeding the level VaR is at most $1 - \alpha$. We denote this α-quantile as

$$\mathrm{VaR}_{S_N}(\alpha) = \inf\{x \geq 0, F_{S_N}(x) \geq \alpha\}.$$

Following the European recommendation of the Solvency II directive, the standard value for α is 0.995, see [55]. It is used by risk managers in banks, insurance companies, and other financial institutions to allocate risk reserves and to determine solvency margins. Also, we have stop-loss premiums (slp's) which are risk measures that are commonly used in reinsurance agreements.

A reinsurance agreement is a common risk management contract between insurance companies, one called the "cedant" and the other the "reinsurer". Its aim is to keep the cedant's long-term earnings stable by protecting the cedant against large losses. The reinsurer absorbs part of the cedant's loss, say $f(S_N)$ where $0 \leq f(S_N) \leq S_N$, leaving the cedant with $I_f(S_N) = S_N - f(S_N)$. In return, the cedant pays a premium linked to

$$\Pi = \mathbb{E}[f(S_N)],$$

under the expected value premium principle.

In practice, there are a variety of reinsurance designs from which an insurer can choose. We focus in this work on the stop-loss reinsurance treaty associated with the following ceded loss function

$$f(S_N) = (S_N - a)_+, \ a \geq 0,$$

where a is referred to as the retention level or priority. The ratemaking of the stop-loss reinsurance policy requires the evaluation of

$$\Pi_a(S_N) = \mathbb{E}\left[(S_N - a)_+\right], \tag{4.1}$$

also known as the usual stop loss premium (slp).

One variation is the limited stop-loss function,

$$f(S_N) = \min[(S_N - a)_+, b], \ b \geq 0, \tag{4.2}$$

where b is called the limit. The limited stop-loss function (4.2) is very appealing in practice because it prevents the cedant from over-estimating their losses and therefore over-charging the reinsurer. We also have the change-loss function defined as

$$f(S_N) = c(S_N - a)_+, \ 0 \leq c \leq 1,$$

which is in between stop-loss and quota-share reinsurance. Note that the ratemaking in each case requires the expectation in (4.1).

From a practical point of view, a reinsurance treaty over the whole portfolio is less expensive to handle than one which involves claim-by-claim management. It also grants protection in the event of an unusual number of claims, triggered for instance by a natural disaster. From a theoretical point of view, it is well known that the stop-loss ceded

function allows one to minimize the variance of the retained loss for a given premium level, see for instance the monograph of Denuit et al. [50]. Recently, it has been shown that stop-loss reinsurance is also optimal when trying to minimize the VaR and the expected shortfall of the retained loss, see the works of Cai et al. [37], Cheung [39], and Chi and Tan [40]. Note that some other ceded loss functions appear in their work, there are however very close to the stop-loss one.

Unfortunately, one is seriously constrained when calculating these quantities analytically, as there are only a few cases where either the p.d.f. or the s.f. is available in a simple tractable form. To estimate the VaR or slp we must find fast and accurate approximations for these functions.

We discuss the use of an approximation of the p.d.f. in terms of the gamma density and its orthonormal polynomials. This method has been studied in the recent works of Goffard et al. [67] and Jin et al. [85]. We emphasize here the computational aspect of this numerical method and detail some practical improvements. An exponential change of measure can be used to recover the p.d.f. of S_N when the claim sizes are governed by a heavy-tailed distribution. This refinement has been successfully applied in the work of Asmussen et al. [17] to recover the density of the sum of lognormally distributed random variables.

This method is compared to a numerical inversion of the Laplace transform which is known to be efficient to recover the survival function of a compound distribution. The critical step in Laplace inversion is to select which numerical integration technique to apply. We implement a method inspired by the work of Abate and Whitt [2] which is very similar to the method of Rolski et al. [119, Chapter 5, Section 5]. An approximation of the slp is then proposed relying on the connection with the survival function of the equilibrium distribution of S_N. Note that Dufresne et al. [54] successfully applied a Laplace inversion based technique to the evaluation of slp.

To close, we want to emphasize the fact that the numerical methods also apply in a risk theory framework. The infinite-time ruin probability in the compound Poisson ruin model is equal to the survival function of a compound geometric distribution. The polynomial approximation and the Laplace inversion methods have been employed, and compared to solve this particular problem in the work of Goffard et al. [68]. We add a more original application by noting that the finite-time non-ruin probability with no initial reserves, again under the classical risk model assumptions, may be rewritten as the slp associated with a compound Poisson distribution where the priority is expressed in terms of the premium rate and the time horizon.

The rest of the paper is organized as follows. Section 4.2 introduces compound distributions, and details their role in risk theory. Section 4.3 presents the approximation method based on orthogonal polynomials. Section 4.4 presents the approximation through the

numerical inversion of the Laplace transform. Section 4.5 is devoted to numerical illustrations where the performances of the two methods are compared; the MATHEMATICA code used in this section is available online [66].

4.2 Compound distributions and risk theory

After setting up some notational conventions for Laplace transforms, compound distributions are introduced along with a brief account of their importance in risk modeling.

4.2.1 Laplace transforms

Definition 4.2.2. *For a function $f : \mathbb{R}_+ \to \mathbb{R}_+$, we define*

$$\mathcal{L}\{f\}(t) \stackrel{\text{def}}{=} \int_0^\infty \mathrm{e}^{-tx} f(x)\mathrm{d}x\,, \quad \textit{for } t \in \mathbb{C} \textit{ with } \Re(t) \geq 0\,,$$

to be the corresponding Laplace transform. *For a positive random variable X with p.d.f. f_X, we write $\mathcal{L}_X(t) \stackrel{\text{def}}{=} \mathcal{L}\{f_X\}(t) = \mathbb{E}\,\mathrm{e}^{-tX}$.* ◇

We have the result that for $t > 0$

$$\mathcal{L}\{F_X\}(t) = \frac{\mathcal{L}\{f_X\}(t)}{t} = \frac{\mathcal{L}_X(t)}{t}\,, \text{ and}$$

$$\mathcal{L}\{\overline{F}_X\}(t) = \frac{1}{t} - \mathcal{L}\{F_X(x)\}(t) = \frac{1 - \mathcal{L}_X(t)}{t}\,.$$

4.2.3 Compound distribution

Let $S_N = \sum_{k=1}^N U_k$ be the aggregated claim amounts associated with a non-life insurance portfolio over a fixed time period. The number of claims, also called the claim frequency, is modeled by a counting random variable N having a probability mass function f_N. The claim sizes form a sequence $\{U_k\}_{k\in\mathbb{N}_+}$ of i.i.d. non-negative random variables with common p.d.f. f_U. We further assume that the claim sizes are independent from the claim frequency, so the random variable S_N follows the compound distribution (f_N, f_U).

As $S_N = 0$ whenever $N = 0$ (assuming this occurs with positive probability), the distribution of S_N is the sum of a singular part (the probability mass $\mathbb{P}(S_N = 0) = f_N(0) > 0$)

and a continuous part (describing S_N where $N > 0$) with a defective p.d.f. $f^+_{S_N}$ and c.d.f. $F^+_{S_N}$. From the law of total probability, we have

$$f^+_{S_N}(x) = \sum_{n=1}^{\infty} f_N(n) f_U^{\top n}(x),\ x \geq 0. \tag{4.3}$$

This density is intractable because of the infinite series. Furthermore, the summands are defined by repeated convolution of f_U with itself which are rarely straightforward to evaluate. The methods presented in this work rely on the knowledge of the Laplace transform of S_N, given by

$$\mathcal{L}_{S_N}(t) = \mathcal{G}_N[\mathcal{L}_U(t)]\,,$$

where $\mathcal{G}_N(t) \stackrel{\text{def}}{=} \mathbb{E}(t^N)$ is the *probability generating function* of N. The simple expression of the Laplace transform has made possible the use of numerical methods involving the moments or transform inversion to recover the distribution of S_N. The distribution is typically recovered using Panjer's algorithm or a Fast Fourier Transform algorithm based on the inversion of the discrete Fourier transform; these two methods are compared in the work of Embrechts and Frei [57]. Our orthogonal polynomial method involves the standard integer moment sequence for S_N, in contrast to more exotic types of moments used by some recent methods. Gzyl and Tagliani [72] uses the fractional moments within a max-entropic based method, while Mnatsakanov and Sarkisian [105] performs an inversion of the scaled Laplace transform via the exponential moments. In addition to proposing an approximation for the survival function of S_N, we provide an efficient way to compute the usual slp (4.1) for reinsurance applications.

4.2.4 Risk theory

In the classical risk model, the financial reserves of a non-life insurance company are modeled by the risk reserve process $\{R(t), t \geq 0\}$, defined as

$$R(t) = u + ct - \sum_{k=1}^{N(t)} U_k.$$

The insurance company holds an initial capital of amount $R(0) = u \geq 0$, and collects premiums at a constant rate of $c > 0$ per unit of time. The number of claims up to time $t \geq 0$ is governed by a homogeneous Poisson process $\{N(t), t \geq 0\}$ with intensity λ. The claim sizes are i.i.d. non-negative random variables independent from $N(t)$.

One of the goals of risk theory to evaluate an insurer's ruin probability, that is, the probability that the financial reserves eventually fall below zero. Of interest are both the

finite-time ruin probability $\psi(u,T)$ and the infinite-time ruin probability, also called the *probability of ultimate ruin*, $\psi(u)$, which are defined as

$$\psi(u,T) = \mathbb{P}\Big(\inf_{0\le t\le T} R(t) \le 0\Big),$$

and

$$\psi(u) = \mathbb{P}\Big(\inf_{t\ge 0} R(t) \le 0\Big).$$

These probabilities are often reformulated (for mathematical convenience) in terms of the associated claims surplus process $\{S(t), t \ge 0\}$,

$$S(t) = \sum_{k=1}^{N(t)} U_k - ct,\ t \ge 0,$$

specifically,

$$\psi(u,T) = \mathbb{P}\Big(\sup_{0\le t\le T} S(t) \ge u\Big) \quad \text{and} \quad \psi(u) = \mathbb{P}\Big(\sup_{t\ge 0} S(t) \ge u\Big).$$

For a general background on risk theory and the evaluation of ruin probabilities, we refer the reader to the monograph of Asmussen and Albrecher [13].

The first connection between compound distributions and ruin probabilities is the following. If the net benefit condition is satisfied, i.e. if the premium rate exceeds the average cost of aggregated claims per unit of time, then the infinite-time ruin probability is given by the survival function of a geometric compound distribution. More precisely,

$$\psi(u) = \mathbb{P}\left(S_N \overset{\text{def}}{=} \sum_{k=1}^{N} U_k^{\top} > u\right) = (1-\rho)\sum_{n=1}^{\infty} \rho^n \overline{F}_{U^{\top}}^{\top n}(u),$$

with $N \sim \mathsf{Geom}_0(\rho)$, $\rho = \lambda\mathbb{E}(U)/c < 1$, and with i.i.d. $U_k^{\top}$ with p.d.f. $f_{U^{\top}}(x) = \overline{F}_U(x)/\mathbb{E}(U)$. This result is known as the Pollaczeck–Khinchine formula, see for instance Asmussen and Albrecher [13, Chapter IV, (2.2)]. Thus it is possible to evaluate the infinite-time ruin probability via Panjer's algorithm. If we are able to determine the Laplace transform of V then we can also apply the polynomial method of Goffard et al. [67], the fractional moment based method of Gzyl et al. [71], and the exponential moments based method of Mnatsakanov et al. [106].

The second connection links the finite-time ruin probability with no initial reserves to the slp associated with a compound distribution. If $N(t) \sim \mathsf{Poisson}(\lambda t)$ (i.e. claims arrive as a homogeneous Poisson process) then the finite-time ruin probability is given by

$$\psi(0,T) = 1 - \frac{1}{cT}\int_0^{cT} \mathbb{P}\Big(\sum_{i=1}^{N(T)} U_i \le x\Big)\mathrm{d}x\,. \tag{4.4}$$

Writing $S_{N(T)} = \sum_{i=1}^{N(T)} U_i$ we can see that (4.4) says that $\psi(0,T) = \mathbb{E}[\min\{S_{N(T)}, cT\}]/cT$ and hence

$$\psi(0,T) = (cT)^{-1}\Big[\mathbb{E}[N(T)]\,\mathbb{E}[U_1] - \Pi_{cT}(S_{N(T)})\Big]. \tag{4.5}$$

Lefèvre and Picard [97, Corollary 4.3] show that equations (4.4) and (4.5) hold in the more general case where the claim arrival process forms a *mixed Poisson process*. This connection has been exploited recently in Lefèvre et al. [98] where the influence of the claim size distribution on the ruin probabilities is studied via stochastic ordering considerations.

4.3 Orthogonal polynomial approximations

4.3.1 Approximating general density functions

Let X be an arbitrary random variable with p.d.f. [1] f_X with respect to some measure λ (typically Lebesgue measure on an interval or counting measure on a subset of $\mathbb{Z}$). We assume that the density is unknown and we propose an approximation of the form

$$\widehat{f}_X(x) = \sum_{k=0}^{K} q_k Q_k(x) f_\nu(x). \tag{4.6}$$

where f_ν is the reference or basis density, associated to a probability measure ν absolutely continuous with respect λ. The sequence $\{Q_k, k \geq 0\}$ is made of polynomials, orthonormal with respect to ν in the sense that

$$\langle Q_k, Q_l \rangle_\nu = \int Q_k(x) Q_l(x) \mathrm{d}\nu(x) = \delta_{k,l},\ \ k,l \in \mathbb{N}.$$

This sequence is generated by the Gram–Schmidt orthogonalization procedure which is only possible if ν admits moments of all orders. If additionally there exists $s > 0$ such that

$$\int \mathrm{e}^{sx} \mathrm{d}\nu(x) < \infty$$

then the sequence of polynomials $\{Q_k, k \geq 0\}$ forms an orthonormal basis of $L^2(\nu)$ which is the space of all square integrable functions with respect to ν, see the monograph by Nagy [132, Chapter 7]. Therefore, if $f_X/f_\nu \in L^2(\nu)$ then the polynomial representation of the density of X with respect to ν follows from orthogonal projection, namely we have

$$f_X(x)/f_\nu(x) = \sum_{k=0}^{\infty} \langle f_X/f_\nu, Q_k \rangle_\nu\, Q_k(x). \tag{4.7}$$

[1] This section is written from the perspective of approximating a p.d.f., however the main results also hold if applied to a defective density.

We label the coefficients of the expansion as $\{q_k, k \geq 0\}$, noting that

$$q_k \stackrel{\text{def}}{=} \langle f_X/f_\nu, Q_k \rangle_\nu = \int Q_k(x) f_X(x) \frac{\mathrm{d}\nu(x)}{f_\nu(x)} = \mathbb{E}\left[Q_k(X)\right], \ k \in \mathbb{N},$$

and thus we can rearrange (4.7) to be

$$f_X(x) = \sum_{k=0}^{\infty} q_k Q_k(x) f_\nu(x). \tag{4.8}$$

The approximation (4.6) follows by simply truncating the series to $K+1$ terms.

The Parseval relationship, $\sum_{k=1}^{\infty} q_k^2 = ||f_X/f_\nu||_\nu^2$, ensures that the sequence $\{q_k, k \geq 0\}$ tends toward 0 as k tends to infinity. The accuracy of the approximation (4.6), for a given order of truncation K, depends on how swiftly these coefficients decay. The L^2 loss associated with the approximation of f_X/f_ν is $\sum_{k=K+1}^{\infty} q_k^2$.

Typical choices of reference distributions are ones that belong to a Natural Exponential Family with Quadratic Variance Function (NEF-QVF) which includes the normal, gamma, hyperbolic, Poisson, binomial, and Pascal distributions. This family of distributions is convenient as the associated orthogonal polynomials are classical, see the characterization by Morris [107]. The polynomials are known explicitly, thus we avoid the time-consuming Gram–Schmidt orthogonalization procedure. Furthermore, it has been shown in a paper by Provost [114] that the recovery of unknown densities from the knowledge of the moments of the distribution naturally leads to approximation in terms of the gamma density and Laguerre polynomials when X admits $\mathbb{R}_+$ as support, and in terms of the normal density and Hermite polynomials when X has $\mathbb{R}$ as support.

4.3.2 Approximating densities of positive random variables

To approximate the p.d.f. for positive X, a natural candidate for the reference density is the gamma density. It has been proven to be efficient in practice, see the work of Goffard et al. [67, 68], and Jin et al. [85]. The work of Papush et al. [113] showed that among the gamma, normal and lognormal distribution, the gamma distribution seems to be better suited to model certain aggregate losses. The lognormal distribution is a problematic choice. Even though the orthogonal polynomials are available in a closed form see Asmussen et al. [17], they do not provide a complete orthogonal system of the L^2 space. The case of the inverse Gaussian as basis received a treatment in the work of Nishii [110], where it is shown that the only way to get a complete system of polynomials is by using the Gram–Schmidt orthogonalization procedure. Differentiating the density (as it is done in the case of NEF-QVF) does not lead to an orthogonal polynomial system, and starting from the Laguerre polynomials leads to a system of orthogonal functions which

is not complete. A solution might be to exploit the bi-orthogonality property pointed out in the work of Hassairi and Zarai [77]. To close this review of reference densities, we mention the work of Nadarajah et al. [108] where Weibull and exponentiated exponential distributions are considered as reference density.

The $\mathsf{Gamma}(r, m)$ distribution, where r is the shape parameter and m is the scale parameter, has a p.d.f.

$$f_\nu(x) \stackrel{\text{def}}{=} \gamma(r, m, x) = \frac{x^{r-1}\mathrm{e}^{-\frac{x}{m}}}{\Gamma(r)m^r},\ x \in \mathbb{R}^+,$$

where $\Gamma(\cdot)$ denotes the gamma function.[1] The associated orthonormal polynomials are given by

$$Q_n(x) = (-1)^n \binom{n+r-1}{n}^{-\frac{1}{2}} L_n^{r-1}\Big(\frac{x}{m}\Big) = (-1)^n \left(\frac{\Gamma(n+r)}{\Gamma(n+1)\Gamma(r)}\right)^{-\frac{1}{2}} L_n^{r-1}\Big(\frac{x}{m}\Big),$$

where $\{L_n^{r-1}, n \geq 0\}$ are the generalized Laguerre polynomials,

$$L_n^{r-1}(x) = \sum_{i=0}^{n} \binom{n+r-1}{n-i} \frac{(-x)^i}{i!} = \sum_{i=0}^{n} \frac{\Gamma(n+r)}{\Gamma(n-i+1)\Gamma(r+i)} \frac{(-x)^i}{i!},\ n \geq 0,$$

cf. the classical book by Szegö [131].

Lemma 4.3.3. *If ν is $\mathsf{Gamma}(r, m)$, the polynomial expansion* (4.8) *may be rewritten as*

$$f_X(x) = \sum_{i=0}^{\infty} p_i \gamma(r+i, m, x), \tag{4.9}$$

where

$$p_i = \sum_{k=i}^{\infty} q_k \frac{(-1)^{i+k}}{i!\,(k-i)!} \sqrt{\frac{k!\Gamma(k+r)}{\Gamma(r)}}, \tag{4.10}$$

and the function $\gamma(r, m, x)$ is the p.d.f. of the $\mathsf{Gamma}(r, m)$ distribution.

Proof. If we change the sum in (4.8) from iterating over Laguerre polynomials to iterating over monomials we get

$$f_X(x) = \sum_{k=0}^{\infty} q_k Q_k(x) \gamma(r, m, x) = \sum_{i=0}^{\infty} c_i x^i \gamma(r, m, x)\,,$$

[1] For the distributions in this chapter, we use MATHEMATICA's parametrization, e.g. the exponential and Erlang distributions are $\mathsf{Exp}(\lambda) = \mathsf{Gamma}(1, 1/\lambda)$ and $\mathsf{Erlang}(n, m) = \mathsf{Gamma}(n, 1/m)$.

where

$$c_i = \sum_{k=0}^{\infty} \text{Coefficient}(x^i, q_k Q_k(x)) = \frac{(-1)^i}{m^i i!} \sum_{k=i}^{\infty} q_k (-1)^k \binom{k+r-1}{k}^{-\frac{1}{2}} \binom{k+r-1}{k-i}.$$

We also note that

$$x^i \gamma(r, m, x) = m^i \frac{\Gamma(r+i)}{\Gamma(r)} \gamma(r+i, m, x),$$

so

$$f_X(x) = \sum_{i=0}^{\infty} c_i m^i \frac{\Gamma(r+i)}{\Gamma(r)} \gamma(r+i, m, x) = \sum_{i=0}^{\infty} p_i \gamma(r+i, m, x),$$

where we have set $p_i = c_i m^i \Gamma(r+i)/\Gamma(r)$. □

Remark 4.3.4. For $r = 1$, the formula for p_i, (4.10), simplifies to

$$p_i = \sum_{k=i}^{\infty} q_k (-1)^{i+k} \binom{k}{i}.$$

The expression of the p.d.f. in (4.9) resembles the one of an Erlang mixture, which are extensively used for risk modeling purposes, cf. Willmot and Woo [137], Lee and Lin [96], and Willmot and Lin [136]. However, the p_i's defined in (4.10) do not form a proper probability mass function as they are not always positive. Hence our approximation cannot be considered as an approximation through an Erlang mixture although it enjoys the same features when it comes to approximating the survival function and the slp as shown in the following result.

Proposition 4.3.5. *Letting $\Gamma_u(r, m, x)$ be the s.f. of the* Gamma(r, m) *distribution, we have:*

(i) the s.f. of X is given by

$$\overline{F}_X(x) = \sum_{i=0}^{\infty} p_i \Gamma_u(r+i, m, x) \quad \text{for } x \geq 0\,, \tag{4.11}$$

(ii) the usual slp of X with priority $a \geq 0$ is given by

$$\mathbb{E}\left[(X-a)_+\right] = \sum_{i=0}^{\infty} p_i \left[m(r+i)\Gamma_u(r+i+1, m, a) - a\Gamma_u(r+i, m, a)\right]. \tag{4.12}$$

Proof. If $f_X/f_\nu \in L^2(\nu)$ then Lemma 4.3.3 allows us to write f_X as in (4.9), and integrating this over $[x,\infty)$ yields the formula (4.11). Now consider the usual slp of X, and note that

$$\begin{aligned}\mathbb{E}\left[(X-a)_+\right] &= \int_a^\infty (x-a) f_X(x)\mathrm{d}x \\ &= \int_a^\infty x f_X(x)\mathrm{d}x - a\overline{F}_X(a). \end{aligned} \tag{4.13}$$

Then notice that for every $i \in \mathbb{N}$, we have that

$$\begin{aligned}\int_a^\infty x\,\gamma(r+i,m,x)\mathrm{d}x &= \int_a^\infty x \frac{x^{r+i-1}\mathrm{e}^{-x/m}}{\Gamma(r+i)m^{r+i}}\mathrm{d}x \\ &= m\frac{\Gamma(r+i+1)}{\Gamma(r+i)}\int_a^\infty \frac{x^{r+i}\mathrm{e}^{-x/m}}{\Gamma(r+i+1)m^{r+i+1}}\mathrm{d}x \\ &= m(r+i)\Gamma_u(r+i+1,m,a). \end{aligned} \tag{4.14}$$

Therefore substituting (4.9) and (4.11) into (4.13) and simplifying with (4.14) yields (4.12).

□

Let us make the connection between our approach and Erlang mixture more precise. Assuming that $f_X/f_\nu \in L^2(\nu)$ then taking the Laplace transform on both side of (4.9) yields

$$\mathcal{L}_X(s) = \sum_{i=0}^\infty p_i \left(\frac{1}{1+sm}\right)^{r+i} = \left(\frac{1}{1+sm}\right)^r \mathcal{P}\left(\frac{1}{1+sm}\right),$$

where $\mathcal{P}(z) = \sum_{i=1}^\infty p_i z^i$ denotes the generating function of the sequence of coefficient $\{p_i\ ,\ i \geq 1\}$. Now setting $z = \frac{1}{1+sm}$ allows to express the generating function $\mathcal{P}(z)$ in terms of the Laplace transform of X as

$$\mathcal{P}(z) = z^{-r}\mathcal{L}_X\left(\frac{1-z}{zm}\right).$$

Remark 4.3.6. The approximation through an Erlang mixture consists in approximating the p.d.f. of a nonnegative random variable X as

$$f_X(x) = \sum_{i=1}^\infty p_i \gamma(i,m,x), \text{ for } x \geq 0.$$

The function $\mathcal{P}(z)$ becomes then the probability generating function (p.g.f.) of a counting random variable M, where $p_i = \mathbb{P}(M=i)$, for $i \geq 1$.

The next example is designed to shed light on the link between our polynomial expansion and an Erlang mixture.

Example 4.3.7. *Suppose that we are interested in approximating the p.d.f. of an exponential random variable* $\mathsf{Gamma}(1, \beta)$. *The generating function of the coefficients is then*

$$\mathcal{P}(z) = z^{1-r} \frac{m}{\beta + z(m - \beta)}.$$

If one takes $r = 1$ *and* $m = \beta$ *then* $\mathcal{P}(z) = 1$ *and the polynomial representation reduces to the exponential p.d.f.. Choosing* $0 < m < \beta$ *leads to* $\mathcal{P}(z) = \frac{m/\beta}{1-z(1-m/\beta)}$, *which is the p.g.f. of a geometric random variable; this recovers the fact that an exponential r.v. can be represented by a zero-truncated geometric sum of exponential r.v.s. For* $m > \beta$, *we have* $\mathcal{P}(z) = \frac{m/\beta}{1+z(1-\beta/m)}$ *which is an alternating sequence that decreases geometrically fast. Recall that our polynomial expansion is valid only if* $m > \beta/2$, *which means that when* $\beta/2 < m \leq \beta$ *our approach coincides with the Erlang mixture technique. It does not when* $m > \beta$. *When* $m \leq \beta/2$, *the Erlang mixture representation holds even though the integrability condition, which is a sufficient one, does not hold.*

The coefficients of the polynomials could be derived by differentiating the generating function $\mathcal{P}(z)$ as

$$p_i = \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}z^k} \mathcal{P}(z)\Big|_{z=0} = \text{Coefficient}(k, \text{MaclaurinSeries}(\mathcal{P}(z))),$$

for $i \geq 0$. In practice, the singularities of the function $\mathcal{P}(z)$ at zero mean this procedure is not viable. Instead, the p_i's are approximated by computing the q_k's and truncating their expression (4.10) up to a given order. The practical evaluation of the q_k's is discussed in Section 4.3.9.

A sufficient condition for $f_X/f_\nu \in L^2(\nu)$ is

$$f_X(x) = \begin{cases} \mathcal{O}(\mathrm{e}^{-x/\delta}) & \text{as } x \to \infty \text{ with } m > \delta/2, \\ \mathcal{O}(x^\beta) & \text{as } x \to 0 \text{ with } r < 2(\beta + 1). \end{cases}$$

When X has a well-defined moment generating function one can typically choose r and m so this integrability condition is satisfied. Define the radius of convergence of the random variable X as

$$\rho_X = \sup\{s > 0 \ , \ \mathcal{L}\{f_X\}(-s) < +\infty\},$$

and consider the following result.

Proposition 4.3.8. *Let* X *be a non-negative random variable having a p.d.f.* f_X, *having a well defined moment generating function, then*

$$f_X(x) = \mathcal{O}(\mathrm{e}^{-x\rho_X}) \text{ as } x \to \infty.$$

Proof. The result follows from applying Chernoff bound on the survival function, then De L'Hôpital's rule enables us to conclude. □

The integrability condition is satisfied if $m > \rho_X^{-1}/2$. When we consider heavy-tailed distributions, which is a desirable model characteristic in the applications, the integrability condition cannot be satisfied. The work-around is to use the expansion

$$\mathrm{e}^{-\theta x} f_X(x) = \sum_{k=0}^{\infty} q_k Q_k(x) f_\nu(x),$$

for some $\theta > 0$. Thus, we can use

$$f_X(x) = \mathrm{e}^{\theta x} \sum_{k=0}^{\infty} q_k Q_k(x) f_\nu(x) = \mathrm{e}^{\theta x} \sum_{i=0}^{\infty} p_i \gamma(r+i, m, x)$$

and since, when $1 - m\theta > 0$,

$$\mathrm{e}^{\theta x} \gamma(r+i, m, x) = (1-m\theta)^{-(r+i)} \gamma\Big(r+i, \frac{m}{1-m\theta}, x\Big)$$

we have

$$f_X(x) = \sum_{i=0}^{\infty} p_i (1-m\theta)^{-(r+i)} \gamma\Big(r+i, \frac{m}{1-m\theta}, x\Big) = \sum_{i=0}^{\infty} \widetilde{p}_i \gamma\Big(r+i, \widetilde{m}, x\Big),$$

where

$$\widetilde{p}_i = \frac{p_i}{(1-m\theta)^{r+i}} \quad \text{and} \quad \widetilde{m} = \frac{m}{1-m\theta}\,.$$

Calculating the q_i's and p_i's, topic covered in Section 4.3.9, requires a Laplace transform of $\mathrm{e}^{-\theta x} f_X(x)$ which is given by

$$\mathcal{L}\{\mathrm{e}^{-\theta x} f_X(x)\}(t) = \mathcal{L}\{f_X(x)\}(t+\theta).$$

The method described above is the same (up to some constants) as approximating the exponentially tilted distribution. This idea has been used in Asmussen et al. [17]. It is easily seen that taking $m > \theta^{-1}/2$ implies that $(\mathrm{e}^{-\theta x} f_X(x))/f_\nu(x) \in L^2(\nu)$.

4.3.9 Approximating densities of positive compound distributions

We now focus on variables S_N which admit a compound distribution. Since these distributions have an atom at 0, we put aside this singularity and focus on the defective p.d.f.

$f^+_{S_N}$. The discussion in Sections 4.3.1 and 4.3.2 also apply to defective densities. Namely, if $f^+_{S_N}/f_\nu \in L^2(\nu)$ then the expansion in Lemma 4.3.3 is valid, we have

$$f^+_{S_N}(x) = \sum_{k=0}^{\infty} q_k Q_k(x)\, \gamma(r, m, x) = \sum_{i=0}^{\infty} p_i \gamma(r+i, m, x), \text{ for } x > 0,$$

where $q_k = \int_0^\infty Q_k(x) f^+_{S_N}(x)\mathrm{d}x$ and p_i is given by (4.10). Truncating the first summation yields

$$f^+_{S_N}(x) \approx \sum_{k=0}^{K} q_k Q_k(x)\, \gamma(r, m, x) = \sum_{i=0}^{K} \widehat{p}_i \gamma(r+i, m, x),$$

where $\widehat{p}_i = \sum_{k=i}^K q_k (-1)^{i+k}/[i!\,(k-i)!]\sqrt{k!\Gamma(k+r)/\Gamma(r)}$ for $i \leq K$. The survival function $\overline{F}_{S_N}$ and the slp $\mathbb{E}\left[(S_N - a)_+\right]$ follows from Proposition 4.3.5. If the integrability condition is not satisfied then the exponentially tilted version of the defective p.d.f. is expanded.

Choice of r and m

The parameters for the polynomial approximations are set differently for the light-tailed and heavy-tailed cases. In the light-tailed cases moment matching of order 2 is the natural procedure to set the values of r and m. We need to take into account the result in Proposition 4.3.8 and make sure that $m > \rho_X^{-1}/2$, where $\rho_X = \sup\{s > 0\ ;\ \mathcal{L}\{f^+_{S_N}\}(-s) < \infty\}$. Hence, the value of ρ_X depends on the distributions of N and U. The two distributions we use for modeling the claim frequency N are the *Poisson* and the *Pascal* distributions. The Poisson distribution is denoted by $\mathsf{Poisson}(\lambda)$ with p.m.f.

$$f_N(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{for } k = 0, 1 \ldots,$$

where $\lambda > 0$. We define the Pascal r.v. to be the number of failures counted before observing $\alpha \in \mathbb{N}_+$ successes, denoted $\mathsf{Pascal}(\alpha, p)$ with p.m.f.

$$f_N(k) = \binom{\alpha + k - 1}{k} p^\alpha q^k, \quad \text{for } k = 0, 1, \ldots.$$

Example 4.3.10. *Let N be Poisson distributed, the moment generating function of S_N is then given by*

$$\mathcal{L}_{S_N}(-s) = \exp\left[\lambda(\mathcal{L}_U(-s) - 1)\right].$$

The radius of convergence of S_N coincides with the one of U, $\rho_{S_N} = \rho_U$. In that case, we can set $r = 1$ and $m{=}\lambda\mathbb{E}(U)$ which corresponds to a moment matching procedure of order 1 or set $r = \lambda\mathbb{E}(U)^2/\mathbb{E}\left(U^2\right)$ and $m = \mathbb{E}\left(U^2\right)/\mathbb{E}(U)$ which, in turns, matches the two first moments.

Example 4.3.11. *Let N be Pascal distributed, the moment generating function of S_N is then given by*

$$\mathcal{L}_{S_N}(-s) = \left[\frac{p}{1 - q\mathcal{L}_U(-s)}\right]^{\alpha}.$$

The radius of convergence ρ_{S_N} is the positive solution of the equation $\mathcal{L}_U(-s) = q^{-1}$. We set $r = 1$ and $m = \rho_{S_N}^{-1}$.

The parametrization proposed in Example 4.3.11 is linked to the fact that it leads to the exact defective p.d.f. in the case of a compound Pascal model with exponentially distributed claim sizes. First, we need to introduce the binomial distribution denoted by $\mathsf{Binomial}(n, p)$ with p.m.f.

$$f_N(k) = \binom{n}{k} p^k q^{n-k}, \quad \text{for } k = 0, 1, \ldots, n,$$

where $p \in (0, 1)$, $n \in \mathbb{N}_+$, and $p + q = 1$. The following lemma, adapted from [112], shows a useful correspondence between the Pascal and binomial distributions when used in compound sums with the exponential distribution.

Lemma 4.3.12. *Consider the random sums $X = \sum_{i=1}^{N_1} U_i$ and $Y = \sum_{i=1}^{N_2} V_i$, where*

$$N_1 \sim \mathsf{Pascal}(\alpha, p), \quad U_i \overset{\text{i.i.d.}}{\sim} \mathsf{Gamma}(1, \beta), \quad N_2 \sim \mathsf{Binomial}(\alpha, q), \quad V_i \overset{\text{i.i.d.}}{\sim} \mathsf{Gamma}(1, p^{-1}\beta),$$

where $p \in (0, 1)$, $\alpha \in \mathbb{N}_+$, $p + q = 1$, and where $\beta > 0$. Then we have $X \overset{\mathcal{D}}{=} Y$.

Proof. Both X and Y have the same Laplace transform, so $X \overset{\mathcal{D}}{=} Y$. □

Corollary 4.3.13. *Consider the compound sum $S_N = \sum_{i=1}^{N} U_i$ where $N \sim \mathsf{Pascal}(\alpha, p)$ and the $U_i \overset{\text{i.i.d.}}{\sim} \mathsf{Gamma}(1, \beta)$. Then the s.f. of S_N is given by*

$$\overline{F}_{S_N}(x) = \sum_{i=1}^{\alpha} \binom{\alpha}{i} q^i p^{\alpha-i} \Gamma_u\left(i, p^{-1}\beta, x\right),$$

and its slp is given by

$$\mathbb{E}\left[(S_N - a)_+\right] = \sum_{i=1}^{\alpha} \binom{\alpha}{i} q^i p^{\alpha-i} \left[\frac{i\beta}{p}\Gamma_u\Big(i + 1, p^{-1}\beta, a\Big) - a\Gamma_u\left(i, p^{-1}\beta, a\right)\right].$$

Proof. By Lemma 4.3.12 we treat S_N as if defined for $N \sim \mathsf{Binomial}(\alpha, q)$ and with $U_i \overset{\text{i.i.d.}}{\sim} \mathsf{Gamma}(1, p^{-1}\beta)$. The result follows by noting $S_n = U_1 + \cdots + U_n \sim \mathsf{Gamma}(n, p^{-1}\beta)$. □

One conclusion of Corollary 4.3.13 is that the exact solution coincides with our approximation when $r = 1$ and $m = p^{-1}\beta$ (and with $K \geq \alpha - 1$). Note that $p\beta^{-1}$ is the solution of the equation $\mathcal{L}_U(-s) = q^{-1}$ which is consistent with the parametrization proposed in Example 4.3.11.

In the heavy-tailed cases (i.e. when exponential tilting is required) we set $\theta = 1$, $m = \theta/2 = 1/2$ (at the lower limit for m; this gives $\widetilde{m} = 1$), and choose $r = \mathbb{E}[U]$.

Computation of the q_k's

The inherent challenge of the implementation of the polynomial method remains the evaluation of the coefficients $\{q_k, k \geq 0\}$. Recall that

$$q_k = \int_0^\infty Q_k(x) f_{S_N}^+(x) \mathrm{d}x, \ k \geq 0.$$

We propose an evaluation based on the Laplace transform $\mathcal{L}\{f_{S_N}^+\}$. Define the generating function of the sequence $\{q_k c_k, k \geq 0\}$ as $\mathcal{Q}(z) = \sum_{k=0}^\infty q_k c_k z^k$, where

$$c_k = \left(\frac{\Gamma(k+r)}{\Gamma(k+1)\Gamma(r)} \right)^{1/2}, \quad \text{for } k \geq 0\,.$$

The following result establishes a link between the Laplace transform of $f_{S_N}^+$ and the generating function $\mathcal{Q}(z)$.

Proposition 4.3.14. *Assume that $f_{S_N}^+ / f_\nu \in L^2(\nu)$, then we have*

$$\mathcal{Q}(z) = (1+z)^{-r} \mathcal{L}\{f_{S_N}^+\}\Big[\frac{-z}{m(1+z)}\Big]. \tag{4.15}$$

Proof. As $f_{S_N}^+ / f_\nu \in L^2(\nu)$, the polynomial representation of $f_{S_N}^+$ follows from the application of Lemma 4.3.3 with

$$f_{S_N}^+(x) = \sum_{k=0}^\infty \sum_{i=0}^k q_k \frac{(-1)^{i+k}}{i!\,(k-i)!} \sqrt{\frac{k!\Gamma(k+r)}{\Gamma(r)}} \gamma(r+i, m, x). \tag{4.16}$$

Taking the Laplace transform in (4.16) yields

$$\begin{aligned}
\mathcal{L}\{f^+_{S_N}\}(s) &= \left(\frac{1}{1+sm}\right)^r \sum_{k=0}^{\infty} q_k \sum_{i=0}^{k} (-1)^{k+i} \left(\frac{\Gamma(k+r)}{\Gamma(k+1)\Gamma(r)}\right)^{1/2} \binom{k}{i} \left(\frac{1}{1+sm}\right)^i \\
&= \left(\frac{1}{1+sm}\right)^r \sum_{k=0}^{\infty} q_k c_k (-1)^k \sum_{i=0}^{k} \binom{k}{i} \left(\frac{-1}{1+sm}\right)^i \\
&= \left(\frac{1}{1+sm}\right)^r \sum_{k=0}^{\infty} q_k c_k (-1)^k \left(\frac{sm}{1+sm}\right)^k \\
&= \left(1 - \frac{sm}{1+sm}\right)^r \mathcal{Q}\left(-\frac{sm}{1+sm}\right).
\end{aligned}$$

Thus (4.15) follows from letting $z = -sm/(1+sm)$. □

The Laplace transform of the defective p.d.f. $f^+_{S_N}$ is given by

$$\mathcal{L}\{f^+_{S_N}\}(s) = \mathcal{L}_{S_N}(s) - \mathbb{P}(N = 0).$$

The coefficients of the polynomials can be derived after differentiation of the generating function $\mathcal{Q}(z)$ as

$$q_k = \frac{1}{c_k} \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}z^k} \mathcal{Q}(z)\Big|_{z=0} = \frac{1}{c_k} \text{Coefficient}(k, \text{MaclaurinSeries}(\mathcal{Q}(z))).$$

4.4 Laplace transform inversion approximations

We present in this section a method inspired from the work of Abate and Whitt [2] to recover the survival function of a compound distribution from the knowledge of its Laplace transform. The methodology is further applied to the computation of slp's by taking advantage of the connection between the slp of S_N and the survival function of the equilibrium distribution of S_N. We begin by stating some useful transform relations, then discuss the general Laplace inversion framework that we will use, and will apply the method to the compound distribution problem.

4.4.1 Numerical Laplace inversion

A function f can be recovered from its Laplace transform by a standard Bromwich integral. We assume $f : \mathbb{R}_+ \to \mathbb{R}_+$, is a measurable function with locally bounded variation. To

define the Bromwich integral, first select a $\gamma > 0$ (we discuss this choice later), then

$$f(x) = \frac{2\mathrm{e}^{\gamma x}}{\pi} \int_0^\infty \cos(xs) \Re\left[\mathcal{L}\{f\}(\gamma + is)\right] \mathrm{d}s.$$

We apply a basic numerical integration system to this integral by first *discretizing* the integral and then *truncating* the resulting infinite sum. In both steps, we follow the steps of Abate and Whitt [2].

Discretization

We will use a semi-infinite trapezoidal rule, despite the apparent simplicity of the method. With a grid size $h > 0$, this discretization yields

$$f(x) \approx f_{\mathrm{disc}}(x) \stackrel{\mathrm{def}}{=} \frac{2\mathrm{e}^{\gamma x}}{\pi} \cdot h \left\{ \frac{1}{2}\mathcal{L}\{f\}(\gamma) + \sum_{j=1}^{\infty} \cos(x \cdot hj) \Re\left[\mathcal{L}\{f\}(\gamma + i \cdot hj)\right] \right\},$$

since $\Re\left[\mathcal{L}\{f\}(\gamma)\right] = \mathcal{L}\{f\}(\gamma)$. We simplify this by choosing $h = \pi/(2x)$ and $\gamma = a/(2x)$ for an $a > 0$, achieving

$$f_{\mathrm{disc}}(x) = \frac{\mathrm{e}^{a/2}}{2x}\mathcal{L}\{f\}\left(\frac{a}{2x}\right) + \frac{\mathrm{e}^{a/2}}{x}\sum_{k=1}^{\infty}(-1)^k \Re\left[\mathcal{L}\{f\}\left(\frac{a + i \cdot 2\pi k}{2x}\right)\right]. \tag{4.17}$$

From Theorem 5.5.1 of [119] we have that the *discretization error* (also called *sampling error*) is simply

$$f_{\mathrm{disc}}(x) - f(x) = \sum_{k=1}^{\infty} \mathrm{e}^{-ak} f\Big((2k+1)x\Big). \tag{4.18}$$

In particular, if $0 \le f(x) \le 1$, then

$$f_{\mathrm{disc}}(x) - f(x) \le \frac{\mathrm{e}^{-a}}{1 - \mathrm{e}^{-a}}. \tag{4.19}$$

There are no absolute value signs here — the discretization introduces a systematic overestimate of the true function value. Also, (4.18) implies a should be as large as possible (limited eventually by finite-precision computation). The benefit of knowing this result is slightly offset by the requirement that h and γ now be functions of x rather than constants.

Truncation

Due to the infinite series, the expression in (4.17) cannot be directly computed, thus it has to be truncated. The arbitrary-seeming choice of h and γ in Section 4.4.1 not only allows for calculation of the discretization error, but also benefits the truncation step. This is because the sum in (4.17) is (nearly) of alternating sign, and thus *Euler series acceleration* can be applied to decrease the truncation error. Define for $\ell = 1, 2, \ldots$

$$s_\ell(x) \stackrel{\text{def}}{=} \frac{\mathrm{e}^{a/2}}{2x}\mathcal{L}\{f\}\left(\frac{a}{2x}\right) + \frac{\mathrm{e}^{a/2}}{x}\sum_{k=1}^{\ell}(-1)^k \Re\left[\mathcal{L}\{f\}\left(\frac{a + i\cdot 2\pi k}{2x}\right)\right].$$

Then, for some positive integers M_1 and M_2,

$$f(x) \approx f_{\text{disc}}(x) \approx f_{\text{approx}}(x) \stackrel{\text{def}}{=} \sum_{k=0}^{M_1}\binom{M_1}{k}2^{-M_1}s_{M_2+k}(x)\,. \tag{4.20}$$

4.4.2 Estimators of survival function and stop-loss premium for compound distributions

For a random sum S_N, we consider using the technique above to evaluate the s.f. $\overline{F}_{S_N}$ and the slp's from their Laplace transform. We invert $\mathcal{L}\{\overline{F}_{S_N}\}$, but note that inverting $\mathcal{L}\{F_{S_N}\}$ produces almost identical results.

This inversion easily gives estimates of $\overline{F}_{S_N}$, though evaluating the slp's requires extra thought. As noted in Dufresne et al. [54], we have that

$$\mathbb{E}\left[(S_N - d)_+\right] = \mathbb{E}(S_N)\overline{F_{S_N^\top}}(d), \tag{4.21}$$

where $S_N^\top$ is a random variable under the *equilibrium distribution* with density

$$f_{S_N^\top}(x) = \begin{cases} \overline{F}_{S_N}(x)/\mathbb{E}(S_N), & \text{for } x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and Laplace transform

$$\mathcal{L}_{S_N^\top}(s) = \frac{1 - \mathcal{L}_{S_N}(s)}{s\mathbb{E}(S_N)}.$$

The slp is then obtained, replacing in (4.21) the s.f. of $S_N^\top$ by its approximation through (4.20).

4.5 Numerical illustrations

In this section, we illustrate the performance of the two proposed numerical procedures. Section 4.5.1 focuses on approximating the s.f. and the slp associated to aggregated claim sizes, while Section 4.5.6 considers the approximation of the finite-time ruin probability with no initial reserves using formula (4.5).

For each test case we compare the orthogonal polynomial approximation, the Laplace inversion approximation, and for the crude Monte Carlo approximation. For the cases when U is gamma distributed, we use the fact that S_n is Erlang distributed to produce an approximate distribution for S_N by truncating N to be less than some large level.

The parameters for the polynomial approximations has been discussed in Section 4.3.9, the calibration is depending on the assumptions over the claim frequency and claim sizes distribution. The parameters for the Laplace inversion technique are set to $M_1 = 11$, $M_2 = 15$ and $a = 18.5$ following the example of Rolski et al. [119, Chapter 5, Section 5]; note, this choice of a implies that the discretization error is less than 10^{-8}, as derived from (4.19).

In each plot, the first subplot shows the estimates each estimator produces, and the second shows the *approximate absolute error*. We define this, for estimator $i \in \{1, \ldots, I\}$, as

$$\begin{aligned}\text{ApproximateAbsoluteError}(\widehat{f}_i, x) &:= \widehat{f}_i(x) - \text{Median}\Big\{\widehat{f}_1(x), \ldots, \widehat{f}_I(x)\Big\} \\ &\approx \widehat{f}_i(x) - f(x) =: \text{AbsoluteError}(\widehat{f}_i, x)\,.\end{aligned}$$

When the different estimators cross each other, the median obtains an unrealistically jagged character. We therefore use as reference a slightly smoothed version of the median, achieved in Mathematica using `GaussianFilter[Medians, 2]`. As noted earlier, all of the code used is available online [66].

4.5.1 Survival function and stop-loss premium computations

To ensure both estimators were implemented correctly, we applied the estimators to the case where $N \sim \mathsf{Pascal}(\alpha = 10, p = 3/4)$ and $U \sim \mathsf{Gamma}(r = 1, m = 1/6)$. Corollary 4.3.13 tells us the orthogonal approximation (with $r = 1$, $m = \lambda/p = 2/9$ and $K = \alpha - 1 = 9$) is equivalent to the true function, which we verified, and the Laplace inversion errors in Tables 4.1 and 4.2 are acceptably small.

Table 4.1: Relative errors for the Laplace inversion s.f. estimator

x	0.5	1	1.5	2	2.5
Error	7.27e-7	1.92e-6	5.86e-6	1.78e-5	4.01e-5

Table 4.2: Relative errors for the Laplace inversion slp estimator

a	0.5	1	1.5	2	2.5
Error	8.68e-7	2.27e-6	5.92e-6	1.12e-5	-2.12e-5

Test 4.5.2. $N \sim \mathsf{Poisson}(\lambda = 2)$, *and* $U \sim \mathsf{Gamma}(r = 3/2, m = 1/3)$

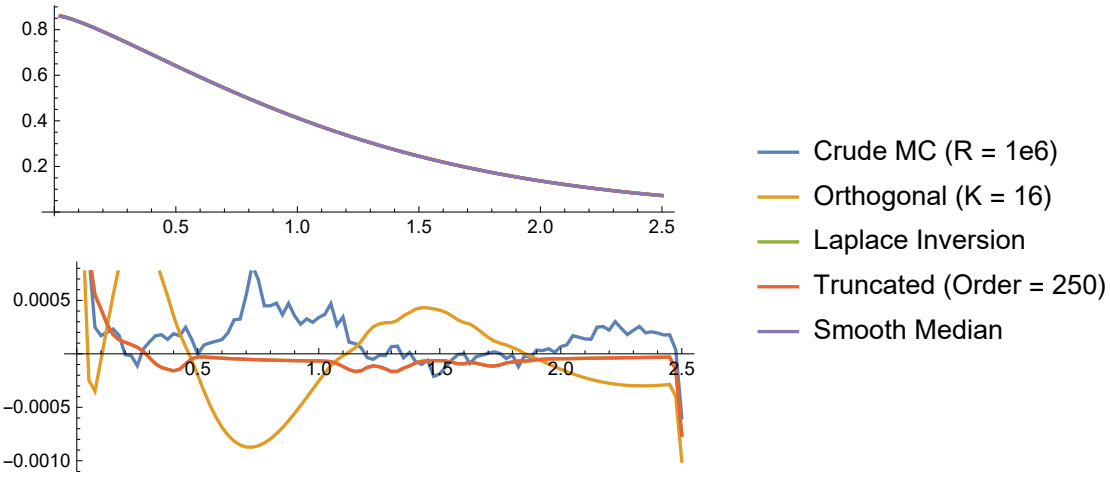


Figure 4.1: Survival function estimates and approximate absolute error.

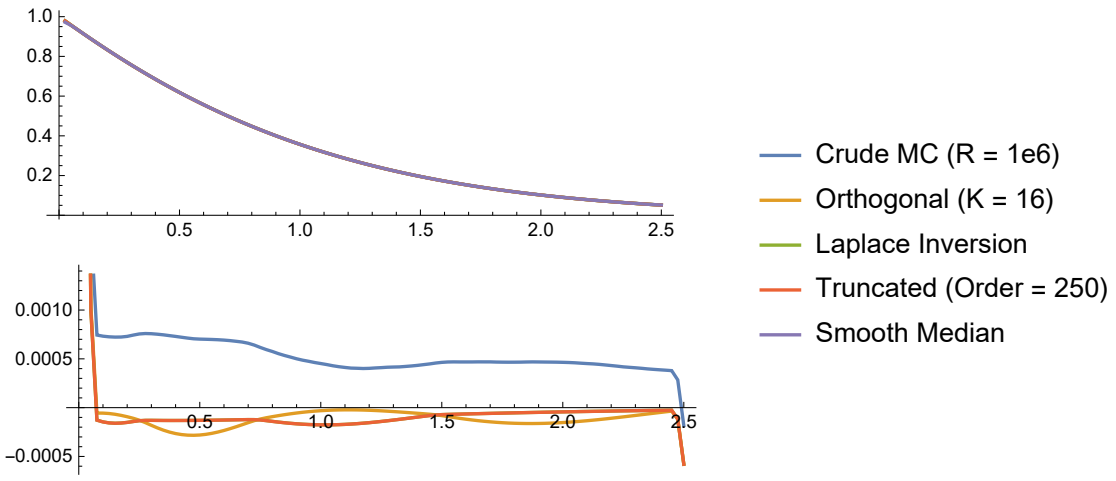


Figure 4.2: Stop-loss premium estimates and approximate absolute error.

Test 4.5.3. $N \sim \mathsf{Pascal}(\alpha = 10, p = 1/6)$, *and* $U \sim \mathsf{Gamma}(r = 3/2, m = 1/75)$

This test case (up to the scaling constant) has been considered by Jin et al. [85, Example 3]. In the plots for this test case, the orthogonal estimator, the Laplace inversion estimator, and the truncated estimator all give the same values and hence are hidden underneath the red line for the truncated estimator.

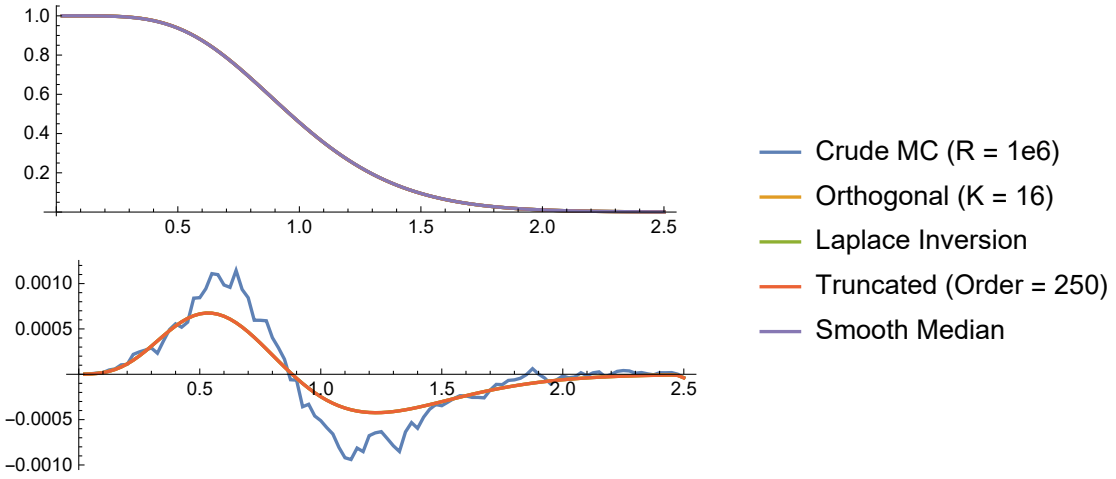


Figure 4.3: Survival function estimates and approximate absolute error.

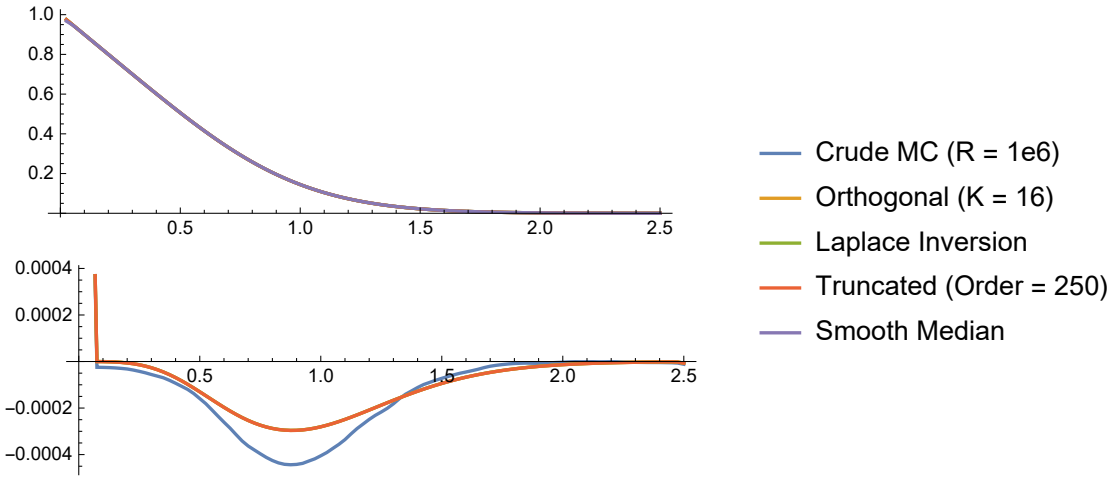


Figure 4.4: Stop-loss premium estimates and approximate absolute error.

Test 4.5.4. *$N \sim \text{Poisson}(\lambda = 4)$, and $U \sim \text{Pareto}(a = 5, b = 11, \theta = 0)$*

The survival function for U, given $x \geq \theta = 0$, is

$$\overline{F}_U(x) = \left(1 + \frac{x - \theta}{a}\right)^{-b} = \left(1 + \frac{x}{5}\right)^{-11}.$$

We note that the Laplace inversion estimator breaks down for small values of x or a in this test case. The specific error given is an "out of memory" exception when MATHEMATICA is attempting to do some algebra with extremely large numbers. It is unclear whether a different implementation or selection of parameters would fix this behaviour.

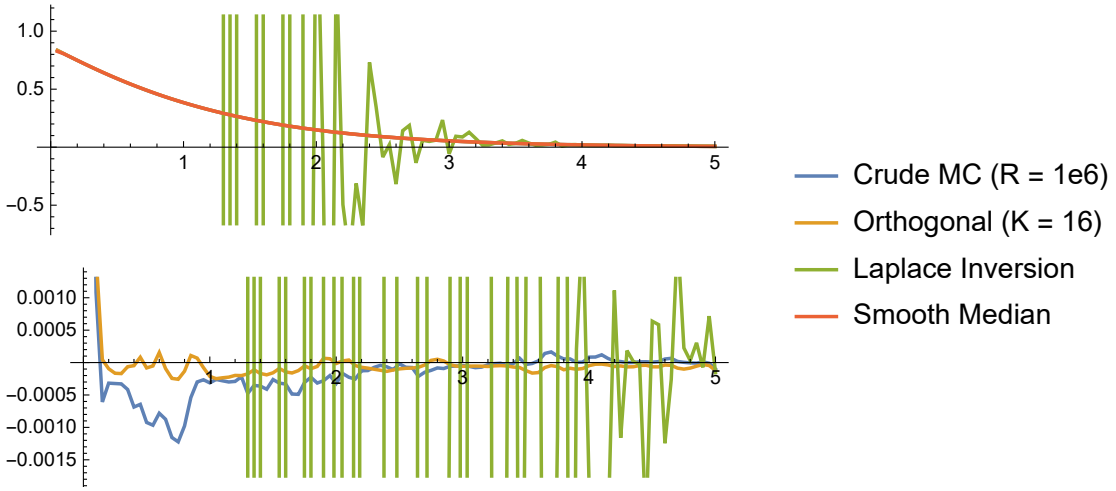


Figure 4.5: Test 3: Survival function estimates and approximate absolute error.

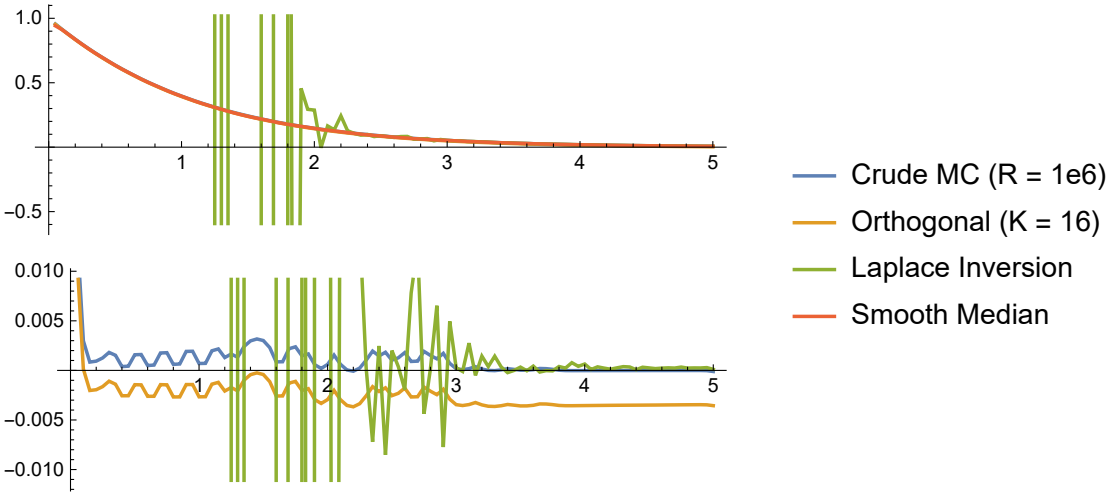


Figure 4.6: Test 3: Stop-loss premium estimates and approximate absolute error.

Test 4.5.5. $N \sim \text{Pascal}(\alpha = 2, p = 1/4)$, *and* $U \sim \text{Weibull}(\beta = 1/2, \lambda = 1/2)$

The survival function for U, given $x \geq 0$, is

$$\overline{F}_U(x) = \exp\left\{-\left(\frac{x}{\lambda}\right)^{\beta}\right\} = \exp\left\{-\sqrt{2x}\right\} .$$

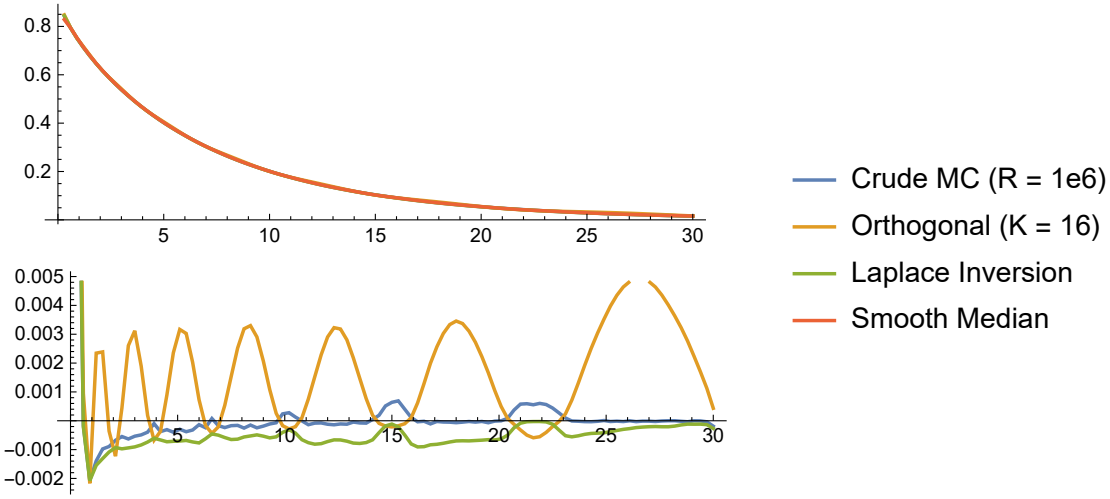


Figure 4.7: Test 4: Survival function estimates and approximate absolute error.

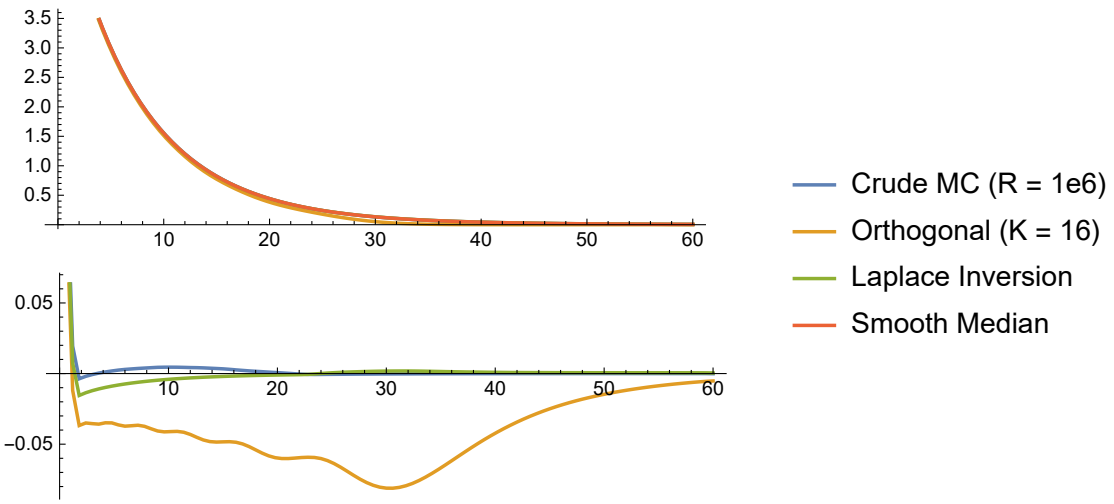


Figure 4.8: Test 4: Stop-loss premium estimates and approximate absolute error.

4.5.6 Finite-time ruin probability with no initial reserve

In this paper we have used common random numbers for all crude Monte Carlo estimators to smooth their estimates. However in the case of ruin probabilities, the distribution from which we are simulating Poisson(λt) is changing for each point, so the technique cannot be applied in the traditional way. Thus the crude Monte Carlo estimates in following plots are not as smooth as above.

Test 4.5.7. $\lambda = 4$ *and* $U \sim \mathsf{Gamma}(r = 2, m = 2)$ *and* $c = 1$

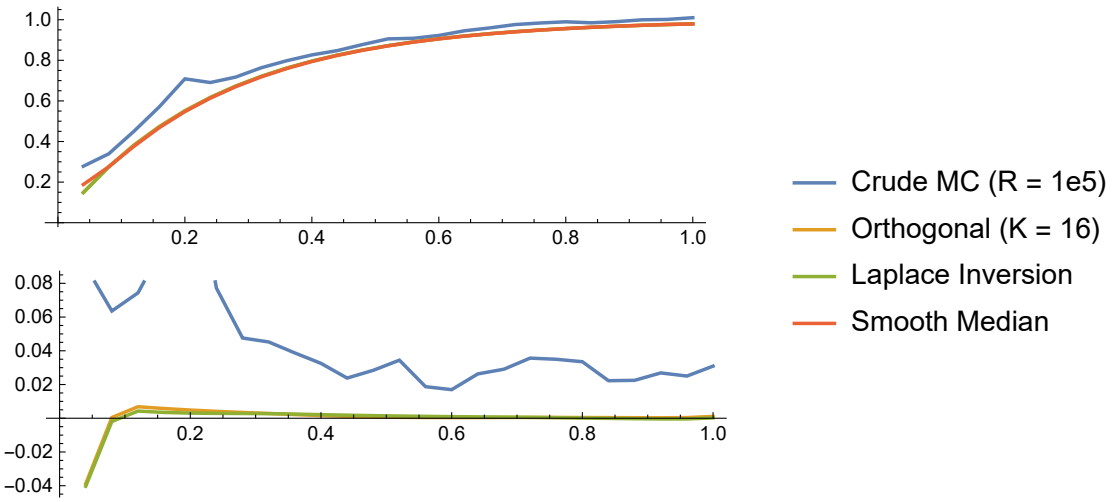


Figure 4.9: Test 5: Ruin probability $\psi(0,t)$ estimates and approximate absolute error.

Test 4.5.8. $\lambda = 2$ *and* $U \sim \mathsf{Pareto}(a = 5, b = 11, \theta = 0)$ *and* $c = 1$

See the discussion of Test 3 for a description of the Laplace inversion estimator's poor behaviour when Pareto variables are involved.

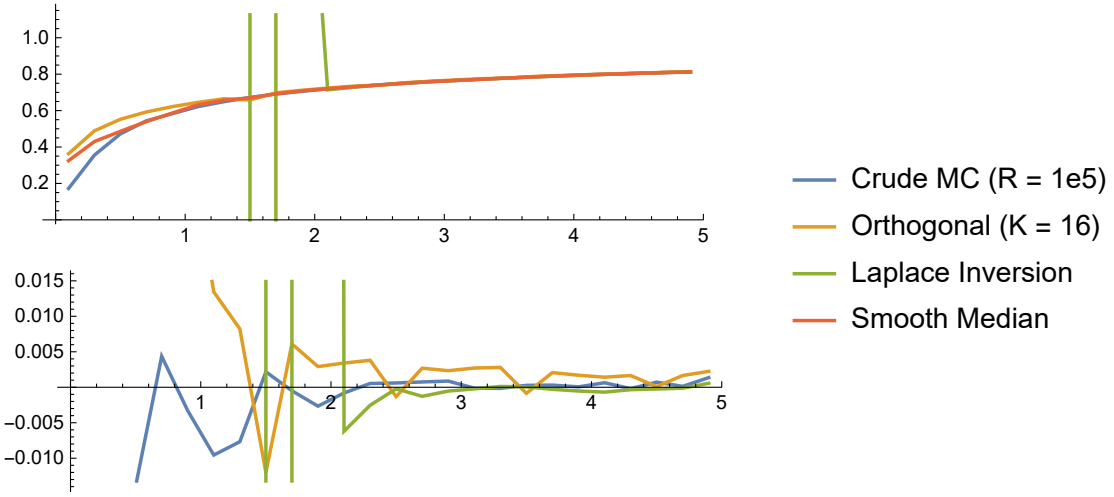


Figure 4.10: Test 6: Ruin probability $\psi(0,t)$ estimates and approximate absolute error.

4.5.9 Concluding remarks

The orthogonal polynomial method has performed well across all the test cases studied. The accuracy is acceptable even with a rather small order of truncation $K = 16$. It produces an approximation having an analytical expression, which is desirable, and in a timely manner. The precision may be improved by adding more terms in the expansions. The main drawback is probably the need for a parametrization tailored to the case studied.

The Laplace transform inversion method yields outstanding result in terms of accuracy. It failed to provide a stable approximation for Pareto distributed claim sizes. The parametrization is automatic and seems to fit the different case studied (except the Pareto one).

Both of the methods are easy to implement and beat a simple truncation or a crude Monte-Carlo approach, which is the main conclusion of our work.

Chapter 5

Tail asymptotics of light-tailed Weibull-like sums

5.1 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. [59] or [13]). 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. [84]. 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 5.8). 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 [120] which was later followed up by the mathematically deeper and somewhat general study of Balkema, Klüppelberg, & Resnick [24], 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 [120] 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 5.4).

5.2 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. [13, 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,

$$\Lambda\Big(x + e(x)y\Big) \;\sim\; \Lambda(x) + y \tag{5.1}$$

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, (5.1) is a necessary and sufficient condition for F to be in $\mathsf{GMDA}(e)$, the maximum domain of attraction of the Gumbel distribution [59]. 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,

$$\begin{aligned}\mathbb{P}(X_1 + X_2 > x) &= (f \top \overline{F})(x) = \int_{-\infty}^{\infty} \lambda(z) \exp\Big\{-\Lambda(z) - \Lambda(x - z)\Big\}\, \mathrm{d}z \\ &= \int_{-\infty}^{\infty} \frac{e(x/2)}{e\Big(x/2 + e(x/2)y\Big)} \exp\left\{-\Lambda\Big(\frac{x}{2} + e\Big(\frac{x}{2}\Big)y\Big) - \Lambda\Big(\frac{x}{2} - e\Big(\frac{x}{2}\Big)y\Big)\right\} \mathrm{d}y, \end{aligned} \tag{5.2}$$

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

$$\Lambda\Big(\frac{x}{2} \pm e\Big(\frac{x}{2}\Big)y\Big) \sim \Lambda\left(\frac{x}{2}\right) \pm y + \frac{y^2}{4\sigma^2\left(\frac{x}{2}\right)}. \tag{5.3}$$

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 (5.1). Combining (5.3) and the self-neglecting property with (5.2) gives us

$$\begin{aligned}\mathbb{P}(X_1 + X_2 > x) &\sim \int_{-\infty}^{\infty} 1 \cdot \exp\left\{-2\Lambda\left(\frac{x}{2}\right) - \frac{y^2}{2\sigma^2(\frac{x}{2})}\right\} \mathrm{d}y \\ &= \sqrt{2\pi\sigma^2(x/2)}\, \exp\Big\{-2\Lambda(x/2)\Big\}\,. \end{aligned} \tag{5.4}$$

In summary, rewriting (5.4) gives

$$\overline{F^{*2}}(x) \;=\; \mathbb{P}(X_1 + X_2 > x) \;\sim\; \overline{F}(x/2)^2 \sqrt{\pi \frac{\lambda(x/2)^2}{\lambda'(x/2)}}\,. \tag{5.5}$$

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 5.2.1. The procedure to arrive at (5.5) 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 5.2.2. 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. [15, V.1, VI.2]. $\diamond$

5.3 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

$$f_i(x) \;\sim d_i x^{\alpha_i + \beta - 1} \mathrm{e}^{-c_i x^\beta}\,, \quad x \to \infty,\ i = 1, 2 \tag{5.6}$$

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 (5.6), we define

$$\eta = c_1^{1/(\beta-1)} + c_2^{1/(\beta-1)}, \quad \theta_1 = c_2^{1/(\beta-1)}/\eta, \quad \theta_2 = c_1^{1/(\beta-1)}/\eta, \quad \kappa = \frac{\eta^{\beta-1}}{\beta c_1 c_2}\,. \tag{5.7}$$

Note that

$$\overline{F}_i(x) \sim \frac{d_i}{\beta c_i} x^{\alpha_i} \mathrm{e}^{-c_i x^\beta} \tag{5.8}$$

(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 5.3.1. *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*

$$e_1(\theta_1 x) \sim e_2(\theta_2 x) \sim \frac{\kappa}{x^{\beta-1}} = \frac{1}{\beta c_1 \theta_1^{\beta-1} x^{\beta-1}} = \frac{1}{\beta c_2 \theta_2^{\beta-1} x^{\beta-1}}. \tag{5.9}$$

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

$$\begin{aligned} c &= c_1\theta_1^{\beta-1}\theta_1 + c_2\theta_2^{\beta-1}\theta_2 = \frac{c_1c_2\theta_1}{\eta^{\beta-1}} + \frac{c_1c_2\theta_2}{\eta^{\beta-1}} = \frac{c_1c_2}{\eta^{\beta-1}} \\ &< \frac{c_1c_2}{\left[c_2^{1/(\beta-1)}\right]^{\beta-1}} = c_1. \end{aligned} \tag{5.10}$$

Similarly, $c < c_2$. □

Lemma 5.3.2. $(1+h)^\beta = 1 + h\beta + \dfrac{h^2}{2}\beta(\beta-1)\omega(h)$ *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 5.3.5 below.

Theorem 5.3.3. *Under assumption* (5.6), $\mathbb{P}(X_1 + X_2 > x) \sim kx^\gamma \mathrm{e}^{-cx^\beta}$ *as* $x \to \infty$, *where* $\gamma = \alpha_1 + \alpha_2 + \beta/2$ *and* $k = d_1d_2\theta_1^{\alpha_1}\theta_2^{\alpha_2}\kappa\eta^{1-\beta}(2\pi\sigma^2)^{1/2}/\beta$, *with* $\theta_1, \theta_2, \kappa, \eta$ *as in* (5.6), *the constant* c *as in Lemma 5.3.1, 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 5.3.4. 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 5.2. $\diamond$

Proof. By Lemma 5.3.1, 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

$$\mathbb{P}\Big(X_1 + X_2 > x, a_- x < X_1 < a_+ x\Big) \ = \ \int_{a_- x}^{a_+ x} f_1(z)\overline{F}_2(x-z)\,\mathrm{d}z \tag{5.11}$$

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 (5.9) holds with equality.

Now

$$\begin{aligned}\mathbb{P}\Big(X_1 + X_2 > x, a_- < X_1 < a_+ x\Big) \ &= \ \int_{a_- x}^{a_+ x} f_1(z)\overline{F}_2(x-z)\,\mathrm{d}z \\ &= \ \int_{a_- x}^{a_+ x} \frac{d_1 d_2}{\beta c_2} z^{\alpha_1+\beta-1}(x-z)^{\alpha_2} \exp\Big\{-c_1 z^\beta - c_2(x-z)^\beta\Big\}\,\mathrm{d}z. \end{aligned} \tag{5.12}$$

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

$$c_1 z^\beta + c_2(x-z)^\beta \ = \ c_1\theta_1^\beta x^\beta\Big(1 + h_1(x,y)\Big)^\beta + c_2\theta_2^\beta x^\beta\Big(1 - h_2(x,y)\Big)^\beta \tag{5.13}$$

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 5.3.2 and using (5.9), the first order term of (5.13) 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 $\omega_1(x,y) = \omega\Big(h_1(x,y)\Big)$, $\omega_2(x,y) = \omega\Big(-h_2(x,y)\Big)$, (5.12) becomes

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

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

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

Similarly, $-h_2(x,y) \geq -a_+/\theta_2 - 1 > 1$. Using Lemmas 5.3.1 and 5.3.2 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 (5.11) is the same as that of

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

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

Corollary 5.3.5. *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*

$$\overline{F^{*n}}(x) \;=\; \mathbb{P}(S_n > x) \;\sim\; k(n) x^{\alpha(n)} \mathrm{e}^{-c(n)x^\beta}\,, \tag{5.14}$$

$$f^{*n}(x) \;\sim\; \beta c(n) k(n) x^{\alpha(n)+\beta-1} \mathrm{e}^{-c(n)x^\beta} \tag{5.15}$$

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

$$k(n) = \frac{d^n}{\beta c} \Big[\frac{2\pi}{\beta(\beta-1)c}\Big]^{(n-1)/2} n^{\frac{1}{2}(\beta - n(2\alpha+\beta)-1)}. \tag{5.16}$$

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 5.3.3 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 5.3.3 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

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

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 $n \geq 2$

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

□

Note that (5.15) is already in Rootzén [120] (see his equations (6.1)–(6.2)). We point out later that the assumptions on the density can be relaxed to $\overline{F}(x) \sim k x^{\alpha} \mathrm{e}^{-cx^{\beta}}$ where $k = d/c\beta$.

5.4 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

$$\lim_{x\to\infty} \lambda(x) = \infty, \tag{5.17}$$

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

$$\lambda'\Big(x + y/\sqrt{\lambda'(x)}\Big) \;\sim\; \lambda'(x). \tag{5.18}$$

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

$$\lim_{x\to\infty} \frac{\gamma\Big(x + y/\sqrt{\lambda'(x)}\Big)}{\gamma(x)} = 1. \tag{5.19}$$

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 (5.18) and (5.19) 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 5.4.1. (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*

$$q_1 + q_2 = x, \quad \lambda_1(q_1) = \lambda_2(q_2) \tag{5.20}$$

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 5.4.1 is in Appendix 5.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 5.4.2. Letting $\tau(y) = \lambda_1^{\leftarrow}(y) + \lambda_2^{\leftarrow}(y)$, the solution of (5.20) can be written

$$q_1(x) \;=\; \lambda_1^{\leftarrow}\Big(\tau^{\leftarrow}(x)\Big), \quad q_2(x) \;=\; \lambda_2^{\leftarrow}\Big(\tau^{\leftarrow}(x)\Big) \tag{5.21}$$

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

5.5 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 5.5.1. *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. 5.5.1 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. 5.5.1 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. 5.5.1 to the following slightly more general form.

Proposition 5.5.2. *Let $\{X_i\}_{i=1}^n$ be independent random variables with density*

$$f_{X_i}(x) = \frac{\beta k^{\gamma_i/\beta} x^{\gamma_i - 1} \mathrm{e}^{-kx^\beta}}{\Gamma(\gamma_i/\beta)} \quad \textit{for } x > 0$$

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

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

5.6 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 5.6.1. *As $x \to \infty$, it holds that*

$$\widehat{F}\big[\lambda(x)\big] \;\sim\; \sqrt{\frac{2\pi}{\lambda'(x)}}\gamma(x)\mathrm{e}^{(\beta-1)x^\beta}, \tag{5.22}$$

$$\mathbb{E}_{\lambda(x)}\, X \;\sim\; x. \tag{5.23}$$

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

$$\sqrt{\lambda'(x)}(X - x) = \sqrt{\beta(\beta-1)x^{\beta-2}}(X - x) \;\xrightarrow{\mathcal{D}}\; N\Big(0, 1\Big)\,. \tag{5.24}$$

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 5.4.1 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

$$\lim_{x\to\infty} \frac{\gamma'(x)}{\sqrt{\lambda'(x)}\gamma(x)} \quad = \; 0. \tag{5.25}$$

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

$$\begin{aligned}
\mathbb{E}[X^k \mathrm{e}^{\lambda(x)X}] &= g(0) + \int_0^\infty g'(z)\overline{F}(z)\mathrm{d}z \\
&= \mathbb{I}\{k=0\} + \int_0^{c_1 x} g'(z)\overline{F}(z)\mathrm{d}z + \int_{c_1 x}^\infty g'(z)\overline{F}(z)\mathrm{d}z \\
&= \mathcal{O}(\mathrm{e}^{\widetilde{c}_1 x^\beta}) + \int_{c_1 x}^\infty \Big[kz^{k-1} + \lambda(x) z^k\Big] \mathrm{e}^{\lambda(x)z} \frac{\gamma(z)}{\lambda(z)} \mathrm{e}^{-z^\beta}\mathrm{d}z
\end{aligned} \tag{5.26}$$

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

Consider integrals of the form $\int_{c_1 x}^\infty z^k \mathrm{e}^{\lambda(x)z}\gamma(z)\mathrm{e}^{-z^\beta}\mathrm{d}z$ 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 5.2 and 5.3, 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 (5.26) we get

$$\begin{aligned}
\mathbb{E}[X^k \mathrm{e}^{\lambda(x)X}] &= \mathcal{O}(\mathrm{e}^{\widetilde{c}_1 x^\beta}) + k\int_{c_1 x}^\infty z^{k-1} \frac{\gamma(z)}{\lambda(z)} \mathrm{e}^{\lambda(x)z - z^\beta}\mathrm{d}z \qquad (5.27) \\
&\quad + \lambda(x) \int_{c_1 x}^\infty z^k \frac{\gamma(z)}{\lambda(z)} \mathrm{e}^{\lambda(x)z - z^\beta}\mathrm{d}z \\
&= \mathcal{O}(\mathrm{e}^{\widetilde{c}_1 x^\beta}) + \sqrt{\frac{2\pi}{\lambda'(x)}} \gamma(x) \mathrm{e}^{(\beta-1)x^\beta} \Big(x^k + \frac{k}{\lambda(x)} x^{k-1}\Big), \qquad (5.28)
\end{aligned}$$

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 (5.22)–(5.23) 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

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

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 (5.22) gives $\mathbb{P}_{\lambda(x)}\Big(\sqrt{\lambda'(x)}(X-x) \le v\Big) \to \Phi(v)$ which is (5.24). □

Remark 5.6.2. Asymptotic normality for the general case $\overline{F}(x) = \mathrm{e}^{-\psi(x)}$ similar to the result of Proposition 5.6.1 is derived in [25]. ◇

Remark 5.6.3. 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

$$\mathbb{P}(S_n > x) \;\;=\;\; \widehat{F}[\theta]^n\,\mathbb{E}_\theta\Big[\mathrm{e}^{-\theta S_n};\, S_n > x\Big] \tag{5.29}$$

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. $\diamond$

5.7 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 [13], and can be investigated using classical saddlepoint techniques. The relevant asymptotic is the classical Esscher approximation :

$$\mathbb{P}(S_N > x) \sim \frac{\left(\widehat{F_{S_N}}[\theta] - \mathrm{e}^{-\mu}\right) \exp\{-\theta x\}}{\theta \sigma_c(\theta)} B_0(\ell)\,, \tag{5.30}$$

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 [84], where also further refinements and variants are given. The issue with implementing (5.30) 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, (5.22) 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 (5.28). Take

$$\widehat{F}^{(k)}[\theta] = \mathbb{E}[X^k \mathrm{e}^{\theta X}] \sim y^k \widehat{F}[\theta], \quad \text{for } k \in \mathbb{N} \tag{5.31}$$

where we've written $\theta = \lambda(y)$ as in Section 5.6. Thus if we set θ as the solution to $\mu y \widetilde{F}[\lambda(y)] = x$ then we get

$$y = 2^{-1/\beta}\left[\frac{(\beta+2)}{(\beta-1)\beta}\mathcal{W}\left(\frac{(\beta-1)\beta}{(\beta+2)}\left(\frac{2^{\frac{1}{\beta}+\frac{1}{2}}x}{c_1}\right)^{\frac{2\beta}{\beta+2}}\right)\right]^{1/\beta} \tag{5.32}$$

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 (5.30) gives us

$$\mathbb{P}(S_N > x) \approx_{\log} \frac{\mathrm{e}^{-\mu}\Big(\exp\{\mu x/y\} - 1\Big)\exp\{-\theta x\}}{\lambda(y)\sqrt{\mu xy}} B_0(\ell)\,. \tag{5.33}$$

Preliminary numerical work indicates that (5.33) 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 $\gamma(x)\mathrm{e}^{-\psi(x)}$.

5.8 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. [117, Prop. 1.4])

$$\overline{F}(x) \sim \overline{H}(x) = C\exp\Big(-\int_0^x \frac{1}{u(t)}\mathrm{d}t\Big), \quad x \to \infty \tag{5.34}$$

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 ([44], [134]) that the class of distributions $\mathcal{L}(\gamma)$ is closed under convolution. In the particular case that the X_i have tails

$$\overline{F}_i(x) = \ell_i(x)x^{\gamma_i - 1}\mathrm{e}^{-kx^\beta}, \quad 1 \le i \le n, \tag{5.35}$$

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 [76] (see also Theorem 6.4 ii) in [5])

$$\mathbb{P}\left(S_n > x\right) \sim \frac{k^{n-1}}{\Gamma(\gamma_0)} x^{\gamma_0 - 1} \prod_{i=1}^{n} \ell_i(x) \mathrm{e}^{-kx^\beta}. \tag{5.36}$$

where $\gamma_0 = \sum_{i=1}^n \gamma_i$. If (5.36) holds with $\beta > 1$, then for non-negative X_i's using the β-norm argument we have as in Section 5.5

$$\mathbb{P}\left(S_n > x\right) \le \mathbb{P}\left(X_1^\beta + \cdots + X_n^\beta > x^\beta / n^{\beta-1}\right) \tag{5.37}$$

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 (5.36) and Theorem 5.4.1

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

and thus the upper bound in (5.37) is logarithmic asymptotically exact.

5.9 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 [15].

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

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

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 5.9.2 below.

One problem that arises is how to simulate from f_θ. Proposition 5.6.1 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 5.9.1. The acceptance ratio can be improved for small x by locally searching for the optimal proposal, that is, the distribution with parameters

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

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. ◇

Proposition 5.9.2. *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 5.3.5 and (5.22),

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

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

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

completing the proof. □

Some estimators based on conditional Monte Carlo ideas are discussed in [12] 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 [12] to more general tails.

Proposition 5.9.3. *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* (5.38) *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 5.3.3. 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

$$c_n \;= c_1\theta_1^\beta + c_2\theta_2^\beta \;=\; \frac{2 + 2^{\beta/(\beta-1)}(n-1)}{\Big(1 + 2^{1/(\beta-1)}(n-1)\Big)^\beta}\,. \tag{5.38}$$

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 $a > 1$

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

and take $a = 2^{1/(\beta-1)}$. □

We finally consider the so-called Asmussen–Kroese estimator

$$Z_{\mathrm{AK}}(x) \;=\; n\,\overline{F}\Big(M_{n-1} \vee (x - S_{n-1})\Big)\,. \tag{5.39}$$

where $M_{n-1} = \max(X_1, \ldots, X_{n-1})$. It was initially developed in [21] with heavy tails in mind, but it was found empirically in [12] that it also provides some variance reduction for light tails, in fact more than $Z_{\mathrm{Cd}}(x)$. We have:

Proposition 5.9.4. *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. □

5.A Proof of Theorem 5.4.1

For the proof of Theorem 5.4.1, we first note that, as shown in BKR, that as $x \to \infty$

$$\frac{\lambda'(x)}{\lambda(x)^2} \;\to\; 0. \tag{5.40}$$

$$\frac{\gamma'(x)}{\sqrt{\lambda'(x)}\gamma(x)} \;\to\; 0. \tag{5.41}$$

In view of Proposition 3.2 in BKR, (5.41) 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

$$\lambda \text{ is flat for } \psi. \tag{5.42}$$

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 (5.40). Using further (5.40) we have that $e = 1/\lambda$ is self-neglecting.

Proof of Theorem 5.4.1 (i). Write $\overline{H}(x) = \gamma(x)\mathrm{e}^{-\psi(x)}/\lambda(x)$. Then

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

Here the last term in $[\cdot]$ goes to 0 according to (5.40). This together with (5.41) 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 $\overline{H}(x) \sim \overline{F}(x)$. □

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

Proof of Theorem 5.4.1 (i). Using integrations by parts yields

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

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 5.A.1. *For any two pairs (γ_1, ψ_1), (γ_2, ψ_2) satisfying the assumptions of Section 5.1, it holds that*

$$\int_{-\infty}^\infty \gamma_1(z)\mathrm{e}^{-\psi_1(z)} \cdot \gamma_2(x-z)\mathrm{e}^{-\psi_2(x-z)}\,\mathrm{d}z \tag{5.43}$$

has the asymptotics given by Theorem 5.4.1(ii).

Proof of Theorem 5.4.1 (ii). This is a reformulation of Theorem 1.1 in BKR. Since by (5.20) $q_1' + q_2' = 1$ we have the claimed relation between λ and λ_1, λ_2, namely

$$\lambda(x) \;=\; \lambda_1\Big(q_1(x)\Big)q_1'(x) + \lambda_2\Big(q_2(x)\Big)q_2'(x) \;=\; \lambda_1(q_1) \;=\; \lambda_2(q_2) \tag{5.44}$$

establishing the proof. □

Proof of Theorem 5.4.1 (iii). We have that $\mathrm{e}^{-\psi_i(x)}, i = 1, 2$ is a von-Mises function (see (5.34)) 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 (5.40) $r_i(x) = \sqrt{\lambda_i(x)}/\lambda_i(x) \to 0$ 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 [58] 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

$$\overline{F_1 * F_2}(x) \;=\; \int_{-\infty}^{\infty} \gamma_1(z)\mathrm{e}^{-\psi_1(x)} \cdot \frac{\gamma_2(x - z)}{\lambda_2(x - z)}\mathrm{e}^{-\psi_2(x-z)}\,\mathrm{d}z \tag{5.45}$$

But by (5.42), γ_2/λ_2 is flat for ψ_2, so using Lemma 5.20 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 (5.44) this is the same as $\gamma(x)\mathrm{e}^{-\psi(x)}/\lambda(x)$. This completes the proof. □

Chapter 6

Rare tail approximation using asymptotics and polar coordinates

TODO

Chapter 7

Efficient simulation for dependent rare events with applications to extremes

(pjl) Add in the closed form soln for ind. X_i's

7.1 Introduction

The estimators in this paper apply to quite general problems, so we will first introduce them in the framework of our main example, namely, as estimators relating to rare maxima of dependent random vectors. For a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with maximum $M = \max_i X_i$, the first problem we consider is estimating

$$\alpha(\gamma) = \mathbb{P}(M > \gamma)\,.$$

This problem has many applications in many areas, for example in actuarial science (e.g. default probabilities [13]), finance (e.g. probability of 'knock-out' in a barrier option [45]), survival analysis, reliability [115] and engineering (e.g. failure probability of a series circuit).

We construct estimators for this probability, which are in terms of

$$E(\gamma) = \sum_{i=1}^{d} \mathbb{I}\{X_i > \gamma\}\,,$$

the random variable which counts the number of X_i which exceed γ.[1] Our two main

[1] We use $\mathbb{I}\{\cdot\}$ to denote the indicator function, and $\mathbb{I}\{\emptyset\} = 1$.

estimators in this setting are

$$\widehat{\alpha}_1 = \sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) + \frac{1}{R}\sum_{r=1}^{R}(1 - E_r(\gamma))\mathbb{I}\{E_r(\gamma) \geq 2\}\,, \text{ and} \tag{7.1}$$

$$\begin{aligned}\widehat{\alpha}_2 = &\sum_{i=1}^{d} \mathbb{P}(X_i > \gamma) - \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} \mathbb{P}(X_i > \gamma, X_j > \gamma) \\ &+ \frac{1}{R}\sum_{r=1}^{R}\left[1 - E_r(\gamma) + \frac{E_r(\gamma)(E_r(\gamma)-1)}{2}\right]\mathbb{I}\{E_r(\gamma) \geq 3\}\,.\end{aligned} \tag{7.2}$$

where $R \in \mathbb{N}$ and the $E_r(\gamma)$s are derived from i.i.d. samples of $\boldsymbol{X}$. The fact that these are unbiased estimators of $\alpha(\gamma)$ follows from Proposition 7.2.2 below. Estimation of $\mathbb{P}(M > \gamma)$ is a difficult problem and treatments in the literature make distributional assumptions on $\boldsymbol{X}$. One such example is Adler et al. [7] where $\boldsymbol{X}$ is assumed to be multivariate normal. In this case, our estimator $\widehat{\alpha}_1$, with appropriate importance sampling, is the same as one of the estimators from [7].

The next problem we consider is estimating

$$\beta_n(\gamma) := \mathbb{E}[Y\mathbb{I}\{E(\gamma) \geq n\}]$$

for $n = 1, \ldots, d$ and some random variable Y. We do not make any assumptions of independence between the $\{X_i > \gamma\}$ events themselves or between the events and Y.

The subcase of $Y = 1$ a.s. has some interesting examples:

$$\beta_1(\gamma) = \mathbb{P}(M > \gamma) = \alpha(\gamma)\,, \quad \text{and} \quad \beta_n(\gamma) = \mathbb{P}(X_{(n)} > \gamma)$$

where $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(d)}$ are the order statistics of $\boldsymbol{X}$. The probability of a parallel circuit failing is a simple application for $\mathbb{P}(X_{(n)} > \gamma)$.

Our main β_1 estimator uses the fact that

$$\{M > \gamma\} := \bigcup_{i=1}^{d}\{X_i > \gamma\} = \bigcup_{i=1}^{d}\{X_1 \leq \gamma, \ldots, X_{i-1} \leq \gamma, X_i > \gamma\} \tag{7.3}$$

where the events in the union on the right are disjoint. This supplies a form of β_1 which is amenable to efficient Monte Carlo estimation:

$$\beta_1 = \sum_{i=1}^{d} \mathbb{E}[Y\,\mathbb{I}\{X_1 \leq \gamma, \ldots, X_{i-1} \leq \gamma\} \mid X_i > \gamma]\,\mathbb{P}\left(X_i > \gamma\right)\,. \tag{7.4}$$

As previously mentioned, while they are main example and motivation, the extremes considered so far are a very specific instance of estimators. We now turn our attention to the general set-up treated in the paper.

Let $A(\gamma) = \cup_{i=1}^{d} A_i(\gamma)$ be the union of events $A_1(\gamma), \ldots, A_d(\gamma)$ for an index parameter $\gamma \in \mathbb{R}$. We consider the problem of estimating $\mathbb{P}(A(\gamma))$ when the events are rare, that is, $\mathbb{P}(A(\gamma)) \to 0$ as $\gamma \to \infty$. Define

$$\alpha(\gamma) := \mathbb{P}(A(\gamma)) \quad \text{and} \qquad E(\gamma) := \sum_{i=1}^{d} \mathbb{I}\{A_i(\gamma)\}\,.$$

Note that we recover our introductory example by having $A_i(\gamma) = \{X_i > \gamma\}$. Aside from this example, $A(\gamma)$ is quite general (a union of arbitrary events) and many interesting events arising in applied probability and statistics can be formulated as a union. The quantity $\beta_n(\gamma)$ is reminiscent of *expected shortfall* from risk management [103].

Traditional Monte Carlo methods are unreliable in the rare-event setting. We will use standard techniques from the *rare-event simulation methodology*, such as importance sampling for variance reduction and applicable measures of efficiency: *bounded relative error* and *logarithmic efficiency*, cf. [15, 65, 123]. The resulting estimators are among the most efficient possible under the most general assumptions.

The paper is structured as follows. In Sections 7.2 and 7.3 we formally introduce our estimators for $\alpha(\gamma)$ and $\beta_n(\gamma)$ respectively, we prove their validity, and show how to combine them with some existing variance reduction techniques; the efficiency properties for the general estimators are analysed in Section 7.4, in addition we further investigate the efficiency for certain important dependence structures. Finally, we evaluate the numerical performance of the estimators in Section 7.5.

7.2 Estimators of α

In the following, we first explain the construction of our estimators of α, then discuss possible variance reduction schemes. As the γ notation can be cumbersome, we simply write $A = A(\gamma)$, $A_i = A_i(\gamma)$, $E = E(\gamma)$, $\alpha = \alpha(\gamma)$ and $\beta_n = \beta_n(\gamma)$. Similarly, we often write $\sum_i$, $\sum_{i<j}$, $\cup_i$, $\cap_i$ for $\sum_{i=1}^{d}$,$\sum_{1=i<j}^{d}$, $\cup_{i=1}^{d}$ and $\cap_{i=1}^{d}$.

7.2.1 Proposed estimators of α

The inclusion–exclusion formula (IEF) provides a representation of α as a summation whose terms are decreasing in size. The formula is

$$\alpha = \mathbb{P}(A) = \sum_{i=1}^{d} (-1)^{i+1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i\right). \tag{7.5}$$

The IEF can rarely be used as its summands are increasingly difficult to calculate numerically. The $\mathbb{P}(A_i)$ terms are typically known, and the $\mathbb{P}(A_i, A_j)$ terms can frequently be calculated, however the remaining higher-dimensional terms are normally intractable for numerical integration algorithms (cf. the *curse of dimensionality* [15, Chapter IX]). Truncating the summation leads to bias, and indeed by the Bonferroni inequalities we have:

$$\alpha \leq \sum_{i=1}^{k} (-1)^{i-1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \quad \text{if } 1 \leq k < d \text{ and } k \text{ is odd,} \tag{7.6}$$

$$\alpha \geq \sum_{i=1}^{k} (-1)^{i-1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \quad \text{if } 1 < k < d \text{ and } k \text{ is even.} \tag{7.7}$$

This higher-order intractability motivates our estimators which use the IEF rewritten in terms of $E = \sum_i \mathbb{I}\{A_i\}$.

Proposition 7.2.2. *For $i = 1, \ldots, d$,*

$$\sum_{|I|=i} \mathbb{I}\left\{\cap_{i \in I} A_i\right\} = \binom{E}{i} \mathbb{I}\{E \geq i\}\,. \tag{7.8}$$

Proof.

$$\sum_{|I|=i} \mathbb{I}\{\cap_{i \in I} A_i\} = \sum_{k=i}^{d} \sum_{|I|=i} \mathbb{I}\{\cap_{i \in I} A_i, E = k\} = \sum_{k=i}^{d} \binom{k}{i} \mathbb{I}\{E = k\} = \binom{E}{i} \mathbb{I}\{E \geq i\}\,.$$

□ □

Taking the expectation of (7.8) gives

$$\sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \mathbb{E}\left[\binom{E}{i} \mathbb{I}\{E \geq i\}\right] \quad \text{for } i = 1, \ldots, d\,.$$

So the following has mean α, and forms the nucleus of our $\widehat{\alpha}_i$ estimators:

$$\sum_{i=1}^{d} (-1)^{i-1} \binom{E}{i} \mathbb{I}\{E \geq i\}\,. \tag{7.9}$$

We present estimators which deterministically *calculate* the first larger terms of the IEF (7.5) and Monte Carlo (MC) *estimate* the remaining smaller terms using sample means of

(7.8). We begin by constructing the single-replicate estimator $\widehat{\alpha}_1$ where the first summand is calculated and the remaining terms are estimated:

$$\begin{aligned}\widehat{\alpha}_1 :&= \sum_i \mathbb{P}(A_i) + \sum_{i=2}^{d} \left[(-1)^{i-1}\binom{E}{i}\mathbb{I}\{E \geq i\}\right] \\ &= \sum_i \mathbb{P}(A_i) + (1-E)\mathbb{I}\{E \geq 2\}\,, \quad \text{using} \quad \sum_{k=0}^{n}(-1)^{k-1}\binom{n}{k} = 0\,.\end{aligned}$$

In identical fashion, the single-replicate estimator calculating the first two terms from the IEF is

$$\begin{aligned}\widehat{\alpha}_2 :=& \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i, A_j) + \sum_{i=3}^{d} \left[(-1)^{i-1}\binom{E}{i}\mathbb{I}\{E \geq i\}\right] \\ =& \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i, A_j) + \left[1 - E + \frac{E(E-1)}{2}\right]\mathbb{I}\{E \geq 3\}\,.\end{aligned}$$

Thus, for $n \in \{1, \ldots, d-1\}$,[1]

$$\widehat{\alpha}_n := \sum_{i=1}^{n}(-1)^{i-1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i\in I} A_i\right) + \left[\sum_{i=0}^{n}(-1)^{i}\binom{E}{i}\right]\mathbb{I}\{E \geq n+1\}\,. \tag{7.10}$$

Thus, $\{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{d-1}\}$ is a collection of estimators which allows the user to control the computational division of labour between numerical integration and Monte Carlo estimation. We will furthermore let $\widehat{\alpha}_0$ be the crude Monte Carlo estimator $\mathbb{I}\{E \geq 1\}$, and note that this falls under the definition in (7.10) if we interpret the empty sum as zero.

The $\widehat{\alpha}_n$ estimators are of decreasing variance in n, however each estimator carries the assumption that one can perform accurate numerical integration for 1 up to n dimensions. As numerical integration can be slow and unreliable in high dimensions we focus on $\widehat{\alpha}_1$, and also show the numerical performance of $\widehat{\alpha}_2$.

In practice, theses estimators will exhibit very modest improvements when compared against their truncated IEF counterparts (i.e., the right side of (7.6) and (7.7)). When combined with importance sampling, as in Section 7.2.5, the improvement is marked. Furthermore, we will show that these estimators possess desirable efficiency properties which are preserved after combining with importance sampling.

[1]Note that by the IEF, we have $\widehat{\alpha}_d := \alpha$, so this possibility is ignored.

7.2.3 Discussion of $\widehat{\alpha}_1$ estimator

The estimator $\widehat{\alpha}_1$ has some nice interpretations. Recall the Boole–Fréchet inequalities

$$\max_i \mathbb{P}(A_i) \leq \alpha = \mathbb{P}(A) \leq \sum_i \mathbb{P}(A_i) =: \overline{\alpha}\,. \tag{7.11}$$

The stochastic part of $\widehat{\alpha}_1$ is an unbiased estimate of $\overline{\alpha} - \alpha \leq 0$. That is to say, $\widehat{\alpha}_1$ MC estimates the difference between the target quantity α and its upper bound given by the Boole–Fréchet inequalities, $\overline{\alpha}$. Similarly, we often have

$$\alpha(\gamma) \sim \sum_i \mathbb{P}(A_i(\gamma))\,,^{1}$$

for example when the A_i exhibit a weak dependence structure. In this case, we can say that $\widehat{\alpha}_1$ MC estimates the difference between α and its (first-order) asymptotic expansion.

7.2.4 Relation of $\widehat{\alpha}_n$ estimators to control variates

An alternative construction of $\{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{d-1}\}$ is to add *control variates* to the crude Monte Carlo estimator $\widehat{\alpha}_0$. We begin by adding the control variate E to $\widehat{\alpha}_0$ with weight $\tau \in \mathbb{R}$:

$$\widehat{\alpha}_1^{\tau} := \mathbb{I}\{E \geq 1\} - \tau\Big[E - \sum_i \mathbb{P}(A_i)\Big]\,.$$

Setting $\tau = 1$ means this estimator simplifies to $\widehat{\alpha}_1$. Next, we add the control variates E and $-\frac{1}{2}E(E-1)$ to $\widehat{\alpha}_0$, and setting the corresponding weights to 1 gives $\widehat{\alpha}_2$. This pattern goes on.

7.2.5 Combining $\widehat{\alpha}_1$ with importance sampling

The family of estimators $\widehat{\alpha}_n$ can be combined with the variance reduction technique called *importance sampling* (IS), cf. [15, 65]. Standard IS theory suggests that we should focus on IS distributions where the event of interest $A = \cup_i A_i = \{E \geq 1\}$ occurs almost surely. A convenient way of constructing such a distribution is as a *mixture distribution*. Say that we condition on A_i with probability

$$p_i := \frac{\mathbb{P}(A_i)}{\sum_j \mathbb{P}(A_j)} = \frac{\mathbb{P}(A_i)}{\overline{\alpha}}\,, \qquad \text{for } i = 1, \ldots, d\,.$$

[1] Using the standard notation that $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$.

A heuristic motivation for this selection comes from a rare-event setting where the asymptotic relationship $\mathbb{P}(A_i(\gamma), A_j(\gamma)) = \mathrm{o}(\mathbb{P}(A_i(\gamma)))$ often occurs for all $i \neq j$. In such a case

$$\mathbb{P}\left(A_i(\gamma) \mid A(\gamma)\right) = \frac{\mathbb{P}(A_i(\gamma))}{\sum_j \mathbb{P}(A_j(\gamma))(1+\mathrm{o}(1))} \sim p_i(\gamma)\,, \qquad \text{as } \gamma \to \infty\,.$$

Now consider the measure

$$\mathbb{Q}^{[1]}(\mathscr{A}) = \sum_i p_i\, \mathbb{P}(\mathscr{A} \mid A_i) \qquad \forall \mathscr{A} \in \mathcal{F}\,,$$

which induces the likelihood ratio of $L^{[1]} := \mathrm{d}\,\mathbb{Q}^{[1]}/\mathrm{d}\,\mathbb{P} = \overline{\alpha}/E$. As

$$\overline{\alpha} + (1-E)\mathbb{I}\{E \geq 2\}L^{[1]} = \overline{\alpha}\Big(1 + \frac{1-E}{E}\Big) = \frac{\overline{\alpha}}{E} \quad \text{ under } \mathbb{Q}^{[1]}\,,$$

we can see that $\widehat{\alpha}_1$ under this change of measure, with $R \in \mathbb{N}$ replicates, is

$$\widehat{\alpha}_1^{[1]} := \frac{1}{R}\sum_{r=1}^{R} \frac{\overline{\alpha}}{E_r^{[1]}}\,, \tag{7.12}$$

where the superscript "[1]" indicates that the $E_r^{[1]}$ are (independently) sampled under $\mathbb{Q}^{[1]}$. This estimator corresponds to one from the paper of Adler et al. [6], though applied in a more general way (they consider rare maxima of normally distributed vectors).

Continuing in the same pattern, consider the *second-order* IS distributions where $\{E \geq 2\}$ occurs almost surely, to be applied to $\widehat{\alpha}_2$. Say that we choose to condition on $A_i \cap A_j$ with probability

$$p_{ij} := \frac{\mathbb{P}(A_i, A_j)}{\sum_{m<n} \mathbb{P}(A_m, A_n)} = \frac{\mathbb{P}(A_i, A_j)}{q}\,, \qquad \text{for } 1 \leq i < j \leq d\,,$$

defining $q := \sum_{i<j} \mathbb{P}(A_i, A_j)$. Now consider the measure

$$\mathbb{Q}^{[2]}(\mathscr{A}) = \sum_{i<j} p_{ij}\, \mathbb{P}(\mathscr{A} \mid A_i, A_j) \qquad \forall \mathscr{A} \in \mathcal{F}\,,$$

which induces a likelihood ratio of

$$L^{[2]} := \frac{\mathrm{d}\,\mathbb{Q}^{[2]}}{\mathrm{d}\,\mathbb{P}} = \frac{q}{\sum_{i<j} \mathbb{I}\{A_i A_j\}} = \frac{q}{\binom{E}{2}} = \frac{2q}{E(E-1)}\,.$$

Thus, after simplifying, the estimator $\widehat{\alpha}_2$ under $\mathbb{Q}^{[2]}$ is

$$\widehat{\alpha}_2^{[2]} := \overline{\alpha} - \frac{2q}{R}\sum_{r=1}^{R} \frac{1}{E_r^{[2]}}. \tag{7.13}$$

Remark 7.2.6. As the $\mathbb{Q}^{[2]}$-mean of $\frac{2}{E}$ is less than 1, this fraction can be seen as a correction term for the two-term truncation of (7.5). We know from (7.7) that $\alpha \geq \overline{\alpha} - q$.

$\diamond$

Both of these IS algorithms have some extra requirements for their use. The first-order estimators require that we can simulate from $\mathbb{P}(\,\cdot \mid A_i)$ and can calculate the $\mathbb{P}(A_i)$. The second-order estimator requires that we can simulate from $\mathbb{P}(\,\cdot \mid A_i, A_j)$ and that we can calculate the $\mathbb{P}(A_i)$ and $\mathbb{P}(A_i, A_j)$. In the rare maxima case, integration routines in MATHEMATICA or MATLAB can usually calculate these probabilities; it is simulating from the conditional distributions which can be the prohibitive requirement, particularly for $\widehat{\alpha}_2^{[2]}$.

7.3 Estimators of β_n

Now, we turn our attention to the estimation of

$$\beta_n := \mathbb{E}[Y \mathbb{I}\{E \geq n\}] \,.$$

We start with β_1, and rewrite the partition (7.3) in terms of the general A_i:

$$A := \bigcup_{i=1}^{d} A_i = A_1 \cup (A_1^{\mathsf{c}} A_2) \cup \cdots \cup (A_1^{\mathsf{c}} \ldots A_{d-1}^{\mathsf{c}} A_d) \,. \tag{7.14}$$

This gives us (the generalised version of (7.4))

$$\begin{aligned} \beta_1 = \mathbb{E}[Y \mid A_1]\, \mathbb{P}(A_1) + \mathbb{E}[Y \mathbb{I}\{A_1\} \mid A_2]\, \mathbb{P}(A_2) \\ + \cdots + \mathbb{E}[Y \mathbb{I}\{A_1^{\mathsf{c}} \ldots A_{d-1}^{\mathsf{c}}\} \mid A_d]\, \mathbb{P}(A_d) \,. \end{aligned}$$

If we assume it is possible to sample from the $\mathbb{P}(\,\cdot \mid A_i)$ conditional distributions—the same assumption required to use the first-order IS estimator $\widehat{\alpha}_1^{[1]}$ from Section 7.2.5—then each of these conditional expectations can be estimated by sample means:

$$\widehat{\beta}_1 := \sum_{i=1}^{d} \frac{\mathbb{P}(A_i)}{\lceil R/d \rceil} \sum_{r=1}^{\lceil R/d \rceil} Y_{i,r} \mathbb{I}\{A_1^{\mathsf{c}} \ldots A_{i-1}^{\mathsf{c}}\}_{i,r} \,. \tag{7.15}$$

Here, the $Y_{i,r}$ and $\mathbb{I}\{\cdot\}_{i,r}$ are sampled independently and conditional on A_i. The following proposition gives the partition of the event $\{E \geq i\}$:

Proposition 7.3.1. *Consider a finite collection of events* $\{A_1, \ldots, A_d\}$ *and for each subset* $I \subset \{1, 2, \ldots, d\}$ *define* [1]

$$B_I := \bigcap_{j \in I} A_j, \qquad C_I := \bigcap_{\substack{k \notin I, \\ k < \max I}} A_k^c.$$

Then

$$\{E \geq m\} = \bigcup_{|I|=m} B_I = \bigcup_{|I|=m} B_I C_I. \tag{7.16}$$

Moreover, the collection of sets $\{B_I C_I : |I| = m\}$ *is disjoint.*

Proof. The first equality in (7.16) is straightforward from the definition of the random variable E. For the second equality we note that the relation $\supseteq$ follows trivially; to prove the opposite relation $\subseteq$ it remains to show that if ω is such that $\omega \in B_I$ and $\omega \notin C_I$, then there exists I' such that $|I'| = m$ and $\omega \in B_{I'} C_{I'}$. Notice that if $\omega \notin C_I$, then there exists a nonempty set J satisfying $\max J < \max I$, with $j \in J$ if and only if $\omega \notin A_j^c$. Select I' as the set formed by the smaller m elements of $I \cup J$. In consequence,

$$\omega \in \Big(\bigcap_{j \in I \cup J} A_j \Big) \Big(\bigcap_{\substack{k \notin I \cup J, \\ k \leq \max I}} A_k^c \Big) \subseteq \Big(\bigcap_{j \in I'} A_j \Big) \Big(\bigcap_{\substack{k \notin I', \\ k \leq \max I'}} A_k^c \Big) = B_{I'} C_{I'}.$$

This completes the proof of the second equivalence in (7.16).

Next we show that the collection of sets $\{B_I C_I : |I| = m\}$ is disjoint. Consider two sets of indexes I_1 and I_2 such that $|I_1| = |I_2| = m$ and $I_1 \neq I_2$. Take i such that $i \in I_1$, $i \notin I_2$ and w.l.o.g. further assume that $i < \max I_2$. Then $B_{I_1} \subseteq A_i$ while $C_{I_2} \subseteq A_i^c$. □ □

This proposition implies that

$$\beta_n = \mathbb{E}\left[Y \mathbb{I}\Big\{ \bigcup_{|I|=n} B_I \Big\} \right] = \mathbb{E}\left[Y \mathbb{I}\Big\{ \bigcup_{|I|=n} B_I C_I \Big\} \right] = \sum_{|I|=n} \mathbb{E}\left[Y \mathbb{I}\left\{ C_I \right\} \Big| B_I \right] \mathbb{P}\left(B_I \right).$$

Therefore, if (i) reliable estimates of $\mathbb{P}\left(B_I\right)$ are available, and (ii) it is possible to simulate from the conditional measures $\mathbb{P}\left(\cdot \mid B_I \right)$, then the following is an unbiased estimator of $\mathbb{E}[Y \mathbb{I}\{E \geq n\}]$:

$$\widehat{\beta}_n := \sum_{|I|=n} \frac{\mathbb{P}(B_I)}{\lceil R / \binom{d}{n} \rceil} \sum_{r=1}^{\lceil R/\binom{d}{n} \rceil} Y_{I,r} \mathbb{I}\{C_I\}_{I,r}\,. \tag{7.17}$$

Here, similar to before, $Y_{I,r}$ and $\mathbb{I}\{\cdot\}_{I,r}$ denote independent sampling conditioned on B_I.

Notice that a permutation of the sets $A_1, \ldots, A_d$ will result in a different collection of events C_I, and also a slightly different estimator.

[1] Using the convention that $\cap_\emptyset = \Omega$.

7.3.2 Applying $\widehat{\beta}_i$ to estimate α

The $\widehat{\beta}_i$ estimators can be used in various ways to estimate the probability $\alpha = \mathbb{P}(A)$. The simplest way is to set $Y = 1$ a.s. in $\widehat{\beta}_1$ (7.17), leading to the estimator

$$\widehat{(\beta_1 \ddagger \alpha)} := \mathbb{P}(A_1) + \sum_{i=2}^{d} \frac{\mathbb{P}(A_i)}{\lceil R/(d-1) \rceil} \sum_{r=1}^{\lceil R/(d-1) \rceil} \mathbb{I}\{A_1^{\mathsf{c}} \dots A_{i-1}^{\mathsf{c}}\}_{i,r}\,, \tag{7.18}$$

using the notation from (7.15). Note, we achieve minor improvement in (7.18) over (7.17) when $Y = 1$ a.s. as $\mathbb{E}[1 \mid A_1] = 1$ does not require estimation.

More effective estimators can be constructed if we use $\widehat{\beta}_n$ to estimate terms from $\widehat{\alpha}_{n-1}$ (7.10). We label the random terms in $\widehat{\alpha}_n$ as

$$R_n := \Big[\sum_{i=0}^{n} (-1)^i \binom{E}{i}\Big] \mathbb{I}\{E \geq n+1\}, \quad \text{and say} \quad \mathcal{R}_n := \mathbb{E}[R_n]\,. \tag{7.19}$$

Now, if we choose $Y := \sum_{i=0}^{n-1} (-1)^i \binom{E}{i}$ then it is obvious that

$$\beta_n := \mathbb{E}\left\{\Big[\sum_{i=0}^{n-1} (-1)^i \binom{E}{i}\Big] \mathbb{I}\{E \geq n\}\right\} = \mathcal{R}_{n-1}.$$

This leads to the set of estimators

$$\begin{aligned}\widehat{(\beta_n \ddagger \alpha)} := &\sum_{i=1}^{n-1} (-1)^{i-1} \sum_{|I|=i} \mathbb{P}\Big(\bigcap_{i \in I} A_i\Big) \\ &+ \sum_{|I|=n} \frac{\mathbb{P}(B_I)}{\lceil R/\binom{d}{n} \rceil} \sum_{r=1}^{\lceil R/\binom{d}{n} \rceil} \Big[\sum_{i=0}^{n-1} (-1)^i \binom{E_{I,r}}{i}\Big] \mathbb{I}\{E \geq n\}_{I,r}\,,\end{aligned}$$

for $n = 2, \dots d-1$. In particular, for $n = 2$

$$\widehat{(\beta_2 \ddagger \alpha)} := \sum_i \mathbb{P}(A_i) + \sum_{i<j} \frac{\mathbb{P}(A_i, A_j)}{\lceil R/\binom{d}{2} \rceil} \sum_{r=1}^{\lceil R/\binom{d}{2} \rceil} (1 - E_{ij,r}) \mathbb{I}\{E \geq 2\}_{ij,r}\,, \tag{7.20}$$

where the ij subscript indicates sampling conditional on $A_i A_j$, similar to before.

7.4 Efficiency results

In this section we analyse the performance of the estimators in a rare-event setting. Recall that in such a setting, $\{A_1(\gamma), \dots, A_d(\gamma)\}$ denotes an indexed collection of not

necessarily independent rare events and our objective is to calculate $\alpha(\gamma) = \mathbb{P}(\bigcup_i^d A_i(\gamma))$ as $\gamma \to \infty$. For such a *rare-event* estimation problem there are specialised concepts of efficiency. In Section 7.4.3 these definitions of efficiency are introduced. In addition, we provide efficiency criteria for the proposed estimators under very general assumptions.

In Sections 7.4.9 and 7.4.13 we specialise in rare events associated with extremes. In such a framework, we show when the estimator $\widehat{\alpha}_1$ is efficient for: i) a vast array of multivariate distributions with identical marginals in Section 7.4.9, and ii) the specific cases of normal and elliptical distributions in Section 7.4.13. For this section we take the number of replicates R to be 1.

7.4.1 Variance Reduction

First we compare the efficiency of our proposed estimator $\widehat{\alpha}_1$ against that of the crude Monte Carlo (CMC) estimator $\widehat{\alpha}_0(\gamma)$ of $\alpha(\gamma) := \mathbb{P}(A(\gamma))$. An upper bound for $\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_0(\gamma)$ is

$$\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_0(\gamma) = \mathbb{P}(A(\gamma))[1 - \mathbb{P}(A(\gamma)] < \mathbb{P}(A(\gamma)) \leq \sum_i \mathbb{P}(A_i(\gamma))\,.$$

This implies that the variance of the CMC estimator is of order $\mathcal{O}(\max_i \mathbb{P}(A_i(\gamma)))$, which is the best possible without making any further assumptions. In contrast an upper bound of $\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma) = \mathbb{V}\mathrm{ar}\,R_1$, where $R_1 = (1-E)\mathbb{I}\{E \geq 2\}$ from (7.19), is

$$\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma) \leq \mathbb{E}[R_1^2] < 2\,\mathbb{E}\left[\binom{E}{2}\mathbb{I}\{E \geq 2\}\right] \underset{(7.8)}{=} 2\sum_{i<j}\mathbb{P}(A_i(\gamma), A_j(\gamma))\,. \tag{7.21}$$

Thus the variance of our estimator $\widehat{\alpha}_1(\gamma)$ is of order $\mathcal{O}(\max_{i<j}\mathbb{P}(A_i(\gamma), A_j(\gamma)))$, so we can conclude that $\widehat{\alpha}_1(\gamma)$ is asymptotically superior to CMC.

Next we turn our attention to the estimator $\widehat{\beta}_n$. The following proposition shows that the reduction of variance of the estimator $\widehat{\beta}_n$ is of at least of a factor $\max_{|I|=n}\mathbb{P}(B_I)$ with respect to the non-conditional (crude) version estimator $\widehat{\beta}_n^{[0]}$ defined as

$$\widehat{\beta}_n^{[0]} := \sum_{|I|=n} \frac{1}{\lceil R/\binom{d}{n}\rceil} \sum_{r=1}^{\lceil R/\binom{d}{n}\rceil} Y_{Ir}\mathbb{I}\{B_I C_I\} \tag{7.22}$$

Proposition 7.4.2.

$$\mathbb{V}\mathrm{ar}(\widehat{\beta}_n) \leq \max_{|I|=n}\mathbb{P}(B_I)\,\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]})\,.$$

Proof. Let $W_I := Y\mathbb{I}\{C_I\}$. By independence of the W_I we can write the variance of $\widehat{\beta}_n$ as

$$\begin{aligned}\mathbb{V}\mathrm{ar}(\widehat{\beta}_n) = \mathbb{V}\mathrm{ar}\left(\sum_{|I|=n} W_I\,\mathbb{P}(B_I)\,\middle|\,B_I\right) &= \sum_{|I|=n}\mathbb{P}(B_I)^2\,\mathbb{V}\mathrm{ar}(W_I \mid B_I)\\ &\le \max_{|I|=n}\mathbb{P}(B_I)\sum_{|I|=n}\mathbb{P}(B_I)\,\mathbb{V}\mathrm{ar}(W_I \mid B_I)\,.\end{aligned}$$

Now, observe that

$$\begin{aligned}\mathbb{P}(B_I)\,\mathbb{V}\mathrm{ar}(W_I \mid B_I) &\le \mathbb{P}(B_I)\,\mathbb{E}[W_I^2 \mid B_I] - \mathbb{P}(B_I)^2\,\mathbb{E}[W_I \mid B_I]^2\\ &= \mathbb{E}[W_I^2\mathbb{I}\{B_I\}] - \mathbb{E}[W_I\mathbb{I}\{B_I\}]^2 = \mathbb{V}\mathrm{ar}\left[W_I\mathbb{I}\{B_I\}\right].\end{aligned}$$

Thus we have proven that

$$\mathbb{V}\mathrm{ar}(\widehat{\beta}_n) \le \max_{|I|=n}\mathbb{P}(B_I)\sum_{|I|=n}\mathbb{V}\mathrm{ar}(W_I\mathbb{I}\{B_I\}) = \max_{|I|=n}\mathbb{P}(B_I)\sum_{|I|=n}\mathbb{V}\mathrm{ar}(\widehat{\beta}_0)\,.$$

□ □

7.4.3 Efficiency criteria

We now ask if and when $\widehat{\alpha}_1$ and $\widehat{\beta}_n$ are efficient in the rare-event sense. We must first define efficiency, as there are several common benchmarks for the efficiency of a rare-event estimator.

Definition 7.4.4. *An estimator $\widehat{p}_\gamma$ of some rare probability p_γ which satisfies $\forall\varepsilon > 0$*

$$\limsup_{\gamma\to\infty}\frac{\mathbb{V}\mathrm{ar}\,\widehat{p}_\gamma}{p_\gamma^{2-\varepsilon}} = 0 \quad (7.23a) \qquad \limsup_{\gamma\to\infty}\frac{\mathbb{V}\mathrm{ar}\,\widehat{p}_\gamma}{p_\gamma^{2}} < \infty \quad (7.23b) \qquad \limsup_{\gamma\to\infty}\frac{\mathbb{V}\mathrm{ar}\,\widehat{p}_\gamma}{p_\gamma^{2}} = 0 \quad (7.23c)$$

has logarithmic efficiency *(LE)* (7.23a)*,* bounded relative error *(BRE)* (7.23b)*, or* vanishing relative error *(VRE)* (7.23c) *respectively.*

The levels of efficiency in Definition 7.4.4 are given in increasing order of strength, that is, VRE ⇒ BRE ⇒ LE. As VRE is often too difficult a goal, we focus on BRE and LE. The following proposition gives an alternative form of the conditions in (7.23) for the specific case of our estimator $\widehat{\alpha}_1$.

Proposition 7.4.5. *The estimator $\widehat{\alpha}_1$ has LE iff it holds that $\forall\varepsilon > 0$*

$$\limsup_{\gamma\to\infty}\frac{\max_{i<j}\,\mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k\,\mathbb{P}(A_k(\gamma))^{2-\varepsilon}} = 0\,, \tag{7.24}$$

and has BRE iff

$$\limsup_{\gamma\to\infty} \frac{\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k \mathbb{P}(A_k(\gamma))^2} < \infty\,. \tag{7.25}$$

Proof. We prove the LE claim (7.24). Proof of the BRE claim follows the same arguments.
($\Rightarrow$) We can see that

$$\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma) \geq \mathbb{V}\mathrm{ar}\,\mathbb{I}\{E \geq 2\} = \mathbb{P}(E \geq 2)\ \mathbb{P}(E \leq 1)\,, \quad \mathbb{P}(E \leq 1) \to 1\,, \tag{7.26}$$

and

$$\mathbb{P}(E \geq 2) \geq \binom{d}{2}^{-1} \sum_{n=2}^{d} \binom{n}{2} \mathbb{P}(E = n) \underset{(7.8)}{=} \binom{d}{2}^{-1} \sum_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))\,. \tag{7.27}$$

So, $\forall \varepsilon > 0$,

$$\begin{aligned} 0 &\underset{(7.23a)}{=} \limsup_{\gamma\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma)}{\mathbb{P}(A)^{2-\varepsilon}} \underset{(7.11)\,\&\,(7.26)}{>} \limsup_{\gamma\to\infty} \frac{\mathbb{P}(E \geq 2)}{\left(\sum_k \mathbb{P}(A_k(\gamma))\right)^{2-\varepsilon}} \\ &\underset{(7.27)}{\geq} \left[d^{2-\varepsilon}\binom{d}{2}\right]^{-1} \limsup_{\gamma\to\infty} \frac{\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k \mathbb{P}(A_k(\gamma))^{2-\varepsilon}} \end{aligned}$$

which implies (7.24).
($\Leftarrow$) We can see that, $\forall \varepsilon > 0$,

$$\begin{aligned} \limsup_{\gamma\to\infty} \frac{\mathbb{V}\mathrm{ar}\,\widehat{\alpha}_1(\gamma)}{\mathbb{P}(A)^{2-\varepsilon}} &\underset{(7.11)\,\&\,(7.21)}{<} \limsup_{\gamma\to\infty} \frac{2\sum_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{(\max_k\ \mathbb{P}(A_k(\gamma)))^{2-\varepsilon}} \\ &\leq 2\binom{d}{2} \limsup_{\gamma\to\infty} \frac{\max_{i<j} \mathbb{P}(A_i(\gamma), A_j(\gamma))}{\max_k \mathbb{P}(A_k(\gamma))^{2-\varepsilon}} \underset{(7.24)}{=} 0\,, \end{aligned}$$

which implies (7.23a). □ □

Example 7.4.6. *If the A_i events are independent then the estimator $\widehat{\alpha}_1$ has BRE.*

For the efficiency of our $\widehat{\beta}_n$ estimators, the following proposition provides a very simple yet non-trivial condition for BRE.

Proposition 7.4.7. *The estimator $\widehat{\beta}_n(\gamma)$ has BRE if*

$$\limsup_{\gamma\to\infty} \frac{\max_{|I|=n} \mathbb{P}(B_I)}{\beta_n(\gamma)} < \infty.$$

Proof. By Proposition 7.4.2 and the hypothesis we have

$$\begin{aligned}\limsup_{\gamma\to\infty}\frac{\mathbb{V}\mathrm{ar}(\widehat{\beta}_n(\gamma))}{\beta_n^2(\gamma)} &\leq \limsup_{\gamma\to\infty}\frac{\max_{|I|=n}\mathbb{P}(B_I)\,\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]}(\gamma))}{\beta_n^2(\gamma)}\\ &\leq c\limsup_{\gamma\to\infty}\frac{\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]}(\gamma))}{\beta_n(\gamma)}\,.\end{aligned}$$

Since $\widehat{\beta}_n^{[0]}$ is an estimator in crude form then $\mathbb{V}\mathrm{ar}(\widehat{\beta}_n^{[0]}(\gamma)) = \mathcal{O}(\beta_n(\gamma))$ as $\gamma\to\infty$, so the proof is complete. □ □

Corollary 7.4.8. *The estimator* $\widehat{(\beta_1 \ddagger \alpha)}$ *from* (7.18) *has BRE.*

7.4.9 Efficiency for identical marginals and dependence

In this and the following subsections, we concentrate on rare events associated to extremes. More precisely, we let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an arbitrary random vector and define $M = \max_i X_i$. Therefore, we define $A_i(\gamma) = \{X_i > \gamma\}$ implying that the event of interest A is equivalent to $\{M > \gamma\}$.

In this subsection, we assume the X_i have identical marginal distributions. This simplifies the condition for BRE of $\widehat{\alpha}_1$, (7.25), so that it is now solely determined by the *copula* of $\boldsymbol{X}$. We investigate some common tail dependence measures of copulas (tail dependence parameter and residual tail index) and also some common families of copulas (Archimedean copulas) to see when the estimator $\widehat{\alpha}_1$ exhibits efficiency.

Asymptotic dependence

The most basic measurement of tail dependence between a pair (X_i, X_j) with common marginal distribution F and copula C_{ij} (cf. [86, 109]) is

$$\lambda_{ij} = \lim_{\upsilon\to 1}\mathbb{P}(X_i > \upsilon \mid X_j > \upsilon) = \lim_{\upsilon\to 1}\frac{1 - 2\upsilon + C_{ij}(\upsilon,\upsilon)}{1-\upsilon}$$

where $\lambda_{ij} \in [0,1]$ is called the *(upper) tail dependence parameter (or coefficient)* [86, 103]. We say the (X_i, X_j) pair exhibit *asymptotic independence* (AI) when $\lambda_{ij} = 0$, or *asymptotic dependence* (AD) when $\lambda_{ij} > 0$. The canonical examples given for each case are the (non-degenerate) bivariate normal distribution for AI, and the bivariate Student t distribution for AD [128].

For $\widehat{\alpha}_1$ to have BRE, all pairs in $\boldsymbol{X}$ must exhibit AI. This is a necessary but not sufficient condition, therefore we will employ a more refined tail dependence measurement.

Residual tail index

We must first define two classes of functions:

- $L(x)$ is *slowly-varying* (at ∞) if $L(cx)/L(x) \to 1$ as $x \to \infty$ for all $c > 0$,
- $f(x)$ is *regularly-varying* (at ∞) with index $\tau > 0$ if it takes the form $f(x) = L(x)x^{-\tau}$ for some $L(x)$ which is slowly-varying (cf. [32, 117]).

We will assume, w.l.o.g., the marginals of $\boldsymbol{X}$ to be unit Fréchet distributed (i.e., $F_1(x) = \exp(-x^{-1}) \sim 1 - x^{-1}$). Ledford and Tawn [93, 94, 95] first noted that the joint survivor functions for a wide array of bivariate distributions satisfy

$$\mathbb{P}(X_i > \gamma, X_j > \gamma) \sim L(\gamma)\gamma^{-1/\eta} \qquad \text{as } \gamma \to \infty \tag{7.28}$$

for a slowly-varying $L(\gamma)$ and an $\eta \in (0, 1]$. In other words, (7.28) says that $\mathbb{P}(X_i > \gamma, X_j > \gamma)$ is regularly-varying with index $1/\eta$.

The index is called the *residual tail index* [49, 111].[1] When (X_i, X_j) exhibit AD (AI) then we typically have $\eta = 1$ ($\eta < 1$).[2] For independent components we have $\eta = 1/2$, so Ledford and Tawn [93] describe bivariate distributions with $\eta = 1/2$ as having *near independence*. When $\eta < 1/2$ the random pair take large values together less frequently than they would if independent.

Returning to our original problem of estimating $\alpha(\gamma)$, let us label the residual tail index for every (X_i, X_j) pair of $\boldsymbol{X}$ as η_{ij}. Also, let $\eta = \max_{ij} \eta_{ij}$ and L be the associated slowly varying function. The following proposition outlines how these values relate to efficiency of $\widehat{\alpha}_1$:

Proposition 7.4.10. *If* (7.28) *is satisfied for the maximal pair of* $\boldsymbol{X}$*, that is,*

$$\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) \sim L(\gamma)\gamma^{-1/\eta} \qquad \text{as } \gamma \to \infty\,,$$

then the estimator $\widehat{\alpha}_1$ *has: i) BRE if* $\eta < 1/2$ *or if* $\eta = 1/2$ *and* $L(\gamma) \not\to \infty$ *as* $\gamma \to \infty$*, ii) LE if* $\eta = 1/2$*.*

Proof. Label the components of $\boldsymbol{X}$ such that

$$\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) = \mathbb{P}(X_1 > \gamma, X_2 > \gamma)$$

[1] The older (and less insightful) name for η is the *coefficient of tail dependence* [93, 116].

[2] Hashorva [75] has found a case where an elliptically distributed (X_i, X_j) has $\eta = 1$ and AI.

Table 7.1: Residual tail dependence index η and $L(x)$ for various copulas. This is a subset of Table 1 of [79] (their row numbers are preserved).

(a) Copulas with BRE.

#	**Name**	η	$L(x)$
1	Ali-Mikhail-Haq	0.5	$1+\tau$
2	BB10 in Joe	0.5	$1+\theta/\tau$
3	Frank	0.5	$\delta/(1-\mathrm{e}^{-\delta})$
4	Morgenstern	0.5	$1+\tau$
5	Plackett	0.5	δ
6	Crowder	0.5	$1+(\theta-1)/\tau$
7	BB2 in Joe	0.5	$\theta(\delta+1)+1$
8	Pareto	0.5	$1+\delta$
9	Raftery	0.5	$\delta/(1-\delta)$

(b) Copulas without BRE.

#	**Name**	η	$L(x)$
11	Joe	1	$2-2^{1/\delta}$
12	BB8 in Joe	1	$2-2(1-\delta)^{\theta-1}$
13	BB6 in Joe	1	$2-2^{1/(\delta\theta)}$
14	Extreme value	1	$2-V(1,1)$
15	B11 in Joe	1	δ
16	BB1 in Joe	1	$2-2^{1/\delta}$
17	BB3 in Joe	1	$2-2^{1/\theta}$
18	BB4 in Joe	1	$2^{-1/\delta}$
19	BB7 in Joe	1	$2-2^{1/\theta}$

then the condition for LE becomes, $\forall \varepsilon > 0$

$$\limsup_{\gamma\to\infty} \frac{\max_{i<j} \mathbb{P}(X_i \geq \gamma, X_j \geq \gamma)}{\max_k \mathbb{P}(X_k \geq \gamma)^{2-\varepsilon}} = \limsup_{\gamma\to\infty} \frac{L(\gamma)\gamma^{-1/\eta}}{(\gamma^{-1})^{2-\varepsilon}} = \limsup_{\gamma\to\infty} L(\gamma)\gamma^{2-\frac{1}{\eta}-\varepsilon} = 0$$

which is equivalent to $\eta \in (0, 1/2]$; the $\eta = 1/2$ case has LE as $\gamma^{-\varepsilon}L(\gamma) \to 0$ for all $\varepsilon > 0$ (see Proposition 1.3.6 part (v) of [32]). Similarly we have BRE for $\eta \in (0, 1/2)$, but for the $\eta = 1/2$ case we also require that $L(\gamma) \not\to \infty$. □ □

Heffernan [79] has conveniently compiled a directory of η and $L(x)$ for many copulas which satisfy (7.28). A summary of these results is given in Table 7.1. In reading Heffernan's directory, one can spot two trends: normally $\eta \in \{1/2, 1\}$ and L is a constant. The oft-cited Gaussian copula is the only exception for both of these trends in Heffernan's directory, having $\eta = (1+\rho)/2$ and $L(x) \propto (\log x)^{-\rho/(1+\rho)}$; Section 7.4.13 deals with the Gaussian case in detail.

Archimedean Copulas

Some of the most frequently used copulas are in the family of *Archimedean copulas*. These are very general models and are widely used in applications due to their flexibility. A copula is Archimedean if there exists a function ψ such that the copula C can be written as

$$C(u_1, \ldots, u_n) = \psi^{\leftarrow}(\psi(u_1) + \cdots + \psi(u_n)).$$

The function ψ, called the *generator* of the copula, defines a copula if its functional inverse is the Laplace transform of a non-negative random variable. For Archimedean copulas we can restate the BRE condition (7.25) in terms of the generator ψ.

Theorem 7.4.11 (Thm. 3.4 of [38])**.** *Let $(U_1, \ldots, U_n) \sim C$ where C is an Archimedean copula with generator ψ. If $\psi^{\leftarrow}$ is twice continuously differentiable and its second derivative is bounded at 0 then $\forall\, i \neq j$*

$$\lim_{u \to 0} \frac{\mathbb{P}(U_i \geq 1 - ux_1, U_j \geq 1 - ux_2)}{u^2} < \infty$$

for any $0 < x_1,\, x_2 < \infty$.

Corollary 7.4.12. *Consider using $\widehat{\alpha}_1$ for a distribution with common marginal distributions and a copula C. If C satisfies the conditions of Theorem 7.4.11 then $\widehat{\alpha}_1$ has BRE.*

Charpentier and Segers [38] have helpfully created a directory of Archimedean copulas from which we can see if the BRE conditions from Corollary 7.4.12 are satisfied. Using this information, we provide a summary of the efficiency status of many Archimedean copulas in Table 7.2.

The efficiency of $\widehat{\alpha}_1$ can be proved without the assumption of identical marginal distributions, but the efficiency must be shown case-by-case for each family of distributions. The next section does this for the multivariate normal distribution and for some elliptical distributions.

7.4.13 Efficiency for the case of normal and elliptical distributions

The efficiency characteristics of normally and elliptically distributed random vectors are very similar. This section defines these distributions, outlines their asymptotic properties, then shows the conditions in which $\widehat{\alpha}_1$ exhibits levels of asymptotic efficiency.

Definitions and categories of elliptical distributions

Let $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the multivariate normal distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^d$ and positive-definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$. Denote the corresponding density $\varphi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\cdot, \cdot)$, and write $\sigma_i^2 := \boldsymbol{\Sigma}_{ii}$, $\rho_{ij} := \boldsymbol{\Sigma}_{ij}/(\sigma_i \sigma_j)$. The normal distribution belong to the class of *elliptical*

Table 7.2: Examples of Archimedean copula families. Names (if they are named) and generator functions are listed, as are the ranges for which θ is valid and the subset of θ which ensures that $\widehat{\alpha}_1$ has BRE. A Θ in the final column means that all valid θ ensure BRE. The families listed appear in Table 4.1 of [109] and Table 1 of [38].

#	**Name**	**Generator** $\psi(t)$	**Valid** θ	**Efficient** θ
1	Clayton	$\frac{1}{\theta}(t^{-\theta}-1)$	$[-1,\infty)$	Θ
2		$(1-t)^{\theta}$	$[1,\infty)$	$\{1\}$
3	Ali–Mikhail–Haq	$\log\frac{1-\theta(1-t)}{t}$	$[-1,1)$	Θ
4	Gumbel–Hougaard	$(-\log t)^{\theta}$	$[1,\infty)$	$\{1\}$
5	Frank	$-\log\frac{\mathrm{e}^{-\theta t}-1}{\mathrm{e}^{-\theta}-1}$	$\mathbb{R}$	$\Theta\setminus\{0\}$
6		$-\log[1-(1-t)^{\theta}]$	$[1,\infty)$	$\{1\}$
7		$-\log[\theta t+(1-\theta)]$	$(0,1]$	Θ
8		$\frac{1-t}{1+(\theta-1)t}$	$[1,\infty)$	Θ
9		$\log(1-\theta\log t)$	$(0,1]$	Θ
10		$\log(2t^{-\theta}-1)$	$(0,1]$	Θ
11		$\log(2-t^{\theta})$	$(0,1/2]$	Θ
12		$(\frac{1}{t}-1)^{\theta}$	$[1,\infty)$	$\{1\}$
13		$(1-\log t)^{\theta}-1$	$(0,\infty)$	Θ
14		$(t^{-1/\theta}-1)^{\theta}$	$[1,\infty)$	$\{1\}$
15		$(1-t^{1/\theta})^{\theta}$	$[1,\infty)$	$\{1\}$
16		$(\frac{\theta}{t}+1)(1-t)$	$[0,\infty)$	Θ
17		$-\log\frac{(1+t)^{-\theta}-1}{2^{-\theta}-1}$	$\mathbb{R}$	$\Theta\setminus\{0\}$
18		$\mathrm{e}^{\theta/(t-1)}$	$[2,\infty)$	$\emptyset$
19		$\mathrm{e}^{\theta/t}-\mathrm{e}^{\theta}$	$(0,\infty)$	Θ
20		$\mathrm{e}^{t^{-\theta}}-\mathrm{e}$	$(0,\infty)$	Θ
21		$1-[1-(1-t)^{\theta}]^{1/\theta}$	$[1,\infty)$	$\{1\}$
22		$\arcsin(1-t^{\theta})$	$(0,1]$	Θ

distributions, which we denote $\mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$, where F is the c.d.f. of a positive r.v. We define $\boldsymbol{X} \sim \mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ as

$$\boldsymbol{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + R\,\boldsymbol{C}\,U \tag{7.29}$$

where $R \sim F$ is called the *radial component*, U is (independent of R and) distributed uniformly on the d-dimensional unit hypersphere, and $\boldsymbol{C} \in \mathbb{R}^{d\times d}$ satisfies $\boldsymbol{C}\boldsymbol{C}^\top = \boldsymbol{\Sigma}$. For background on elliptical distributions, see [11]. The efficiency of $\widehat{\alpha}_1$ turns out to be related with max-domain of attraction (MDA) of the radial component. The MDA is known from standard extreme value theory, see [49].

We consider some subclasses of elliptical distributions depending on the MDA of the radial distribution:

- $F \in$ MDA(Fréchet), then Theorem 4.3 of [83] implies that $\boldsymbol{X}$ has asymptotic dependence and $\widehat{\alpha}_1$ is never efficient (see Section 7.4.9).
- $F \in$ MDA(Weibull), then components of $\boldsymbol{X}$ are light-tailed and uninteresting (in a rare-event context).
- $F \in$ MDA(Gumbel), this is the interesting case which includes the normal distribution. Hashorva [74] label these the *type I elliptical random vectors*.

Efficiency for type I elliptical distributions

Take $\boldsymbol{X} \sim \mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ where the radial distribution $F \in$ MDA(Gumbel) has support $(0, x_F)$, for some $x_F \in \overline{\mathbb{R}}$, and where $\{\sigma_1, \ldots, \sigma_d\}$ are in decreasing order. By definition of the Gumbel MDA, one can find a scaling function $w(x)$ satisfying

$$\lim_{x\to x_F} \frac{\overline{F}(x + t/w(x))}{\overline{F}(x)} = \mathrm{e}^{-t}.$$

One frequently takes $w(x) := \overline{F}(x)/\int_x^{x_F} \overline{F}(s)\mathrm{d}s$. Also, define $\upsilon_i(\gamma) := (\gamma - \mu_i)/\sigma_i$ and $a_{ij} := \sigma_j/\sigma_i$. If $\rho_{ij} \geq a_{ij}$ then set

$$\mu_{ij} := \mu_j \quad \text{and} \quad \kappa_{ij} := \sigma_j$$

otherwise for $\rho_{ij} < a_{ij}$

$$\mu_{ij} := \frac{\mu_i - a_{ij}\rho_{ij}(\mu_1 + \mu_2) + a^2\mu_j}{\alpha_{ij}(1-\rho_{ij}^2)} \quad \text{and} \quad \kappa_{ij} := \frac{\sigma_i^2\sigma_j^2(1-\rho_{ij}^2)}{\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j + \sigma_j^2}\,.$$

We now apply the asymptotic properties outlined in the Appendix to assess the efficiency of $\widehat{\alpha}_1$ for type I elliptical distributions.

Theorem 7.4.14. *Consider* $\boldsymbol{X} \sim \mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ *where* $F \in$ *MDA(Gumbel), and let*

$$\kappa := \max_{i<j} \kappa_{ij}\,, \quad \mu := \max_{i<j\,:\,\kappa=\kappa_{ij}} \mu_{ij}\,, \quad \textit{and} \quad \upsilon(\gamma) := (\gamma-\mu)/\kappa + \mathrm{o}(1)\,.$$

If $\kappa > \sigma_1$,[1] *then* $\widehat{\alpha}_1$ *has LE if*

$$\forall \varepsilon > 0 \quad \limsup_{\gamma \to x_F} \frac{w(\upsilon(\gamma))\overline{F}(\upsilon(\gamma))}{w(\upsilon_1(\gamma))\overline{F}(\upsilon_1(\gamma))^{2-\varepsilon}} < \infty\,. \tag{7.30}$$

Moreover, if (7.30) *holds for* $\varepsilon = 0$ *then* $\widehat{\alpha}_1$ *has BRE.*

Proof. It follows from (7.33) and Theorem 7.A.3 in the Appendix. □ □

Example 7.4.15 (Kotz Type III)**.** *One family of type I elliptical distributions, is the* Kotz Type III *distributions, defined by*

$$\overline{F}(\gamma) = (K + \mathrm{o}(1))\gamma^N \exp(-ru^\delta), \quad w(\gamma) = r\delta\gamma^{\delta-1}\,, \quad \textit{for } \gamma > 0,$$

with $K, \delta, N > 0$*. In this case it is clear that*

$$\lim_{\gamma\to\infty} \frac{w(\upsilon(\gamma))}{w(\upsilon_1(\gamma))} = \left(\frac{\sigma_1}{\kappa}\right)^{\delta-1} < \infty,$$

while

$$\begin{aligned}
&\limsup_{\gamma\to\infty} \frac{\overline{F}(\upsilon(\gamma))}{\overline{F}(\upsilon_1(\gamma))^2} \\
&= \limsup_{\gamma\to\infty} \left(\frac{\sigma_1^2}{\kappa\gamma}\right)^N \exp\left\{-r\left(\left(\frac{\gamma-\mu}{\kappa}\right)^\delta - 2\left(\frac{\gamma-\mu_1}{\sigma_1}\right)^\delta\right)\right\}, \\
&= \limsup_{\gamma\to\infty} \left(\frac{\sigma_1^2}{\kappa\gamma}\right)^N \exp\left\{-r\left(\frac{\gamma^\delta - \delta\mu\gamma^{\delta-1} + \mathrm{o}(\gamma^{\delta-1})}{\kappa^\delta} - \frac{\gamma^\delta - \delta\mu_1\gamma^{\delta-1} + \mathrm{o}(\gamma^{\delta-1})}{\sigma_1^\delta/2}\right)\right\}.
\end{aligned}$$

Hence, $\widehat{\alpha}_1$ *has BRE in the following cases*

- $\sigma_1^\delta > 2\kappa^\delta$*, or*
- $\sigma_1^\delta = 2\kappa^\delta$*,* $\delta > 1$ *and* $\mu_1 > \mu$*.*

The estimator $\widehat{\alpha}_1$ *has LE if* $\sigma_1^\delta = 2\kappa^\delta$*, and is inefficient when* $\sigma_1^\delta < 2\kappa^\delta$*.*

Example 7.4.16 (Normal distributions)**.** *The normal distribution is a Kotz III type distribution with* $\delta = 2$*. Hence,* $\widehat{\alpha}_1$ *has BRE if* $\sigma_1^2 > 2\kappa^2$*, or* $\sigma_1^2 = 2\kappa^2$ *and* $\mu_1 > \mu$*. The estimator* $\widehat{\alpha}_1$ *has LE if* $\sigma_1^2 = 2\kappa^2$*, and is inefficient when* $\sigma_1^2 < 2\kappa^2$*.*

[1] This implies that the Savage condition (see Appendix) is fulfilled at least for one pair.

Frequently, a set of random variables represents as a stochastic process $\{X_n\}_{n\geq 0}$. The value of $\mathbb{P}(M > \gamma)$, with $M := \max_{1\leq n\leq d} X_n$, in such cases usually valuable. The simplest case to take is when all X_n have identical marginals such as in stationary processes; one such example is the autoregressive (AR) process.

Example 7.4.17 (AR(1) processes)**.** *Say $X_t = \varphi X_{t-1} + \varepsilon_t$, where $|\varphi| < 1$ and ε_t are i.i.d. $\mathcal{N}_1(0, \sigma_\varepsilon^2)$, and we start the process in stationarity. We have that each X_i has the same marginal distribution, $X_i \sim \mathcal{N}_1(0, \sigma_\varepsilon^2/(1-\varphi^2))$, and*

$$\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma) = \begin{cases} \mathbb{P}(X_\bullet > \gamma, X_{\bullet+1} > \gamma) & \text{if } \varphi > 0 \\ \mathbb{P}(X_\bullet > \gamma, X_{\bullet+2} > \gamma) & \text{if } \varphi < 0 \\ \mathbb{P}(X_\bullet > \gamma)^2 & \text{if } \varphi = 0 \end{cases}.$$

For $\varphi \neq 0$ we know that

$$(X_{\bullet+1} \mid X_\bullet = \gamma) \sim \mathcal{N}_1(\varphi\gamma, \sigma_\varepsilon^2), \text{ and } (X_{\bullet+2} \mid X_\bullet = \gamma) \sim \mathcal{N}_1(\varphi^2\gamma, \sigma_\varepsilon^2(1-\varphi^4)/(1-\varphi^2)).$$

When $\varphi = 0$ the X_i are independent and $\widehat{\alpha}_1$ is trivially efficient, and when $\varphi \in (-1,1)\backslash\{0\}$ we have (noting that $\{X_\bullet > \gamma\} \to \{X_\bullet = \gamma\}$) that

$$\begin{aligned} \lim_{\gamma\to\infty} \frac{\max_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma)}{\max_i \mathbb{P}(X_i > \gamma)^2} &= \lim_{\gamma\to\infty} \frac{\mathbb{P}(X_\bullet > \gamma, X_{\bullet+(1\text{ or }2)} > \gamma)}{\mathbb{P}(X_\bullet > \gamma)^2} \\ &= \lim_{\gamma\to\infty} \frac{\mathbb{P}(X_{\bullet+(1\text{ or }2)} > \gamma \mid X_\bullet = \gamma)}{\mathbb{P}(X_\bullet > \gamma)} \\ &= 0 \end{aligned}$$

as $\sigma_\varepsilon^2 < \sigma_\varepsilon^2(1-\varphi^4)/(1-\varphi^2)) < \sigma_\varepsilon^2/(1-\varphi^2)$. Therefore, we have BRE of $\widehat{\alpha}_1$ for all stationary AR(1) processes.

7.5 Numerical experiments

We explore the performance of the estimators for the problem of $\mathbb{P}(M > \gamma)$ for $M = \max_i X_i$, where $\boldsymbol{X}$ is multivariate normal and multivariate Laplace distributed. The following notation is used: $\boldsymbol{X}_{-i}$ ($\boldsymbol{X}_{-i,-j}$) is the random vector $\boldsymbol{X}$ with X_i (X_i and X_j) removed, $\boldsymbol{0}$ is the vector of zeros, $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{x}^\top$ is the transpose of $\boldsymbol{x}$, and $X \perp\!\!\!\perp Y$ means X and Y are independent. We use some standard distributions: $\mathcal{E}(\lambda)$ for exponential ($f(x) \propto \mathrm{e}^{-\lambda x}$), $\mathcal{IG}(\mu, \lambda)$ for inverse Gaussian ($f(x) \propto x^{-3/2}\mathrm{e}^{-\lambda(x-\mu)^2/(2\mu^2 x)}$), $\mathcal{L}$ for Laplace (defined in Case 2 below). The MATLAB and MATHEMATICA code used to generate them are available online [9].

Case 1: Multivariate Normal distributions

Let $\boldsymbol{X} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I} + \rho$; that is, each $X_i \sim \mathcal{N}_1(0,1)$ and $\mathbb{C}\mathrm{orr}(X_i, X_j) = \rho$. We implement the first- and second-order IS regimes. The necessary conditional distributions are well-known and simple; both $\boldsymbol{X}_{-i} \mid X_i$ and $\boldsymbol{X}_{-i,-j} \mid (X_i, X_j)$ are normally distributed [10]. Sampling from $X_i \mid X_i > \gamma$ can be easily done by acceptance–rejection with shifted exponential proposals [118] (or by inverse transform sampling [15, Remark 2.4], though this can be problematic using only double precision arithmetic). To simulate $(X_i, X_j) \mid \min\{X_i, X_j\} > \gamma$ we use Botev's MATLAB library [33], but also remark that a Gibb's sampler is a commonly used alternative [35, 118].

Case 2: Multivariate Laplace distributions

Let $\boldsymbol{X} \sim \mathcal{L}$. We can define this distribution by

$$\boldsymbol{X} \overset{\mathcal{D}}{=} \sqrt{R}\boldsymbol{Y}\,, \quad \text{where } \boldsymbol{Y} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{I}), R \sim \mathcal{E}(1), \boldsymbol{Y} \perp\!\!\!\perp R\,.$$

The distribution has been applied in a financial context [82], and is examined in [56, 89]. From the former we have that the density of $\mathcal{L}$ is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = 2(2\pi)^{-d/2} K_{(d/2)-1}\Big(\sqrt{2\boldsymbol{x}^\top \boldsymbol{x}}\Big)\left(\sqrt{\tfrac{1}{2}\boldsymbol{x}^\top \boldsymbol{x}}\right)^{1-(d/2)}$$

where $K_n(\cdot)$ denotes the modified Bessel function of the second kind of order n.

To implement the first-order IS algorithm we need the conditional distributions $X_i \mid X_i > \gamma$ and $\boldsymbol{X}_{-i} \mid \boldsymbol{X}_i$. Assuming $\gamma > 0$ we can derive that $(X_i \mid X_i > \gamma) \sim \mathcal{E}(\sqrt{2})$. Further calculation gives

$$\boldsymbol{X}_{-1} \mid X_1 \;\overset{\mathcal{D}}{=}\; \frac{X_1}{Y_1}\boldsymbol{Y}_{-1} \mid (\sqrt{R}Y_1 = X_1) \;\overset{\mathcal{D}}{=}\; \frac{X_1}{Y_{1,X_1}}\boldsymbol{Y}_{-1}\,,$$

where $Y_{1,X_1} \sim (Y_1 \mid \sqrt{R}Y_1 = X_1)$, noting that $Y_{1,X_1} \perp\!\!\!\perp \boldsymbol{Y}_{-1}$ because of the independence between the entries of $\boldsymbol{Y}$. Direct calculation gives

$$f_{Y_i \mid \sqrt{R}Y_i}(y_i \mid x_i) = 2\,|y_i| \exp\left\{-x_i^2/y_i^2 - x_i^2/2 + \sqrt{2}\,|x_i|\right\}/(\sqrt{\pi}y_i^2)$$

which is the density of $\sqrt{X}$ where $X \sim \mathcal{IG}(\sqrt{2}|x_i|, 2x_i^2)$. This is summarised in the following algorithm.

Algorithm 1 Sampling $\boldsymbol{X}_{-i} \mid X_i > \gamma$ for the Laplace distribution

1: $X_i \leftarrow \mathcal{E}(\sqrt{2})$
2: $Y_{i,X_i} \leftarrow \mathcal{IG}(\sqrt{2}|X_i|, 2X_i^2)$.
3: $\boldsymbol{Y}_{-i} \leftarrow \mathcal{N}_{d-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1})$.
4: **return** $X_i \boldsymbol{Y}_{-i} / Y_{i,X_i}$.

7.5.1 Test setup

The estimators tested are $\widehat{\alpha}_0$ (crude Monte Carlo) and $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\alpha}_1^{[1]}$, $\widehat{\alpha}_2^{[2]}$, $\widehat{(\beta_1 \ddagger \alpha)}$, $\widehat{(\beta_2 \ddagger \alpha)}$, defined in (7.1), (7.2), (7.12), (7.13), (7.18) and (7.20) respectively. As a reference, we show the true value α (calculated by numerical integration using MATHEMATICA), and the first two truncations of the IEF: $\overline{\alpha}(\gamma) := \sum_i \mathbb{P}(X_i > \gamma)$ and $\overline{\alpha}(\gamma) - q(\gamma) := \sum_i \mathbb{P}(X_i > \gamma) - \sum_{i<j} \mathbb{P}(X_i > \gamma, X_j > \gamma)$. Each estimator is given $R = 10^6$, and an asterisk is placed in table entries where the corresponding estimate had 0 variance (i.e., the estimator had degenerated).

7.5.2 Results

Estimators	γ			
	2	4	6	8
α	5.633e-02	1.095e-04	3.838e-09	2.481e-15
$\widehat{\alpha}_0$	5.651e-02	1.140e-04	0*	0*
$\overline{\alpha}$	9.100e-02	1.267e-04	3.946e-09	2.488e-15
$\overline{\alpha}-q$	4.000e-02	1.055e-04	3.827e-09	2.480e-15
$\widehat{\alpha}_1$	5.650e-02	1.047e-04	3.946e-09*	2.488e-15*
$\widehat{\alpha}_2$	5.605e-02	1.075e-04	3.827e-09*	2.480e-15*
$\widehat{\alpha}_1^{[1]}$	5.637e-02	1.096e-04	3.837e-09	2.481e-15
$\widehat{\alpha}_2^{[2]}$	5.633e-02	1.095e-04	3.838e-09	2.481e-15
$\widehat{(\beta_1 \ddagger \alpha)}$	5.634e-02	1.095e-04	3.838e-09	2.480e-15
$\widehat{(\beta_2 \ddagger \alpha)}$	5.631e-02	1.095e-04	3.838e-09	2.481e-15

Table 7.3: Estimates of $\mathbb{P}(M > \gamma)$ where $M = \max_i X_i$ and $\boldsymbol{X} \sim \mathcal{N}_4(\mathbf{0}_4, \boldsymbol{\Sigma})$, $\rho = 0.75$.

Estimators	γ			
	2	4	6	8
$\widehat{\alpha}_0$	3.109e-03	4.075e-02	1*	1*
$\overline{\alpha}$	6.154e-01	1.566e-01	2.822e-02	3.142e-03
$\overline{\alpha}-q$	2.899e-01	3.665e-02	2.827e-03	1.147e-04
$\widehat{\alpha}_1$	2.977e-03	4.429e-02	2.822e-02*	3.142e-03*
$\widehat{\alpha}_2$	5.077e-03	1.839e-02	2.827e-03*	1.147e-04*
$\widehat{\alpha}_1^{[1]}$	6.918e-04	4.639e-04	1.747e-04	2.192e-05
$\widehat{\alpha}_2^{[2]}$	7.838e-08	8.647e-05	1.237e-05	4.010e-08
$\widehat{(\beta_1 \ddagger \alpha)}$	6.564e-05	7.046e-05	6.227e-05	4.362e-05
$\widehat{(\beta_2 \ddagger \alpha)}$	3.493e-04	1.593e-05	6.883e-06	3.340e-07

Table 7.4: Absolute relative errors of the estimates in Table 7.3.

Estimators	γ			
	2	4	6	8
$\widehat{\alpha}_0$	2.309e-01	1.068e-02	0	0
$\widehat{\alpha}_1$	2.557e-01	5.099e-03	0	0
$\widehat{\alpha}_2$	1.885e-01	1.414e-03	0	0
$\widehat{\alpha}_1^{[1]}$	2.817e-02	3.071e-05	4.650e-10	9.972e-17
$\widehat{\alpha}_2^{[2]}$	9.901e-03	4.244e-06	1.908e-11	8.575e-19
$\widehat{(\beta_1 \ddagger \alpha)}$	1.929e-02	2.089e-05	3.197e-10	6.994e-17
$\widehat{(\beta_2 \ddagger \alpha)}$	1.306e-02	5.265e-06	2.310e-11	1.035e-18

Table 7.5: Standard deviations of the estimates in Table 7.3.

Estimators	γ			
	6	8	10	12
α	4.093e-04	2.435e-05	1.442e-06	8.526e-08
$\widehat{\alpha}_0$	3.910e-04	2.000e-05	2.000e-06	0*
$\overline{\alpha}$	4.130e-04	2.441e-05	1.443e-06	8.527e-08
$\overline{\alpha}-q$	4.093e-04	2.435e-05	1.442e-06	8.526e-08
$\widehat{\alpha}_1$	4.120e-04	2.441e-05*	1.443e-06*	8.527e-08*
$\widehat{\alpha}_2$	4.093e-04*	2.435e-05*	1.442e-06*	8.526e-08*
$\widehat{\alpha}_1^{[1]}$	4.093e-04	2.435e-05	1.442e-06	8.526e-08
$\widehat{(\beta_1 \ddagger \alpha)}$	4.093e-04	2.435e-05	1.442e-06	8.526e-08

Table 7.6: Estimates of $\mathbb{P}(M > \gamma)$ where $M = \max_i X_i$ and $\boldsymbol{X} \sim \boldsymbol{\mathcal{L}}$, $d = 4$.

Estimators	γ			
	6	8	10	12
$\widehat{\alpha}_0$	4.472e-02	1.786e-01	3.873e-01	1*
$\overline{\alpha}$	8.959e-03	2.473e-03	6.987e-04	2.003e-04
$\overline{\alpha}-q$	8.067e-05	8.266e-06	8.757e-07	9.506e-08
$\widehat{\alpha}_1$	6.516e-03	2.473e-03*	6.987e-04*	2.003e-04*
$\widehat{\alpha}_2$	8.067e-05*	8.266e-06*	8.757e-07*	9.506e-08*
$\widehat{\alpha}_1^{[1]}$	8.470e-06	1.023e-05	3.019e-05	1.577e-05
$\widehat{(\beta_1 \ddagger \alpha)}$	4.515e-05	2.948e-05	2.151e-06	2.833e-06

Table 7.7: Absolute relative errors of the estimates in Table 7.6.

Estimators	γ			
	6	8	10	12
$\widehat{\alpha}_0$	1.977e-02	4.472e-03	1.414e-03	0
$\widehat{\alpha}_1$	1.000e-03	0	0	0
$\widehat{\alpha}_2$	0	0	0	0
$\widehat{\alpha}_1^{[1]}$	2.735e-05	8.581e-07	2.752e-08	8.189e-10
$\widehat{(\beta_1 \ddagger \alpha)}$	1.937e-05	6.086e-07	1.908e-08	5.990e-10

Table 7.8: Standard deviations of the estimates in Table 7.6.

7.5.3 Discussion

We begin with some trends which we expected to find in the results:

- all estimators outperform crude Monte Carlo $\widehat{\alpha}_0$,
- the estimators which calculate $\mathbb{P}(X_i > \gamma)$ outperform those which do not,
- the estimators which calculate $\mathbb{P}(X_i > \gamma, X_j > \gamma)$ outperform those which only use the univariate $\mathbb{P}(X_i > \gamma)$,
- the importance sampling estimators improve upon their original counterparts,
- the second-order IS improves upon the first-order IS.

Also noticed in the performance of the $\widehat{\alpha}$ estimators:

- the $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ estimators often degenerated (i.e. had zero variance) to $\overline{\alpha}$ and $\overline{\alpha}-q$ respectively,
- the degeneration begin for smaller γ when the $\boldsymbol{X}$ had a weaker dependence structure.

Table 7.9 shows the degeneration of the estimators in various examples involving multivariate normal distributions.

The fact that the estimators degenerate is not wholly undesirable, as they degenerate to the deterministic functions $\overline{\alpha}$ and $\overline{\alpha}-q$ which are highly accurate when degeneration occurs. Obviously, for very large γ one would not resort to Monte Carlo methods as the asymptote $\overline{\alpha}$ would be accurate enough for most purposes; one could use the $\widehat{\alpha}$ estimators until the sample variance is below some threshold, then switch to the faster deterministic estimators $\overline{\alpha}$ and $\overline{\alpha}-q$.

Regarding the $\widehat{(\beta_1 \ddagger \alpha)}$ and $\widehat{(\beta_2 \ddagger \alpha)}$ estimators:

- their performance is roughly the same as than their $\widehat{\alpha}_1^{[1]}$ and $\widehat{\alpha}_2^{[2]}$ counterparts,
- they perform better when the dependence between the variables is weak.

One must remember that the $\widehat{\beta}_i$ estimators are valid for a much larger class of problems (estimating expectations, not just probabilities). Also, we would expect that the $\widehat{\beta}_i$-based estimators compare favorably to the $\widehat{\alpha}_i^{[i]}$ IS-based estimators when d is large, as the method involves no likelihood term which can degenerate.

Test cases		γ			
d	ρ	2	4	6	8
3	-0.25	0.00957	1*	1*	1*
	0	0.00255	1*	1*	1*
	0.5	0.00166	1*	1*	1*
	0.75	0.005	0.165	1*	1*
4	-0.25	0.00955	1*	1*	1*
	0	0.0185	1*	1*	1*
	0.5	0.00139	1*	1*	1*
	0.75	0.00484	0.283	1*	1*
Average		0.00663	0.806	1	1

(a) $\widehat{\alpha}_1$ to $\overline{\alpha}$

Test cases		γ			
d	ρ	2	4	6	8
3	-0.25	1*	1*	1*	1*
	0	0.151*	1*	1*	1*
	0.5	0.0764	1*	1*	1*
	0.75	0.0172	0.754	1*	1*
4	-0.25	1*	1*	1*	1*
	0	0.189	1*	1*	1*
	0.5	0.0153	1*	1*	1*
	0.75	0.0175	0.502	1*	1*
Average		0.308	0.907	1	1

(b) $\widehat{\alpha}_2$ to $\overline{\alpha}-q$

Table 7.9: Ratios of absolute relative errors for pairs of estimators. Each row corresponds to a separate distribution for $\boldsymbol{X}$, each being $\mathcal{N}_d$ distributed with standard normal marginals and constant correlation ρ.

7.6 Conclusion

In this paper we presented new estimators for the tail probability of a union of dependent rare events. The key idea in both estimators is that the tail probability of the such a rare event can be well approximated by the Bonferroni approximations:

$$\alpha = \mathbb{P}(A) \approx \sum_{i=1}^{k} (-1)^{i-1} \sum_{|I|=i} \mathbb{P}\left(\bigcap_{i \in I} A_i \right) \text{ for } k = 1, 2\,.$$

We provided conditions which ensure $\widehat{\alpha}_1$ and $\widehat{\beta}_i$ have logarithmic efficiency and bounded relative error. The estimators were tested on the classical example of rare maxima of random vectors. Furthermore, we note the fact that our $\widehat{\beta}_i$ estimators can be applied to a more general setting which could make useful for a larger variety of estimation problems.

7.6.1 Future work

In this paper we did not discuss stratification strategies for $\widehat{\beta}_i$ that could result in further reductions in variance. Nor did we investigate which permutations of the A_i minimise

the variance of $\widehat{\beta}_i$. Further investigation into the use of $\widehat{\beta}_i$ to estimate tail probabilities of order statistics would be of value.

7.A Elliptical distribution asymptotics

7.A.1 Asymptotic properties of normal distributions

In general, for an $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma})$, Theorem 2.6.1 of Bryc [36] states that for all measurable $A \subset \mathbb{R}^d$ the

$$\lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}(\boldsymbol{X} \geq nA) = -\inf_{\boldsymbol{x}\in A} \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}. \tag{7.31}$$

The asymptotic properties of elliptical distributions also relate to this quadratic programming problem, which Hashorva [73, 74] denotes as

$$\mathcal{P}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{t}) := \text{minimise } \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x} \text{ under the linear constraint } \boldsymbol{x} \geq \boldsymbol{t}. \tag{7.32}$$

The program $\mathcal{P}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{t})$ is usually minimised at the boundary $\boldsymbol{t}$, and hence the asymptotic form (7.31) is very simple. This occurs when $\boldsymbol{\Sigma}^{-1}\boldsymbol{t} > \boldsymbol{0}$ (componentwise), a condition often called the *Savage condition* after Richard Savage [126]. For the cases when the Savage condition fails, the asymptotics change as some components of $\boldsymbol{X}$ become irrelevant in the limit. Figure 7.1 graphically shows some contours of $\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}$ for some $\boldsymbol{\Sigma}$ which do and do not satisfy the Savage condition.

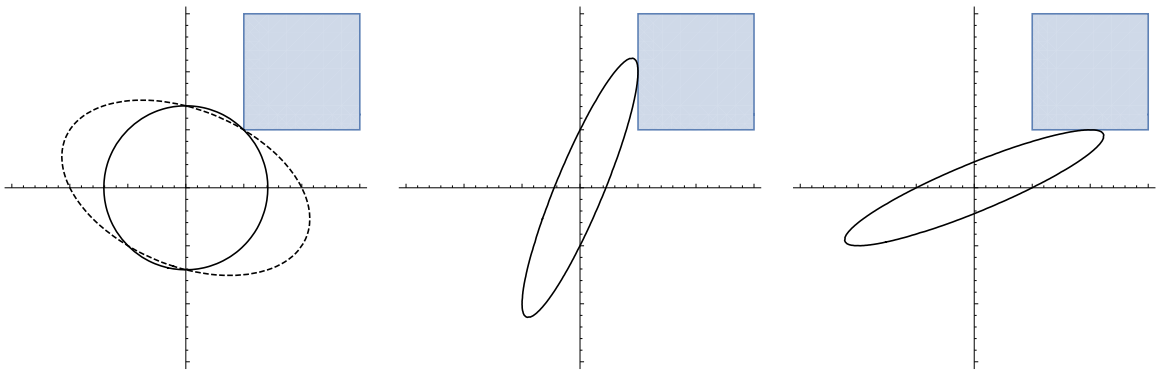


Figure 7.1: Contours of $\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}$ for example $\boldsymbol{\Sigma}$ which: (a) satisfy the Savage condition (i.e., $\boldsymbol{\Sigma}^{-1}\boldsymbol{1} > \boldsymbol{0}$), and (b)–(c) do not satisfy the condition. The covariance matrices, in MATLAB notation, are: (a) $\boldsymbol{\Sigma} = \boldsymbol{I}$ and $\boldsymbol{\Sigma} = [2, -1/2; -1/2, 1]$, (b) $\boldsymbol{\Sigma} = [1, 2; 2, 5]$, and (c) $\boldsymbol{\Sigma} = [5, 2; 2, 1]$.

7.A.2 Asymptotic properties of type I elliptical distributions

Take $\boldsymbol{X} \sim \mathcal{ELL}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ where the radial distribution $F \in$ MDA(Gumbel) has support $(0, x_F)$, for some $x_F \in \overline{\mathbb{R}}$, and where $\{\sigma_1, \ldots, \sigma_d\}$ are in decreasing order. The univariate and bivariate asymptotics, $\mathbb{P}(X_i > \gamma)$ and $\mathbb{P}(X_i > \gamma, X_j > \gamma)$, can be written in terms of the scaling function $w(\gamma)$ and of $\overline{F}((\gamma - \mu)/\kappa)$ for some particular μ and κ. Theorem 12.3.1 of Berman [31] gives the univariate case,

$$\mathbb{P}(X_i > \gamma) = (1 + \mathrm{o}(1)) \frac{\overline{F}(\upsilon_i(\gamma))}{\sqrt{2\pi \upsilon_i(\gamma) w(\upsilon_i(\gamma))}} \quad \text{as } \gamma \to x_F \tag{7.33}$$

where $\upsilon_i(\gamma) = (\gamma - \mu_i)/\sigma_i$. The bivariate case, i.e. $\mathbb{P}(X_i > \gamma, X_j > \gamma)$, relies on the following constants. Define $a_{ij} := \sigma_j/\sigma_i$. If $\rho_{ij} \geq a_{ij}$ then define

$$\mu_{ij} := \mu_j \quad \text{and} \quad \kappa_{ij} := \sigma_j$$

otherwise for $\rho_{ij} < a_{ij}$

$$\mu_{ij} := \frac{\mu_i - a_{ij}\rho_{ij}(\mu_1 + \mu_2) + a^2 \mu_j}{\alpha_{ij}(1 - \rho_{ij}^2)} \quad \text{and} \quad \kappa_{ij} := \frac{\sigma_i^2 \sigma_j^2 (1 - \rho_{ij}^2)}{\sigma_i^2 - 2\rho_{ij}\sigma_i\sigma_j + \sigma_j^2}.$$

Theorem 7.A.3. *Let (X_i, X_j) be a pair from a type I elliptical random vector $\boldsymbol{X} \sim E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F)$ and consider $\gamma \nearrow x_F$. Then with $\upsilon_{ij}(\gamma) = (\gamma - \mu_{ij})/\kappa_{ij} + c_{ij}(\gamma)$ for some $c_{ij}(\gamma) \in \mathrm{o}(1)$,*

$$\mathbb{P}(X_i > \gamma, X_j > \gamma) = \overline{F}(\upsilon_{ij}(\gamma)) \times \begin{cases} \Big(2\pi \upsilon_{ij}(\gamma) w(\upsilon_{ij}(\gamma))\Big)^{-1/2} (1 + \mathrm{o}(1)), & \text{if } \rho_{ij} > a_{ij}, \\ \Big(2\pi \upsilon_{ij}(\gamma) w(\upsilon_{ij}(\gamma))\Big)^{-1} (C_{a,\rho} + \mathrm{o}(1)), & \text{if } \rho_{ij} < a_{ij}, \end{cases}$$

for a $C_{a,\rho} \in \mathbb{R}_+$. Furthermore, if either $\mu_i \geq \mu_j$ or $\lim_{\gamma \to x_F} w(\gamma)/\gamma < \infty$, then there exists a $C_\rho \in \mathbb{R}_+$ such that

$$\mathbb{P}(X_i > \gamma, X_j > \gamma) = \overline{F}(\upsilon_{ij}(\gamma)) \Big(2\pi \upsilon_{ij}(\gamma) w(\upsilon_{ij}(\gamma))\Big)^{-1/2} (C_\rho + \mathrm{o}(1)), \quad \text{if } \rho_{ij} = a_{ij}.$$

Proof. Use Theorem 2 of Hashorva [74]. First we consider the case $a_{ij} < \rho_{ij}$. In such a case it holds that

$$\begin{aligned} \lim_{\gamma \to x_F} \sqrt{\frac{w(\upsilon_j(\gamma))}{\upsilon_j(\gamma)}} \Big(\upsilon_i(\gamma) - \rho_{ij}\upsilon_j(\gamma)\Big) &= \lim_{\gamma \to x_F} \sqrt{w(\upsilon_j(\gamma))\upsilon_j(\gamma)} \left(\frac{\upsilon_i(\gamma)}{\upsilon_j(\gamma)} - \rho_{ij}\right) \\ &= \lim_{\gamma \to x_F} \sqrt{w(\upsilon_j(\gamma))\upsilon_j(\gamma)} \Big(a_{ij} - \rho_{ij}\Big) = -\infty. \end{aligned}$$

Hence, the hypotheses of Case i) of Theorem 2 of Hashorva [74] hold and the first result follows. In the case where $a_{ij} = \rho_{ij}$ then

$$\lim_{\gamma \to x_F} \sqrt{\frac{w(\upsilon_j(\gamma))}{\upsilon_j(\gamma)}} \Big(\upsilon_i(\gamma) - \rho_{ij}\upsilon_j(\gamma)\Big) = \lim_{\gamma \to x_F} \sqrt{\frac{w(\upsilon_j(\gamma))}{\upsilon_j(\gamma)}} \frac{(\mu_j - \mu_i)}{\sigma_i}.$$

The last limit remains bounded from above if either $\mu_i > \mu_j$ or $\lim_{\gamma\to\infty} w(\gamma)/\gamma < \infty$. For the case $a_{ij} > \rho_{ij}$ we define $a_{ij}(\gamma) := \upsilon_i(\gamma)/\upsilon_j(\gamma)$ so $\lim_{\gamma\to\infty} a_{ij}(\gamma) = a_{ij}$.

We let

$$\tau_{ij}(\gamma) = \sqrt{\frac{1 - 2a_{ij}(\gamma)\rho_{ij} + a_{ij}^2(\gamma)}{1 - \rho_{ij}^2}}, \quad \tau_{ij} := \lim_{\gamma\to\infty} \tau_{ij}(\gamma) = \sqrt{\frac{1 - 2a_{ij}\rho_{ij} + a_{ij}^2}{1 - \rho_{ij}^2}}.$$

The results follows by noting that

$$\upsilon_j(\gamma)\tau_{ij}(\gamma) = \upsilon_{ij}(\gamma), \qquad \upsilon_{ij}(\gamma) = \frac{\gamma - \mu_{ij}}{\tau_{ij}} + \mathrm{o}(1).$$

□ □

Chapter 8

Approximating the Laplace transform of the sum of dependent lognormals

8.1 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 [8, 48, 53, 87, 99]. 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 [52, 53]. Indeed, many of the approximations for the Laplace transform of sums of independent lognormals originated from the wireless communications community [27]. 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) [104]. 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 [51], [60], and the new Chapter 1 in the recently revised volume of McNeil et al. [103]. Comprehensive surveys of applications and numerical methods for the LN and SLN distributions are in [70, 18, 19].

There exist many approximations to the density of the SLN distribution. Many approximations work from the premise [28] 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 $\approx$ LN approximation*. Some well-known SLN $\approx$ LN approximations are the Fenton–Wilkinson [61] and Schwartz–Yeh [127] approaches. These were originally specified for sums of *independent* lognormals, but have since been generalised to the dependent case [4]. A more recent procedure (for the independent case) is the minimax approximation of Beaulieu and Xie [29], calculating the values of μ_L and σ_L which minimise the maximum difference between the densities of S_n and L. However, [29] 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 [28] 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 [22] characterised the right-tail asymptotics. Next, Gao et al. [64] gave the asymptotic form of the left tail for $n = 2$. Gulisashvili and Tankov [70] 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 [70, 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. [19].

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. [18, 19] 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 [18, 19] to arbitrary n and dependence. The defining integral for the Laplace transform of S_n is

$$\mathcal{L}(\theta) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \int_{\mathbb{R}^n} \exp\left\{-\theta \sum_{i=1}^{n} \mathrm{e}^{\mu_i} \mathrm{e}^{x_i} - \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{D} \boldsymbol{x}\right\} \mathrm{d}\boldsymbol{x} \tag{8.1}$$

where $\boldsymbol{D} \stackrel{\text{def}}{=} \boldsymbol{\Sigma}^{-1}$ (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 [18, 19], 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.

8.2 Approximating the Laplace transform

Although the definition (8.1) 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 8.2.1 below) are

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

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 8.2.1. *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 $\mathrm{diag}(\cdot)$ converts vectors to matrices and vice versa, like the MATLAB function.

Proposition 8.2.2. *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 $\boldsymbol{\Lambda} \stackrel{\text{def}}{=} \theta\mathrm{diag}(\mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})$.

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

$$-\theta \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*} = \boldsymbol{D}\boldsymbol{x}^* \text{ which implies } -\theta(\mathbf{e}^{\boldsymbol{\mu}})^\top \mathbf{e}^{\boldsymbol{x}^*} = \mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}^*. \tag{8.2}$$

Therefore the expansion becomes

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

□ □

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

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

We need a suitable error or correction term in order to assess the accuracy of this approximation, so we will decompose the original integral (8.1) into $\mathcal{L}(\theta) = \widetilde{\mathcal{L}}(\theta) I(\theta)$. In the integral of (8.1) change variables such that $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{H}^{-1/2}\boldsymbol{y}$. Then by applying (8.2), multiplying by $\exp\{\mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}^* - \mathbf{1}^\top \boldsymbol{D}\boldsymbol{x}^*\}$, and rearranging, we arrive at

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

where

$$I(\theta) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n}} \exp\left\{(\boldsymbol{x}^*)^\top \boldsymbol{D}\Big(\mathbf{e}^{\boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}} - \mathbf{1} - \boldsymbol{H}^{-\frac{1}{2}}\boldsymbol{y}\Big) - \frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{\Sigma}\boldsymbol{H})^{-1}\boldsymbol{y}\right\} \mathrm{d}\boldsymbol{y}. \tag{8.3}$$

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

Proposition 8.2.3. *We have that*

$$I(\theta) = \mathbb{E}\Big[g(\boldsymbol{H}^{-\frac{1}{2}}Z)\Big] = \sqrt{\det(\boldsymbol{\Sigma}\boldsymbol{H})}\ \mathbb{E}\Big[v(\boldsymbol{\Sigma}^{\frac{1}{2}}Z)\Big] \tag{8.4}$$

where

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

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

Proof. To show that $I(\theta)$ can be written as the first expectation in (8.4), 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 (8.3).

To prove $I(\theta)$ equals the second expectation of (8.4), change variables in (8.3) so that $\boldsymbol{y} = (\boldsymbol{\Sigma}\boldsymbol{H})^{1/2}\boldsymbol{z}$, giving

$$I(\theta) = \sqrt{\det(\boldsymbol{\Sigma}\boldsymbol{H})}\int_{\mathbb{R}^n} \frac{1}{\sqrt{(2\pi)^n}} \exp\left\{(\boldsymbol{x}^*)^\top \boldsymbol{D}\Big(\mathrm{e}^{\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{z}} - \mathbf{1} - \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{z}\Big) - \frac{1}{2}\boldsymbol{z}^\top\boldsymbol{I}\boldsymbol{z}\right\}\mathrm{d}\boldsymbol{z}. \tag{8.5}$$

□ □

Remark 8.2.4. When $n = 1$, $\boldsymbol{\Sigma} = \sigma^2$ and $\mu = 0$, (8.5) 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 [46] 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 [18] equation (2.3).

8.3 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

$$\theta \mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*} + \boldsymbol{D}\boldsymbol{x}^* = \mathbf{0}. \tag{8.6}$$

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

$$x_i^* = \sum_{j=1}^{n} \beta_{i,j} \log_j \theta - \mu_i + c_i + r_i(\theta) \tag{8.7}$$

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

$$x_i^* = -\log\theta + \log_2\theta - \mu_i + \log a_i + r_i(\theta) \tag{8.8}$$

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 (8.8) in (8.6) 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 (8.8) and its derivation is much more intricate.

Theorem 8.3.2. *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 $1 < k_i \le n$. *All* x_i^* *in* $\mathcal{F}_-$ *follow the general form of* (8.7). *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 $1 < k_i \le n$. *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 8.3.3 below.*

See Remark 8.3.7 for some further remarks on the role of the signs of the row sums.

Algorithm 8.3.3.

1. *Let* $\boldsymbol{\beta}_{\bullet,1}$ *be the value of* $\boldsymbol{w}$ *that minimises* $\boldsymbol{w}^\top\boldsymbol{D}\boldsymbol{w}$ *over the set* $\{\boldsymbol{w} : w_i \le -1\}$. *It will be proved in the appendix that the solution has* $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,1} \le 0$ *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* $k = 2,\dots,n$ *recursively calculate* $\boldsymbol{\beta}_{\bullet,k}$ *as the value of* $\boldsymbol{w}$ *that minimises* $\boldsymbol{w}^\top\boldsymbol{D}\boldsymbol{w}$ *whilst satisfying*

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

   It will be proved in the appendix that the solution has $\boldsymbol{D}_{i,\bullet}\boldsymbol{\beta}_{\bullet,k} \le 0$ *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*

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

   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*

$$\boldsymbol{c}_- = -\boldsymbol{D}_{-,-}^{-1}\boldsymbol{D}_{-,+}(\boldsymbol{c}_+ - \boldsymbol{\mu}_+) + \boldsymbol{\mu}_-. \tag{8.9}$$

Proof of Theorem 8.3.2. We propose a solution of the form (8.7) and show that when the $\beta_{i,j}$ are constructed from Algorithm 8.3.3, the remainder term r_i is 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 $1 < k_i \le n$ such that

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

Insertion in (8.6) 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 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 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 8.3.1 we have $c_i = \log a_i$, and in Theorem 8.3.2 (8.9) 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 $\mathcal{F}_\sim \stackrel{\text{def}}{=} \mathcal{F}_*^{\mathrm{c}}$; 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 8.3.4. *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

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

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 $\overline{\boldsymbol{D}}\boldsymbol{b}_* = \overline{\boldsymbol{D}}\boldsymbol{1} = \overline{\boldsymbol{a}} = \mathrm{e}^{\boldsymbol{c}_*} > \boldsymbol{0}$; thus $\overline{\boldsymbol{D}}$ has all row sums positive and $\boldsymbol{c}_* = \log(\overline{\boldsymbol{D}}\boldsymbol{b}_*) = \log\overline{\boldsymbol{a}}$. □

There are some simple forms of $\boldsymbol{\Sigma}$ which fall into the case where all $a_i > 0$. These include the case where all diagonal elements of $\boldsymbol{\Sigma}$ are identical, and all non-diagonal elements are identical. Note, by positive definiteness of $\boldsymbol{\Sigma}$ 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 $a_i \le 0$ are not difficult to find. For the case $n = 2$ with variances $\sigma_1^2 \le \sigma_2^2$ 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. [64] 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 $\boldsymbol{\Sigma}$ which have some non-positive row sums of $\boldsymbol{\Sigma}^{-1}$.

Example 8.3.5. *Consider $\boldsymbol{\mu} = (-10, 0, 10)^\top$ and*

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

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

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

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 8.3.7. The importance of the sign of the row sums of $\boldsymbol{D}$, as illustrated by Proposition 8.3.1, perplexed us for quite some time. However Gulisashvili and Tankov [70] 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

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

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 [101], 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 $\boldsymbol{A} \stackrel{\text{def}}{=} \boldsymbol{D} - \operatorname{diag}(\boldsymbol{D})$ and write the defining equation as

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

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.

8.4 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 8.2. Define $\boldsymbol{\sigma} \stackrel{\text{def}}{=} \operatorname{diag}(\boldsymbol{H})^{-1/2} \in (0,\infty)^n$ and $\boldsymbol{M} \stackrel{\text{def}}{=} \operatorname{diag}(\boldsymbol{\sigma})\boldsymbol{H}\operatorname{diag}(\boldsymbol{\sigma}) \in \mathbb{R}^{n\times n}$. In (8.3), substitute $\boldsymbol{H}^{-1/2}\boldsymbol{y} = \boldsymbol{\sigma} \circ \boldsymbol{z}$, so

$$I(\theta) = \int_{\mathbb{R}^n} \frac{\exp\{-\frac{1}{2}\boldsymbol{z}^\top \boldsymbol{M}\boldsymbol{z}\}}{\sqrt{(2\pi)^n \det(\boldsymbol{M}^{-1})}} \exp\left\{-\theta(\mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})^\top \left(\mathbf{e}^{\boldsymbol{\sigma}\circ\boldsymbol{z}} - \mathbf{1} - \boldsymbol{\sigma}\circ\boldsymbol{z} - \frac{1}{2}(\boldsymbol{\sigma}\circ\boldsymbol{z})^{\mathbf{2}}\right)\right\} \mathrm{d}\boldsymbol{z}. \quad (8.10)$$

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

$$\theta \mathrm{e}^{\mu_i + x_i^*} \sigma_i^\ell = \theta \mathrm{e}^{\mu_i + x_i^*} (\theta \mathrm{e}^{\mu_i + x_i^*} + D_{i,i})^{-\frac{\ell}{2}} = \mathrm{o}(1). \quad (8.11)$$

Consider the second exponent of (8.10). 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 (8.11) we have

$$\theta(\mathbf{e}^{\boldsymbol{\mu}+\boldsymbol{x}^*})^\top \Big(\mathbf{e}^{\boldsymbol{\sigma}\circ\boldsymbol{z}} - \mathbf{1} - \boldsymbol{\sigma}\circ\boldsymbol{z} - \frac{1}{2}(\boldsymbol{\sigma}\circ\boldsymbol{z})^{\mathbf{2}}\Big) = \mathrm{o}(1). \tag{8.12}$$

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 $\boldsymbol{M} \to \boldsymbol{M}^* \stackrel{\text{def}}{=} \mathrm{diag}(\boldsymbol{I}_{n_+}, \boldsymbol{F})$ 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 (8.10) is a non-degenerate multivariate normal density.

Proposition 8.4.1. $\lim_{\theta\to\infty} I(\theta) = 1$.

Proof. We use the dominated convergence theorem. By (8.12) 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 (8.12). Next, convexity implies that the exponent is bounded by $-g_2\|\boldsymbol{z}\|$ for $\|\boldsymbol{z}\| > 1$. In total we have the bound

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

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

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

The simplest approach is to numerically integrate the original expression in (8.1). 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

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

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 8.2.3 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 [18]. 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 $Z_1, \ldots, Z_R \overset{\text{i.i.d.}}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$,

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

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 $(\boldsymbol{x}^*)^\top \boldsymbol{D}\boldsymbol{Z}_r^{\boldsymbol{2}}$ (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 [65] or [15]. Both individually achieved significant variance reduction, and together provided the best estimator. Specifically,

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

where $\boldsymbol{q}_r \overset{\text{def}}{=} \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Phi}^{-1}(\boldsymbol{u}_r)$, using $\boldsymbol{\Phi}^{-1}(\cdot)$ as the (element-wise) standard normal inverse c.d.f., and where $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots\}$ is the n-dimensional Sobol sequence started at the same point for every θ. Therefore, $\widehat{\mathcal{L}}_{\mathrm{Q}}(\theta)$ 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.

8.6 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 [92] for the software implementation used to create these results.

Table 8.1: Relative error for various approximations of $\mathcal{L}(\theta)$ for $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = [1, 0.5; 0.5, 1]$. 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
$\widehat{\mathcal{L}}_{\mathrm{CMC}}$	1.29e-2	$*$	$*$	$*$	$*$
$\widehat{\mathcal{L}}_{\mathrm{IS}}$	3.36e-4	2.96e-4	2.57e-4	2.31e-4	2.11e-4
$\widehat{\mathcal{L}}_{\mathrm{Q}}$	-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 [129] and so-called power algorithms [23] can be used. We report on the results of using the Gaver–Stehfest algorithm as implemented by Mallet [100].

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 [15]), and kernel density estimation. The following estimators are reported: the conditional Monte Carlo estimator $\widehat{f}_{\mathrm{Cond}}$, $\widetilde{f} \stackrel{\mathrm{def}}{=} \mathcal{L}^{-1}(\widetilde{\mathcal{L}}(\cdot))$, $\widehat{f}_{\mathrm{IS}} \stackrel{\mathrm{def}}{=} \mathcal{L}^{-1}(\widehat{\mathcal{L}}_{\mathrm{IS}}(\cdot))$ and $\widehat{f}_{\mathrm{Q}} \stackrel{\mathrm{def}}{=} \mathcal{L}^{-1}(\widehat{\mathcal{L}}_{\mathrm{Q}}(\cdot))$.

Table 8.2: 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 $\widehat{f}_{\mathrm{Cond}}$, $\widehat{f}_{\mathrm{IS}}$ and $\widehat{f}_{\mathrm{Q}}$.

x	0.01	1	1.5	2	3
$\widehat{f}_{\mathrm{Cond}}$	-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
$\widehat{f}_{\mathrm{IS}}$	1.94e-3	1.43e-2	-6.13e-3	4.00e-2	3.68e-3
$\widehat{f}_{\mathrm{Q}}$	2.90e-4	1.11e-2	-9.04e-3	3.70e-2	2.44e-3

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 [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).

8.7 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 8.3 gives an accurate starting value for the iterations.

8.A Remaining steps in the proof of Theorem 8.3.2

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. □

Chapter 9

Conclusion

TODO

Bibliography

[1] Joseph Abate, G.L. Choudhury, and Ward Whitt. On the Laguerre method for numerically inverting Laplace transforms. *INFORMS Journal on Computing*, 8(4):413–427, 1995.

[2] Joseph Abate and Ward Whitt. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10(1):5–87, Jan 1992.

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

[4] 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.

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

[6] Robert J Adler. An introduction to continuity, extrema, and related topics for general gaussian processes. *Lecture Notes-Monograph Series*, 12:i–155, 1990.

[7] Robert J Adler, Jose Blanchet, Jingchen Liu, et al. Efficient monte carlo for high excursions of Gaussian random fields. *The Annals of Applied Probability*, 22(3):1167–1214, 2012.

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

[9] Lars Nørvang Andersen, Patrick J. Laub, and Leonardo Rojas-Nandayapa. *Online accompaniment for "Rare-event simulation for extrema of dependent random variables"*, 2016. Available at `https://github.com/Pat-Laub/RareMaxima`.

[10] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition, 1984.

[11] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.

[12] 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`.

[13] Søren Asmussen and Hansjörg Albrecher. *Ruin probabilities*, volume 14 of *Advanced Series on Statistical Science and Applied Probability*. World Scientific Publishing Co Pte Ltd, River Edge, NJ, 2 edition, 2010. Advanced series on statistical science & applied probability; v. 14.

[14] Søren Asmussen, Jose Blanchet, Sandeep Juneja, and Leonardo Rojas-Nandayapa. Efficient simulation of tail probabilities of sums of correlated lognormals. *Annals of Operations Research*, 189(1):5–23, 2011.

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

[16] Søren Asmussen, Pierre-Olivier Goffard, and Patrick J. Laub. *Online accompaniment for "Orthonormal polynomial expansions and lognormal sum densities"*, 2016. Available at `https://github.com/Pat-Laub/SLNOrthogonalPolynomials`.

[17] Søren Asmussen, Pierre-Olivier Goffard, and Patrick J. Laub. Orthonormal polynomial expansion and lognormal sum densities. *arXiv preprint arXiv:1601.01763*, 2016.

[18] 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.

[19] Søren Asmussen, Jens Ledet Jensen, and Leonardo Rojas-Nandayapa. Exponential family techniques in the lognormal left tail. *Scandinavian Journal of Statistics*, 2016. To appear, DOI 10.1111/sjos.12203.

[20] Søren Asmussen and Dirk P. Kroese. Improved algorithms for rare event simulation with heavy tails. *Adv. in Appl. Probab.*, 38(2):545–558, June 2006.

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

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

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

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

[25] 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.

[26] Ole E Barndorff-Nielsen and David Roxbee Cox. *Asymptotic Techniques for use in Statistics*. Chapman & Hall, 1989.

[27] 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.

[28] 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.

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

[30] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic Analysis on Semigroups*, volume 100 of *Graduate Texts in Mathematics*. Springer–Verlag, 1984.

[31] Simeon Berman. *Sojourns and Extremes of Stochastic Processes*. CRC Press, 1992.

[32] Nicholas H Bingham, Charles M Goldie, and Jozef L. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge university press, Cambridge, 1989.

[33] Zdravko I. Botev. The normal law under linear restrictions: Simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society, Series B*, 79:pp. 1–24.

[34] Zdravko I. Botev, Joseph F. Grotowski, and Dirk P. Kroese. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010.

[35] JA Breslaw. Random sampling from a truncated multivariate normal distribution. *Applied Mathematics Letters*, 7(1):1–6, 1994.

[36] Wlodzimierz Bryc. *The Normal Distribution: Characterizations with Applications*, volume 100. Springer Science & Business Media, 2012.

[37] Jun Cai, Ken Seng Tan, Chengguo Weng, and Yi Zhang. Optimal reinsurance under VaR and CTE risk measures. *Insurance: Mathematics and Economics*, 43(1):185 – 196, 2008.

[38] Arthur Charpentier and Johan Segers. Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis*, 100(7):1521–1537, 2009.

[39] Ka Chun Cheung. Optimal reinsurance revisited — a geometric approach. *ASTIN Bulletin*, 40(1):221–239, 005 2010.

[40] Yichun Chi and Ken Seng Tan. Optimal reinsurance under VaR and CVaR risk measures: a simplified approach. *ASTIN Bulletin*, 41(2):487–509, 11 2011.

[41] T. S. Chihara. On generalized Stieljes-Wigert and related orthogonal polynomials. *Journal of Computational and Applied Mathematics*, 5(4):291–297, 1979.

[42] T. S. Chihara. *An Introduction to Orthogonal Polynomials*. Courier Corporation, 2011.

[43] J. S. Christiansen. The moment problem associated with the Stieljes-Wigert polynomials. *Journal of Mathematical Analysis and Applications*, 277(1):218–245, 2003.

[44] Daren BH Cline. Convolution tails, product tails and domains of attraction. *Probability Theory and Related Fields*, 72(4):529–557, 1986.

[45] Rama Cont. Encyclopedia of quantitative finance. 2010.

[46] 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.

[47] H. Cramer. *Mathematical Methods of Statistics*, volume 9. Princeton University Press, 1999.

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

[49] Laurens De Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer Science & Business Media, 2007.

[50] M. Denuit, J. Dhaene, M. J. Goovaert, and R. Kaas. *Actuarial Theory for Dependent Risk: Measures, Orders and Models*. John Wiley & Sons, Inc., 2006.

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

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

[53] Daniel Dufresne. Sums of lognormals. *Proceedings of the 43rd Actuarial Science Research Conference*, 2008.

[54] Daniel Dufresne, Jose Garrido, and Manuel Morales. Fourier inversion formulas in option pricing and insurance. *Methodology and Computing in Applied Probability*, 11(3):359–383, 2009.

[55] EIOPA. Quantitative Impact Studies V: Technical Specifications. Technical report, European comission, Brussels, 2010.

[56] Torbjørn Eltoft, Taesu Kim, and Te-Won Lee. On the multivariate Laplace distribution. *IEEE Signal Processing Letters*, 13(5):300–303, 2006.

[57] Paul Embrechts and M. Frei. Panjer recursion versus FFT for compound distribution. *Mathematical Methods of Operations Research*, 69:497–508, July 2009.

[58] Paul Embrechts, Enkelejd Hashorva, and Thomas Mikosch. Aggregation of log-linear risks. *Journal of Applied Probability*, 51(A):203–212, 2014.

[59] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

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

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

[62] Carlo Fischione, Fabio Graziosi, and Fortunato Santucci. Approximation for a sum of on-off lognormal processes with wireless applications. *IEEE Transactions on Communications*, 55(10):1984–1993, 2007.

[63] Nial Friel and Jason Wyse. Estimating the evidence– a review. *Statistica Neerlandica*, 66(3):288–308, August 2012.

[64] 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.

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

[66] Pierre-Olivier Goffard and Patrick J. Laub. *Online accompaniment for "Two numerical methods to evaluate stop-loss premiums"*, 2017. Available at `https://github.com/Pat-Laub/ActuarialOrthogonalPolynomials`.

[67] Pierre-Olivier Goffard, Stéphane Loisel, and Denys Pommeret. Polynomial approximations for bivariate aggregate claims amount probability distributions. *Methodology and Computing in Applied Probability*, pages 1–24, 2015.

[68] Pierre-Olivier Goffard, Stéphane Loisel, and Denys Pommeret. A polynomial expansion to approximate the ultimate ruin probability in the compound Poisson ruin model. *Journal of Computational and Applied Mathematics*, 296:499 – 511, 2016.

[69] John A Gubner. *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, 2006.

[70] Archil Gulisashvili and Peter Tankov. Tail behavior of sums and differences of lognormal random variables. *Bernoulli*, 22(1):444–493, 2015. To appear, accessed online on 26th August 2015 at `http://www.e-publications.org/ims/submission/BEJ/user/submissionFile/17119?confirm=ef609013`.

[71] Henryk Gzyl, P.L. Novi-Inverardi, and Aldo Tagliani. Determination of the probability of ultimate ruin probability by maximum entropy applied to fractional moments. *Insurance: Mathematics and Economics*, 53(2):457–463, 2013.

[72] Henryk Gzyl and Aldo Tagliani. Determination of the distribution of total loss from the fractional moments of its exponential. *Applied Mathematics and Computation*, 219(4):2124 – 2133, 2012.

[73] Enkelejd Hashorva. Asymptotics and bounds for multivariate Gaussian tails. *Journal of Theoretical Probability*, 18(1):79–97, 2005.

[74] Enkelejd Hashorva. Asymptotic properties of type I elliptical random vectors. *Extremes*, 10(4):175–206, 2007.

[75] Enkelejd Hashorva. On the residual dependence index of elliptical distributions. *Statistics & Probability Letters*, 80(13):1070–1078, 2010.

[76] Enkelejd Hashorva and Jinzhu Li. Asymptotics for a discrete-time risk model with the emphasis on financial risk. *Probability in the Engineering and Informational Sciences*, 28(4):573–588, 2014.

[77] Abdelhamid Hassairi and Mohammed Zarai. Characterization of the cubic exponential families by orthogonality of polynomials. *The Annals of Probability*, 32(3):2463–2476, 2004.

[78] Marwane Ben Hcine and Ridha Bouallegue. Highly accurate log skew normal approximation to the sum of correlated lognormals. *In the Proc. of NeTCoM 2014*, 2015.

[79] Janet E Heffernan. A directory of coefficients of tail dependence. *Extremes*, 3(3):279–290, 2000.

[80] C.C. Heyde. On a property of the lognormal distribution. *J. Roy. Statist.*, 5(Ser. B):392–393, 1963.

[81] Chris C. Holmes and Leonhard Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian analysis*, 1(1):145–168, 2006.

[82] Zhi Huang and P Shahabuddin. Rare-event, heavy-tailed simulations using hazard function transformations, with applications to value-at-risk. In *Proceedings of the 2003 Winter Simulation Conference, 2003.*, 2003.

[83] Henrik Hult, Filip Lindskog, et al. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34(3):587–608, 2002.

[84] Jens Ledet Jensen. *Saddlepoint approximations*. Number 16. Oxford University Press, 1995.

[85] Tao Jin, Serge B. Provost, and Jiandong Ren. Moment-based density approximations for aggregate losses. *Scandinavian Actuarial Journal*, 2016(3):216–245, 2016.

[86] Harry Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

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

[88] Stuart A Klugman, Harry H Panjer, and Gordon E Willmot. *Loss models: from data to decisions*, volume 715. John Wiley & Sons, 2012.

[89] Samuel Kotz, Tomaz J Kozubowski, and Krzysztof Podgórski. Asymmetric multivariate Laplace distribution. In *The Laplace Distribution and Generalizations*, pages 239–272. Springer, 2001.

[90] Dirk P. Kroese, Thomas Taimre, Zdravko I. Botev, and Reuven Y. Rubinstein. *Solutions manual to accompany Simulation and the Monte Carlo method*. Wiley-Interscience, 2007.

[91] Patrick J. Laub, Søren Asmussen, Jens Ledet Jensen, and Leonardo Rojas-Nandayapa. Approximating the Laplace transform of the sum of dependent lognormals. *Advances in Applied Probability.*, 2016. In Festschrift for Nick Bingham. To appear.

[92] Patrick J. Laub, Søren Asmussen, Jens Ledet Jensen, and Leonardo Rojas-Nandayapa. *Online accompaniment for "Approximating the Laplace transform of the sum of dependent lognormals"*, 2016. Available at `https://github.com/Pat-Laub/SLNLaplaceTransformApprox`.

[93] Anthony W Ledford and Jonathan A Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

[94] Anthony W Ledford and Jonathan A Tawn. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499, 1997.

[95] Anthony W Ledford, Jonathan A Tawn, et al. Concomitant tail behaviour for extremes. *Advances in Applied Probability*, 30(1):197–215, 1998.

[96] Simon C. K. Lee and X. Sheldon Lin. Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal*, 14(1):107–130, 2010.

[97] C. Lefèvre and P. Picard. A new look at the homogeneous risk model. *Insurance: Mathematics and Economics*, 49(3):512–519, 2011.

[98] C. Lefèvre, J. Trufin, and P. Zuyderhoff. Some comparison results for finite-time ruin probabilities in the classical risk model. *Insurance: Mathematics and Economics*, 77(Supplement C):143–149, 2017.

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

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

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

[102] NA Marlow. A normal limit theorem for power sums of independent random variables. *Bell System Technical Journal*, 46(9):2081–2089, 1967.

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

[104] 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.

[105] Robert M. Mnatsakanov and Khachatur Sarkisian. A note on recovering the distributions from exponential moments. *Applied Mathematics and Computation*, 219(16):8730 – 8737, 2013.

[106] Robert M. Mnatsakanov, Khachatur Sarkisian, and A. Hakobyan. Approximation of the ruin probability using the scaled Laplace transform inversion. *Applied Mathematics and Computation*, 268:717–727, 2015.

[107] Carl N. Morris. Natural exponential families with quadratic variance functions. *The Annals of Statistics*, 10(1):65–80, 1982.

[108] Saralees Nadarajah, Jeffrey Chu, and Xiao Jiang. On moment based density approximations for aggregate losses. *Journal of Computational and Applied Mathematics*, 298:152 – 166, 2016.

[109] Roger B Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 2nd edition, 2006.

[110] Ryuei Nishii. *Orthogonal Functions of Inverse Gaussian Distributions*, pages 243–250. Springer US, Boston, MA, 1996.

[111] Natalia Nolde. Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95, 2014.

[112] Harry H Panjer and Gordon E Willmot. Finite sum evaluation of the negative binomial-exponential model. *ASTIN Bulletin: The Journal of the IAA*, 12(2):133–137, 1981.

[113] D.E Papush, G. S. Patrik, and F. Podgaits. Approximations of th aggregate loss distribution. *CAS Forum (Winter)*, pages 175–186, 2001.

[114] Serge B. Provost. Moment-based density approximants. *Mathematica Journal*, 9(4):727–756, 2005.

[115] Marvin Rausand and Arnljot Høyland. *System Reliability Theory: Models, Statistical Methods, and Applications*, volume 396. John Wiley & Sons, 2004.

[116] Sidney I Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336, 2002.

[117] Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.

[118] Christian P Robert. Simulation of truncated normal variables. *Statistics and computing*, 5(2):121–125, 1995.

[119] Tomasz Rolski, Hanspeter Schmidli, Volker Schmidt, and Jozef L. Teugels. *Risk Processes for finance and insurance*, pages 147–204. John Wiley & Sons, Inc., 2008.

[120] Holger Rootzen. A ratio limit theorem for the tails of weighted sums. *The Annals of Probability*, pages 728–747, 1987.

[121] Jeffrey S. Rosenthal. *A first look at rigorous probability theory*. World Scientific Publishing Co Inc, 2006.

[122] Reuven Y. Rubinstein. Sensitivity analysis of discrete event systems by the "push out" method. *Annals of Operations Research*, 39(1):229–250, December 1992.

[123] Reuven Y. Rubinstein and Dirk P. Kroese. *Simulation and the Monte Carlo method*, volume 707. John Wiley & Sons, 2011.

[124] Ludger Rüschendorf. *Mathematical risk analysis*. Springer, 2013.

[125] R. Salomone, Patrick J. Laub, and Zdravko I. Botev. *Source code and supplementary material for "Monte Carlo Estimation of the Density of the Sum of Dependent Random Variables"*, 2018. Available at `https://github.com/Pat-Laub/PushoutDensityEstimation`.

[126] I Richard Savage. Mills' ratio for multivariate normal distributions. *J. Res. Nat. Bur. Standards Sect. B*, 66:93–96, 1962.

[127] 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.

[128] Masaaki Sibuya. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210, 1960.

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

[130] T. J. Stieljes. *Recherches sur les fractions continues*. Annales de la Faculté des sciences de Toulouse: Mathématiques, 1894.

[131] G. Szegö. *Orthogonal Polynomials*, volume XXIII. American Mathematical Society Colloquium Publications, 1939.

[132] B. Szökefalvi-Nagy. *Introduction to Real Functions and Orthogonal Expansions*. Akadémiai Kiadó, 1965.

[133] Olof Thorin and Nils Wikstad. Calculation of ruin probabilities when the claim distribution is lognormal. *Astin Bulletin*, 9(1-2):231–246, 1977.

[134] Toshiro Watanabe. Convolution equivalence and distributions of random sums. *Probability Theory and Related Fields*, 142(3):367–397, 2008.

[135] S. Wigert. Sur les polynomes orthogonaux et l'approximation des fonctions continues. *Almqvist and Wiksell*, 1923.

[136] Gordon E. Willmot and X. Sheldon Lin. Risk modelling with the mixed Erlang distribution. *Applied Stochastic Models in Business and Industry*, 27(1):2–16, 2011.

[137] Gordon E. Willmot and Jae-Kyung Woo. On the class of erlang mixtures with risk theoretic applications. *North American Actuarial Journal*, 11(2):99–115, 2007.