HEAVY TAILS AND DEPENDENCE
WITH APPLICATIONS IN INSURANCE

by

Jae Hoon Jho

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Faculty of Actuarial Science and Insurance)

at the

Cass Business School, City University

2008, December

DISCARD THIS PAGE

TABLE OF CONTENTS

Page

DISCARD THIS PAGE

LIST OF TABLES

Table Page

DISCARD THIS PAGE

LIST OF FIGURES

Appendix
Figure Page

ACKNOWLEDGMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. My utmost gratitude goes to my supervisor, Dr. Vladimir Kaishev for his kindness, his expertise, and his patience. I believe that I learned not only how to do research but also how to work together. I also would like to thank my joint supervisor Professor Richard Verrall for his kind advices throughout my research in London.

I especially want to thank Dr. Jiwook Jang who encouraged me to go ahead with my thesis. Professor Tzvetan Ignatov, Dr. Andreas Tsanaka, deserve a special thanks as my thesis examiners and I would like to thank Professor Celia Glass, the chair of my viva voce.

All my friends at 4th and 5th floor in Cass building made it a convivial place to work. In particular, I would like to thank Alexandre Rubesam and his firends for their friendship and help in the past four years. I was also delighted to live together with Shivasharan Nadavulakere sharing a flat. I was happy to play music at Cross Street Baptist Church on every Sunday with worship team. I wish that God be always with my brothers and sisters, Dennis Philip, who led me to Cross Street, Pastor Mark Richard Shelton, his wife and worship leader Tina Shelton, excellent singers Marcia Roswell-Joseph, Benjamin Mills, Sapphira Gordon, and Valerie Dock.

Above all, I would like to give my special thanks to my parents whose love enabled me to complete this work.

ABSTRACT

In this thesis we study the tail behavior of a random variable and sum of dependent random variables using the extreme value theory. We examine the tail behavior of a single random variable by mixture distribution models, and the asymptotic properties of the value-at-risk measure of dependent regularly varying random variables.

In order to obtain a flexible fit not only on the tail but also on the body of the underlying distribution, mixture distributions are introduced with finite or infinite number of thresholds, where the consistency of the heavy-tailedness is preserved by the conditional layer mixture. Hazard rate functions of the conditional layer mixture distributions are studied and the mixture of the hazard rate functions can be used in modeling the mixture distributions equivalently.

Impact of heavy-tailedness and dependence on the value-at-risk measure is examined for the sum of regularly varying random variables under quite general dependence structure and we conclude that the extreme value index completely determines the tail behavior of the compound sum of regularly varying random variables with respect to the value-at-risk measure.

In addition, a hierarchical structure composed of maximal Markov sequences is introduced to simplify a given pool of risks under arbitrary dependence and we propose a computational method of the aggregate distribution of each maximal Markov sequence.

Chapter 1

Introduction

1.1 Background

Sound mathematical models are necessary nowadays to solve actuarial problems such as pricing, reserving, and optimal retention levels in reinsurance contract. Appropriate mathematical models of insurance risks are gaining attentions and more often emphasized especially when risks are heavy-tailed or dependent, which are two main focuses of this dissertation.

For independent risks, a number of risk models have been developed and applied in practice. Continuous or discrete time stochastic processes, numerical analysis, and simulation techniques have been widely used for this purpose. If dependence intervenes among risks, however, we fail to apply or generalize the independent risk models more often than not. For instance, Picard and Lefèvre (1997) considered a continuous time compound Poisson process of an insurance company and proposed an elegant expression of the ruin probability in finite time assuming discrete claim size distributions and claims independence. Panjer's recursion (Panjer, 1981) for the aggregate loss distribution also assumes the independence of integer-valued losses, whereas the recursion cannot be applied when losses are dependent. In finance multivariate analysis has been applied to address the dependence of multiple risks for an optimal portfolio selection. Despite the lack of flexibility, multivariate normal distributions and the family of elliptical distributions have been used to model the dependence structure of financial and insurance risks where correlations play important roles. Multivariate regular variation is another multivariate analysis approach to dependence modeling especially when the marginals are heavy-tailed. Copula has been also popularly used to implement the dependence structure among risks. For example, if we assume Markov property

among a number of risks, a chain of bivariate copulas becomes a very flexible tool to capture the dependence. According to the principles proposed by Basel II (Basel Committee on Banking Supervision, 2004), operational risks are classified into eight business lines and seven event types. Based on the standardized classification matrix of operational risks, the total aggregate loss can be written as a sum of aggregate losses of each business line (or event type), i.e. $Y_1 + \cdots + Y_d$ where each Y_i represents for the aggregate loss of i-th line of business. Under this aggregate risk process, it is of our special interest to study the impact of dependence among the lines of business (or event types) on the total aggregate loss distribution and its tail behavior.

Heavy-tailedness of a risk is as important as dependence structure and they shouldn't be studied separately. Moscadelli (2004) mentioned the empirical observation of the heavy-tailed operational risks. Similar discussion can be found in Dutta and Perry (2004). Since heavy-tailed risks usually have very low frequency of loss occurrences, empirical approaches such as parametric estimation and Bayesian method often fail to measure the correct tail behavior of a risk. Therefore, one has to appeal to purely mathematical models such as extreme value theory. The study of heavy-tailed distributions requires advanced knowledge of mathematics and statistics. It stems from basic questions such as how to define heavy-tailedess and how heavy-tailed a distribution is. The questions can be answered in terms of the family of distributions the loss random variables belong to, such as long-tailed distributions, subexponential distributions, or regular variations. In classical extreme value theory, a limit law for the maxima has been used for the heavy-tailedness of distributions since Fisher and Tippett (1928), which classifies three families of extreme value distributions, Frechet, Gumbel, and Weibull distributions. Considering the total aggregate loss of multiple lines of business or products in insurance and operational risk, the individual aggregate loss of one business line (or event type), Y_i, can be considered as a member of certain distribution family such as subexponential, regularly varying distributions, or certain maximum domain of attraction to reflect the existence of heavy-tailedness of the risk. The study of heavy-tailed distributions is important because heavy-tailed risks often violate many properties we expect in general risk modeling, an example of which is the value-at-risk measure. It is well known that the value-at-risk is not a coherent measure (Artzner et al., 1999), and it is not clear whether the subadditivity holds even under

the independence assumption. It naturally raises questions; which distribution families behave differently from what we expect in the tail and how dependence interacts with heavy-tailedness with respect to a given risk measure.

In sum, if we consider a risk process of sum of losses in insurance or operational risks, dependence and heavy-tailedness should be taken into account together to achieve a sound mathematical model of the tail behavior. We approach this problem in the following ways. First we introduce mixture distribution models for a risk consisting of multiple sub-risks each of which has distinct characteristics. We also study the properties of the mixture distributions inherited from the distributional components, such as maximum domain of attractions and the hazard rate functions of the mixture. Secondly, we introduce a maximal Markov sequence decomposition of a pool of risks into disjoint sub-pools of risks and propose a computational method of the aggregate distribution of each sub-pool of risks. Lastly, we examine the asymptotic tail behavior of sum of regularly varying random variables under quite general dependence structure and discuss the impact of dependence and heavy-tailedness on the value-at-risk measure.

1.2 Outline of Thesis

In chapter 2 we introduce the classical theory of extreme value distributions and its properties. We propose three types of mixture of k underlying distributions with a sequence of thresholds $0 = u_0 < u_1 < \cdots < u_k$ where a positive integer k is possibly infinity. Main goal of this chapter is the maximum domain of attractions of the mixture distributions and the properties inherited from the distributional components. We first construct a mixture of finite number of distributions and investigate the maximum domain of attraction and its normalizing constants. We prove that there is a equivalent representation between the layer mixture model and the linear mixture model. Secondly, we generalize the finite mixture models we proposed to the infinite mixture model of infinite number of layers and thresholds. The properties of the mixture models inherited from the distributional components are provided, among which the hazard function representation is emphasized: the hazard rate function of the conditional layer mixture distribution can be written as a mixture of the hazard functions of the distributional components, which is a unique property

of the conditional mixture among all mixture models. Limiting distributions of the conditional layer mixture model is also considered and we shortly discuss the Lorenz curves of the conditional layer mixtures. An application of the mixture model to the g-and-h distributions is provided with numerical examples where we use the threshold estimate calculated in Appendix B.

In chapter 3 we introduce a decomposition of a pool of risks under arbitrary dependence into disjoint maximal Markov sequences. Considering the random variables in a maximal Markov sequence, we present a computational method for the aggregate loss distribution of non-identical and dependent random variables. We derive formulae for the aggregate density and aggregate distribution function and introduce an application on a chain of pairwise bivariate copulas to implement Markovian dependence structure among losses. Numerical examples are also given when the marginal distributions are exponential with Falie-Gumbel-Morgenstern copulas, and Pareto distributed with Gaussian copulas. We also provide actuarial applications of the proposed computational method; Bayesian premiums conditioning on the sum of the past observations and stop-loss premium calculation of dependent claims.

In chapter 4 we focus on the value-at-risk measure of finite sum of loss random variables or compound sum generated by an appropriate counting process with finite expectation. We discuss the classical convolution theorem for the regularly varying random variables, and the closure property and max-stability of regular variation. Introducing the concept of negligible joint tail probability, we provide a sufficient condition of the joint distribution to generalize the convolution theorem when losses are dependent with regularly varying tails. It follows that the shape parameter of the regularly varying distribution, the common distribution of the loss random variables, completely determine the asymptotic super(sub)additivity of the value-at-risk of the aggregate loss. Two numerical examples are given to illustrate the super- and subadditivity of the value-at-risk of Pareto distributed dependent losses when they satisfy the negligible joint tail condition.

Results of the thesis were presented in the conferences and submitted to the journals as follows:

- J. Jang and J. Jho. *Recursive procedure for the aggregate distribution of dependent variables*. Radon Workshop on Financial and Actuarial Mathematics for Young Researchers, Linz, Austria, May 30 - 31, 2007.

- J. Jang and J. Jho. *Asymptotic super(sub)additivity of the value-at-risk of regularly varying dependent variables*. 11th International Congress on Insurance: Mathematics and Economics, Piraeus, Greece, July 10-12, 2007.

- J. Jang and J. Jho. *Asymptotic super(sub)additivity of the value-at-risk of regularly varying dependent variables*. Finance, Stochastics, Insurance, Bonn, Germany, February 25-29, 2008.

- J. Jang and J. Jho. *Asymptotic super(sub)additivity of the value-at-risk of regularly varying random variables*. Submitted to Journal of Applied Probability, 2008.

- J. Jho and V. Kaishev. *On some mixture distributions and their extreme value behavior*. Actuarial Reseach Paper, No.185, Cass Business School, ISBN 978-1-905752-14-0, 2008.

- J. Jho and V. Kaishev. *On some mixture distributions and their extreme value behavior*. 2nd International Workshop on Computational and Financial Econometrics, Neuchâtel, Switzerland, 2008.

- J. Jho and V. Kaishev. *On some mixture distributions and their extreme value behavior*. Submitted to Computational Statistics and Data Analysis, 2008.

1.3 Some Conventions

The list of notations and symbols is included in the Appendix C. We assume loss random variables are continuous and the probability density functions exist unless specified. The counting process of loss occurrence is assumed to have finite mean $E[N_t]$ for all $t > 0$. The italics is used to emphasize some ideas and the symbol $\square$ is used for the end of proofs and examples.

Chapter 2

On Some Mixture Distributions and their Extreme Value Behavior

This chapter is based on the paper by Jho and Kaishev (2008). Mixture models are widely used as a flexible modeling tool in lifetime data analysis and reliability engineering in estimating distribution functions of failure times. The most common mixture distribution, so called finite mixture distribution, is often defined by a mixture of density functions such as

$$f(x|\theta_1, \ldots, \theta_n) = \sum_{j=1}^{n} c_j f_j(x|\theta_j) \tag{2.1}$$

where $\theta_1, \ldots, \theta_n$ are vectors of parameters and c_j are appropriate positive weights such that $\sum_{j=1}^{n} c_j = 1$. The finite mixture distribution (2.1) has been widely used in reliability analysis to fit the probability density function of failure times, see for example Bucar et al. (2004). A number of practical applications of (2.1) can be found in Al-Hussaini and El-Adll (2004) and the references therein. Under the finite mixture distribution model, Nurmi (2004) introduced a Bayesian method to estimate the parameters of the underlying distributions and the optimal number of sub-poppulations. Computational implementations can also be found in Bettina and Friedrich (2007), where the authors fitted the finite mixtures of generalized linear regressions using the R software package (R Development Core Team, 2005).

In this chapter we consider some specific types of mixture models encountered in insurance and reinsurance applications. Mixtures of discrete and continuous random variables naturally arise in modeling individual claim amounts covered by the ceding and reinsurance companies

in excess of loss, and *ECOMOR* reinsurance contracts. For example, in an excess of loss contract with a limiting threshold(level), L and a retention level, M, $M \leq L$, the amount, $W_i^c = \min(W_i, M) + \max(0, W_i - L)$, covered by the cedent from each individual claim W_i, $i = 1, 2, ...$, with generic distribution $F_W(x)$, has a distribution function

$$F_{W^c}(x) = I_{\{x<M\}} F_W(x) + I_{\{M \leq x\}} F_W(L - M + x) \tag{2.2}$$

which is a mixture distribution. It is not difficult to see that (2.2) can be equivalently expressed as

$$\overline{F}_{W^c}(x) = I_{\{x<M\}} \overline{F}_W(x) + I_{\{M \leq x\}} \overline{F}_W(L) \overline{H}(x) \tag{2.3}$$

where $H(x)$ is the conditional distribution of the exceedances $M + W_i - L$ under the condition $W_i > L$ defined by

$$H(x) = \frac{F_W(L - M + x) - F_W(L)}{1 - F_W(L)}.$$

For brevity, we will denote by $H(x)$ the distribution of the exceedances $M + W_i - L \,|\, W_i > L$. Clearly, with the transformation $y = x - M$, the latter distribution function can be re-written as

$$H(y + M) = \frac{F_W(L + y) - F_W(L)}{(1 - F_W(L))}, \quad y \geq 0$$

which is the distribution of the exceedances $W_i - L \,|\, W_i > L$. Following a result, due to Balkema and de Haan (1974), and Pickands (1975) we can conclude that, for a high enough threshold L, the distribution $H(x + M)$ can be well approximated with a generalized Pareto distribution. This suggests that, for high enough threshold L, one can assume that individual claim amounts below L come from a generic distribution $F_W(x)$, whereas claim amounts above L come from a generalized Pareto distribution with appropriate parameters.

The interpretation of model (2.2) in the form of the mixture distribution in (2.3) motivates us to consider slightly more general (re)insurance applications in which individual claim amounts X have mixture distribution, i.e. claims below a threshold u are assumed to come from a distribution F_1, whereas claim amounts above u are assumed to come from a different distribution F_2, usually with heavier tail than F_1. A model of this type, has been considered by Behrens et al. (2004) in the context of Bayesian threshold estimation of the extreme value distributions. The authors adopt the generalized Pareto distribution for the approximation of F_2, the tail part of the mixture distribution. The corresponding mixture distribution, $F^{(2)}$ can be formally written as

$$\overline{F}^{(2)}(x) = I_{\{x<u\}}\overline{F}_1(x) + I_{\{u\leq x\}}\overline{F}_1(u)\overline{F}_2(x) \tag{2.4}$$

where $\overline{F}(x) = 1 - F(x)$, for any distribution function $F(x)$. Obviously, $F^{(2)}(x)$ is a distribution with a jump discontinuity at $x = u$, unless $\overline{F}_2(u) = 1$. A slight variation of (2.4) is the model

$$\overline{F}^{(2)}(x) = I_{\{x<u\}}\overline{F}_1(x) + I_{\{u\leq x\}}\overline{F}_1(u)\overline{F}_2(x-u) \tag{2.5}$$

if there exists a density of $F_2(x)$ on $[0,\infty)$. Often reinsurance arrangements involve complicated reinsurance programs with more than two levels which require considering even a more complex mixture of k distributional componets $\{F_i\}_{i=1,\ldots,k}$ such that $F_i(0) = 0$ and the density $f_i(x)$ exists on $[0,\infty)$ for each i. Existence of $f_i(x)$ is required since in the sequel we will be dealing with hazard rate functions of the mixture distributions. The mixture model is of the form

$$F^{(k)}(x) = \begin{cases} H_1(x) & \text{, if } x < u_1 \\ H_2(x) & \text{, if } u_1 \leq x < u_2 \\ \vdots & \\ H_k(x) & \text{, if } u_{k-1} \leq x \end{cases} \tag{2.6}$$

where $\overline{H}_i(x) = \overline{H}_{i-1}(u_{i-1})\overline{F}_i(x-u_{i-1})$ for each $1 < i \leq k$ and $\overline{H}_1 = \overline{F}_1$.

It is interesting to note that the mixture distribution of the type (2.6) can be written as the distribution of a sum of layer random variables, which is a probabilistic interpretation of this mixture model. Let X_i, $i = 1, \ldots, k$ be the loss random variables, defined on an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with distribution functions H_i, $i = 1, \ldots, k$ as in (2.6), and define the layer random variable $L_i(\omega)$ induced from X_i by

$$L_i(\omega) = \begin{cases} 0 & \text{if} \quad X_i(\omega) < u_i \\ X_i(\omega) & \text{if} \quad u_i \leq X_i(\omega) < u_{i+1} \\ \infty & \text{if} \quad u_{i+1} \leq X_i(\omega) \end{cases}$$

for $\omega \in \Omega$. Then the sum of the layer random variables, $L^{(k)} = \sum_{i=1}^{k} L_i$, has the distribution function $F^{(k)}$ which is the mixture of $\{F_i\}_{i=1,\ldots,k}$, since

$$F_{L^{(k)}}(x) = \sum_{i=1}^{k} I_{[u_{i-1},u_i)}(x) H_i(x) = F^{(k)}(x) \tag{2.7}$$

where $I_{[u_i,u_{i+1})}$ denotes the indicator function on $[u_i, u_{i+1})$. In this sense, it is natural to call the mixture distribution of the type (2.6) or (2.7) *a layer mixture*.

Motivated by the modification of (2.4) as (2.5), we can consider a similar mixture distribution with continuity at each threshold by conditioning instead of shifting as follows

$$\overline{F}^{(2)}(x) = I_{\{x<u\}} \overline{F}_1(x) + I_{\{u \leq x\}} \overline{F}_1(u) \overline{F}_2(x|x \geq u). \tag{2.8}$$

where we denote by $\overline{F}_2(x|x \geq u)$ the conditional tail probability $1 - F_2(x|x \geq u)$. It is also natural that we call the mixture (2.8) a *conditional layer mixture*. The conditional mixture of k distributions can be defined in a similar way. The formal definitions of the mixtures of k distributions can be found in the next section. Although (2.5) and (2.8) possess similar mathematical forms, there is a fundamental difference between the layer mixture and the conditional layer mixture: The hazard rate function of the conditional mixture distribution is a simple mixture of the hazard rate

functions, which does not hold for the layer mixture given in (2.5). We revisit this unique property and other useful applications of the conditional layer mixture in section 2.3.

In general, the layer mixture (2.6) and the conditional layer mixture (2.8) appear not only in reinsurance applications but also in the context of general insurance. For example, the layer or conditional layer mixture distribution can be applied to model any pool of risks composed of multiple number, k of heterogeneous risks, $\{S_i\}_{i=1,\ldots,k}$. The total risk process, S can be expressed as

$$S = \sum_{j=1}^{k} S_j, \quad \text{where} \quad S_j = \sum_{i=1}^{N_j(t)} X_i^{(j)},$$

for some appropriate claims occurrence processes $N_j(t)$, $j = 1, \ldots, k$. Since S_j are not identical, we have k distinct distribution functions, $\{F_{S_j}\}_{j=1,\ldots,k}$, and hence the risk process S can be modelled by the sum of k risk processes $\{S_j\}_{j=1,\ldots,k}$ through the mixture of k distributions, $\{F_{S_j}\}_{j=1,\ldots,k}$. Pools of operational risks have been considered recently in operational risk modelling by means of the so called standardized operational risk classification matrix which classifies risks into eight business lines and seven event types according to the principles proposed by Basel II (Basel Committee on Banking Supervision, 2004). For more details, see Moscadelli (2004); Dutta and Perry (2004). If S_j takes values in the interval $[u_{i-1}, u_i)$ for each $i = 1, \ldots, k$ where u_k can be possibly infinity, we can model the risk S by the layer or conditional layer mixture of k distributions as follows,

$$\mathbb{P}(S \leq x) = F^{(k)}(x) = \sum_{j=1}^{k} I_{[u_{j-1},u_j)} H_j(x)$$

where $\overline{H}_j(x) = \overline{H}_{j-1}(u_{j-1})\overline{F}_{S_j}(x - u_{j-1})$ or $\overline{H}_{j-1}(u_{j-1})\overline{F}_{S_j}(x|x \geq u_{j-1})$ as in (2.6) and (2.8).

In all such models, where the distribution of the individual claim or the aggregate claim amounts can be represented as an appropriate mixture of various distributions, it is essential to be able to derive conclusions about the asymptotic behavior of the extreme claim amounts as the size of the claims tends to infinity. As is well known such large claims often result from catastrophic events and cause huge financial losses and even bankruptcy of insurance and reinsurance

companies. This has justified the considerable interest in the asymptotic behavior of the maximum of n, $(n > 1)$ claims with a continuous generic distribution, which has been addressed in a number of recent publications, among which the monograph by Embrechts et al. (2002), Kotz and Nadarajah (2000), papers by Klüppelberg (2006) and Cebrian et al. (2003). Relatively little attention has been devoted to studying the asymptotic behavior of maxima in the case of claims modelled by a mixture distribution of the kind described above. A non-actuarial paper, dealing with asymptotic properties of maxima of mixtures applied to the context of parallel processing of a task is the paper by Kang and Serfozo (1999).

The aim of this chapter is to study the asymptotic behavior of the maximum of a series of claim amounts modelled by a mixture distribution of the risk described above and their actuarial applications of fitting a loss distribution by a mixture of certain distributional components. The outline of the chapter is as follows. In section 2.1 we present the formal definitions of three types of mixture distributions with a sequence of thresholds $0 = u_0 < u_1 < \ldots$. In section 2.2 we study the maximum domain of attraction and the normalizing constants of the mixture distributions. In section 2.3 we generalize the definition of the conditional layer mixture in (2.8) or (2.10) by considering the case $k \to \infty$, i.e. considering mixtures of infinitely many layers with infinitely many thresholds, which we call the *infinite layer mixture*. Such models are very interesting because they allow flexible modelling of the extreme behavior on the entire domain without strict threshold restriction, by assuming appropriate extreme value index for each layer. We examine the unique property of the conditional layer mixture model and its hazard rate function representation. Another useful aspect of the conditional layer mixture model is its convenient hazard rate function representation. In particular, as we will show in section 2.3.1, in order to define a conditional layer mixture with differentiable cdf one needs to define a continuous simple mixture of the hazard rate functions of the distributional components. We also discuss the limiting distribution of the infinite mixture model as the size of each layer gets arbitrarily small, which enables us to approximate any distribution with continuous hazard rate function by a infinite conditional layer mixture of exponential distributions. Numerical application of the conditional layer mixture on g-and-h distribution is given in section 2.4.

2.1 Three Mixture Distribution Models

In this section we present the formal definitions of the three mixture distribution models, the layer mixture, the linear mixture, and the conditional layer mixture of k distributions introduced in $(2.5) \sim (2.8)$. We also discuss the characteristics of the mixture models and give some illustrations.

2.1.1 Layer Mixture Distributions

The mixture of two distributions in (2.5) leads us to defining recursively the mixture of k, $(k > 0)$ distributions, as follows.

Definition 2.1 (Layer Mixture) Given distributions $\{F_i\}_{i=1,2,...}$ and thresholds $0 = u_0 < u_1 < \ldots$, the layer mixture of the first k distributions denoted by $F^{(k)}$ is defined recursively as

$$\overline{F^{(k)}}(x) = I_{\{x<u_{k-1}\}}(x)\overline{F^{(k-1)}}(x) + I_{\{x\geq u_{k-1}\}}(x)\overline{F^{(k-1)}}(u_{k-1})\overline{F}_k(x-u_{k-1}) \tag{2.9}$$

for any integer $k > 1$ and $\overline{F}^{(1)}(x) = \overline{F}_1(x)$.

The mixture distribution $F^{(k)}$ is simply an extension of the mixture of two distributions. It is well-defined and continuous at each threshold u_i. Figure 2.1 is an illustration of the mixture distribution of light-tailed exponential distributions and a heavy-tailed Pareto distribution.

2.1.2 Linear Mixture Distributions

As we already mentioned in the introduction, the layer mixture distribution has its motivation from practical applications, for example, in reinsurance modelling with multiple layers with more than one thresholds. The second mixture model we are going to consider here represents a generalization of the model (2.1) since we now release the restriction on the weights c_j to be strictly positive. Thus, we will consider mixtures with positive or negative coefficients which sum to one.

Definition 2.2 (Linear Mixture) Given distributions $\{F_i\}_{i=1,2,...}$, the linear mixture of k distributions denoted by $F^{(k)}$ is defined by

$$F^{(k)}(x) = \sum_{i=1}^{k} c_i F_i(x)$$

where c_i are negative or positive constants that sum to 1.

Kang and Serfozo (1999) considered more general linear mixture of countable collection of distributions and examined the maximum domain of attraction of the mixture. It is important to consider such linear mixture models since, as we will show in the next section, any layer mixture distribution can be represented as a linear mixture. Therefore, all of the important asymptotic results for the linear mixture model, see Kang and Serfozo (1999), can be applied to the layer mixture model. Next we consider the conditional layer mixture, introduced earlier, and illustrate its useful properties.

2.1.3 Conditional Layer Mixture Distributions

The conditional layer mixture model is a modification of the layer mixture model, which improves the behavior of the mixture distribution in the neighborhood of each threshold. The layer mixture model in Definition 2.1 has a drawback in that each distribution F_i in $F^{(k)}$ behaves at each threshold u_i as at the origin since each F_i is shifted by u_i. We can see a clear evidence of this in the left panel of Figure 2.1. At the threshold u_3, Pareto distribution F_4 is mixed with $F^{(3)}$. The distributional components F_1, F_2, and F_3 are exponentially distributed and hence F_4 is heavier-tailed than $F^{(3)}$. However, the graph shows that there is a slight bump at $x = u_3$, which can be explained by the fact that, near the threshold u_3, the mixture distribution is no longer consistent in terms of increasing heavy-tailedness. This is mainly due to the underestimating of the scaling effect of the Pareto distribution in the mixture model. In the neighborhood of $x = 0$, even heavy-tailed Pareto is viewed less heavy-tailed than the mixture of three exponential distributions in the neighborhood of u_3. In order to solve this problem, we introduce the following mixture distribution. The key point of this mixture is that, for a given interval $[u_i, u_{i+1})$, we adopt the i-th distribution exactly on the

same interval without loss of continuity at each threshold. This is possible by way of conditional survival functions.

Definition 2.3 (Conditional Layer Mixture) Given distributions $\{F_i\}_{i=1,2,\ldots}$ and thresholds $0 = u_0 < u_1 < \ldots$, the conditional layer mixture of the first k distributions denoted by $F^{(k)}$ is defined recursively as

$$\overline{F^{(k)}}(x) = I_{\{x<u_{k-1}\}}(x)\overline{F^{(k-1)}}(x) + I_{\{x\geq u_{k-1}\}}(x)\overline{F^{(k-1)}}(u_{k-1})\overline{F}_k(x|x \geq u_{k-1}) \tag{2.10}$$

for any integer $k > 1$ and $\overline{F}^{(1)}(x) = \overline{F}_1(x)$, where $\overline{F}_k(x|x \geq u_{k-1}) = 1 - F_k(x|x \geq u_{k-1})$.

Figure 2.1 illustrates the consistency of increasing heavy-tailedness of the conditional layer mixture distribution. Since there is no scaling of thresholds any more, higher order mixture distribution possesses heavier tail, i.e.

$$\overline{F^{(i)}}(x) \leq \overline{F^{(j)}}(x)$$

for $i < j$ if $\overline{F}_i(x) \leq \overline{F}_j(x)$ on $[u_{j-1}, \infty)$. It is very clear that the slight bump at threshold u_3 in the left disappears in the right panel of Figure 2.1.

2.2 Maximum Domain of Attractions of Mixture Distributions

Here we are interested in the maximum domain of attraction of each mixture model we introduced in the previous section. The fundamental theorem by Fisher and Tippett (1928) classifies the possible limit laws of the maxima of *i.i.d* random variables X_i, $M_n = \max(X_1, \ldots, X_n)$, as $n \to \infty$, introducing appropriate normalizing constants a_n and b_n. If there exist normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$ and some non-degenerating distribution H such that

$$\frac{M_n - b_n}{a_n} \longrightarrow H$$

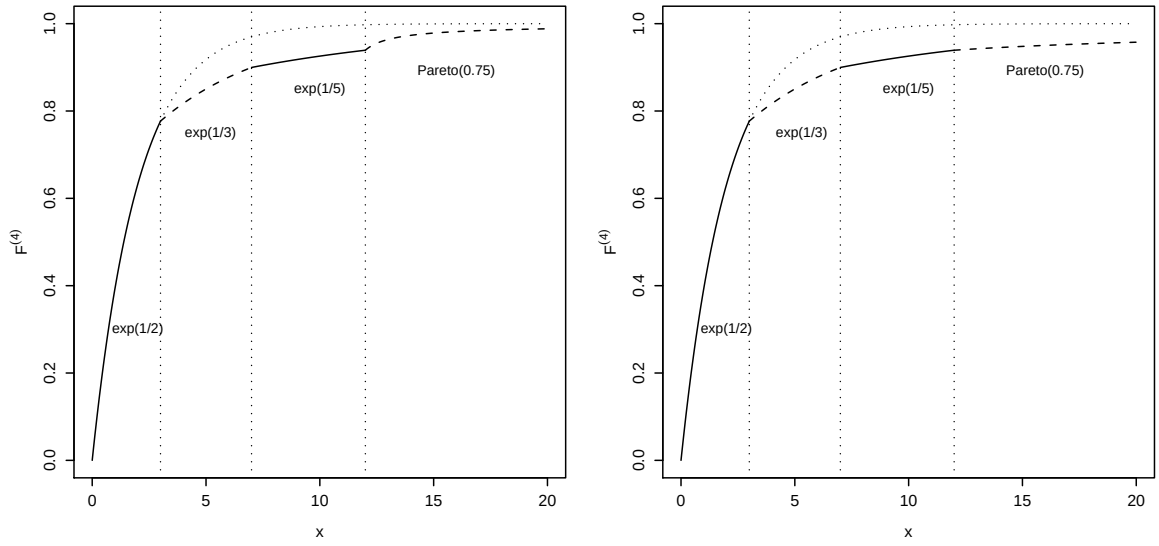


Figure 2.1 Layer mixture distribution and Conditional layer mixture distribution $(k = 4)$.

in distribution, then H is equal to one of the three distributions,

$$\begin{aligned}
\text{Fréchet} \quad \Phi_\alpha(x) &= \begin{cases} 0, & \text{if } x \le 0 \\ \exp\big(-x^{-\alpha}\big), & \text{if } x > 0 \end{cases} \\
\text{Weibull} \quad \Psi_\alpha(x) &= \begin{cases} \exp\big(-(-x)^{\alpha}\big), & \text{if } x \le 0 \\ 1, & \text{if } x > 0 \end{cases} \\
\text{Gumbel} \quad \Lambda(x) &= \begin{cases} \exp\big(-e^{-x}\big), & \text{if } x \in \mathbb{R} \\ 1, & \text{if } x > 0 \end{cases}
\end{aligned} \tag{2.11}$$

for $\alpha > 0$ and this is usually expressed as $F \in MDA(H)$ where F is the common distribution function for X_i and H is either one of the extreme value distribution in (2.11). Here we will be concerned with the asymptotic behavior of the maxima, M_n, in the case when X_i has a mixture distribution, $F^{(k)}$, whose k-th component, H_k belongs to one of the maximum domain of attraction $MDA(H)$, Φ, Λ, or Ψ. Intuitively we would expect that the mixture distribution would belong to the same MDA, i.e.

$$F^{(k)} \in MDA(\cdot) \quad \text{if} \quad F_k \in MDA(\cdot) \tag{2.12}$$

since MDA is about the limiting distribution of the maxima M_n, which is governed by the tail of the distribution if n is large enough. For similar asymptotic properties of the family of heavy-tailed distributions, see Cai and Tang (2004). Some results for the MDA of mixture distributions can be found in the following papers. Mladenović (1999) found the normalizing constants in special examples such as normal mixtures and Cauchy mixtures. Kang and Serfozo (1999) derived general formula of the normalizing constants under the existence of tail-dominant distribution.

Before we show that (2.12) holds for all the three mixture models introduced in section 2.1, for convenience we state two lemmas which will be used repeatedly throughout this chapter.

Lemma 2.4 (Convergence Criterion) Let F be a distribution function and $a_n > 0, b_n \in \mathbb{R}$ for $n = 1, 2, \ldots$. Then the following two statements are equivalent.

1. For an extreme value distribution H of the type in (2.11) with normalizing constants a_n and b_n,
$$F \in MDA(H)$$
2. For all $x \in \mathbb{R}$, as $n \to \infty$,
$$n\overline{F}(a_n x + b_n) \to -\log H(x)$$

Lemma 2.5 (Kang and Serfozo (1999)) The following statements are equivalent for an extreme value distribution H of the type in (2.11). For $\gamma > 0$,

1. $F \in MDA(H)$ with normalizing constants a_n and b_n.
2. $n\overline{F}(a_n^* x + b_n^*) \sim -\frac{1}{\gamma}\log H(x) \quad \text{as} \quad n \to \infty$,

where the normalizing constants are related as follows

$$\begin{cases} a_n^* = \gamma^{1/\alpha} a_n & b_n^* = b_n = 0 & \text{if } H \text{ is Fréchet,} \\ a_n^* = a_n & b_n^* = b_n + a_n \log\gamma & \text{if } H \text{ is Gumbel,} \\ a_n^* = \gamma^{-1/\alpha} a_n & b_n^* = b_n = 0 & \text{if } H \text{ is Weibull.} \end{cases}$$

The above lemmas provide us with a method to determine the maximum domain of attractions and the normalizing constants for the mixtures of k distributions introduced in section 2.1. For more details, see Theorem 1.5.1, Leadbetter et al. (1983) and Theorem 2, Kang and Serfozo (1999).

2.2.1 MDA of Layer Mixture Distributions

Let $F^{(k)}$ be a layer mixture distributions given by Definition 2.1. Then the following proposition implies that $F^{(k)}$ belongs to the maximum domain of attraction of F_k with modified normalizing constants.

Proposition 2.6 Let $F^{(k)}$ be of the type as in Definition 2.1. If $F_k \in MDA(H)$ for an extreme value distribution H of the type in (2.11), then $F^{(k)} \in MDA(H)$ and the normalizing constants are

$$\begin{cases} a_n = \gamma^{1/\alpha} a_n^*, \quad b_n = u_{k-1}, & \text{if } H \text{ is Fréchet,} \\ a_n = a_n^*, \quad b_n = b_n^* + u_{k-1} + a_n^* \log \gamma, & \text{if } H \text{ is Gumbel,} \\ a_n = \gamma^{-1/\alpha} a_n^* \quad b_n = u_{k-1}, & \text{if } H \text{ is Weibull.} \end{cases}$$

where a_n^* and b_n^* are the normalizing constants of $F_k(x)$, $x_n^* = a_n^* x + b_n^*$, and

$$\gamma = \prod_{i=1}^{k-1} r_i \quad \text{where} \quad r_i = \overline{F}_i(u_i - u_{i-1}).$$

Proof: Let $x_n^* = a_n^* x + b_n^*$ be the normalizing constants for $F_k(x)$, i.e.

$$n\overline{F_k}(x_n^*) \to -\log H(x).$$

By Definition 2.1 and Lemma 2.4

$$\begin{aligned} n\overline{F^{(k)}}(x_n^* + u_{k-1}) &= n\{I_{\{x_n^* < u_{k-1}\}}(x_n^*)\overline{F^{(k-1)}}(x_n^* + u_{k-1}) \\ &\qquad + I_{\{x_n^* \geq u_{k-1}\}}(x_n^* + u_{k-1})\overline{F^{(k-1)}}(u_{k-1})\overline{F_k}(x_n^*)\} \\ &\sim n\overline{F^{(k-1)}}(u_{k-1})\overline{F_k}(x_n^*) \end{aligned}$$

as $n \to \infty$. It follows that, by Lemma 2.5,

$$n\overline{F^{(k)}}(a_n^* x + b_n^* + u_{k-1}) \sim -\overline{F^{(k-1)}}(u_{k-1}) \log H(x),$$

$$n\overline{F^{(k)}}(a_n x + b_n) \sim -\log H(x)$$

where the normalizing constants for H of Fréchet type,

$$
\begin{aligned}
a_n &= \overline{F^{(k-1)}}(u_{k-1})^{1/\alpha} a_n^* \\
&= \overline{F^{(k-2)}}(u_{k-2})^{1/\alpha} \overline{F}_{k-1}(u_{k-1} - u_{k-2})^{1/\alpha} a_n^* \\
&\vdots \\
&= \overline{F}_1(u_1 - u_0)^{1/\alpha} \cdots \overline{F}_{k-1}(u_{k-1} - u_{k-2})^{1/\alpha} a_n^* \\
&= a_n^* \prod_{i=1}^{k-1} \overline{F}_i(u_i - u_{i-1})^{1/\alpha}
\end{aligned}
$$

and $b_n = b_n^* + u_{k-1} = u_{k-1}$. For the extreme value distribution of Gumbel and Weibull type, similar argument holds in the same manner. Applying the convergence criterion in Lemma 2.4 again, we conclude that $F^{(k)} \in MDA(H)$ with normalizing constants a_n and b_n, which completes the proof. □

2.2.2 MDA of Linear Mixture Distributions

Consider the linear mixture model in Definition 2.2 and suppose that there exists a distribution F^* satisfying

$$\lim_{x \to \infty} \frac{\overline{F_i}(x)}{\overline{F^*}(x)} = r_i \tag{2.13}$$

for some $r_i \geq 0$ and this limit is uniform in i in case I is an infinite set. Then we say that the tail of the distribution F^* dominates those of $\{F_i : i \in I\}$. Theorem 2 in Kang and Serfozo (1999) provides the relationship between the maximum domain of attraction and normalizing constants of the mixture distribution $F^{(k)}$ and the tail-dominating distribution F^*. We state this theorem here for convenience.

Theorem 2.7 (Kang and Serfozo (1999)) Suppose $F^{(k)}$ is a linear mixture of the type in Definition 2.2 and satisfies (2.13) for each i. Let $\gamma = \sum_{i \in I} c_i r_i$ and assume γ is positive. Then the following statements are equivalent.

1. $F \in \text{MDA}(H)$ with normalizing constants a_n, b_n.

2. $F^* \in \text{MDA}(H)$ with normalizing constants a_n^*, b_n^*.

When these statements hold, the normalizing constants are related as follows

$$\begin{cases} a_n = \gamma^{1/\alpha} a_n^*, & b_n = b_n^* = 0, & \text{if } H \text{ is Fréchet,} \\ a_n = a_n^*, & b_n = b_n^* + a_n^* \log \gamma, & \text{if } H \text{ is Gumbel,} \\ a_n = \gamma^{-1/\alpha} a_n^*, & b_n = b_n^* = 0, & \text{if } H \text{ is Weibull.} \end{cases}$$

where α is the extreme value index of each type of extreme value distribution defined in (2.11).

The linear mixture of k distributions, therefore, belongs to the maximum domain of attraction of F_k if there exists a tail-dominating distribution F^* with $\gamma > 0$. In practice, the existence of the distribution F^* is assumed without loss of generality since higher layer is often modelled by heavier-tailed distribution and we may set $F^* = F_k$ which reduces to $\lim_{x\to\infty} \overline{F_i}(x)/\overline{F^*}(x) = r_i \geq 0$ for all $i = 1, \ldots, k$.

As we mentioned in the previous section, the layer mixture model in Definition 2.1 can be written as a linear mixture model with appropriate positive or negative weights. This is an interesting result since it allows us to interpret the layer mixture models, (2.3)∼(2.6) and (2.9), with their interesting (re)insurance applications (see the introduction) as a linear mixture model and apply the known results such as Theorem 2 of Kang and Serfozo (1999) to the layer mixture model to investigate their asymptotic behavior. Let $\{F_i\}_{i=1,\ldots,k}$ be the components of the layer mixture distribution $F^{(k)}$ with thresholds $0 = u_0 < u_1 < \cdots < u_{k-1}$ such that

$$\overline{F^{(k)}}(x) = I_{\{x<u_{k-1}\}}(x)\overline{F^{(k-1)}}(x) + I_{\{x\geq u_{k-1}\}}(x)\overline{F^{(k-1)}}(u_{k-1})\overline{F}_k(x-u_{k-1}).$$

It is equivalent to

$$F^{(k)}(x)\Big|_{[u_{i-1},u_i)} = H_i(x) \tag{2.14}$$

where $\overline{H_i}(x) = \overline{H_{i-1}}(u_{i-1})\overline{F}_i(x-u_{i-1})$ for $i = 2, 3, \cdots, k$ and $H_1 = F_1$. The following theorem shows that a layer mixture distribution can be written as a linear combination of $H_i I_{\{\cdot\}}$ with constants $c_i = (-1)^{i-1}$.

Theorem 2.8 The layer mixture distribution can be written as a linear mixture of the form

$$F^{(k)}(x) = c_1 H_1(x) + \sum_{j=1}^{k-1} \sum_{i=2j}^{2j+1} c_i H_{i-j}(x) I_{\{u_j \le x\}}(x)$$

where $c_i = (-1)^{i-1}$ for $i = 1, 2, \cdots, 2k-1$ and $\overline{H_i}(x) = \overline{H_{i-1}}(u_{i-1})\overline{F_i}(x - u_{i-1})$ for $i = 2, 3, \cdots, k$ and $\overline{H}_1 = \overline{F}_1$.

Proof: By induction on k. For $k = 2$, it is immediate to see

$$H_1(x) - H_1(x) I_{\{u_1 < x\}}(x) + H_2(x) I_{\{u_1 < x\}}(x) = F^{(2)}(x)$$

by direct substitution. Note that (2.14) holds for the layer mixtures, $F^{(2)}, \ldots, F^{(k-1)}$. Denote the indicator function by $I_{\{\cdot\}} = I_{\{\cdot\}}(x)$, then

$$\begin{aligned}
F^{(k)}(x) &= F^{(k-1)}(x) I_{x<u_{k-1}} + F_k(x) I_{u_{k-1} \le x} \\
&= \big\{ H_1(x) - H_1(x) I_{u_1 \le x} + H_2(x) I_{u_1 \le x} \\
&\qquad\qquad \vdots \\
&\qquad\qquad - H_{k-2}(x) I_{u_{k-2} \le x} + H_{k-1}(x) I_{u_{k-2} \le x} \big\} \cdot I_{x<u_{k-1}} \\
&\quad + H_k(x) I_{u_{k-1} \le x} \\
&= H_1(x) I_{x<u_{k-1}} - H_1(x) I_{u_1 \le x < u_{k-1}} + H_2(x) I_{u_1 \le x < u_{k-1}} \\
&\qquad\qquad \vdots \\
&\qquad\qquad - H_{k-2}(x) I_{u_{k-2} \le x < u_{k-1}} + H_{k-1}(x) I_{u_{k-2} \le x < u_{k-1}} \\
&\quad + H_k(x) I_{u_{k-1} \le x}
\end{aligned}$$

since $I_{a \le x} \cdot I_{x<b} = I_{a \le x<b}$ for $a < b$, hence

$$
\begin{aligned}
= \; & H_1(x) - H_1(x)I_{u_{k-1}\le x} \\
& -H_1(x)I_{u_1\le x} + H_1(x)I_{u_{k-1}\le x} + H_2(x)I_{u_1\le x} - H_2(x)I_{u_{k-1}\le x} \\
& \vdots \\
& -H_{k-2}(x)I_{u_{k-2}\le x} + H_{k-2}(x)I_{u_{k-1}\le x} + H_{k-1}(x)I_{u_{k-2}\le x} - H_{k-1}(x)I_{u_{k-1}\le x} \\
& +H_k(x)I_{u_{k-1}\le x}
\end{aligned}
$$

by substituting $H_i(x)I_{u_{i-1}\le x<u_{k-1}}$ with $H_i(x)I_{u_{i-1}\le x} - H_i(x)I_{u_{k-1}\le x}$ for each $i = 1, 2, \cdots, k-1$, it reduces to

$$
\begin{aligned}
= \; & H_1(x) - H_1(x)I_{u_1\le x} + H_2(x)I_{u_1\le x} - \cdots - H_{k-2}(x)I_{u_{k-2}\le x} + H_{k-1}(x)I_{u_{k-2}\le x} \\
& +I_{u_{k-1}\le x}\big\{ - H_1(x) + H_1(x) - H_2(x) + \cdots + H_{k-2}(x) - H_{k-1}(x)\big\} \\
& +H_k(x)I_{u_{k-1}\le x} \\
= \; & H_1(x) - H_1(x)I_{u_1\le x} + H_2(x)I_{u_1\le x} - \cdots - H_{k-2}(x)I_{u_{k-2}\le x} + H_{k-1}(x)I_{u_{k-2}\le x} \\
& -H_{k-1}(x)I_{u_{k-1}\le x} + H_k(x)I_{u_{k-1}\le x} \\
= \; & H_1(x) + \sum_{j=1}^{k-1}\sum_{i=2j}^{2j+1}(-1)^{i-1}H_{i-j}(x)I_{u_j\le x}
\end{aligned}
$$

which proves the theorem. □

If the tail of F_k dominates those of $H_i I_{\{\}}$, the above theorem leads to the same result as in Proposition 2.6 by Theorem 2 of Kang and Serfozo (1999) with tail-dominating distribution $F^* = F_k$ as follows.

Proposition 2.9 Let $F^{(k)}$ be a layer mixture distribution as in Definition 2.1. The layer mixture $F^{(k)}$ can be written as a linear mixture of constants c_i and the distributions $H_i I_{\{\cdot\}}$ as in (2.14), tail-dominated by $\overline{F}_k$ with $\gamma > 0$. If $F_k \in MDA(H)$ for an extreme value distribution H of the

type in (2.11), then $F^{(k)} \in MDA(H)$ and the normalizing constants are

$$\begin{cases} a_n = \gamma^{1/\alpha} a_n^*, \quad b_n = u_{k-1}, & \text{if } H \text{ is Fréchet,} \\ a_n = a_n^*, \quad b_n = b_n^* + u_{k-1} + a_n^* \log \gamma, & \text{if } H \text{ is Gumbel,} \\ a_n = \gamma^{-1/\alpha} a_n^* \quad b_n = u_{k-1}, & \text{if } H \text{ is Weibull.} \end{cases}$$

where a_n^* and b_n^* are the normalizing constants of F_k, $x_n^* = a_n^* x + b_n^*$ and

$$\gamma = \sum_{i=1}^{k} c_i r_i = \prod_{i=1}^{k-1} \bar{F}_i(u_i - u_{i-1}).$$

Proof: Let a_n^* and b_n^* be the normalizing constants for F_k. If we set $F^*(x) = F_k(x - u_{k-1})$, a_n^* and $b_n^* + u_{k-1}$ are the normalizing constants for F^*. By the proof of Lemma 2.8, $\sum_{i=1}^{2k-2} c_i r_i = 0$, $c_{2k-1} = 1$, and

$$\begin{aligned} \gamma &= r_{2k-1} = \lim_{x \to \infty} \frac{\overline{H}_k(x) I_{\{u_{k-1} < x\}}(x)}{\overline{F}^*(x)} \\ &= \lim_{x \to \infty} \frac{\overline{H}_{k-1}(u_{k-1}) \overline{F}_k(x - u_{k-1})}{\overline{F}_k(x - u_{k-1})} = \prod_{i=1}^{k-1} \bar{F}_i(u_i - u_{i-1}) > 0. \end{aligned}$$

Therefore the theorem follows immediately by Theorem 2.7. □

Thus, we can see that Proposition 2.9 makes it possible to pick up F_k as a tail-dominating distribution F^* of all distributional components $H_i I_{\{\cdot\}}$ of the linear mixture representation of the layer mixture model (2.9) and apply Theorem 2.7 to examine the extreme behavior of the layer mixture. It is instructive to illustrate this point and consider some examples of how Proposition 2.9 applies to particular tail-dominating distributions F^* with heavy-tails.

Example 2.10 (Pareto) Consider a layer mixture of k distributions with thresholds $0 = u_0 < u_1 < \ldots < u_{k-1}$ for $k > 1$ such that $F_i(x) = 1 - e^{-\frac{1}{\mu_i} x}$ for $i = 1, \ldots, k-1$ and $F_k(x) = 1 - (1 + \xi x)^{-\frac{1}{\xi}}$ for $\xi > 0$ and $1 + \xi x > 0$. Denote $u_i - u_{i-1}$ by δ_i for each $i = 1, \ldots, k$. If we

choose $F^* = F_k$,

$$\begin{aligned}
\gamma &= \lim_{x\to\infty} \frac{\overline{H}_k(x) I_{\{u_{k-1}<x\}}(x)}{\overline{F^*}(x)} \\
&= \lim_{x\to\infty} \frac{\overline{H}_{k-1}(x)\overline{F}_k(x-u_{k-1})}{\overline{F_k}(x)} \\
&= \overline{F}_1(\delta_1)\overline{F}_2(\delta_2)\cdots\overline{F}_{k-1}(\delta_{k-1}) \\
&= e^{-\delta_1/\mu_1} e^{-\delta_2/\mu_2}\cdots e^{-\delta_{k-1}/\mu_{k-1}} = e^{-\sum_{i=1}^{k-1}\frac{\delta_i}{\mu_i}} > 0.
\end{aligned}$$

where H_i are as in (2.14). By Proposition 2.9 we conclude that $F^{(k)}$ belongs to the same maximum domain of attractions $MDA(\Phi_{1/\xi})$ of the generalized Pareto distribution and the normalizing constants are $a_n = \gamma^\xi\, a_n^*$ and $b_n = b_n^*$ where a_n^*, b_n^* are the normalizing constants of the generalized Pareto distribution F_k. We provide the numerical result of the mixture distribution $F^{(k)}$ with $k = 4$, $\mu_i = i$ and $u_i = i$ in Figure 2.2.

Example 2.11 (Lognormal) Consider the layer mixture distribution in Exmaple 2.10 when H_k is a lognormal distribution. Denote the mean and standard deviation of $\log X$ by μ, and σ, respectively. Note that

$$\overline{F_k}(x) = 1 - \Phi\Big(\frac{\log x - \mu}{\sigma}\Big)$$

where Φ is the standard normal distribution function. It is well known that Lognormal distribution belongs to the maximum domain of attraction of Gumbel distribution, $MDA(\Lambda)$. If we choose $F^* = F_k$,

$$\begin{aligned}
\gamma &= \lim_{x\to\infty} \frac{\overline{H}_k(x) I_{\{u_{k-1}<x\}}(x)}{\overline{F^*}(x)} \\
&= \overline{H}_{k-1}(u_{k-1}) \lim_{x\to\infty} \frac{1-\Phi\left(\frac{\log(x-u_{k-1})-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)} \\
&= \overline{H}_{k-1}(u_{k-1}) \lim_{x\to\infty} \frac{-\frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x-u_{k-1})-\mu}{\sigma}\right)^2}}{-\frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}} \\
&= \overline{H}_{k-1}(u_{k-1}) \lim_{x\to\infty} e^{-\frac{1}{\sigma^2}(\log x-\log(x-u_{k-1}))(\log x+\log(x-u_{k-1})-2\mu)}
\end{aligned}$$

where H_i are as in (2.14), and applying $\log x - \log(x-u) = \int_{x-u}^{x} \frac{1}{s}\, ds \le \frac{u}{x-u}$, we get $(\log x + \log(x-u) - 2\mu)(\log x - \log(x-u)) \le (\log x + \log(x-u) - 2\mu)\frac{u}{x-u} \to 0$ as $x \to \infty$ for any $u > 0$ and hence $(\log x - \log(x-u_{k-1}))(\log x + \log(x-u_{k-1}) - 2\mu) \to 0$, which leads to

$$\begin{aligned}
&= \overline{H}_{k-1}(u_{k-1})\, e^0 \\
&= \overline{F}_1(\delta_1)\overline{F}_2(\delta_2)\cdots\overline{F}_{k-1}(\delta_{k-1}) \\
&= e^{-\sum_{i=1}^{k-1}\frac{\delta_i}{\mu_i}} > 0.
\end{aligned}$$

By Proposition 2.9 again, we conclude that $F^{(k)}$ belongs to the same maximum domain of attractions $MDA(\Lambda)$ of the Lognormal distribution and the normalizing constants are $a_n = a_n^*$ and $b_n = b_n^* + a_n^* \log\gamma$ where a_n^* and b_n^* are the normalizing constants of the Lognormal distribution H_k. We also provide a numerical result of the mixture distribution $F^{(k)}$ with $k = 4$, $\mu_i = i$, $\xi = 0.75$ and $u_i = i$ in Figure 2.2.

Remark 2.12 In the examples above we observe that the normalizing constants of the layer mixture distributions $F^{(k)}$ are $(\gamma^{1/\alpha}a_n^*, b_n^*)$ or $(a_n^*, b_n^* + a_n^*\log\gamma)$ without the threshold shift by u_{k-1} since we chose $F^* = F_k$ instead of $F^*(x) = F_k(x-u_{k-1})$. This is always possible when the k-th distribution is long-tailed, i.e.

$$\lim_{x\to\infty} \frac{\overline{F}_k(x-u)}{\overline{F}_k(x)} = 1.$$

for $u > 0$.

2.2.3 MDA of Conditional Layer Mixture Distributions

The normalizing constants and maximum domain of attractions of the conditional mixture distribution can be found in a similar way as in the layer mixture distributions. The following proposition is analogous to Proposition 2.6.

Proposition 2.13 Let $F^{(k)}$ be of the type in Definition 2.3. If $F_k \in MDA(H)$ for an extreme value distribution H of the type in (2.11)., then $F^{(k)} \in MDA(H)$ and the normalizing constants are

$$\begin{cases} a_n = \gamma^{1/\alpha} a_n^*, \quad b_n = 0, & \text{if } H \text{ is Fréchet,} \\ a_n = a_n^*, \quad b_n = b_n^* + a_n^* \log\gamma, & \text{if } H \text{ is Gumbel,} \\ a_n = \gamma^{-1/\alpha} a_n^*, \quad b_n = 0, & \text{if } H \text{ is Weibull.} \end{cases}$$

where a_n^* and b_n^* are the normalizing constants of F_k, $x_n^* = a_n^* x + b_n^*$, and

$$\gamma = \prod_{i=1}^{k-1} r_i \quad \text{where} \quad r_i = \frac{\overline{F_i}(u_i)}{\overline{F_{i+1}}(u_i)}.$$

Proof: Let $x_n^* = a_n^* x + b_n^*$ be the normalizing constants for F_k, i.e.

$$n\overline{F_k}(x_n^*) \to -\log H(x)$$

by Lemma 2.4. It follows that

$$
\begin{aligned}
n\overline{F^{(k)}}(x_n^*) &= n\{I_{\{x_n^*\le u_{k-1}\}}(x_n^*)\overline{F^{(k-1)}}(x_n^*) \\
&\qquad +I_{\{x>u_{k-1}\}}(x_n^*)\overline{F^{(k-1)}}(u_{k-1})\overline{F_k}(x_n^*|x>u_{k-1})\} \\
&\sim n\,\overline{F}^{(k-1)}(u_{k-1})\overline{F}_k(x_n^*|x>u_{k-1}) \\
&= n\,\overline{F}^{(k-2)}(u_{k-2})\overline{F}_{k-1}(u_{k-1}|x>u_{k-2})\overline{F}_k(x_n^*|x>u_{k-1}) \\
&= n\,\overline{F}_1(u_1|x>u_0)\cdots\overline{F}_{k-1}(u_{k-1}|x>u_{k-2})\overline{F}_k(x_n^*|x>u_{k-1}) \\
&= n\,\frac{\overline{F}_1(u_1)}{\overline{F}_1(u_0)}\cdots\frac{\overline{F}_{k-1}(u_{k-1})}{\overline{F}_{k-1}(u_{k-2})}\cdot\frac{\overline{F}_k(x_n^*)}{\overline{F}_k(u_{k-1})} \\
&= n\prod_{i=1}^{k-1}\frac{\overline{F}_i(u_i)}{\overline{F}_{i+1}(u_i)}\cdot\overline{F}_k(x_n^*) = n\prod_{i=1}^{k-1}r_i\cdot\overline{F}_k(x_n^*) \\
&= n\,\gamma\,\overline{F}_k(x_n^*)
\end{aligned}
$$

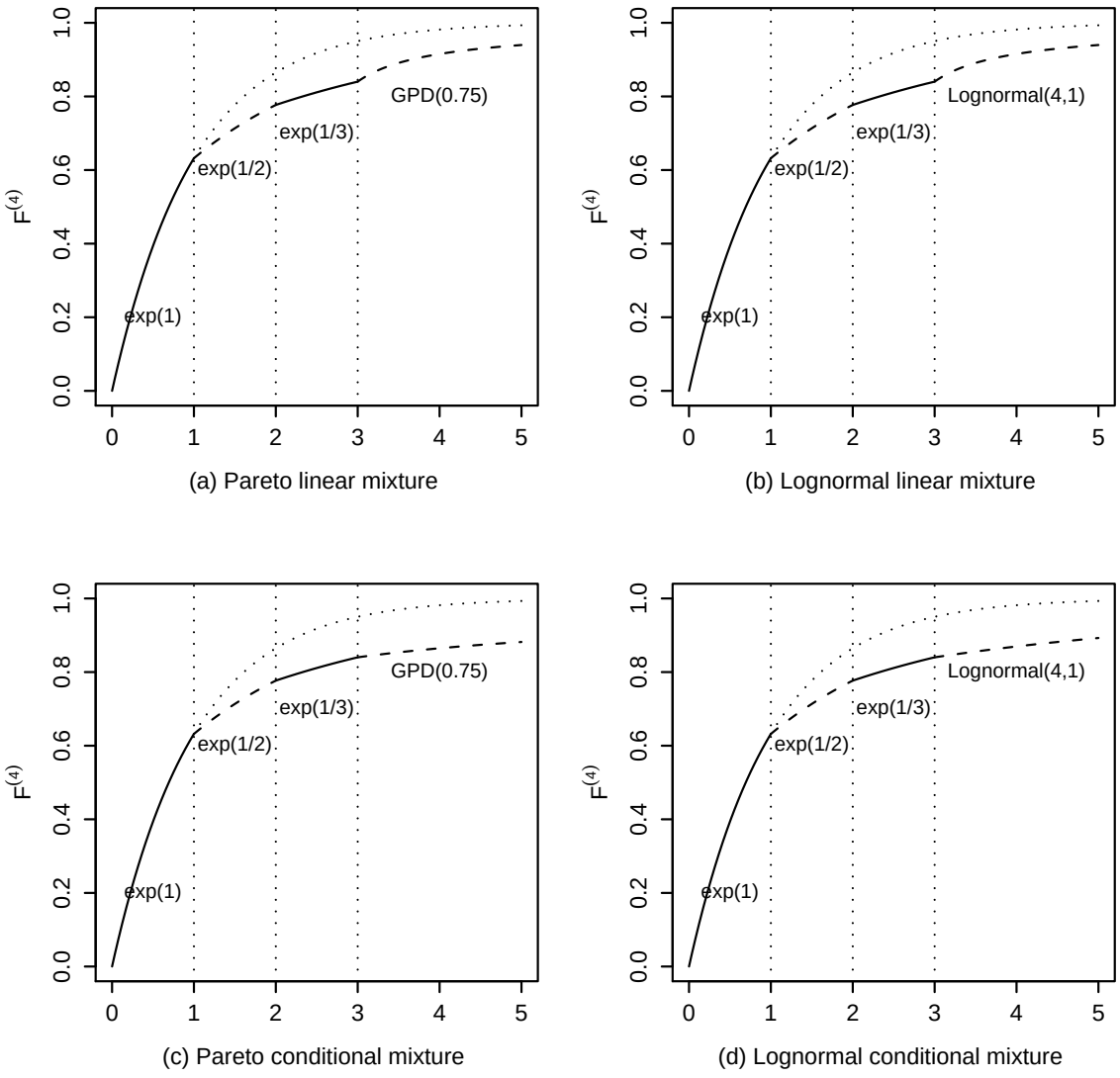
by similar argument as in the proof of Proposition 2.6. It follows that by Lemma 2.5,

$$
\begin{aligned}
n\overline{F^{(k)}}(x_n^*) &\sim -\gamma\log H(x), \\
n\overline{F^{(k)}}(a_nx+b_n) &\sim -\log H(x)
\end{aligned}
$$

where the normalizing constants for H of Fréchet type are

$$
a_n = \gamma^{1/\alpha}a_n^* = a_n^*\prod_{i=1}^{k-1}\Big(\frac{\overline{F}_i(u_i)}{\overline{F_{i+1}}(u_i)}\Big)^{1/\alpha}
$$

and $b_n=0$. For the extreme value distribution of Gumbel and Weibull, similar argument holds in the same manner. Applying the convergence criterion in Lemma 2.4 again, we conclude $F^{(k)}\in MDA(H)$ with normalizing constants a_n and b_n, which completes the proof. □



(a) Pareto linear mixture

(b) Lognormal linear mixture

(c) Pareto conditional mixture

(d) Lognormal conditional mixture

Figure 2.2 Layer mixture distributions vs. conditional layer mixture distributions $(k = 4)$.

Example 2.14 Consider a conditional mixture of k distributions at thresholds $0 = u_0 < u_1, \ldots <$ u_{k-1} such that $F_i(x) = 1 - e^{-\frac{1}{\mu_i}x}$ for $i = 1, \ldots, k-1$ and F_k belogns to some maximum domain of attraction. By Proposition 2.13, $F^{(k)}$ beglongs to the same maximum domain of attraction as F_k and

$$\begin{aligned}
\gamma &= \prod_{i=1}^{k-1} r_i = \prod_{i=1}^{k-1} \frac{\overline{F}_i(u_i)}{\overline{F}_{i+1}(u_i)} \\
&= \frac{e^{-u_1/\mu_1}}{e^{-u_1/\mu_2}} \cdots \frac{e^{-u_{k-2}/\mu_{k-2}}}{e^{-u_{k-2}/\mu_{k-1}}} \frac{e^{-u_{k-1}/\mu_{k-1}}}{\overline{F}_k(u_{k-1})} \\
&= e^{-\sum_{i=1}^{k-1} \frac{\delta_i}{\mu_i}} \Big/ \overline{F}_k(u_{k-1}).
\end{aligned}$$

where $\delta_i = u_i - u_{i-1}$ is the size of each layer. We provide the numerical results of the mixture distribution $F^{(4)}$ with $\mu_i = i$ and $u_i = i$ for $i < 4$ when F_4 is the generalized Pareto distribution and the Lognormal distribution, respectively in Figure 2.2.

2.3 Infinite Mixture Distributions and Hazard Rate Functions

The mixed distributions we proposed in the previous section are flexible enough to cover wide range of higher moments such as skewness and kurtosis. However, it is not easy to determine or estimate the thresholds $u_1, \ldots, u_{k-1}$ and corresponding distributional components $\{F_i\}_{i=1,\ldots,k}$. The latter difficulty might be overcome if the thresholds are given and there exist enough data points for each layer. Unless we have a precise method of thresholds estimation, it is not appropriate to assume specific values for the thresholds. In this sense, we propose a mixture model with infinite number of layers(thresholds) and the limiting distribution of the mixture model, which can be viewed as a mathematical background to solve the threshold selection problem. For example, a non-decreasing sequence of thresholds such as $u_k \to \infty$ defines an infinite number of layers on which appropriate distributions F_i are specified. Since there is no upper limit of the thresholds, any tail of the infinite mixture distribution is explained not by a single distributional component F_k but by an infinite number of components $\{F_i\}_{i \geq k}$ for some $k > 0$. In this way it is not required

to select a fixed threshold high enough to approximate the corresponding tail by a single F_k. In other words, the mixture models of finite number of layers cease to be mixture distributions in the tail whereas the infinite mixture model always has its tail as a infinite mixture distribution again. Since the conditional layer mixture model is superior to the layer mixture model in terms of the consistency of increasing heavy-tailedness, we develop the infinite conditional layer mixture model and its limiting distribution in the sequel.

Another important topic in this section is the hazard rate function representation of the conditional mixture distribution which can also be applied to the infinite conditional layer mixture. Since the existence(continuity) of the hazard function of a distribution is equivalent to the continuity(differentiability) of the distribution, the results in the following section provides us an easy method of creating continuous or differentiable conditional layer mixture distributions.

We also discuss the limiting distribution of the infinite conditional layer mixture distribution as the size of each layer gets arbitrarily small. As a result the limiting distribution possesses a continuously varying heavy-tailedness which can be implemented by a function of certain parameters of heavy-tailedness, for example, the shape parameter of Pareto distribution.

2.3.1 Infinite Mixture Distributions and Hazard Rate Functions

Suppose that each distribution F_i has its density function f_i and the hazard function h_i for each $i = 1, \ldots, k$, i.e.

$$h_i(x) = \frac{f_i(x)}{1 - F_i(x)} \quad \text{and} \quad \overline{F}_i(x) = \exp\big(- \int_0^x h_i(s)\, ds\big).$$

Let us define the simple mixture of the first k hazard functions by

$$h^{(k)}(x) = \sum_{i=1}^{k-1} I_{\{u_{i-1} \le x < u_i\}}(x) h_i(x) + I_{\{u_{k-1} \le x\}}(x) h_k(x).$$

which is integrable as a finite linear combination of integrable functions. The following theorem implies that the hazard function of the conditional mixture distribution is the mixture of the hazard

functions of the distributional components, which is a unique property of the conditional layer mixture distribution among all mixture models.

Theorem 2.15 Given distributions $\{F_i\}_{i=1,2,\ldots}$ and thresholds $0 = u_0 < u_1 < \ldots$, let h_i be the hazard function of F_i for each i. Then the conditional mixture distribution $F^{(k)}$ has the hazard function $h^{(k)}$ which is the mixture of the first k hazard functions, i.e.

$$\overline{F^{(k)}}(x) = \exp\big(-\int_0^x h^{(k)}(s)\,ds\big)$$

Proof: First suppose that $x \leq u_{k-1}$, then there exist thresholds such that $u_{j-1} \leq x < u_j$ for $j \leq k-1$. From the proof of Proposition 2.13, we have

$$\begin{aligned}
\overline{F^{(k)}}(x) &= \Big\{\prod_{i=1}^{j-1} \frac{\overline{F}_i(u_i)}{\overline{F}_i(u_{i-1})}\Big\} \frac{\overline{F}_j(x)}{\overline{F}_j(u_{j-1})} \\
&= \Big\{\prod_{i=1}^{j-1} \exp\big(-\int_{u_{i-1}}^{u_i} h_i(s)\,ds\big)\Big\} \exp\big(-\int_{u_{j-1}}^{x} h_j(s)\,ds\big) \\
&= \exp\big(-\int_0^x h^{(k)}(s)\,ds\big).
\end{aligned}$$

If $x > u_{k-1}$, set $j = k$, which completes the proof. □

Definition 2.16 (Infinite Conditional Layer Mixture) Given distributions $\{F_i\}_{i=1,2,\ldots}$ and thresholds $0 = u_0 < u_1 < \ldots$, the infinite conditional layer mixture distribution denoted by $F^{(\infty)}$ is defined by

$$\overline{F^{(\infty)}}(x) = \sum_{i=1}^{\infty} I_{\{u_{i-1} \leq x < u_i\}}(x)\overline{F^{(i)}}(x)$$

or equivalently

$$\overline{F^{(\infty)}}(x) = \lim_{k\to\infty} \overline{F^{(k)}}(x)$$

where $F^{(k)}$ is the conditional layer mixture of the first k distributions as in Definition 2.3.

Note that the conditional layer mixture of finite number of distributions, $F^{(k)}$, is a special case of the infinite mixture, since if we choose $F_k = F_i$ and $u_k = u_i = \infty$ for $i \geq k$, $F^{(\infty)}$ reduces to $F^{(k)}$. The above definition simply implies that the restriction of $F^{(\infty)}$ on each layer $[u_{i-1}, u_i)$ is equal to the conditional layer mixture distribution $F^{(i)}$ for each i, i.e.

$$\overline{F^{(\infty)}}(x)\Big|_{[u_{i-1},u_i)} = \overline{F^{(i)}}(x)$$

which is well-defined since $\overline{F^{(j)}}(x) = \overline{F^{(i)}}(x)$ for $x \in (u_{i-1}, u_i]$ for any $i \leq j$. Suppose that each distribution F_i has its density function f_i and the hazard function h_i and denote the mixture of the hazard functions by $h^{(\infty)}$ such that

$$h^{(\infty)}(x) = \sum_{i=1}^{\infty} I_{\{u_{i-1}\leq x<u_i\}}(x) h_i(x), \tag{2.15}$$

or equivalently, $h^{(\infty)}(x) = h_i(x)$ for $x \in [u_{i-1}, u_i)$. Then the following corollary is the generalization of Theorem 2.15 allowing $k = \infty$.

Corollary 2.17 Given distributions $\{F_i\}_{i=1,2,\ldots}$ and thresholds $0 = u_0 < u_1 < \ldots$, the infinite conditional layer mixture distribution $F^{(\infty)}$ has the mixture hazard function $h^{(\infty)}$, i.e.

$$\overline{F^{(\infty)}}(x) = \exp\big(-\int_0^x h^{(\infty)}(s)\, ds\big)$$

where $F^{(\infty)}$ and $h^{(\infty)}$ are defined in Definition 2.16 and (2.15) respectively.

Proof: Let $g_N(x) = \sum_{i=1}^{N} I_{\{u_{i-1} \le x < u_i\}}(x) h_i(x)$ for $x < u_i$ and 0 otherwise. Then $g_N \uparrow h^{(\infty)}$ as $N \to \infty$ and hence we have

$$\begin{aligned}
\overline{F^{(\infty)}}(x) &= \lim_{N\to\infty} F^{(N)}(x) \\
&= \lim_{N\to\infty} \exp\big(-\int_0^x h^{(N)}(s)\,ds\big) \quad \text{by Theorem 2.15} \\
&= \exp\big(-\int_0^x \lim_{N\to\infty} h^{(N)}(s)\,ds\big) \quad \text{by monotone convergence theorem} \\
&= \exp\big(-\int_0^x h^{(\infty)}(s)\,ds\big)
\end{aligned}$$

which completes the proof. □

Note that the mixture hazard function in (2.15) takes a form of simple function whereas the mixture distribution takes a recursive form as in Definition 2.16. Therefore, in practice, it is much easier to build a mixture distribution from a mixture hazard function. Moreover, if we model a heavy-tailed distribution with decreasing hazard rate, we can start from appropriate hazard functions $\{h_i\}$ satisfying $h_i(x) \ge h_j(x)$ for $i < j$ and derive the mixture distribution $F^{(\infty)}$ from the mixture hazard function $h^{(\infty)}$. For example, suppose that we want to find a infinite mixture of Pareto distributions each of which has different shape parameter $\alpha_i > 0$ such that

$$\overline{F}_i(x) = (1+x)^{-\alpha_i}.$$

Suppose $\alpha_i \downarrow 0$ as $i \to \infty$, then the mixture distribution $F^{(\infty)}(x)$ can be calculated easily at each $x = u_k$ as follows; for any integer $k > 0$,

$$\begin{aligned}
\overline{F}^{(\infty)}(u_k) &= \exp\big(-\int_0^{u_k} h^{(\infty)}(x)\big) \\
&= \exp\big(-\sum_{i=1}^{k}\int_{u_{i-1}}^{u_i} h_i(x)\big) \\
&= \exp\big(-\sum_{i=1}^{k}\alpha_i \log(\frac{1+u_i}{1+u_{i-1}})\big) \\
&= \prod_{i=1}^{k}(\frac{1+u_i}{1+u_{i-1}})^{-\alpha_i}
\end{aligned}$$

where $h_i(x) = \alpha_i/(1+x)$ for each i. Note that the mixture distribution $F^{(\infty)}$ is heavier-tailed than any Pareto distribution since $\alpha_i \downarrow 0$.

Example 2.18 Consider a continuous mixture hazard function as a polygon such that

$$h^{(\infty)}(x) = \sum_{i=1}^{\infty} I_{\{i-1\le x<i\}} h_i(x) \quad \text{where} \quad h_i(x) = -\frac{1}{i(i+1)}x + \frac{2}{1+i}$$

for each $i = 1, 2, \ldots$. The left graph in Figure 2.3 plots the mixture hazard function $h^{(\infty)}$ and $1/(1+x)$ which is the hazard function of the Pareto distribution $F(x) = 1-(1+x)^{-1}$. By Corollary 2.17, the infinite mixture distribution can be easily calculated at any $x \ge 0$. For example, for any positive integer n,

$$\begin{aligned}
\overline{F}^{(\infty)}(n) &= \exp\big(-\sum_{i=1}^{n}\int_{i-1}^{i} h_i(x)\big) \\
&= \exp\big(-\sum_{i=1}^{n}\frac{2i+1}{2i(i+1)}\big).
\end{aligned}$$

The right panel of Figure 2.3 plots the infinite mixture distribution and the Pareto distribution $F(x) = 1-(x+1)^{-1}$ in log scale. Note that $F^{(\infty)}$ is differentiable since $h^{(\infty)}$ is continuous.

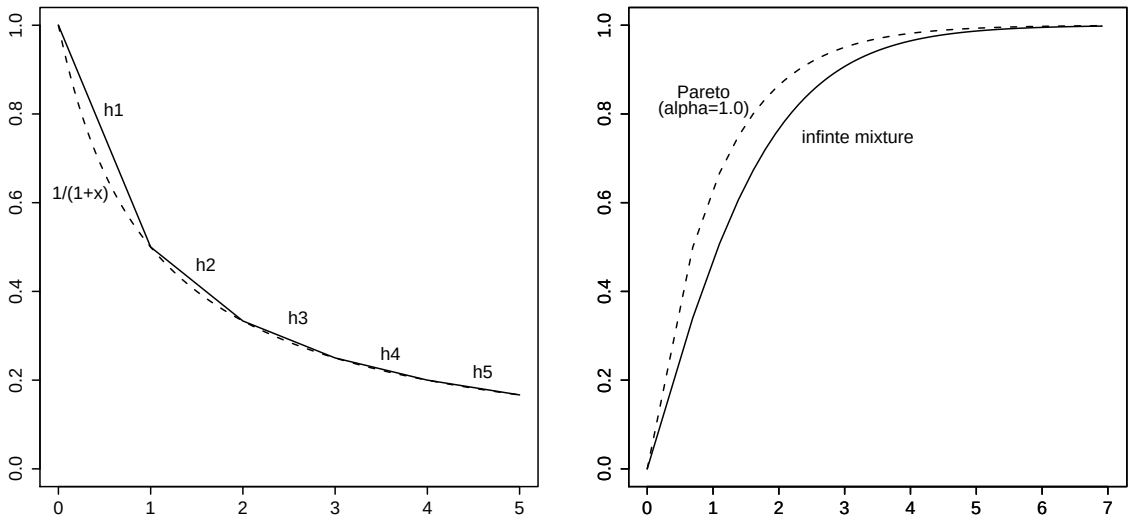


Figure 2.3 Hazard rate functions and the distribution functions of the infinite mixture(solid) and Pareto with $\alpha = 1$ (dotted).

2.3.2 Lorenz Curves of the Conditional Mixture Distributions

In economics the Lorenz curve is used to describe inequality in population's income or wealth. The Lorenz curve of distribution F with finite mean is simply defined by the ratio,

$$L(F(x)) = \frac{\int_{-\infty}^{x} t f(t) dt}{\int_{-\infty}^{\infty} t f(t) dt}$$

where f is the density function of F. In the following we examine the Lorenz curves of mixture distributions with an illustrative example. For a non-negative loss random variables X and its distribution $F^{(\infty)}$, let $\mu^{(i)} = \int_0^\infty t\, dF_i(t)$ and $\mu^{(\infty)} = \int_0^\infty t\, dF^{(\infty)}(t)$ where each F_i has its hazard function h_i. For $x \in [u_{k-1}, u_k)$ for some positive integer k, write $v_i = u_i$ for $i < k$ and $v_k = x$. Then we have

$$\begin{aligned}
L(F^{(\infty)}(x)) &= \frac{\int_0^x t\, dF^{(\infty)}(t)}{\mu^{(\infty)}} \\
&= -\sum_{i=1}^{k} \frac{1}{\mu^{(\infty)}} \frac{\overline{F}^{(\infty)}(v_{i-1})}{\overline{F}_i(v_{i-1})} \int_{v_{i-1}}^{v_i} t\, d\overline{F}_i(t) \\
&= \sum_{i=1}^{k} \frac{\mu^{(i)}}{\mu^{(\infty)}} \frac{\overline{F}^{(\infty)}(v_{i-1})}{\overline{F}_i(v_{i-1})} \big[L(F_i(v_i)) - L(F_i(v_{i-1}))\big] \\
&= \sum_{i=1}^{k} c_i\, \Delta L_i
\end{aligned}$$

where $c_i = \mu^{(i)}\overline{F}^{(\infty)}(u_{i-1})/\mu^{(\infty)}\overline{F}_i(u_{i-1})$ and $\Delta L_i = L(F_i(u_i)) - L(F_i(u_{i-1}))$ for $i < k$ and $\Delta L_k = L(F_i(x)) - L(F_i(u_{k-1}))$. Therefore the Lorenz curve L of the mixture distribution $F^{(\infty)}$ can be viewed as a mixture of the Lorenz curves L_i of distributions F_i.

Example 2.19 Consider a sequence of thresholds and Pareto distribution functions such as

$$u_{i-1} = i - 1 \quad \text{and} \quad \overline{F}_i(x) = (1+x)^{-2-\frac{1}{i}}$$

for $i = 1, 2, \ldots$. The Pareto distributions above have decreasing shape parameters from 3 to 2 as i increases, and hence the mixture distribution $F^{(\infty)}(x)$ has increasing heavy-tailedness as x becomes large. By direct calculation we have

$$\begin{aligned}
\mu^{(i)} &= \frac{i}{i+1}, \\
\mu^{(\infty)} &= \sum_{i=1}^{\infty} \int_{i-1}^{i} \overline{F}^{(\infty)}(x)\, dx \\
&= \sum_{i=1}^{\infty} \{ \frac{\overline{F}^{(\infty)}(i-1)}{\overline{F}_i(i-1)} \int_{i-1}^{i} \overline{F}_i(x)\, dx \} \\
&= \sum_{i=1}^{\infty} \Big[\overline{F}^{(\infty)}(i-1)\, i^{2+\frac{1}{i}} \frac{i}{i+1} \left\{ i^{-\frac{1+i}{i}} - (1+i)^{-\frac{1+i}{i}} \right\} \Big] \\
L(F(u_i)) &= 1 - (1+i)^{-\frac{1+i}{i}}.
\end{aligned}$$

From the paragraph followed by Example 2.18 in the previous section with $\alpha_i = 2 + 1/i$,

$$\overline{F}^{(\infty)}(i) = \prod_{j=1}^{i} \big(\frac{1+j}{j}\big)^{-(2+\frac{1}{j})}.$$

Combining all of the results above we have, for a positive integer $n > 1$,

$$L(F^{(\infty)}(n)) = \frac{1}{\mu^{(\infty)}} \sum_{i=1}^{n} \{ \prod_{j=1}^{i-1} \big(\frac{1+j}{j}\big)^{-(2+\frac{1}{j})}\, i^{2+\frac{1}{i}} \frac{i}{i+1} \big[i^{-\frac{1+i}{i}} - (1+i)^{-\frac{1+i}{i}} \big] \}.$$

Figure 2.4 is an illustration of the Lorenz curves of the mixture distribution and Pareto distribution with shape parameter $\alpha = 2, 2.77$, and 3.

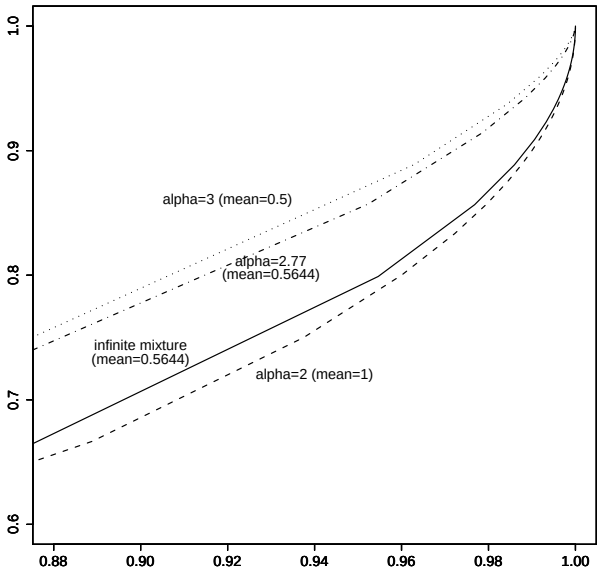


Figure 2.4 Lorenz curves of Pareto distributions with mean values $0.5(\alpha = 3)$, $1.0(\alpha = 2)$, $0.5644(\alpha = 2.77)$, and the mixture distribution in Example 2.19

.

2.3.3 Limiting Distribution of Infinite Mixture

We already observed in the previous section that the conditional layer mixture distribution is superior to the linear or layer mixture model with respect to the consistency of heavy-tailedness. However, the mixture distributions $F^{(k)}$ and $F^{(\infty)}$ are not differentiable at each threshold u_i in general whereas smoothness of loss distributions is often required in risk modelling. It is not difficult to see that this non-differentiability problem is due to the jumps of the mixture hazard functions $h^{(k)}$ or $h^{(\infty)}$ at each threshold. Therefore, if we can approximate a continuous hazard function h by the mixture hazard function $h^{(\infty)}$, the distribution F induced from h can be viewed as a limiting distribution induced from $h^{(\infty)}$.

Theorem 2.20 Consider a sequence of thresholds $u_0 < u_1 < \ldots$ such that $\delta = u_i - u_{i-1} > 0$ for all i and a sequence of distributional components $\{F_i(x;\delta)\}$ with hazard rate functions $\{h_i(x;\delta)\}$. Suppose $h(x) = \lim_{\delta \to 0} h^{(\infty)}(x;\delta)$ exists and denote the distribution induced from $h(x)$ by $F(x) =$

$1 - \exp(-\int_0^x h(s)\,ds)$. If $\overline{F}(x) > 0$ for $x > 0$, the hazard function of the limiting mixture distribution is the limit of the mixture hazard function as $\delta \to 0$, i.e.

$$\lim_{\delta\to 0} \overline{F^{(\infty)}}(x;\delta) = \exp\big(-\int_0^x \lim_{\delta\to 0} h^{(\infty)}(s;\delta)\,ds\big).$$

Moreover, if the limiting hazard function h is continuous, the limiting distribution F is differentiable.

Proof: Since $\overline{F}(x) > 0$ for $x > 0$, $\int_0^x h(s)\,ds < \infty$ and hence there exists $\epsilon > 0$ such that $h^{(\infty)}(s;\delta) < h(s) + \epsilon$ for $0 < s < x$. Then

$$\begin{aligned}\lim_{\delta\to 0} \overline{F^{(\infty)}}(x;\delta) &= \lim_{\delta\to 0} \exp\big(-\int_0^x h^{(\infty)}(s;\delta)\,ds\big) \\ &= \exp\big(-\int_0^x \lim_{\delta\to 0} h^{(\infty)}(s;\delta)\,ds\big)\end{aligned}$$

by dominated convergence theorem, which proves the first assertion. Moreover, if the limiting hazard rate function h is continuous,

$$\begin{aligned}\frac{d}{dx}\overline{F}(x) &= \lim_{c\to 0} \frac{\exp(-\int_0^{x+c} h(s)\,ds) - \exp(\int_0^x h(s)\,ds)}{c} \\ &= -h(x)\,\exp\big(-\int_0^x h(s)\,ds\big)\end{aligned}$$

which completes the proof. □

The following corollary is an immediate result from Theorem 2.20, which provides us with a theoretical background to approximate any smooth distribution by a conditional mixture of *exponential* distributions. It is very interesting to note that the exponential distributions can be seen as a set of basis which reconstructs any differetiable distribution as a conditional mixture.

Corollary 2.21 Any distribution F with a continuous hazard rate function h is a limiting distribution of the infinite conditional layer mixture of exponential distributions, i.e.

$$\overline{F}(x) = \lim_{\delta \to 0} \overline{F^{(\infty)}}(x; \delta)$$

where $\overline{F}_i(x) = \exp(-h(u_i)x)$ for a sequence of thresholds $u_0 < u_1 < \cdots$ and $\delta = u_{i+1} - u_i$.

Proof: By the continuity, the hazard rate function h can be approximated by a step function s as follows.

$$\lim_{\delta \to 0} s(x; \delta) = h(x) \quad \text{where} \quad s(x)\big|_{[u_i, u_{i+1})} = h(u_i)$$

for each i. Therefore, s is an infinite mixture of the hazard functions h_i, each of which is a constant function such that $h_i(x) = h(u_i)$. Since the constant hazard rate functions, h_i, generate exponential distribution functions, we have

$$\overline{F}_i(x) = e^{-h(u_i)x}$$

for each i. By Theorem 2.20, F is a limiting distribution of the infinite mixture of exponential distributions which completes the proof. □

In what follows, we provide two numerical exmaples. The first example is an empirical distribution fit. Without assuming any parametric distribution, we can approximate the empirical distribution as presice as we want by a mixture of exponential distributions, which is a parametric distribution. This method solves problems such as uncertainties of threshold and parametric distribution selection. We no longer need to assume a parametric form or a certain threshold level to fit emprical data. The second example illustrates a reconstruction of parametric distribution which doesn't have an explicit form where we use the g-and-h distribution.

Example 2.22 (Danish Fire Loss Data) The data used in this example is the *Danish fire loss data* with losses over one million Danish Krone. We approximate the density function f and the survival function $\overline{F}$ on $[u_{i-1}, u_i)$ as follows;

$$f(x) \approx \frac{d_i}{(u_i - u_{i-1})\, n} \quad \text{and} \quad \overline{F}(x) \approx \frac{r_{i-1} - \frac{1}{2}d_i}{n} \quad \text{for} \quad x \in [u_{i-1}, u_i).$$

where d_i and r_{i-1} are the number of losses in u_{i-1}, u_i and u_{i-1}, ∞, respectively. Therefore the empirical hazard rate function can be written as

$$\hat{h}(x) = \frac{d_i}{(u_i - u_{i-1})(r_{i-1} - \frac{1}{2}d_i)}, \quad x \in [u_{i-1}, u_i)$$

for each positive integer i, which leads us to identify the components of the mixture hazard function by $\hat{h}$, i.e.

$$h_i(x) = \hat{h}(u_{i-1}) = \frac{d_i}{(u_i - u_{i-1})(r_{i-1} - \frac{1}{2}d_i)}, \quad x \in [u_{i-1}, u_i) \tag{2.16}$$

and the Danish fire loss data can be fitted by the infinite conditional mixture of exponential distributions with mean values $\lambda_i = \hat{h}(u_{i-1})$ by Corollary 2.21. Figure 2.5 plots the empirical hazard rate function in the left panel and the infinite conditional layer mixture distribution fit to the empirical distribution in the right panel. □

Example 2.23 (g-and-h simulation) Dutta and Perry (2004) proposed the parametric g-and-h distributions as a statistical tool to measure operational risk at the enterprise level as well as at the Basel II business line and event type levels (Moscadelli, 2004) based on Loss Distribution Approach (LDA). They found that, with respect to the capital estimates at the enterprise level, the g-and-h distribution resulted in realistic and consistent capital estimates across all of the institutions they considered. The g-and-h distribution with four parameters (a, b, g, h) can be defined as a transformation of the standard normal random variable Z such that

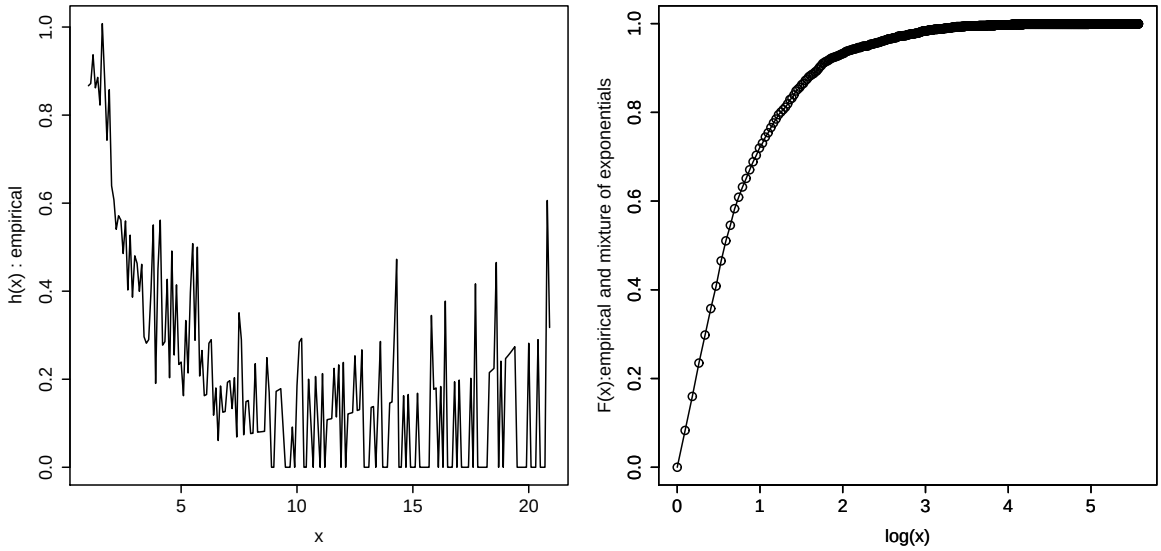


Figure 2.5 Empirical hazard rate function(left) and the mixture distribution fit to the empirical distribution(right): empirical(circled), infinite conditional layer mixture(solid line)

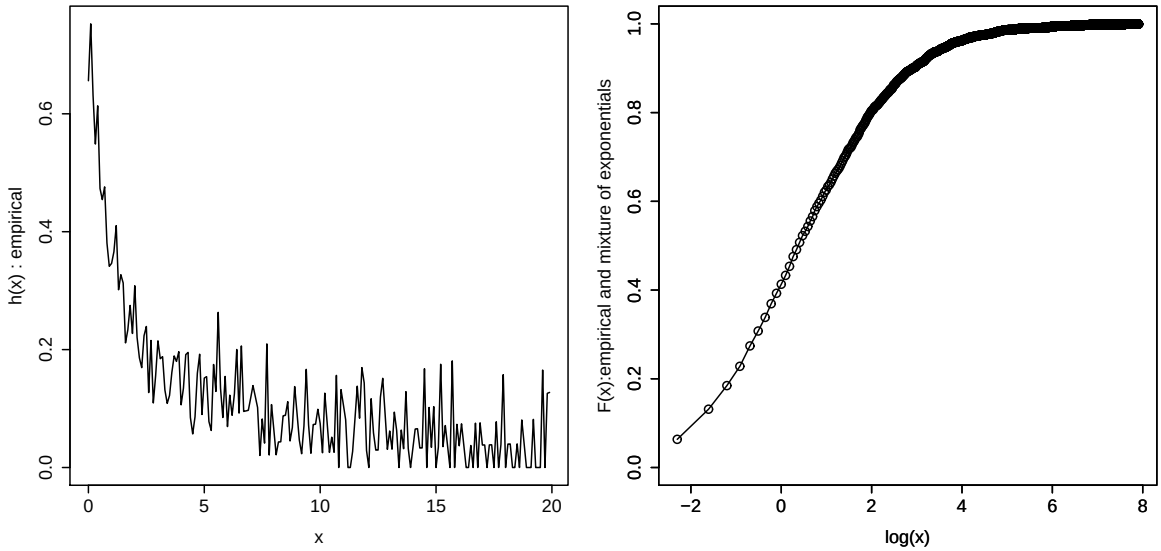


Figure 2.6 Hazard rate function simulation(left) and the mixture distribution fit to the g-and-h distribution(right): simulated(circled), infinite conditional layer mixture(solid line)

$$F(x) = Pr(X \leq x), \quad X = a + b\,\frac{e^{gZ} - 1}{g} e^{\frac{hZ^2}{2}}$$

where g and h can be real valued functions of Z^2. The advantage of using the g-and-h distribution for loss modelling lies on its flexibility. Operational losses are known to vary from low frequency and high severity to high frequency and low severity. More often than not, one or two parameter loss distributions fail to fit such operational risk data because of the short ranges of skewness and kurtosis. In this sense the g-and-h distribution is very useful because the g-and-h distributions cover a wide range of skewness-kurtosis as illustrated in Figure 3 of Dutta and Perry (2004). We simulated $5,000$ samples from positive g-and-h distributed random variable $X|X \geq 0$ with $a = 0, b = 1, g = 2.0, h = 0.2$. The values of g and h are in the empirical range proposed by Dutta and Perry (2004). We calculate the empirical hazard rate function as in (2.16) and deduce the infinite conditional mixture of exponential distributions in the same manner. Figure 2.6 plots the empirical hazard rate function in the left panel and the infinite conditional layer mixture distribution fit to the empirical distribution in the right panel. □

2.4 Applications on the g-and-h Distributions

In this section we provide practical applications of the conditional mixture distributions on the g-and-h distributions. We fit three parametric distributions to Danish fire loss data by the maximum likelihood estimation. In order to assess goodness-of-fit, we use graphical method and statistical goodness-of-fit test. We introduce the family of g-and-h distributions and its properties and examine the maximum domain of attractions of the mixture of the g-and-h distributions.

Dutta and Perry (2004) proposed the parametric g-and-h distributions as a statistical tool to measure operational risk at the enterprise level as well as at the Basel II business line and event type levels (Moscadelli, 2004) based on Loss Distribution Approach (LDA). They found that, with respect to the capital estimates at the enterprise level, the g-and-h distribution resulted in realistic and consistent capital estimates across all of the institutions they considered. The g-and-h distribution with four parameters (a, b, g, h) can be defined as a transformation of the standard normal

random variable Z such that

$$F(x) = Pr(X \leq x), \quad X = a + b\,\frac{e^{gZ} - 1}{g} e^{\frac{hZ^2}{2}}$$

where g and h can be real valued functions of Z^2. The advantage of using the g-and-h distribution for loss modelling lies on its flexibility. Operational losses are known to vary from low frequency and high severity to high frequency and low severity. More often than not, one or two parameter loss distributions fail to fit such operational risk data because of the short ranges of skewness and kurtosis. In this sense the g-and-h distribution is very useful because the g-and-h distributions cover a wide range of skewness-kurtosis as illustrated in Fig. 3 of Dutta and Perry (2004).

Degen et al. (2006) also considered the g-and-h distributions and showed that they belong to the family of regularly varying distributions or subexponential distributions when $a = 0, b = 1, g < 0$, and $h \geq 0$ i.e.

Theorem 2.24 (Degen et al. (2006)) Suppose F is a g-and-h distribution with $g, h > 0$, then $\overline{F} = x^{-\frac{1}{h}} L(x)$ for some slowly varying function $L(x)$. For $h = 0$ and $g > 0$, we have $F \in \mathcal{S} \backslash \mathcal{R}$ where $\mathcal{S}$ and $\mathcal{R}$ denote the class of subexponential distributions and regularly varying distributions, respectively.

We focus on the maximum domain of attractions of the conditional mixture distribution $F^{(k)}$ with k-th g-and-h distribution and the thresholds $0 = u_0 < u_1 < \cdots < u_{k-1}$ when $a = 0, b = 1, g \geq 0$, and $h \geq 0$. The k-th distributional component F_k above the highest threshold u_{k-1} corresponds to one of the types

$$X = \begin{cases} \frac{e^{gZ}-1}{g} e^{\frac{hZ^2}{2}}, & g, h > 0 \\ \frac{e^{gZ}-1}{g}, & g > 0, h = 0 \\ Z e^{\frac{hZ^2}{2}}, & g = 0, h > 0. \end{cases}$$

where the random variable X can be viewed as the standard normal distribution Z when $g = 0$ and $h = 0$.

Suppose $h > 0$ and consider the conditional mixture $F^{(k)}$ of k underlying distributions $\{F_i\}_{i=1,\ldots,k}$ where F_k is a g-and-h distribution and $\overline{F}_k$ dominates all of $\overline{F}_i$ for $i = 1, \ldots, k-1$. We can write $\overline{F}_k \sim x^{-1/h} L(x)$ by Theorem 2.24 and it is immediate to see that it belongs to $MDA(\Phi_{1/h})$. Therefore, $F^{(k)}$ belongs to the same $MDA(\Phi_{1/h})$ by Proposition 2.13 and the normalizing constants can be found as in Example 2.14. If $F^{(k)}$ is of the linear mixture type as in (2.14) and as long as the g-and-h distribution F_k tail-dominates all F_i for $i < k$, we choose $F^* = F_k$ and conclude $F^{(k)} \in MDA(\Phi_{1/h})$ by Theorem 2.7 and Theorem 2.8.

When $h = 0$, the g-and-h distribution becomes a scaled Lognormal distribution (or g-distribution), i.e. $X = e^{gZ-1}/g$, and

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(Z \leq \frac{\log(gx)+1}{g}) = \Phi\Big(\frac{\log x - (\log \frac{1}{g} - 1)}{g}\Big)$$

which is a lognormal distribution with $\mu = \log \frac{1}{g} - 1$ and $\sigma = g$. Although Lognormal distributions are not regularly varying, we already showed that they are long-tailed in Example 2.11. Therefore, the conditional mixture distribution $F^{(k)}$ belongs to the same maximum domain of attraction of the lognormal distribution, $MDA(\Lambda)$ by Theorem 2.13. If $F^{(k)}$ is of the linear mixture type as in (2.14) and tail-dominates F_i for all $i = 1, \ldots, k-1$, we choose $F^* = F_k$ and conclude $F^{(k)} \in MDA(\Lambda)$, since

$$\begin{aligned}
\gamma &= \lim_{x\to\infty} \frac{\overline{F}_k(x)}{\overline{F^*}(x)} \\
&= \overline{F}_{k-1}(u_{k-1}) \lim_{x\to\infty} \frac{\overline{F}_k(x-u_{k-1})}{\overline{F}_k(x)} \\
&= \overline{F}_{k-1}(u_{k-1}) > 0.
\end{aligned}$$

2.4.1 Data

The data used in this section is *Danish fire loss data* with losses over one million Danish Krone. It has been used in numerous papers for the threshold estimations or extreme value distribution estimations such as the shape ξ and scale β parameter estimations of the generalized Pareto

distribution,

$$\overline{F}(x) = \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}}.$$

for $1+\xi/\beta > 0$. The median and mean are 1.778 and 3.385 million respectively, and the number of claims over 20 million and 10 million Danish Krone are respectively 36 and 109, which are 1.66% and 5.03% of the sample observations. The exponential quantile plot shows an upward trend above 45° straight line, which is a strong evidence of the heavy tailed distribution. In Appendix B we provide a maximum likelihood estimator of the shape parameter of the extreme value distribution with a numerical example.

2.4.2 Selected Parametric Models and Estimation Methods

We consider three parametric distributions, the generalized Pareto distribution, the g-and-h distribution, and a mixture of two g-and-h distributions with a threshold $u > 0$ for $g, h \geq 0$ excluding the trivial case $g = h = 0$. The closed form of the distribution function or the transformations of the random variables can be found in the following Table 2.1.

For a given sample $x_1, \ldots, x_n$, the likelihood function of the generalized Pareto distribution is written as a product of the density functions at each x_i for $i = 1, \ldots, n$,

Generalized Pareto	$F(x) = 1 - (1 + \frac{\xi}{\beta}x)^{-\frac{1}{\xi}}, \quad x \geq 0$	$\xi > 0, \beta > 0$
Single g-and-h	$F \sim \frac{e^{gZ}-1}{g} e^{hZ^2/2}$	$g, h \geq 0,$
	$Z \sim \mathcal{N}(0,1)$	$(g,h) \neq (0,0)$
g-and-h mixture	$\overline{F}^{(2)}(x) = I_{\{x<u\}}(x)\overline{H}_1(x)$	$g_i, h_i \geq 0,$
	$+I_{\{x\geq u\}}(x)\overline{H}_1(u)\bar{H}_2(x\|x>u)$	$(g_i,h_i) \neq (0,0)$
	$H_i \sim \frac{e^{g_i Z}-1}{g_i} e^{h_i Z^2/2}, \quad Z \sim \mathcal{N}(0,1)$	$i = 1,2$

Table 2.1 Selected parametric distributions

$$l(x_1,\ldots,x_n|\xi,\beta)=\prod_{i=1}^{n}\frac{1}{\beta}(1+\frac{\xi}{\beta}x_i)^{-1-\frac{1}{\xi}}$$

We denote the maximum likelihood estimates of ξ and β by $\hat{\xi}$ and $\hat{\beta}$, respectively.

In order to calculate the likelihood function of the g-and-h distribution F for a given sample $\{x_1,\ldots,x_n\}$, we need to compute $f(x_i|\Theta)$ for each x_i and given parameter set $\Theta=(g,h)$. Let us denote the generalized inverse of F by Q, i.e. $Q(\cdot)=F^{-1}(\cdot)$. Then $Q(F(x))=x$ and $Q^{'}(F(x))f(x)=1$ by differentiation, which reduces to

$$f(x|g,h)=\frac{1}{Q'(F(x))}. \tag{2.17}$$

Define a real valued function $k(x)$ by

$$k(x)=\begin{cases}\frac{e^{gx}-1}{g}e^{\frac{hx^2}{2}}, & g,h>0\\ \frac{e^{gx}-1}{g}, & g>0,h=0\\ xe^{\frac{hx^2}{2}}, & g=0,h>0.\end{cases}$$

Since k is strictly increasing, the distribution F_X can be written as

$$F_X(x)=\Phi(k^{-1}(x)),$$

where Φ is the standard normal distribution. The quantile function Q can be also written as a composite function of k and the standard normal distribution as follows,

$$F^{-1}(y)=k(\Phi^{-1}(y)).$$

For more details see Remark 2.1, Degen et al. (2006). Moreover, the derivative of Q can be written

as

$$\begin{aligned} Q^{'}(y) &= \frac{d}{dy}F^{-1}(y) = \frac{d}{dy}k(\Phi^{-1}(y)) = k^{'}(\Phi^{-1}(y))\frac{d}{dy}\Phi^{-1}(y) \\ &= \frac{k^{'}(\Phi^{-1}(y))}{\phi(\Phi^{-1}(y))}. \end{aligned} \tag{2.18}$$

Combining (2.17) and (2.18), we obtain

$$f(x_i|g,h) = \frac{\phi(\Phi^{-1}(F(x_i)))}{k'(\Phi^{-1}(F(x_i)))}.$$

Therefore, if we have a method of evaluating $F(x_i|g,h)$ for each x_i, the likelihood function $l(x_1,\ldots,x_n|g,h)$ can be calculated as a product of the above expressions. However, since there is no closed form of $k^{-1}(x)$ in general, $F_X(x_i)$ should be computed numerically for each $x_i > 0$ as follows. For a given $x_i > 0$, let y_i be the solution of the equation $k(y_i) = x_i$. The value of y_i can be found by many root-finding algorithms in numerical analysis and the uniqueness of the solution is guaranteed since $k(x)$ is strictly increasing for $g, h \geq 0$. Then $F(x_i)$ can be approximated by the numerical solution $\hat{y}_i$ and hence we can also approximate the likelihood function of the g-and-h distribution as follows.

$$\begin{aligned} k(\hat{y}_i) &\approx x_i \\ F(x_i) &\approx \Phi(\hat{y}_i) \\ l(x_1,\ldots,x_n|g,h) &\approx \prod_{i=1}^{n} \frac{\phi(\Phi^{-1}(\Phi(\hat{y}_i)))}{k'(\Phi^{-1}(\Phi(\hat{y}_i)))} = \prod_{i=1}^{n} \frac{\phi(\hat{y}_i)}{k'(\hat{y}_i)} \end{aligned}$$

2.4.3 Results

Generalized Pareto Distribution: The negative log likelihood function of the generalized Pareto distribution attains its minimum at $(\hat{\xi}, \hat{\beta}) = (0.61, 0.32)$. Since the extreme value distributions have been widely used for modelling loss distributions in numerous papers, we don't discuss the details in this section. See McNeil (1997) and Cebrian et al. (2003).

The g-and-h Distribution: Figure 2.7 plots the negative log likelihood function of the g-and-h distribution in the neighborhood of the origin $(0, 0)$. We used Monte Carlo method to locate the minimum of the negative log likelihood in this example and the numerical estimates are $(\hat{g}, \hat{h}) = (0, 0.38)$ which implies that the fitted distribution is an h-distribution which is subexponential such that

$$X \sim Ze^{\frac{\hat{h}Z^2}{2}}.$$

Since h parameter in the g-and-h distribution is responsible for the kurtosis, h-distribution can be used to model the heavy-tailedness of the sample. However, it may fail to explain the skewness of the loss distribution due to the lack of g parameter.

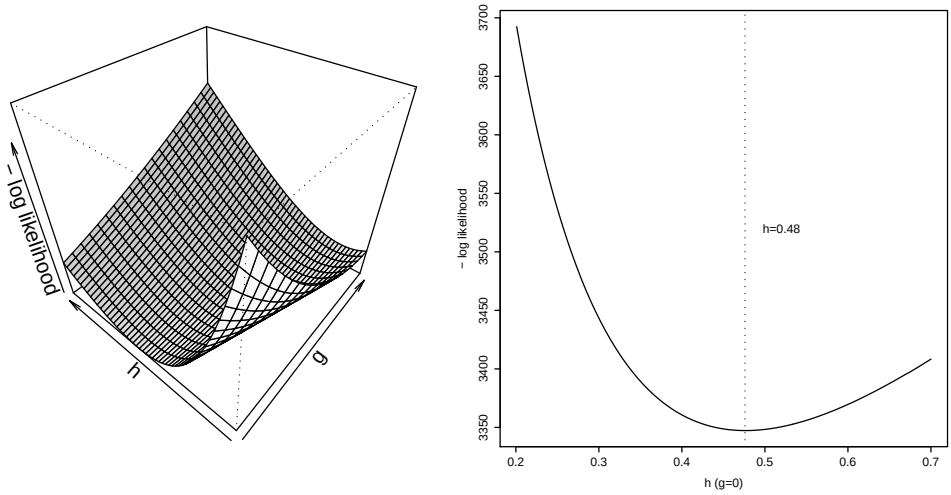


Figure 2.7 Negative log likelihood of the single g-and-h distribution

The Mixture of two g-and-h Distributions: For the conditional mixture of two g-and-h distributions F_1 and F_2 we first introduce the threshold $u > 0$, where the loss random variable satisfy $X|X < u \sim F_1$ and $X|X \geq u \sim F_2$. Since the threshold choice is critical to the parameter estimation of the extreme value distributions, we chose three different thresholds $u = 3, 6.234$, and 10 million where $u = 6.234$ million is the threshold estimate calculated by Bayesian method in Appendix B. The number of losses exceeding each threshold is 532, 177, and 107 respectively.

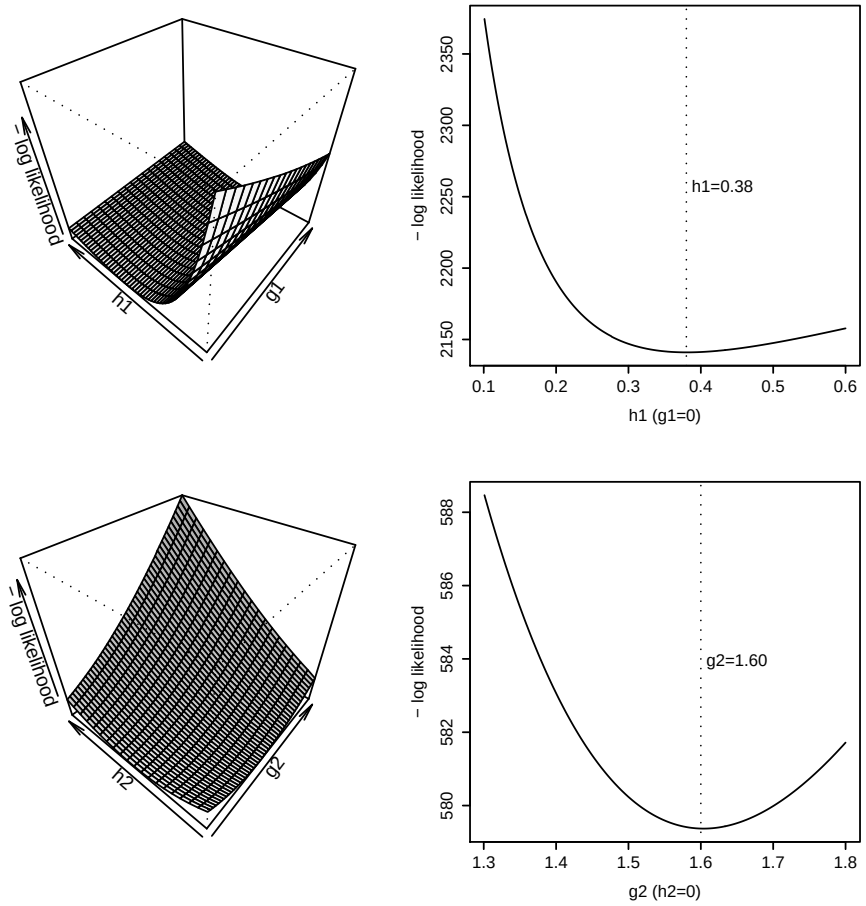


Figure 2.8 Negative log likelihood of the mixture of two g-and-h distributions when $u = 6.21$.

For $u = 3$ million (532 exceedances) the negative log likelihood function attains its minimum at $(\hat{g_1}, \hat{h_1}) = (0.26, 0.02)$ and $(\hat{g_2}, \hat{h_2}) = (0, 0.57)$ respectively. Therefore it is the mixture of the g-and-h and h-distribution such as

$$X_{\{x<u\}} \sim \frac{e^{\hat{g}_1 Z} - 1}{\hat{g}_1} e^{\frac{\hat{h}_1 Z^2}{2}} \quad \text{and} \quad X_{\{x \geq u\}} \sim e^{\frac{\hat{h}_2 Z^2}{2}}$$

which asserts that losses below and above the threshold are responsible for the strong skewness($\hat{g_1} = 0.26$) and the heavy-tailedness ($\hat{h_2} = 0.57$) of the sample data respectively. Figure 2.8 plots the negative log likelihood function in the neighborhood of $(\hat{\xi}, \hat{\beta})$ when $u = 3$.

For $u = 10$ million (107 exceedances), the negative log likelihood attains its minimum at $(\hat{g_1}, \hat{h_1}) = (0, 0.37)$ and $(\hat{g_2}, \hat{h_2}) = (1.04, 0.23)$ respectively. It is the mixture of the h- and the g-and-h distributions such as

$$X_{\{x<u\}} \sim e^{\frac{\hat{h}_1 Z^2}{2}} \quad \text{and} \quad X_{\{x\geq u\}} \sim \frac{e^{\hat{g}_2 Z} - 1}{\hat{g}_2} e^{\frac{\hat{h}_2 Z^2}{2}}$$

which implies that there exists strong skewness $(\hat{g_2} = 1.04)$ and relatively moderate heavy-tailedness $(\hat{h_2} = 0.23)$ in the right tail above the threshold $u = 10$.

It is interesting to observe that for $u = 6.234$ million (177 exceedances), the two underlying distributions F_1 and F_2 resulted in h-distribution and g-distribution respectively. Figure 2.8 plots the negative log likelihood function of the g-and-h distribution below and above the threshold u respectively. The mixture distribution is heavy-tailed $(\hat{h}_1 = 0.37)$ as much as the single g-and-h distribution $(\hat{h} = 0.48)$ from 1 million to 6.234 million and it is highly skewed $(\hat{g}_2 = 1.60)$ in the right tail above the threshold $u = 6.234$ million. In this particular example, with appropriate level of threshold we can fit the mixture distribution to the sample data and observe that each of the distributional components measures the skewness and the heavy-tailedness separately, which is not possible under non-mixture model.

Table 2.2 is the summary of the estimates of g and h for each threshold u. We can observe that the tail inference of the loss distribution is very sensitive to the threshold choice due to the bias-variance trade off. For example, there are few data points above the threshold $u = 10$ which results in parameter uncertainty for F_2. Also $u = 3$ is too low to apply asymptotic properties in extreme value theory. As a matter of fact the estimated $\hat{h}_2(0.23)$ when $u = 10$ reflects less heavy-tailedness than $\hat{h}_2(0.57)$ when $u = 3$. The discrepancy is due to the uncertainty of the parameters of F_2 when $u = 10$ or the lack of theoretical justification of F_2 when $u = 3$.

Goodness of fit: Two methods are applied to assess the goodness-of-fits for each parameter estimation, Kolmogorov-Smirnov test (K-S) and Quantile-Quantile plot (Q-Q). Table 2.2 is the summary of the parameter estimates, K-S statistics, and p-values. The mixture model with $u = 3$ million shows the poorest fit and fails to pass the K-S test. The GPD model and the g-and-h mixture

model with $u = 6.21$ million show the best fits with p-values 0.3571 and 0.1391, respectively. The distribution fits and the Q-Q plots of the selected parametric models are illustrated in Figure 2.9 and 2.10.

Distributions	Parameters	Estimates	K-S Statistics	p-values
single Generalized Pareto	(ξ, β)	$(0.61, 0.32)$	0.0281	0.3571
single g-and-h	(g, h)	$(0, 0.48)$	0.0554	0.0026
g-and-h mixture	(g_1, h_1)	$(0.26, 0.02)$	0.1527	0.0000
$(u = 3.0)$	(g_2, h_2)	$(0, 0.57)$		
g-and-h mixture	(g_1, h_1)	$(0, 0.38)$	0.0351	0.1391
$(u = 6.234)$	(g_2, h_2)	$(1.60, 0)$		
g-and-h mixture	(g_1, h_1)	$(0, 0.37)$	0.0383	0.0832
$(u = 10.0)$	(g_2, h_2)	$(1.04, 0.23)$		

Table 2.2 Maximum likelihood estimates and goodness-of-fits

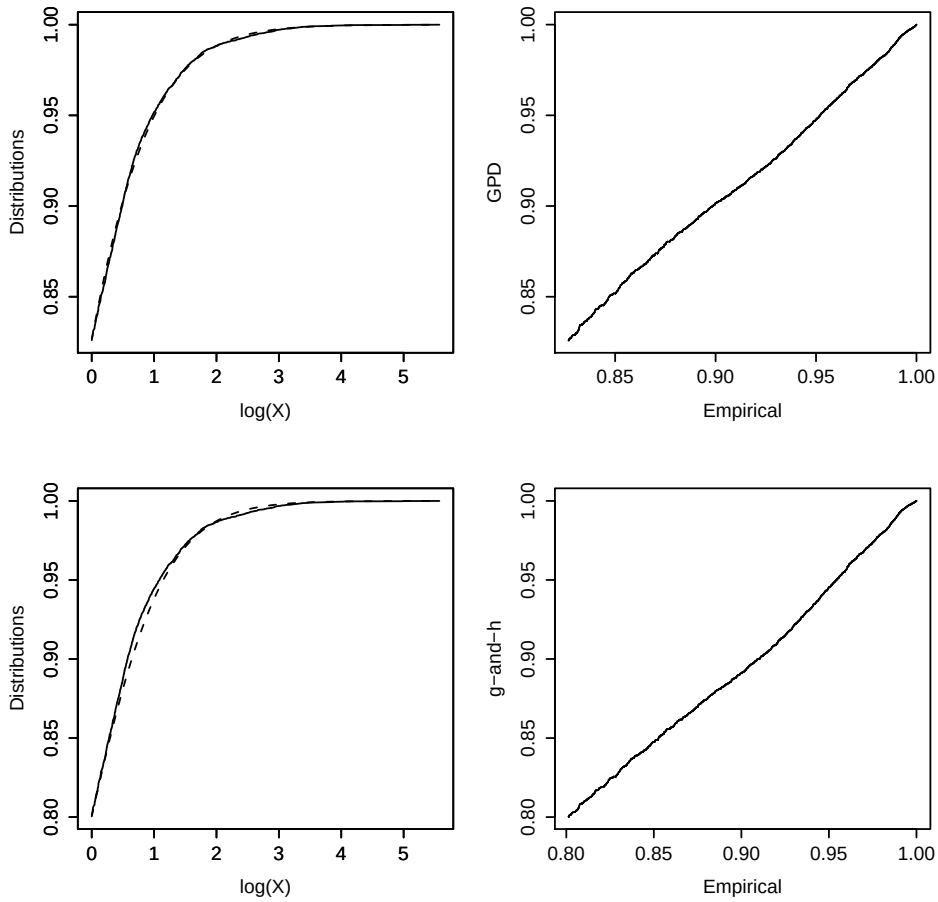


Figure 2.9 Distribution fits and Q-Q plots of the generalized Pareto distribution (Top) and the single g-and-h distribution (Bottom).

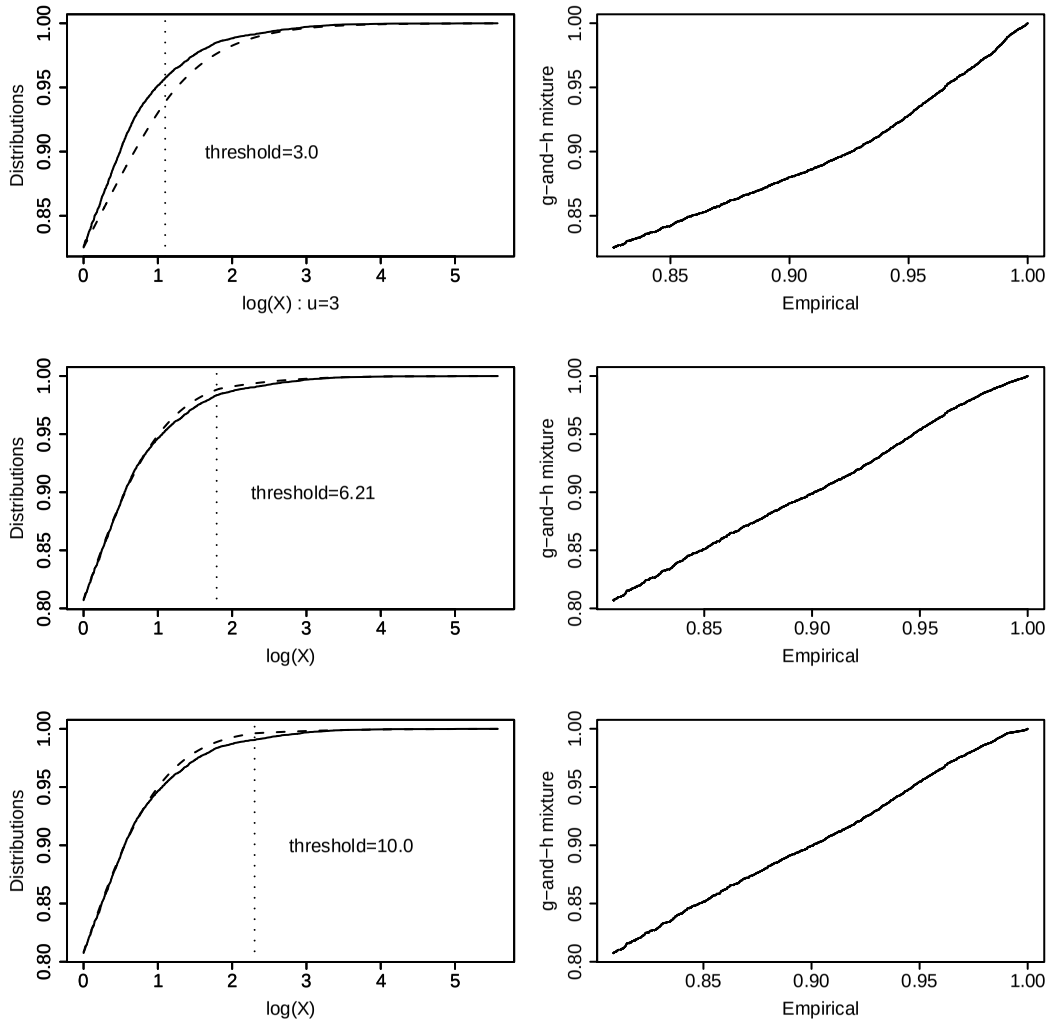


Figure 2.10 Distribution fits and Q-Q plots of the mixture of two g-and-h distributions.

2.5 Conclusions

We have demonstrated that mixture distributions naturally arise in (re)insurance risk modelling and that they have considerable advantages over other non-mixture parametric models in dealing with large claims. The specific mixture distributions we have considered in this chapter are suitable for modelling (re)insurance risks since they capture the far tails of loss distributions due to their layer structure. We have shown that the maximum domain of attraction of the mixture distributions are completely determined by the maximum domain of attraction of the last distributional component and hence the tail behavior of the mixture distribution can be fully explained by the tail behavior or the last component. We also discussed the hazard rate functions of the conditional layer mixture distributions, which is very unique among all mixture models. Since we can build a conditional layer mixture distribution by mixing hazard rate functions in a simple way, the analytic complexity of general mixture models can be overcome by the hazard rate function representation of the conditional layer mixture model. The infinite mixture model and its hazard rate function expressed as a simple mixture, gives a possible solution to the threshold selection problem by modelling the tail above any threshold level as another infinite mixture distribution. Lastly, we have proposed a limiting distribution of the infinite conditional mixture and have shown that any distribution with continuous hazard rate function can be approximated by a mixture of exponential distributions.

Chapter 3

Computation of the Aggregate Distribution of a Maximal Markov Sequence

This chapter is based on the paper by Jang and Jho (2008a). Financial institutions such as banks and insurance companies have a pool of risks. The aggregation of these risks as loss random variables has been used in insurance and operational risk modeling. Let us define a pool of risks by a set of random variables,

$$E = \{X_1, \dots X_d\}$$

for $d > 0$ and denote the sum of the random variables by S_E, where X_i are not necessarily identical. In particular, if X_i are independent, the aggregate loss distribution F_{S_E} can be calculated by the convolution of F_{X_i}, although it is a time-consuming approach. The difficulty arises when the random variables are dependent, since the convolution is an operation on functions not on random variables. An axiomatic approach to this problem can be found in Frank (1991) where the author considered the convolution for dependent random variables as a particular member of distributional counterpart of binary operations, for example, a bivariate copula and introduced the generalized convolutions. Another difficulty lies on the size of the pool of risks, d, with respect to the numerical efficiency. With no specific assumptions on the dependence structure, we propose a method of reducing the size of the risk pool to obtain an equivalent and simpler risk pool and hence faster calculation in the followings.

Consider a sequence E_1 as a subset of E such that

$$E_1 = \{X_{i_1}, \dots, X_{i_k}\} \subset E$$

for $i_1 < \cdots < i_k \leq d$. Then E_1 is called a *maximal Markov sequence* if it satisfies Markov property but $E_1 \cup \{X_i\}$ violates Markov property for any $X_i \in E \backslash E_1$. For the definition of Markov property or Markov process, see Nelsen (1999) or Joe (1997). If $E_1 \neq E$, consider another maximal Markov sequence $E_2 \subset E \backslash E_1$ and repeat this process until E_i exhaust the set E. Eventually we have a decomposition of the pool of risks by disjoint maximal Markov sequences as follows,

$$\cup_{i=1}^{j} E_i = E \tag{3.1}$$

for a positive integer $j < d$. Note that the decomposition in (3.1) may not be unique. Denote the sum of the random variables in each E_i by S_{E_i}, then we can write $S_E = \sum_{i=1}^{j} S_{E_i}$ and

$$F_{S_E}(s) = \mathbb{P}(S_{E_1} + \cdots + S_{E_j} \leq s)$$

For convenience, we rename the random variables $\{S_{E_1}, \ldots, S_{E_j}\}$ by $\{X_1^{(2)}, \ldots, X_j^{(2)}\}$ where j depends on the number of maximal Markov sequences and we will hence use in general the notation $\{X_1^{(2)}, \ldots, X_{d_2}^{(2)}\}$, which is called the second stage risk pool equivalent to the first stage and we have $\sum_{i=1}^{d} X_i = \sum_{i=1}^{d_2} X_i^{(2)}$. Since d is finite, repeating this process, we obtain k-th stage risk pool which is no longer decomposable, and hence satisfies Markov property, as in the following diagram.

$$\begin{gathered}
\{X_1, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, X_d\} \\
\{X_1^{(2)}, \ldots\ldots\ldots\ldots\ldots, X_{d_2}^{(2)}\} \\
\vdots \\
\{X_1^{(k)}, \ldots, X_{d_k}^{(k)}\}.
\end{gathered}$$

Due to this hierarchical structure, $X_i^{(j)}$ is a sum of random variables in a maximal Markov sequence at $(j-1)$-th stage risk pool. Therefore if we have a computational method for the aggregate distribution of random variables satisfying Markov property, the aggregate distribution of $X_1 + \cdots + X_d$ can be calculated by the aggregate distribution of $X_1^{(k)} + \cdots + X_{d_k}^{(k)}$ where each $X_i^{(k)}$ is again a sum of random variables in $(k-1)$ stage maximal Markov sequence and hence $F_{X_i^{(k)}}$ is an aggregate distribution of a $(k-1)$-th stage maximal Markov sequence and so on. The efficiency of the computation depends on how good a decomposition is at each stage, i.e. how large each

maximal Markov sequence is. Therefore it is very crucial to compute the aggregate distribution of each maximal Markov sequence, which is the main goal of this chapter.

In section 3.1 we present a computational method for the distribution of sum of random variables in a maximal Markov sequence and in section 3.2 we introduce a chain of pairwise bivariate copulas to implement the dependence on the maximal Markov sequence. We provide numerical examples of the method in section 3.3, and actuarial applications on Bayesian premiums and stop-loss premiums in section 3.4 and 3.5.

3.1 Computation of the Aggregate Distribution of Non-identical and Dependent Variables in a Maximal Markov Sequence

For a given pool of risks E, choose a maximal Markov sequence $E_1 = \{X_1, \ldots, X_n\} \subset E$ for a positive integer $n \leq d$. Denote the joint distribution of $X_1, \ldots, X_n$ by $H_n(x_1, \ldots, x_n)$ for each positive integer $n > 1$. As a consequence of Markov property, *Chapman-Kolmogorov equation* holds as follows.

$$\mathbb{P}(X_k \leq x_k | X_i = x_i) = \int_0^\infty \mathbb{P}(X_k \leq x_k | X_j = x_j) \frac{\partial \mathbb{P}(X_j \leq x_j | X_i = x_i)}{\partial x_j} \, dx_j$$

for $i < j < k$. If the conditional densities exist, the above equation takes the analogous form,

$$f_{X_k|X_i}(x_k|x_i) = \int_0^\infty f_{X_k|X_j}(x_k|x_j) f_{X_j|X_i}(x_j|x_i) \, dx_j.$$

Now let us denote the aggregate loss by $S_n = \sum_{i=1}^n X_i$ for a positive integer n. Assuming Markov property, we first derive an iterated integral equation of the joint density function f_{S_{n-1},X_n} for each $n > 2$ in the following lemma.

Lemma 3.1 Suppose that X_i are continuous, non-negative, non-identical, and dependent loss random variables. Further we assume $\{X_1, \ldots, X_n\}$ satisfies Markov property. If the joint density function of X_n and S_{n-1} exists for each $n > 2$, the following equation holds.

$$f_{X_n,S_{n-1}}(t,s) = \int_0^s f_{X_n|X_{n-1}}(t \mid w)\, f_{X_{n-1},S_{n-2}}(w, s-w)\, dw. \tag{3.2}$$

Proof: First consider the well-known result of conditional probability,

$$\frac{\partial}{\partial s}\mathbb{P}(X_n \le t, S_{n-1} \le s) = \frac{\partial}{\partial s}\mathbb{P}(S_{n-1} \le s) \cdot \mathbb{P}(X_n \le t \mid S_{n-1} = s).$$

Applying the *Chapman-Kolmogorov equation* on the conditional probability in the right hand side by conditioning on S_{n-2} and X_{n-1}, the above equation can be rewritten as follows.

$$\begin{aligned}\frac{\partial}{\partial s}\mathbb{P}(X_n \le t, S_{n-1} \le s) &= \int_0^s f_{S_{n-1}}(s)\, \mathbb{P}(X_n \le t \mid X_{n-1} = w, S_{n-2} = s-w) \\ &\qquad \times f_{X_{n-1},S_{n-2}\,|\,S_{n-1}}(w, s-w \mid s)\, dw \\ &= \int_0^s \mathbb{P}(X_n \le t \mid X_{n-1} = w)\, f_{X_{n-1},S_{n-2}}(w, s-w)\, dw.\end{aligned}$$

The lemma follows immediately by differentiating the last equation with respect to t. □

Note that the result (3.2) in Theorem 3.1 is of the form analogous to the convolution of X_n and S_{n-1}. The joint density $f_{X_n|S_{n-1}}$ is derived from $f_{X_{n-1},S_{n-2}}$ through the kernel integration by $f_{X_n|X_{n-1}}$. Applying the result of Lemma 3.1, we can derive another integral equation of the aggregate loss density function f_{S_n} in the following theorem. In special, if we assume the independence of losses, it is identical to the classical convolution formula.

Theorem 3.2 Suppose that X_i are continuous, non-negative, non-identical, and dependent loss random variables. Further we assume $\{X_1, \ldots, X_n\}$ satisfies Markov property. If the joint density function of X_n and S_{n-1} exists for each $n > 2$, the density function of the aggregate loss can be written as follows.

$$f_{S_n}(t) = \int_0^t \int_0^s f_{X_n|X_{n-1}}(t-s \mid w)\, f_{X_{n-1},S_{n-2}}(w, s-w)\, dw\, ds \tag{3.3}$$

where $f_{X_{n-1},S_{n-2}}(w, s-w)$ can be derived by Lemma 3.1.

Proof: By conditioning on X_{n-1} and S_{n-2} again,

$$\begin{aligned}
\mathbb{P}(S_n \le t) &= \int_0^t \mathbb{P}(X_n \le t-s \,|\, S_{n-1}=s)\, f_{S_{n-1}}(s)\; ds \\
&= \int_0^t \int_0^s \mathbb{P}(X_n \le t-s \,|\, X_{n-1}=w, S_{n-2}=s-w) \\
&\qquad \times f_{X_{n-1},S_{n-2}\,|\,S_{n-1}}(w, s-w \,|\, s)\, f_{S_{n-1}}(s)\; dw\, ds \\
&= \int_0^t \int_0^s \mathbb{P}(X_n \le t-s \,|\, X_{n-1}=w)\, f_{X_{n-1},S_{n-2}}(w, s-w)\; dw\, ds.
\end{aligned}$$

The density function follows immediately by differentiating the distribution function $\mathbb{P}(S_n \le t)$ with respect to t and applying the *Leibniz Rule*, which completes the proof. □

Lemma 3.1 and Theorem 3.2 provide a numerical algorithm to compute the aggregate density or aggregate distribution function of non-identical dependent losses in a maximal Markov sequence. In the next section we introduce a chain of bivariate copulas to implement the dependence on the maximal Markov sequence.

3.2 Application on Bivariate Copulas

We give a brief overview of bivariate copulas and implement the dependence structure on a maximal Markov sequence $\{X_1, \ldots, X_n\}$ by a chain of bivariate copuals. The formal definition of bivariate copulas can be found in many textbooks or papers and we follow Nelsen (1999).

Definition 3.3 Let $I = [0, 1]$, a unit closed interval. A bivariate copula is a function $C : I^2 \to I$ with the following properties:

1. For every u, v in I

$$C(u, 0) = 0 = C(0, v) \quad \text{and} \quad C(u, 1) = u, \quad C(1, v) = v.$$

2. For every u_1, u_2, v_1, v_2 in I such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Copula is an efficient tool of modeling dependence structure since the dependence and the marginals of random variables can be studied separately. In particular, the copula function related to any multivariate joint distribution with continuous marginal distributions is uniquely determined due to the following Sklar's theorem (Sklar, 1959).

Theorem 3.4 Let H be a joint distribution function with marginal distributions F and G. Then there exists a copula $C : [0,1]^2 \to [0,1]$ such that for all $x, y \in [-\infty, \infty]$,

$$H(x, y) = C(F(x), G(y)). \tag{3.4}$$

If F and G are continuous, then C is unique. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (3.4) is a joint distribution function with marginal distributions F and G.

Note that the copula function is not uniquely determined if the marginals are not continuous. If we define the generalized inverse of F by $F^{-1}(t) = \inf\{x : F(x) \geq t\}$, the equation (3.4) takes the following analogous form,

$$H(F^{-1}(u), G^{-1}(v)) = C(u, v).$$

Moreover, it is not difficult to show that if the joint density $h(x, y)$ of $H(x, y)$ exists,

$$h(x, y) = c\big(F(x), G(x)\big) f(x) g(x)$$

where $c(u, v)$ is called the density of the copula and defined by

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v).$$

For the random variables in a maximal Markov sequence $E_1 = \{X_1, \ldots, X_n\}$, let X_{i-1} and X_i be dependent by bivariate copulas $C^{(i)}$ with continuous partial derivatives for each i such that $H_{i-1,i}(x_{i-1}, x_i) = C^{(i)}\big(F_{i-1}(x_{i-1}), F_i(x_i)\big)$ where $H_{i-1,i}$ is the joint distribution of X_{i-1} and X_i. The joint density function of X_{i-1} and X_i can be viewed as

$$f_{i-1}(x_{i-1}) f_i(x_i)\, c^{(i)}\big(F_{i-1}(x_{i-1}), F_i(x_i)\big) \tag{3.5}$$

and hence the conditional density of $X_i|X_{i-1}$ is equal to $c^{(i)}\big(F_{i-1}(x_{i-1}), F_i(x_i)\big) f(x_i)$. From the equation (3.2) in Lemma 3.1 the joint density function of X_n and S_{n-1} can be written as follows.

$$f_{X_n, S_{n-1}}(t, s) = \int_0^s c^{(n)}\big(F_{n-1}(t), F_n(w)\big) f_{X_n}(t)\, f_{X_{n-1}, S_{n-2}}(w, s-w)\; dw. \tag{3.6}$$

In the same manner, by substituting (3.6) into (3.3) in Theorem 3.2, the density function of S_n can be written as follows,

$$f_{S_n}(t) = \int_0^t \int_0^s c^{(n)}\big(F_{n-1}(w), F_n(t-s)\big) f_{X_n}(t-s)\; f_{X_{n-1}, S_{n-2}}(w, s-w)\; dw\, ds. \tag{3.7}$$

Denote the partial derivative of the copula $C^{(i)}(u, v)$ with respect to u by $C_u^{(i)}(u, v)$, i.e.

$$C_u^{(i)}(u, v) = \frac{\partial}{\partial u} C^{(i)}(u, v).$$

From the last equation of the proof in Theorem 3.2, the aggregate distribution of S_n can be written

as

$$F_{S_n}(t) = \int_0^t \int_0^s C_u^{(n)}\big(F_{n-1}(w), F_n(t-s)\big)\ f_{X_{n-1},S_{n-2}}(w, s-w)\ dw\, ds. \tag{3.8}$$

Therefore (3.6) and (3.8) constitute a numerical algorithm for the aggregate loss distribution F_{S_n} for $n > 2$.

3.3 Numerical Examples

In this section we provide two examples of Lemma 3.1 and Theorem 3.2 with applications on Farlie-Gumbel-Morgenstern copula and Gaussian copula. Farlie-Gumbel-Morgenstern copula has simple algebraic expression and is well-defined on the entire domain, $(u, v) \in [0, 1]^2$. The example in section 3.3.2 can be applied to similar copulas such as Archimedean copulas. Gaussian copulas provide us a standard of copula modeling since it projects the dependence structure of the multivariate standard normal distribution onto any multivariate joint distribution with non-normal marginals. For more details about the construction, properties, and estimations of the copulas, see Frees and Valdez (1998), Genest and MacKay (1986), Genest and Rivest (1993), and Joe (1997). The example in section 3.3.3 can be also applied to t-copulas by simple substitutions of $C_u(u, v)$ and $c(u, v)$. Exponential and Pareto distributions are chosen for marginal distributions in order to illustrate the effect of light or heavy-tailedness on the aggregate distributions.

3.3.1 Algorithms

Given $t, s > 0$ and $n > 2$, we introduce a recursive algorithm of the numerical calculation for the joint density function, $f_{X_n,S_{n-1}}(t, s)$ associated with marginal density functions f_{X_i} and copula densities $c^{(i)}$ for $i = 1, \ldots, n$ as follows.

AL1. For the joint density function $f_{X_n,S_{n-1}}(t, s)$,

PROCEDURE(t, s, n)

IF $n = 2$ THEN RETURN $f_{X_2,X_1}(t, s)$

ELSE

RETURN $\int_0^s c^{(n)}(F_{n-1}(t), F_n(w))\cdot$ PROCEDURE$(w, s-w, n-1)\, dw$

END IF

END PROCEDURE

The aggregate density function $f_{X_n}(t)$ and the distribution function $F_{X_n}(t)$ can be calculated numerically by calling the procedure in AL1 above as follows.

AL2. For the aggregate density function $f_{X_n}(t)$,

STEP 1: $a(w; t, s, n) = c^{(n)}(F_{n-1}(w), F_n(t-s)) \cdot f_{X_n}(t-s)\cdot$ PROCEDURE$(w, s-w, n-1)$

STEP 2: $b(s; t, n) = \int_0^s a(w; t, s, n)\, dw$

STEP 3: $f_{X_n}(t) = \int_0^t b(s; t, n)\, ds$

AL3. For the aggregate distribution function $F_{X_n}(t)$,

STEP 1: $A(w; t, s, n) = C_u^{(n)}(F_{n-1}(w), F_n(t-s))\cdot$ PROCEDURE$(w, s-w, n-1)$

STEP 2: $B(s; t, n) = \int_0^s A(w; t, s, n)\, dw$

STEP 3: $F_{X_n}(t) = \int_0^t B(s; t, n)\, ds$

For the the univariate integrals in AL1, AL2, and AL3, any numerical integration method can be applied, for example Newton-Cotes formulas and Gaussian quadrature. In this chapter we do not discuss about the efficiency of the numerical approximation since main topic of this chapter is focused on the dependence structure of multiple random variables implemented by copulas and an introduction to a numerical method of the aggregate distribution function. For the following numerical examples we apply the trapezoidal rule, a simple brute-force calculation for convenience.

AL0. Trapezoidal rule for $\int_a^b f(x)\,dx$

CHOOSE AN INTEGER $k > 0$.
SET ACCUMULATOR $= 0$ AND $i = 0$
WHILE $i < n$ DO
    ACCUMULATOR = ACCUMULATOR + $\frac{b-a}{2k}\{f(a + i\frac{b-a}{k}) + f(a + (i+1)\frac{b-a}{k})\}$
    $i = i + 1$
END WHILE
RETURN ACCUMULATOR

3.3.2 Farlie-Gumbel-Morgenstern Copula and Exponential Distributions

Consider a maximal Markov sequence $E_1 \subset E$ and denote the sum of the random variable in E_1 by $S_n = X_1 + \ldots + X_n$ where X_i are exponentially distributed such that

$$1 - F_{X_i}(x) = e^{-\frac{x}{\mu}}$$

for $x \geq 0$ and $\mu > 0$. For the dependence of X_{i-1} and X_i for each $i > 1$, we choose Farlie-Gumbel-Morgenstern Copula defined by

$$C(u, v) = uv + \lambda uv(1 - u)(1 - v)$$

for $(u, v) \in [0, 1]^2$ and $-1 \leq \lambda \leq 1$. By the definition of the copula, it is immediate to find the partial derivative and the density of the copula as follows.

$$C_u(u, v) = v + \lambda v(1 - v)(1 - 2u),$$

$$c(u, v) = 1 + \lambda(1 - 2v)(1 - 2u).$$

Suppose that $\{X_1, \ldots, X_n\}$ satisfies Markov property. Then the aggregate loss distribution F_{S_n} and the aggregate density function f_{S_n} for any dimension $n > 2$ can be calculated numerically by (3.5) $\sim$ (3.8). Fig. 3.2 is the sketch of the aggregate distribution F_{S_4} and the aggregate density f_{S_4} computed numerically by (3.5) $\sim$ (3.8) and simulations, respectively when $\mu = 1$ and $\lambda = -1, 1$. We generated 100 million samples for Monte-Carlo simulation. The length of each subinterval is equal to 0.01 for the univariate numerical integrations in the computational method, (3.5) $\sim$ (3.8). It is obvious from Fig. 3.2 that the positive or negative dependence between subsequent losses has substantial influence on both of the tails. The aggregate distribution with $\lambda = 1$ has fatter tails in the left and right than the aggregate distribution with $\lambda = -1$. In other words, large(small) losses are more likely to incur subsequent large(small) losses under the pairwise positive dependence $(\lambda > 0)$. Conversely, the negative dependence $(\lambda < 0)$ reduces the chance of extreme events so that large(or small) losses are less likely to occur subsequently.

Numerical			Simulated			Diff/Sim		
x	$\lambda = -1$	$\lambda = 1$	x	$\lambda = -1$	$\lambda = 1$	x	$\lambda = -1$	$\lambda = 1$
2.1	0.097154	0.245932	2.1	0.097385	0.246678	2.1	0.002376	0.003023
3.6	0.462003	0.508551	3.6	0.462715	0.509805	3.6	0.001539	0.002460
4.8	0.728725	0.678574	4.8	0.729644	0.680003	4.8	0.001260	0.002102
6.3	0.905492	0.830151	6.3	0.906509	0.831611	6.3	0.001122	0.001756
7.2	0.952690	0.890303	7.2	0.953768	0.891792	7.2	0.001130	0.001670
7.8	0.970524	0.919685	7.8	0.971620	0.921178	7.8	0.001128	0.001621
9.3	0.990879	0.965219	9.3	0.992000	0.966690	9.3	0.001130	0.001521
11.7	0.997904	0.991459	11.7	0.999038	0.992937	11.7	0.001135	0.001488

Table 3.1 Numerical and simulated values of the aggregate distributions with exponential marginal distributions and Falie-Gumbel-Morgenstern copula; dimension $n = 4$, exponential mean $\mu = 1$, copula parameter $\lambda = -1, 1$, and the number of simulations is $100,000,000$.

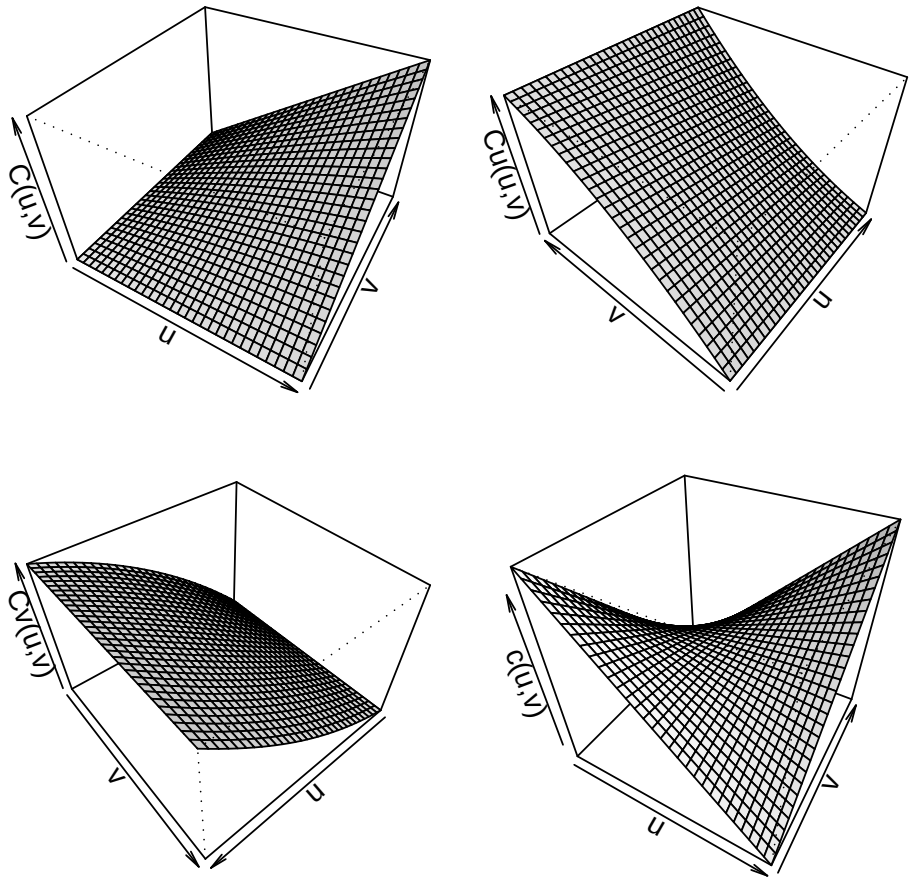


Figure 3.1 $\lambda = 1$. Top left: Farlie-Gumbel-Morgenstern Copula $C(u, v)$ on the unit square $(u, v) \in [0.1]^2$. Top right: partial derivative of $C(u, v)$ with respect to u on $(0, 1)^2$. Bottom left: partial derivative of $C(u, v)$ with respect to v on $(0, 1)^2$. Bottom right: copula density on $(0, 1)^2$

Numerical	F_{S_4}	Numerical	f_{S_4}	Simulation
$h = 0.01$	$h = 0.005$	$h = 0.01$	$h = 0.005$	100 million
499.660000	3993.390000	500.180000	3999.820000	838.470000

Table 3.2 Runtimes(seconds) of the simulation method and the computational method where h is the length of each subinterval in every univariate numerical integration.

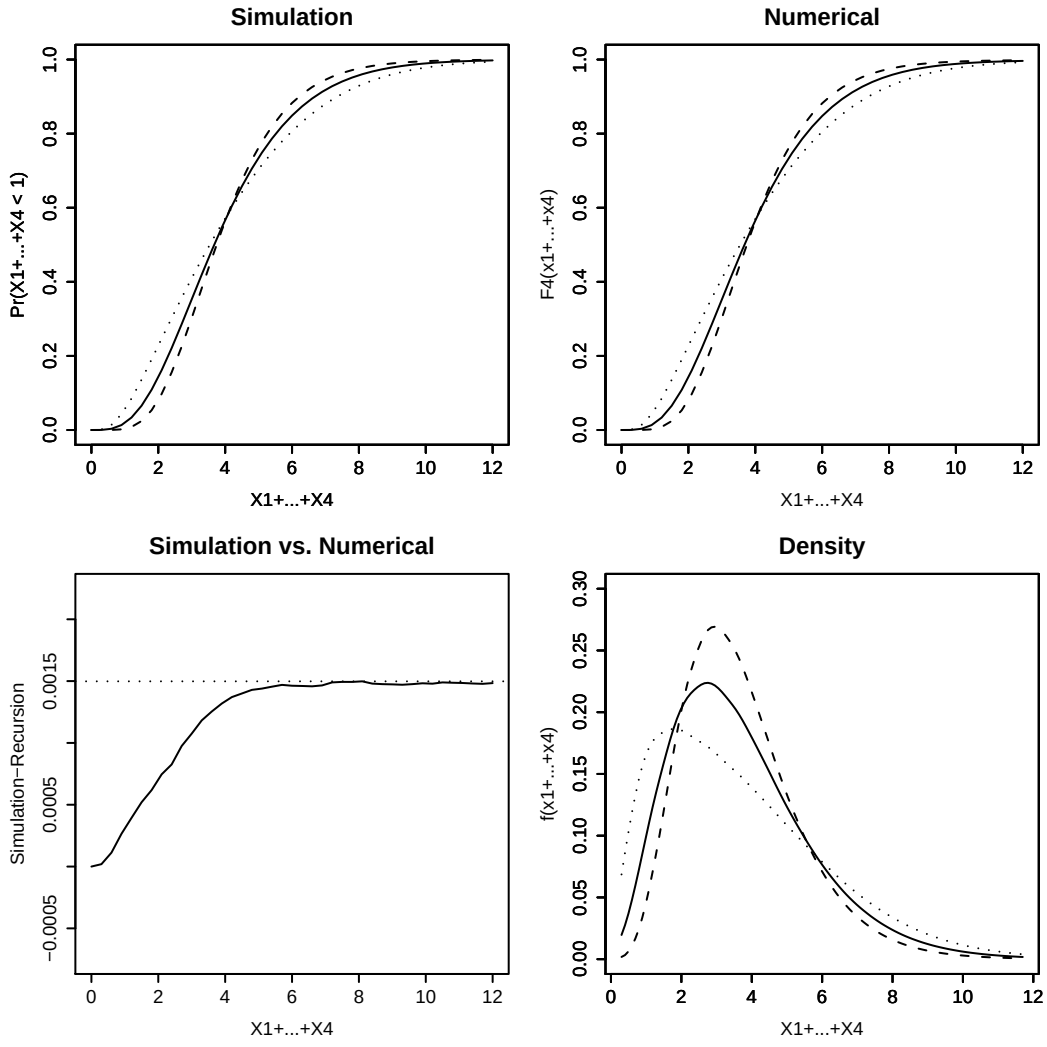


Figure 3.2 $\lambda = -1$ (dashed), $\lambda = 0$ (solid), $\lambda = 1$ (dotted). Top left: aggregate distributions of $X_1 + \cdots + X_4$ with Farlie-Gumbel-Morgenstern copula and exponential ($\mu = 1$) marginal distributions by simulation. Top right: aggregate distributions with Farlie-Gumbel-Morgenstern copula and Exponential ($\mu = 1$) marginal distributions by Theorem 3.2. Bottom left: difference between simulation and numerical approximation. Bottom right: aggregate density functions by Theorem 3.2.

3.3.3 Gaussian Copula and Pareto Distributions

Consider a maximal Markov sequence $E_1 \subset E$ and denote the sum of the random variable in E_1 by $S_n = X_1 + \ldots + X_n$ where X_i are Pareto distributed such that

$$1 - F_{X_i}(x) = \Big(\frac{\beta}{x+\beta}\Big)^{\alpha}$$

for $\alpha, \beta > 0$ and $x \geq 0$. Pareto distribution with $0 < \alpha < 1$ is well known as an extremely heavy-tailed loss distribution. For $\alpha > 1$ and a positive integer $k > \alpha$, k-th moment of Pareto(α, β) is as follows.

$$E[X^k] = \frac{\beta^k k}{(\alpha-1)\cdots(\alpha-k)}.$$

For $0 < \alpha \leq 1$, however, no finite moments exist. The parameter α determines the maximum domain of attraction of the distribution and it is one of the most important indices in the family of extreme value distributions. For more details, see Embrechts et al. (1997)

For the dependence between X_i and X_{i-1} for each $i > 2$, we choose Gaussian copula with correlation coefficient $-1 < \rho < 1$. Let ϕ and Φ be the density and the distribution function of the standard normal distribution $\mathcal{N}(0,1)$, respectively. Denote the density and the distribution function of the bivariate standard normal distribution with correlation coefficient ρ by ϕ_ρ and Φ_ρ, then Gaussian copula C is defined by

$$C(u,v) = \Phi_\rho\big(\Phi^{-1}(u), \Phi^{-1}(v)\big)$$

where $u = F_{i-1}(x_{i-1})$ and $v = F_i(x_i)$ for each $i > 1$. Gaussian copula transforms the random variables to the standard normal random variables and projects them onto the bivariate standard normal dependence structure. As a result the random variables defined by Gaussian copula behave as if they were bivariate standard normal random variables although the marginals are not normally

distributed. In order to apply the numerical procedure of (3.5) $\sim$ (3.8), it is necessary to find the explicit forms of $c^{(i)}(F_{i-1}(x_{i-1}), F_i(x_i))$ and $C_u^{(i)}(F_{i-1}(x_{i-1}), F_i(x_i))$. Moreover, we should take extra care of the well-definedness of the copula derivatives. For example, $C_u^{(i)}(u, v)$ doesn't converge to a finite value as $(u, v) \to (0, v)$ for any $v \in (0, 1)$. Fig.3.3 illustrates the singularities of the copula derivatives on the boundary of the domain $[0, 1]^2$.

It is not difficult to derive the following two results from the definition of the Gaussian copula,

$$C_u(u, v) = \Phi\Big(\frac{\Phi^{-1}(v) - \rho\,\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\Big)$$

$$c(u, v) = \frac{1}{\sqrt{1-\rho^2}} e^{-\frac{\rho^2}{2(1-\rho^2)}\left\{\Phi^{-1}(u)^2 + \Phi^{-1}(v)^2 - \frac{2}{\rho}\Phi^{-1}(u)\Phi^{-1}(v)\right\}}$$

for all $(u, v) \in (0, 1)^2$. Applying these results into (3.5) $\sim$ (3.8), $F_{S_n}(t)$ and $f_{S_n}(t)$ can be calculated numerically with desired precisions for any finite $t \geq 0$ and positive integer $n > 2$. Table 3.3 and Fig. 3.4 are the summary of the numerical values and the sketch of the aggregate distributions F_{S_4}, where the marginal distributions are identically Pareto($\alpha = 0.9$, $\beta = 0.4$) distributed under Gaussian copula framework with different values of correlation coefficients, $\rho = -0.7, 0$, and 0.7. Simulated values are also compared and the number of simulations is again $100,000,000$.

Similar to the case of Farlie-Gumbel-Morgenstern copula, the Gaussian copula with negative correlation $(\rho = -0.7)$ also shows that the aggregate density function is less dispersed and more centered than the positive correlation $(\rho = 0.7)$ and vice versa. As a Pareto distribution is used as a marginal, we can easily see that the aggregate distribution has heavier tail than the aggregate distribution of exponential marginals with Farlie-Gumbel-Morgenstern copula.

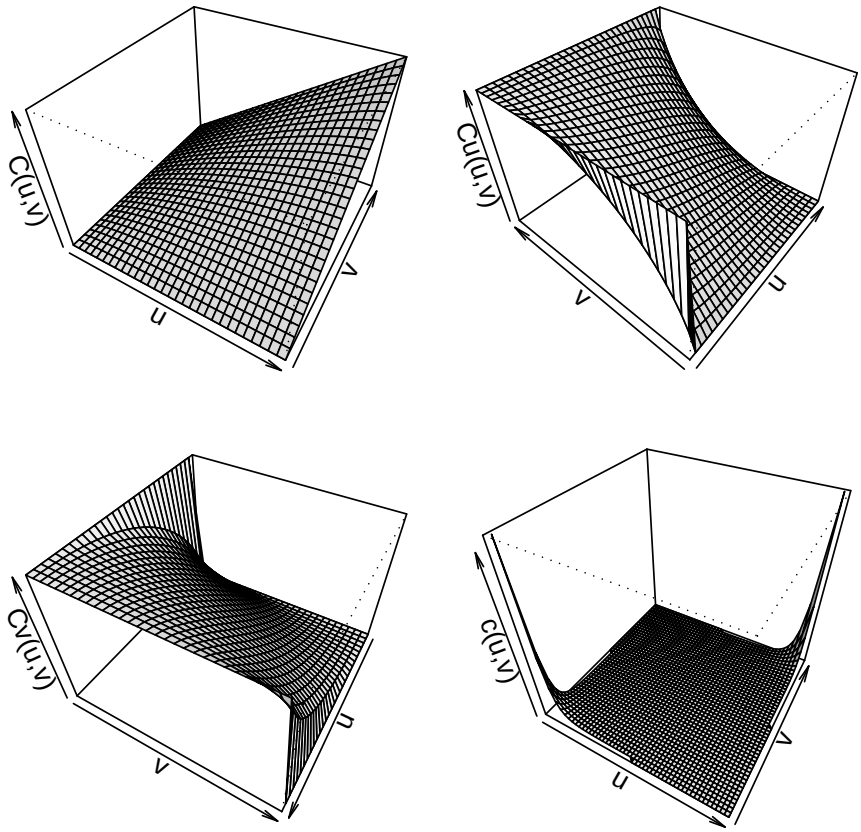


Figure 3.3 $\rho = 0.5$. Top left: Gaussian Copula $C(u, v)$ on the unit square $(u, v) \in [0.1]^2$. Top right: partial derivative of $C(u, v)$ with regard to u on $(0, 1)^2$. Bottom left: partial derivative of $C(u, v)$ with regard to v on $(0, 1)^2$. Bottom right: copula density on $(0, 1)^2$

Numerical			Simulated			Diff/Sim		
x	$\rho = -0.7$	$\rho = 0.7$	x	$\rho = -0.7$	$\rho = 0.7$	x	$\rho = -0.7$	$\rho = 0.7$
2.8	0.293849	0.499795	2.8	0.294012	0.502817	2.8	0.000554	0.006011
4.4	0.511512	0.616991	4.4	0.511531	0.619997	4.4	0.000037	0.004849
5.6	0.606551	0.673484	5.6	0.606537	0.676516	5.6	-0.000024	0.004481
7.6	0.702002	0.737013	7.6	0.701982	0.740036	7.6	-0.000029	0.004085
9.2	0.749585	0.771883	9.2	0.749539	0.774880	9.2	-0.000062	0.003869
11.2	0.790581	0.803865	11.2	0.790514	0.806847	11.2	-0.000085	0.003697
13.2	0.819624	0.827636	13.2	0.819483	0.830641	13.2	-0.000171	0.003618
16.0	0.848561	0.852262	16.0	0.848357	0.855292	16.0	-0.000240	0.003542

Table 3.3 Numerical and simulated values of the aggregate distributions with Pareto marginals and Gaussian copula; dimension $n = 4$, Pareto distribution parameters $\alpha = 0.9$, $\beta = 0.4$, correlation coefficients $\rho = -0.7, 0.7$, and the number of simulations is $100,000,000$.

Numerical F_{S_4} $h = 0.01$	Numerical f_{S_4} $h = 0.01$	Simulation 100 million
7997.310000	8080.630000	1073.870000

Table 3.4 Runtimes(seconds) of the simulation method and the computational method where h is the length of each subinterval in every univariate numerical integration.

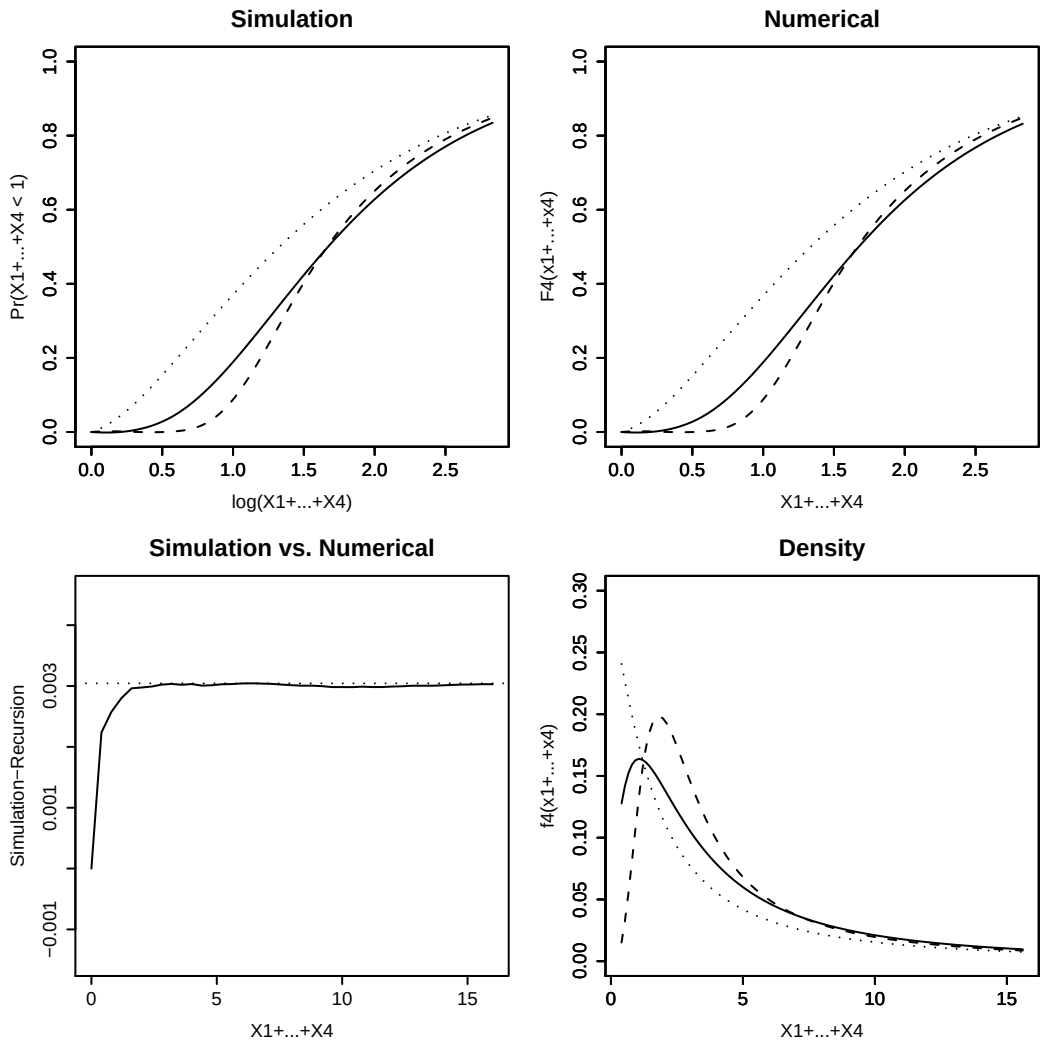


Figure 3.4 $\rho = -0.7$ (dashed), $\rho = 0$ (solid), $\rho = 0.7$ (dotted). Top left: aggregate distributions of $X_1 + \cdots + X_4$ with Gaussian copula and Pareto($\alpha = 0.9$, $\beta = 0.4$) marginal distributions by simulation. Top right: aggregate distributions with Gaussian copula and Pareto($\alpha = 0.9$, $\beta = 0.4$) marginal distributions by Theorem 3.2. Bottom left: difference between simulation and Theorem 3.2. Bottom right: aggregate density functions by Theorem 3.2.

3.4 Applications on Bayesian Premiums

Conditional expectations have been widely used to predict future premiums based on the past claim observations in insurance pricing. Let us denote the pure premium by $\mu_{n+1} = E[X_{n+1}]$ and the hypothetical premium by $\mu_{n+1}(\Theta)$ where Θ is the associated parameter with the policyholder or group of policyholders. Because the parameter Θ or the distribution of Θ is unknown in general, we are often required to use the conditional expectation with respect to the past data, $X_1, \ldots, X_n$. We can use the Bayesian premium $E[X_{n+1}|X_1, \ldots, X_n]$ as a future premium $\mu_{n+1}(\Theta)$ for a particular group of policyholders possessing homogeneous risk. In higher dimension n, however, it is not easy to compute the Bayesian premium because of the functional complexity of the joint distribution, $F_{\mathbf{X}}(x_1, \ldots, x_{n+1})$ in general dependence structure.

In order to avoid this numerical difficulty, Bühlmann (1967) proposed the following credibility model to approximate $\mu_{n+1}(\Theta)$ as a linear function of the past observations $X_1, \cdots, X_n$, $i.i.d.$ conditional on Θ,

$$\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i \tag{3.9}$$

with appropriate coefficients α_i for $i = 0, \cdots, n$. The random variables X_i are assumed to be identically distributed and independent conditional on Θ. By elementary calculus, we can find the coefficients α_i minimizing the squared error in the following manner. Let Q be the expectation of the squared error,

$$Q = E\Big\{ \Big[\mu_{n+1}(\Theta) - \alpha_0 - \sum_{i=1}^{n} \alpha_i X_i\Big]^2 \Big\}.$$

If we denote by $(\tilde{\alpha}_0, \cdots, \tilde{\alpha}_n)$ the values of α_i minimizing Q, we have

$$\frac{\partial Q}{\partial \alpha_i}\Big|_{(\tilde{\alpha}_0, \cdots, \tilde{\alpha}_n)} = 0 \quad \text{for} \quad i = 0, \cdots, n.$$

Solving the system of equations, we obtain so called normal equations,

$$E(X_{n+1}) = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_i E(X_i), \quad Cov(X_i, X_{n+1}) = \sum_{j=1}^{n} \tilde{\alpha}_i Cov(X_i, X_j) \tag{3.10}$$

for $i = 1, \ldots, n$. The above system of n-equations can be viewed as a matrix equation, i.e.

$$\begin{pmatrix} Cov(X_1, X_{n+1}) \\ Cov(X_2, X_{n+1}) \\ \vdots \\ Cov(X_n, X_{n+1}) \end{pmatrix} = \begin{pmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \ldots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & & \ldots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \ldots & Cov(X_n, X_n) \end{pmatrix} \begin{pmatrix} \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \\ \vdots \\ \tilde{\alpha}_n \end{pmatrix}.$$

Therefore, the coefficients $\tilde{\alpha}_i$ can be found as long as the covariance matrix of $(X_1, \cdots, X_n)$ is not singular. For more details, see Klugman et al. (1998).

Bühlmann's linear model (3.9) is very useful when the conditional density $f_{X_{n+1}|\Theta}$ or the conditional expectation $E[X_{n+1}|X_1, \cdots, X_n]$ is hard to calculate, for example, when the random variables X_i are not independent in higher dimension n. However, this method can't be applied to approximate the Bayesian premium when losses are negatively correlated, which is illustrated in Example 3.5 and Table 3.

3.4.1 Conditioning on S_n

In the following we apply Lemma 3.1 and Theorem 3.2 discussed in the previous section to compute the Bayesian premium directly, conditioning on the sum of the past observations $S_n = \sum_{i=1}^{n} X_i$ instead of $\{X_1, \cdots, X_n\}$. Before we proceed to compute the conditional expectation, $E[X_{n+1}|S_n]$, we briefly examine the similarity between Bühlmann's linear model and the following simpler model conditioning on S_n.

Let us assume the Bayesian premium is of the form,

$$\alpha + \beta S_n, \quad \text{where} \quad S_n = \sum_{i=1}^{n} X_i, \quad E[X_i] < \infty \tag{3.11}$$

and α, β need to be specified. Let W be the expectation of the squared error, i.e.

$$W = E\Big\{ \Big[\mu_{n+1}(\Theta) - \alpha - \beta S_n\Big]^2 \Big\}.$$

Similarly, if we denote the values of α, β minimizing W by $\tilde{\alpha}, \tilde{\beta}$, they satisfy the following partial differential equations,

$$\frac{\partial W}{\partial \alpha}|_{(\tilde{\alpha},\tilde{\beta})} = 0, \quad \frac{\partial W}{\partial \beta}|_{(\tilde{\alpha},\tilde{\beta})} = 0$$

which reduce to the analogous normal equations,

$$E(X_{n+1}) = \tilde{\alpha} + \tilde{\beta} E(S_n), \quad Cov(S_n, X_{n+1}) = \tilde{\beta} Var(S_n). \tag{3.12}$$

It is immediate to find the solution of (3.12),

$$\tilde{\alpha} = E[X_{n+1}] - \tilde{\beta} E[S_n], \quad \tilde{\beta} = \frac{\sum_{i=1}^{n} Cov(X_i, X_{n+1})}{Var(S_n)}.$$

It is very interesting that two functional assumptions (3.9) and (3.11) of the credibility premiums provide the same credibility factor under certain dependence structure as in the following example.

Example 3.5 (i) Bühlmann's Credibility Premium: Suppose that $E[X_i] = \mu$, $Var(X_i) = \sigma^2$, and $Cov(X_i, X_j) = \rho\sigma^2$ for $i \neq j$ where the correlation coefficient ρ satisfies $0 < \rho < 1$. Then the

classical credibility premium of $\mu_{n+1}(\Theta)$ can be calculated as a linear combination of $X_1, \cdots, X_n$,

$$\tilde{\alpha_0} + \sum_{i=1}^{n} \tilde{\alpha_i} X_i$$

where the coefficients $\tilde{\alpha_i}$ are as follows.

$$\tilde{\alpha_0} = \frac{(1-\rho)\mu}{1-\rho+n\rho}, \quad \tilde{\alpha_i} = \frac{\rho}{1-\rho+n\rho} \quad \text{for} \quad i = 1, \cdots, n.$$

The credibility premium reduces to

$$(1-Z)\mu + Z\bar{X} \quad \text{where} \quad Z = \frac{n\rho}{1-\rho+n\rho}, \quad \bar{X} = \sum_{i=1}^{n} X_i/n \tag{3.13}$$

which is the desired weighted average of μ and $\bar{X}$ if $0 < \rho < 1$ (Klugman et al., 1998).
(ii) Bayesian premium conditioning on S_n: Now we consider the credibility premium of the form (3.11). The credibility premium in this case is a simple linear combination of S_n only, i.e. $\tilde{\alpha} + \tilde{\beta} S_n$. From the normal equations (3.12), the coefficients $\tilde{\alpha}, \tilde{\beta}$ can be easily calculated as

$$\begin{aligned} \tilde{\alpha} &= \mu(1 - n\tilde{\beta}) = \frac{n(1-\rho)\mu}{1-\rho+n\rho} = \tilde{\alpha}_0 \\ \tilde{\beta} &= \frac{\sum_{i=1}^{n} Cov(X_i, X_{n+1})}{Var(S_n)} = \frac{\rho}{1-\rho+n\rho} = \tilde{\alpha}_i \end{aligned}$$

which gives the identical credibility factor Z and premium as in the previous result of (3.13). Therefore the credibility premium conditioning on the sum of the past losses does not lose any information of the past observations, if $X_1, \cdots, X_n$ possess the same mean value and the correlation coefficients of (X_i, X_j) are identical for $i \neq j$. In general, S_n can be viewed as a linearly sufficient statistic, $T \cdot [X_1, \ldots, X_n]'$, for an appropriate non-ramdom matrix T, which is $[1, \ldots, 1]$ in this case. Therefore

$$\begin{aligned} E[X_{n+1}|X_1,\ldots,X_n] &= E[X_{n+1}\,|\,T\cdot(X_1,\ldots,X_n)'], \\ \tilde{\alpha_0}+\sum_{i=1}^{n}\tilde{\alpha}_i X_i &= \tilde{\alpha}+\tilde{\beta}S_n. \end{aligned}$$

For more about linearly sufficient statistics, see Sundt (1991) and references therein. □

Now consider the following Bayesian premium conditioning on the sum S_n without any functional assumption of the conditional expectation,

$$E[X_{n+1}|S_n]. \tag{3.14}$$

Summing up the past observations, we may lose intrinsic information among X_i more or less. However, by simplifying the conditioning, we have an advantage that it is possible to evaluate the conditional expectation by applying Lemma 3.1 and Theorem 3.2 under the assumption of Markov property among $\{X_1,\ldots,X_n\}$. The conditional expectation (3.14) is the best estimator in the following sense. For any function g of S_n,

$$E\Big\{\big[X_{n+1}-E[X_{n+1}|S_n]\big]^2\Big\} \le E\Big\{\big[X_{n+1}-g(S_n)\big]^2\Big\}.$$

It is easy to write the following asymptotic approximation of the Bayesian premium in terms of the joint density of X_{n+1}, S_n and the aggregate density S_n,

$$E[X_{n+1}|S_n] = \lim_{M\to\infty}\int_0^M t\, f_{X_{n+1}|S_n}(t|s)\,dt = \lim_{M\to\infty}\int_0^M t\,\frac{f_{X_{n+1},S_n}(t,s)}{f_{S_n}(s)}\,dt \tag{3.15}$$

where $f_{X_{n+1},S_n}(x,s)$ and $f_{S_n}(s)$ can be numerically calculated by Lemma 3.1 and Theorem3.2. In sum, if the following conditions are satisfied,

1. Loss random variables X_i are non-negative and continuous.

2. $\{X_1, \ldots, X_n\}$ satisfies Markov property.

3. $f_{X_{i+1}|X_i}(x_{i+1}|x_i)$ is known for each $i = 1, \ldots, n$,

Bayesian premium can be approximated numerically with desired precision by (3.15); For any given $\epsilon > 0$ and positive integer $n > 1$

$$\left| E[X_{n+1}|S_n = s] - \int_0^M t \, \frac{f_{X_{n+1},S_n}(t,s)}{f_{S_n}(s)} \, dt \right| < \epsilon$$

for sufficiently large $M > 0$ where $f_{X_{n+1},S_n}(t,s)$ and $f_{S_n}(s)$ can be calculated by Lemma 3.1 and Theorem 3.2.

3.4.2 Numerical Examples

Table 3.5 and Fig. 3.5 are the summary of the Bayesian premiums $E[X_4|S_3]$ evaluated by (3.5) $\sim$ (3.8) for various values of S_3. We also assume $\{X_1, \ldots, X_n\}$ satisfies Markov property. Top in Table 3.5 and left in Fig. 3.5 are the Bayesian premiums when losses are exponentially distributed with unit mean and subsequent losses (X_{i-1}, X_i) are dependent through Farlie-Gumbel-Morgenstern copula with $\lambda = -1, 1$. Bottom and right in Table 3.5 and Fig. 3.5 are the case of Pareto marginal distributions and Gaussian copula. Marginals follow Pareto$(4.0, 3.0)$ distributions with unit mean and each pair of subsequent losses (X_i, X_{i+1}) are dependent through Gaussian copula with $\rho = -0.7, 0.7$.

It is very interesting to notice the phenomenon against our intuition in this example. In Table 3.5 and Fig. 3.5, Bayesian premium with negative dependence of subsequent losses increases(decreases) as S_3 decreases(increases), which implies that Bayesian premiums are heavily affected by the sign of the correlations under Markov property. The negative dependence structure, $\lambda = -1$ or $\rho = -0.7$ has strong influence on Bayesian premiums so that smaller(larger) the aggregate loss is, the higher(lower) the credibility premium is. In other words, Bayesian premium

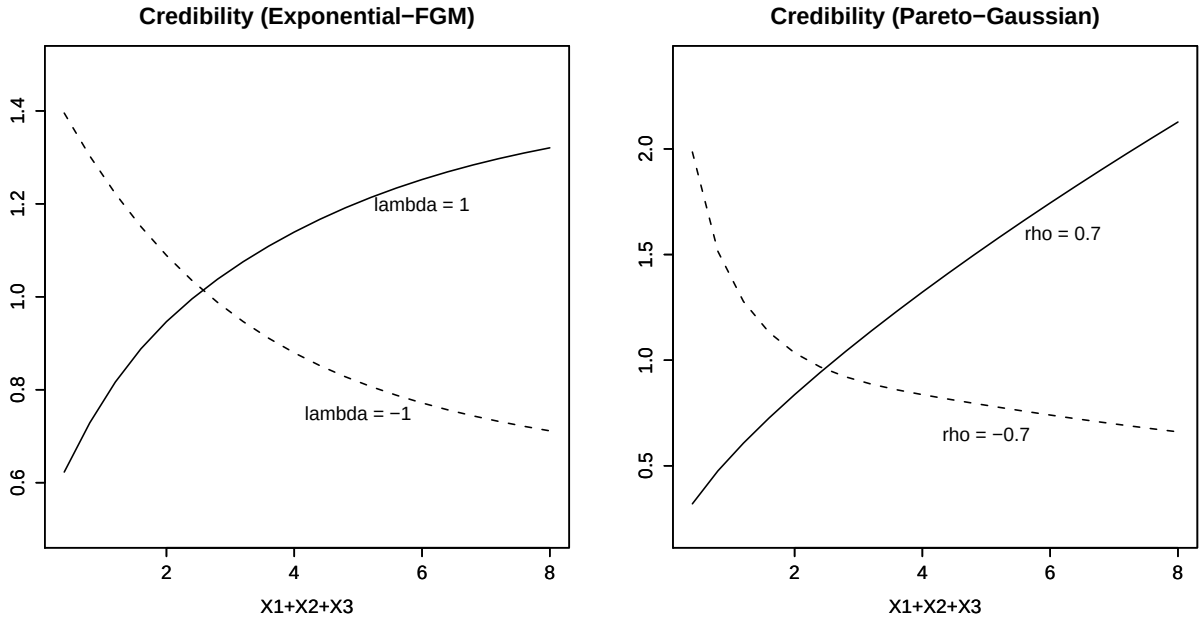


Figure 3.5 Bayesian Premiums $E[X_4|S_3]$. Left: Unit mean exponential marginal distributions and Farlie-Gumbel-Morgenstern copulas with $\lambda = -1, 1$(equivalently $\rho = -0.25, 0.25$). Right: Unit mean Pareto$(4.0, 3.0)$ marginal distributions and Gaussian copulas ($\rho = -0.7, 0.7$).

Exp-FGM	S_3	1.2	2.0	2.8	3.6	4.4	5.2	6.0	6.8	7.6
$E[X_4\|S_3]$	$\lambda = -1$	1.222	1.089	0.988	0.911	0.852	0.807	0.772	0.744	0.721
	$\lambda = 1$	0.817	0.947	1.038	1.109	1.167	1.214	1.252	1.284	1.309
Pareto-Gau	S_3	1.2	2.0	2.8	3.6	4.4	5.2	6.0	6.8	7.6
$E[X_4\|S_3]$	$\rho = -0.7$	1.278	1.035	0.921	0.860	0.816	0.777	0.740	0.706	0.675
	$\rho = 0.7$	0.607	0.837	1.042	1.232	1.410	1.581	1.744	1.901	2.053

Table 3.5 Selected values of Fig.5. Top: Unit mean Exponential marginal distributions and Farlie-Gumbel-Morgenstern copulas. Bottom: Unit mean Pareto marginal distributions and Gaussian copulas.

with the sum of the past losses smaller(larger) than $n\,E[X_1]$ becomes ironically larger(smaller) than $E[X_1]$.

3.5 Application on Stop-Loss Insurance

The computational method of Lemma 3.1 and Theorem 3.2 can be also used to calculate the limited expected value of the aggregate loss,

$$\int_a^b \big(1 - F_{S_n}(s)\big)\, ds \tag{3.16}$$

for any non-negative real numbers $a < b$. Amount of aggregate loss covered by insurer(reinsurer) in stop-loss insurance with retention level $L > 0$ can be numerically computed with desired precision by (3.16) with $a = 0$ and $b = L$ ($a = L$ and $b = \infty$). Since $E[S_n] = \sum_{i=1}^n E[X_i]$ regardless of the dependence structure among X_i, amount covered by reinsurer is simply the difference of $E[S_n]$ and insurer's aggregate loss. In the following we examine how much the dependence structure distort the aggregate loss amount covered by insurer and reinsurer, respectively.

Fig. 3.2 and Fig. 3.4 in section 3.3 illustrate the effect of negative or positive correlations of (X_i, X_{i+1}) on the aggregate distributions. It is very clear that the aggregate density functions with pairwise positive correlations are more dispersed than those with negative correlations. The area between the distribution of independent losses and any other distribution of dependent losses can be considered as loss amount adjustment or risk loading due to the dependence effect. In stop loss insurance, Lemma 3.1 and Theorem 3.2 can be applied to compute the amount of aggregate loss covered by insurer and reinsurer under Markov property among multiple lines of business or product. For example, consider the aggregate loss of section 3.3.2 where we choose Farlie-Gumbel-Morgenstein copula and exponential marginals. Denote the retention level and the loss amount of each line of business or product by L and X_i, respectively, for $i = 1 \cdots n$, then $E_\lambda[\min(S_n, L)]$ and $E_\lambda[\max(S_n - L, 0)]$ can be calculated for $\lambda \in [-1, 1]$. The adjustment amount of aggregate loss due to the dependence can be numerically calculated as

$$P_{ins} = E_\lambda[\min(S_4, L)] - E_0[\min(S_4, L)] \quad \text{for insurer}$$

$$P_{re} = -P_{ins} \quad \text{for reinsurer} \tag{3.17}$$

where $E_\lambda[\cdot]$ and $E_0[\cdot]$ are the expectations with respect to the aggregate distribution functions corresponding to the dependent $(\lambda \neq 0)$ and independent $(\lambda = 0)$ cases. The equation (3.17) holds true since the two areas overlapped by any two curves of the distributions in the Fig. 3.2 are equal.

The expected aggregate loss amount under the independence assumption can be adjusted to the amount under certain dependence between losses by multiplying the adjustment coefficients corresponding to each retention L in the following manner. Suppose that insurer or reinsurer calculated its aggregate loss amount under the independence assumption. The adjustment coefficient, the ratio of the expected aggregate loss amount under dependence to the amount under independence, can be calculated as follows.

$$c_{ins} = \frac{E_\lambda[\min(S_n, L)]}{E_0[\min(S_n, L)]},$$

$$c_{re} = \frac{E[S_n] - E_\lambda[\min(S_n, L)]}{E[S_n] - E_0[\min(S_n, L)]} = \frac{nE[X_1] - E_\lambda[\min(S_n, L)]}{nE[X_1] - E_0[\min(S_n, L)]}$$

for insurer and reinsurer, respectively. Table 3.6 is the summary of the numerical values of the expectations and the amounts of adjustment due to the dependence. It is immediate from Table 3.6 and Fig. 3.6 that if dependence is not accounted for in stop-loss pricing, insurer(reinsurer) is always less(over) charged for all retention level $L > 0$ when each subsequent losses are negatively correlated. In the same manner, insurer(reinsurer) is always over(less) charged for all retention level $L > 0$ when each subsequent losses are positively correlated. The adjustment coefficients in Table 3.6 can be used to correct the discrepancy caused by the dependence for the aggregate loss amount coverd by the insurer and reinsurer if losses are dependent with Markov property.

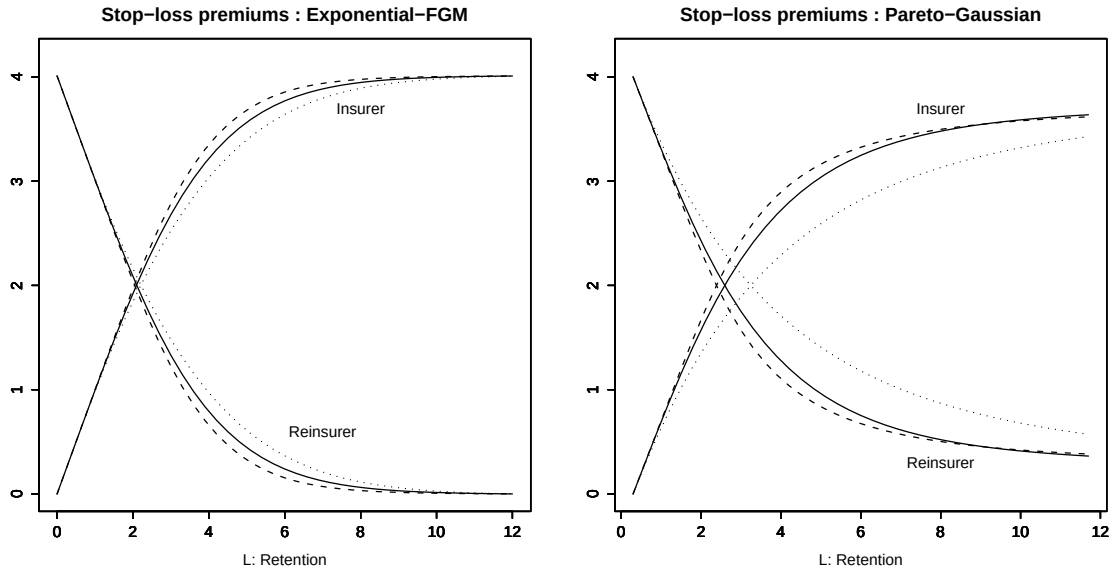


Figure 3.6 Aggregate loss amount covered by insurer and reinsurer in stop-loss insurance at retention level L. Left: Unit mean exponential marginals and Farlie-Gumbel-Morgenstern copulas with $\lambda = -1$ (dashed), $\lambda = 0$ (solid), $\lambda = 1$ (dotted). Right: Unit mean Pareto(4.0,3.0) marginals and Gaussian copulas with $\theta = -0.7$ (dashed) $\theta = 0$ (solid), $\theta = 0.7$ (dotted).

	$\lambda = 0$	$\lambda = -1$			$\lambda = 1$		
L	$E_0[\cdot]$	$E_\lambda[\cdot]$	c_{ins}	c_{re}	$E_\lambda[\cdot]$	c_{ins}	c_{re}
2.1	2.008573	2.059099	1.025155	0.974565	1.922591	0.957192	1.041406
3.6	3.028855	3.156010	1.041981	0.869946	2.851486	0.941440	1.177694
4.8	3.508838	3.631421	1.034935	0.754366	3.335304	0.950544	1.340432
6.3	3.812538	3.887058	1.019546	0.619548	3.695597	0.969327	1.578174
7.2	3.901152	3.948849	1.012227	0.555876	3.819998	0.979197	1.720920
7.8	3.937759	3.971578	1.008588	0.522649	3.876712	0.984497	1.808656
9.3	3.985425	3.997622	1.003061	0.475384	3.958980	0.993365	1.973333
11.7	4.008183	4.008411	1.000057	0.500942	4.003868	0.998923	1.992191
	$\theta = 0$	$\theta = -0.7$			$\theta = 0.7$		
L	$E_0[\cdot]$	$E_\theta[\cdot]$	c_{ins}	c_{re}	$E_\theta[\cdot]$	c_{ins}	c_{re}
2.4	1.870399	2.010100	1.074691	0.934400	1.589323	0.849724	1.131985
3.9	2.684855	2.858933	1.064837	0.867636	2.262259	0.842600	1.321331
5.1	3.067392	3.183525	1.037861	0.875475	2.625897	0.856068	1.473398
6.6	3.343818	3.392542	1.014571	0.925747	2.940841	0.879486	1.614125
7.5	3.445358	3.466240	1.006061	0.962351	3.078798	0.893607	1.660896
8.1	3.495988	3.502803	1.001949	0.986478	3.155172	0.902512	1.676206
9.6	3.583456	3.566245	0.995197	1.041319	3.305100	0.922322	1.668250
11.7	3.652033	3.616677	0.990319	1.101608	3.447319	0.943945	1.588316

Table 3.6 Insurer's and reinsurer's amount of loss and the adjustment coefficients c_{ins}, c_{re} for each retention level L where $E_*[\cdot] = E_*[\min(X_1 + \cdots + X_4, L)]$. Top: Unit mean exponential marginals and Farlie-Gumbel-Morgenstern copulas. Bottom: Unit mean Pareto(4.0,3.0) marginals and Gaussian copulas.

3.6 Conclusions

We introduce a decomposition of a pool of risks into disjoint maximal Markov sequences in order to obtain a smaller size of equivalent risk pool. Considering the random variables in a maximal Markov sequence, we present a computational method for the aggregate loss distribution of non-identical and dependent random variables. We provide an application of the method with a sequence of pairwise bivarate copulas to implement the dependence among losses with numerical examples. Actuarial applications of the computational method are also given; Bayesian premiums are calculated conditioning on the sum of past observations and we examine the impact of dependence on stop-loss insurance.

Chapter 4

Asymptotic Super(Sub)additivity of the Value-at-risk of Regularly Varying Random Variables

This chapter is based on the paper by Jang and Jho (2008b). Value-at-risk has been one of the most popular methods of risk management because of its simplicity and downside risk measurement. Recently, it has gained popularity in quantifying the operational risks of banks and insurance companies since the Basel II proposed advanced measurement approach (AMA) requiring quantitative method based on the company's internal and external empirical data (Basel Committee on Banking Supervision, 2004). Value-at-risk often fails to measure the financial and operational risks because of its lack of subadditivity and stability. In the context of the value-at-risk measure, there are two major issues of concern in this chapter, dependence and heavy-tailedness. According to the principle proposed by Basel II, operational risks are classified into eight business lines and seven event types. Applying the standardized classification matrix of operational risks, the total aggregate loss can be modeled by

$$Y(t) = Y_1(t) + \cdots + Y_d(t) \tag{4.1}$$

where each $Y_i(t)$ represents for the individual aggregate loss of each line of business (or event type) such that

$$Y_i(t) = \sum_{k=1}^{N_i(t)} Y_k^{(i)}, \quad t \geq 0 \tag{4.2}$$

where d is a positive integer and $Y_k^{(i)}$ are the individual losses independent of the appropriate counting processes $N_i(t)$ for $t > 0$.

The above risk model has been widely used in the presence of certain dependence structure among losses in numerous papers. Value-at-risk becomes a coherent measure under certain multivariate structure such as elliptical distributions (Embrechts et al., 2002). Embrechts et al. (2003) introduced the concept of copula to find the bound of the value-at-risk of the aggregate risk. Mesfioui and Quessy (2005) discussed the bounds on the value-at-risk for the sum of possibly dependent risks when only partial information is available about the dependence structure by way of copulas. Chavez-Demoulin et al. (2006a,b) also considered the aggregate loss of compound type and discussed whether the subadditivity of the value-at-risk holds or not when the severity distributions are very skewed or extremely heavy-tailed or there exists a special dependence among losses. Wüthrich (2003) and Alink et al. (2004) considered a finite sum of identically distributed continuous random variables with Archimedean copula the generator of which is regularly varying. The authors showed that the value-at-risk of the aggregate loss is asymptotically proportional to the value-at-risk of the individual loss. Böcker and Klüppelberg (2006, 2005) considered a d-dimensional compound Poisson process $Y(t) = (Y_1(t), \ldots, Y_d(t))_{t\geq 0}$ and developed a multivariate loss distribution under Levy copula framework to model the dependence between different operational risk cells of the classification matrix. They derived asymptotic approximation of the value-at-risk of the total aggregate loss under the assumption of certain multivariate model such that the severity distributions are subexponential, the counting processes are Poisson, and the dependence is modeled by Levy copulas.

In this chapter we consider two risk processes (4.1) and (4.2) with quite general dependence structure among losses with heavy tails. Suppose we are given the parametric forms or the estimates of the distributions of $Y_1, \cdots, Y_d$ for a fixed time $t > 0$. In particular, if Y_i's are independent, the total aggregate loss distribution is simply the convolution of F_{Y_i},

$$F_Y(x) = F_{Y_1} \star \cdots \star F_{Y_d}\ (x).$$

However, we don't have much knowledge about the exact form of the aggregate distribution when losses are dependent. The numerical computation of $\mathbb{P}(Y_1+\cdots+Y_d \leq y)$ of dependent variables is also very difficult in higher dimensions. Most of well-known recursive formulae following Panjer's recursion (Panjer, 1981) assume that losses are independent or integer-valued. The computation of the individual aggregate loss distribution, $F_{Y_i(t)}$, is even more sophisticated since it involves the counting processes $N_i(t)$ instead of the deterministic value for the loss occurrences. In order to avoid time-consuming numerical computation, we may approximate the value-at-risk of the aggregate loss by the sum of the individual value-at-risks. However, this simple method often fails to calibrate correct measure of risks since the value-at-risk is not a coherent risk measure (Artzner et al., 1999). Therefore it is very risky to assume the subadditivity or comonotonicity of the value-at-risk, and it may provide even worse risk measurement when losses are from the family of heavy-tailed distributions such as infinite mean loss random variables (Chavez-Demoulin et al., 2006a).

Danielsson et al. (2005) showed that the value-at-risk satisfies the subadditivity in the tail region when losses follow the multivariate regularly varying distributions with extreme value index $\alpha > 1$. However, the condition that $\alpha > 1$ doesn't include the case of the extremely heavy-tailed losses while it has been observed that insurance and financial risks often show the existence of extreme heavy-tailedness such as subexponential distributions or regularly varying distributions with $0 < \alpha \leq 1$. Moscadelli (2004) mentioned the empirical observation of the heavy-tailedness of the operational risks and Chavez-Demoulin et al. (2006a) also discussed the extreme behavior of the infinite mean distributions in the tail and presented many illustrations. In sum, it is not much known whether the subadditivity of the value-at-risk exists for the risks with extreme heavy-tails. It naturally raises a question that the superadditivity of the value-at-risk may exist in the tail when losses are extremely heavy-tailed or dependent, and if it does, what would be the sufficient conditions.

The following is the outline of this chapter. In section 2, we discuss the classical convolution theorem for regularly varying random variables and introduce the concept of the negligible joint

tail probability. We provide a sufficient condition of the joint distribution to generalize the convolution theorem when the random variables are not necessarily independent with regularly varying tails. Applying the result, we derive the relationship between the extreme value indices of the regularly varying distributions and the asymptotic properties of the value-at-risk in the presence of dependence. In section 3, we present numerical examples of the super(sub)additivity of the value-at-risk of regularly varying random variables under copula framework of dependence. We conclude afterwards.

In this chapter we follow usual functional notations; X_i identical regularly varying random variables not necessarily independent, F_i marginal distribution functions of X_i, H the joint distribution function of $\{X_i\}$, and $\sim$ asymptotic equivalence at $x_0 \in [-\infty, \infty]$ in the following sense,

$$a(x) \sim b(x) \quad \text{if and only if} \quad \lim_{x \to x_0} \frac{a(x)}{b(x)} = 1.$$

for real-valued functions $a(x)$ and $b(x)$.

4.1 Aggregate Distributions of Regularly Varying Random Variables

If X_i are independent, the distribution of the sum is the convolution of the distributions of X_i. However, if X_i are dependent, convolution formula can't be applied to find the aggregate distribution. In this section we provide a sufficient condition for the dependence structure among random variables to generalize Feller's convolution theorem with respect to the sum of random variables.

4.1.1 Convolution Theorem

The convolution of distribution functions F, G is defined by

$$F \star G\ (s) = \int_{-\infty}^{\infty} G(s-x)\, dF(x). \tag{4.3}$$

It is well-known that the convolution of distributions is a distribution function and if G is bounded and continuous, then so is $F \star G$ (Feller, 1971). For any independent random variables X, Y with distributions F, G, the distribution of the sum $X + Y$ can be written as

$$\mathbb{P}(X + Y \leq s) = F \star G\ (s). \tag{4.4}$$

Similarly, n-convolutions of distributions $F_1, \ldots, F_n$ can be defined recursively for $n \geq 2$ and the distribution function of the independent sum $X_1 + \cdots + X_n$ can be written as $F_1 \star \cdots \star F_n$.

A distribution F on $[0, \infty)$ is called subexponential if $\bar{F}(x) > 0$ for every x and

$$\lim_{x \to \infty} \frac{\overline{F^{n\star}}(x)}{\bar{F}(x)} = n \tag{4.5}$$

for each positive integer $n \geq 2$ where $\bar{F}(x)$ is the tail of the distribution F defined by $1 - F(x)$ and $F^{n\star}$ is the n-convolution of the distribution F. We denote the family of distributions satisfying (4.5) by $\mathcal{S}$. The family of subexponential distributions includes the following distributions;

1. Pareto: $F(x) = 1 - \left(\frac{\beta}{x+\beta}\right)^{\alpha}, \qquad \alpha > 0, \quad \beta > 0,$
2. Lognormal: $F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \qquad \mu \in \mathbb{R}, \quad \sigma > 0,$
3. Weibull distribution : $F(x) = 1 - e^{-cx^{\tau}}, \qquad c > 0, \quad 0 < \tau < 1.$

Subexponential distributions have been widely used for extreme loss modeling since it satisfies the max-sum equivalence for $i.i.d$ random variables. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed with distribution function $F \in \mathcal{S}$. Then

$$\begin{aligned} \mathbb{P}(\max(X_1 + \cdots + X_n) > x) &= 1 - F^n(x) \\ &= \overline{F}(x) \sum_{k=0}^{n-1} F^k(x) \\ &\sim n\overline{F}(x), \quad x \to \infty. \end{aligned}$$

Since $\mathbb{P}(X_1 + \cdots + X_n > x) = \overline{F^{n\star}}(x) \sim n\overline{F}(x)$, the following max-sum equivalence holds;

$$\mathbb{P}(X_1 + \cdots + X_n > x) \sim \mathbb{P}(\max(X_1, \cdots, X_n) > x)$$

which can be interpreted that the extreme of the aggregate loss is due to a single extreme loss and other losses are negligible compared to the extreme (Cai and Tang, 2004).

It is easy to verify that the convolution of $i.i.d$ random variables in $\mathcal{S}$ is also subexponential. Embrechts and Goldie (1980) showed that for $F, G \in \mathcal{S}$,

$$F \star G \in \mathcal{S} \quad \text{if} \quad \sup_{x>0} \frac{\overline{F}(x)}{\overline{G}(x)} < \infty.$$

Therefore if we set $F = G$, it is immediate to get $F^{2\star} \in \mathcal{S}$. Since $\sup_{x>0} \frac{\overline{F^{n\star}}(x)}{\overline{F}(x)} < \infty$, for all $n \geq 2$, the closure property of n-convolution of $i.i.d$ subexponential random variables follows by induction. However, the family of subexponential distributions doesn't satisfy the max-sum equivalence in general and it is also well-known since Leslie (1989) that subexponential family is not closed under convolutions, i.e. if X_1 and X_2 are not identical but $F_1, F_2 \in \mathcal{S}$, it is possible that

$$F_1 \star F_2 \notin \mathcal{S}$$

$$\mathbb{P}(X_1 + X_2 > x) \nsim \mathbb{P}(\max(X_1, X_2) > x)$$

for $F_1, F_2 \in \mathcal{S}$.

Let us consider a proper subset of the subexponential family satisfying either convolution closure or max-sum equivalence. Note that there are other classifications of distributions with respect to the asymptotic tail probability as follows.

1. Long-tailed: $F \in \mathcal{L}$ if

$$\lim_{x\to\infty} \frac{\overline{F}(x-t)}{\overline{F}(x)} = 1, \qquad \text{for any} \quad t > 0.$$

2. Subexponential: $F \in \mathcal{S}$ if

$$\lim_{x\to\infty} \frac{\overline{F^{\star n}}(x)}{\overline{F}(x)} = n, \qquad \text{for any} \quad n \geq 2.$$

3. Dominated varying: $F \in \mathcal{D}$ if

$$\limsup_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} < \infty, \qquad t \in (0,1).$$

4. Consistent varying: $F \in \mathcal{C}$ if

$$\lim_{y\uparrow 1} \lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = 1.$$

5. Extended regular varying: $F \in \mathcal{E}$ if

$$\liminf_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} \geq t^{-\alpha} \quad \text{for some} \quad \alpha \geq 0, \quad \text{all} \quad t \geq 1.$$

6. Regularly varying: $F \in \mathcal{R}_{-\alpha}$ if

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha} \quad \text{for a parameter} \quad \alpha \geq 0 \quad \text{and every} \quad t > 0.$$

We restrict the distributions in this chapter to the family of regularly varying distributions, $\mathcal{R}_{-\alpha}$ for $\alpha > 0$. A positive function L defined on $[0,\infty)$ is said to vary slowly at infinity if and only if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \quad \text{for every} \quad t > 0.$$

A distribution function F varies regularly with the extreme value index α if and only if it is of the form $\bar{F}(x) = x^{-\alpha}L(x)$ for some slowly varying function $L(x)$. The family of regularly varying distributions is a proper subset of the subexponential family and as a matter of fact the following is true. For every $\alpha > 0$,

$$\mathcal{R}_{-\alpha} \subset \mathcal{E} \subset \mathcal{C} \subset \mathcal{D} \cap \mathcal{L} \subset \mathcal{S} \subset \mathcal{L}. \tag{4.6}$$

The max-sum equivalence is valid in $\mathcal{R}_{-\alpha}$ for $\alpha > 0$ if we assume independence. Cai and Tang (2004) proved that it holds for the larger classes $\mathcal{C}$ and $\mathcal{D} \cap \mathcal{L}$. Regularly varying distributions, $\mathcal{R}_{-\alpha}$, also satisfy the convolution closure property under the independent assumption due to Feller (1971). Therefore the family of regularly varying distributions is a proper subset of the subexponetial family satisfying both of the convolution closure and the max-sum equivalence for independent random variables. Here we state, for the convenience, Feller's outstanding result of convolutions of regularly varying distributions (Feller, 1971).

Theorem 4.1 If F_1 and F_2 are two distribution functions such that

$$1 - F_i(x) = x^{-\alpha} L_i(x)$$

with $\alpha > 0$ and L_i slowly varying for each $i = 1, 2$, then the convolution $G = F_1 \star F_2$ varies regularly such that

$$1 - G(x) \sim x^{-\alpha}\big(L_1(x) + L_2(x)\big) \tag{4.7}$$

as $x \to \infty$.

For the proof of the theorem, see Feller (1971) or Embrechts et al. (1997). The n-dimensional convolution theorem is an immediate consequence of (4.7), which can be stated as $1 - G(x) \sim x^{-\alpha}(L_1(x) + \cdots + L_n(x))$ where $G = F_1 \star \cdots \star F_n$.

4.1.2 Dependent Regularly Varying Random Variables

In this section we discuss the analogous properties of the convolution closure and max-sum equivalence for regularly varying random variables not necessarily independent. We present a sufficient condition of the joint distribution H of $X_1, \cdots, X_n$ for the following two properties; If $X_i \in \mathcal{R}_{-\alpha}$ for $\alpha > 0$,

1. Max-sum equivalence of regularly varying random variables not necessarily independent:

$$\mathbb{P}(X_1 + \cdots + X_n > s) \sim \mathbb{P}(\max(X_1, \ldots, X_n) > s), \tag{4.8}$$

as $s \to \infty$ and

2. Closure property of the distributions of regularly varying random variables not necessarily independent:

$$\mathbb{P}(X_1 + \cdots + X_n \leq s) \in \mathcal{R}_{-\alpha}. \tag{4.9}$$

Definition 4.2 Let X_1 and X_2 be random variables not necessarily independent with regularly varying tails. Denote the marginals and joint distribution functions by F_1, F_2, and H, respectively. Denote the joint tail probability of H by $\hat{H}$ such that

$$\hat{H}(x_1, x_2) = \mathbb{P}(X_1 > x_1, X_2 > x_2)$$

Then the tail probability of the joint distribution is called *negligible* compared to those of marginal distributions if and only if

$$\frac{\hat{H}(t,t)}{\bar{F}_1(t) + \bar{F}_2(t)} \to 0 \tag{4.10}$$

as $t \to \infty$. If (4.10) holds, the random variables $\{X_1, X_2\}$ or the joint distribution H are called to satisfy the *negligible joint tail condition.*

The above definition (4.10) is equivalent to the sufficient condition

$$\lim_{x,y\to\infty} \frac{1 - F_X(x) - F_Y(y) + H_{X,Y}(x,y)}{1 - H_{X,Y}(x,y)} = 0.$$

for the asymptotic independence of $\max(X_1, \cdots, X_n)$ and $\max(Y_1, \cdots, Y_n)$, even though X_i and Y_i are not independent (Johnson and Kotz, 1972).

If we assume the independence of X_1 and X_2, the tail probability of the joint distribution H is always negligible since $\hat{H}(x,y) = \bar{F}_1(x)\bar{F}_2(y)$. The family of distributions satisfying the negligible joint tail condition includes many well-known distribution families such as

1. The bivariate normal distribution with $|\rho| < 1$.

2. Bivariate distribution of type

$$H(x_1, x_2) = F_1(x_1)F_2(x_2)(1 + \lambda\bar{F}_1(x_1)\bar{F}_2(x_2)).$$

3. Bivariate exponential distributions of type

$$H(x_1, x_2) = 1 - e^{-x_1} - e^{-x_2} - \exp(-x_1 - x_2 - \theta x_1 x_2)$$

for $x_1, x_2 > 0$, and $0 \leq \theta \leq 1$.

4. The bivariate logistic distribution

$$H(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}.$$

For more details of the examples, see Johnson and Kotz (1972).

Remark 4.3 If X_i are regularly varying with the same extreme value index $\alpha > 0$ such that $\overline{F}_i = x^{-\alpha} L_i(x)$, the negligible joint tail condition in Definition 4.2 is equivalent to the following asymptotic tail condition.

$$\frac{\hat{H}(\delta_1 t, \delta_2 t)}{\bar{F}_1(c_1 t) + \bar{F}_2(c_2 t)} \to 0 \tag{4.11}$$

as $x \to \infty$ for any positive real numbers δ_1, δ_2, c_1, and c_2.

Before we show that two statements (4.8) and (4.9) hold if the joint distribution satisfies the negligible joint tail condition, we give an example of a bivariate distribution with a copula satisfying the negligible joint tail condition in Definition 4.10. Definitions and properties of copulas can be found in many textbooks and papers such as Nelsen (1999).

Example 4.4 Let X_1 and X_2 be the regularly varying random variables not necessarily independent. Suppose that there exists a copula C such that $H(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ and the copula density $c(u, v)$ is bounded by some constant $M > 0$ on $[1 - \delta, 1] \times [1 - \delta, 1]$ for some real $0 \leq \delta < 1$. A simple example is Farlie-Gumbel-Morgenstern copula defined by $C(u, v) = uv + \lambda uv(1 - u)(1 - v)$ for $-1 \leq \lambda \leq 1$. Due to the boundedness of the copula density, the joint tail probability is also bounded by the product of two marginal tail probabilities

up to constant, i.e.

$$\hat{H}(t,t) \;=\; \int_t^\infty\int_t^\infty c\big(F_1(x_1),F_2(x_2)\big)f_1(x_1)f_2(x_2)\;dx_1dx_2 \le M\bar{F}_1(t)\bar{F}_2(t)$$

where M is the upper bound of the copula density. Then

$$\frac{\bar{H}(t,t)}{\bar{F}_1(t)+\bar{F}_2(t)} \;\le\; \frac{M\bar{F}_1(t)\bar{F}_2(t)}{\bar{F}_1(t)+\bar{F}_2(t)} \le M\frac{\epsilon^2}{2\epsilon} = \frac{M}{2}\epsilon$$

which vanishes as $t \to \infty$ since ϵ is arbitrarily small. Therefore the tail probability of the joint distribution H is negligible compared to those of marginal distributions. □

Suppose we have dependent random variables belonging to the same family of regularly varying distributions, $\mathcal{R}_{-\alpha}$ for some $\alpha > 0$. If the tail behavior of the joint distribution is not extraordinary, we may assume the negligible joint tail condition. By Theorem 4.7 and Corollary 4.8 in the following, we prove two statements (4.8) and (4.9) hold in $\mathcal{R}_{-\alpha}$, if the joint distribution satisfies (4.10). Beforehand, we need the following lemma.

Lemma 4.5 Suppose that the random variables X_1 and X_2 are not necessarily independent and the distributions F_1 and F_2 vary regularly with the same extreme value index $\alpha \ge 0$ such that

$$1 - F_i(x) = x^{-\alpha}L_i(x)$$

with L_i slowly varying for each $i = 1, 2$. If the joint distribution satisfies the negligible joint tail condition in (4.10), the distribution of $X_1 + X_2$ denoted by G varies regularly with the same extreme value index α such that

$$1 - G(x) \sim x^{-\alpha}(L_1(x) + L_2(x)) \tag{4.12}$$

as $x \to \infty$.

Proof: We follow Feller with a slight modification which is necessary in case of dependent random variables. Put $t^{'}$=$(1+\delta)t$. Let E be the event of $\{X_1+X_2>t\}$. Then, E will include the set $\{E_1 \cup E_2\}$ where E_1=$\{X_1>t^{'},X_2>-\delta t\}$, and E_2=$\{X_2>t^{'},X_1>-\delta t\}$. Now let E_3=$E_1 \cap E_2$=$\{X_1>t^{'},X_2>t^{'}\}$. Then, for a given $\epsilon > 0$ and $\delta > 0$, there exists a sufficiently large t such that

$$\mathbb{P}(E_1) + \mathbb{P}(E_2) = \hat{H}(t^{'}, -\delta t) + \hat{H}(-\delta t, t^{'}) \geq (1 - \frac{\epsilon}{2})(\bar{F}_1(t^{'}) + \bar{F}_2(t^{'})).$$

Moreover, $\hat{H}(t^{'}, t^{'})$ is bounded by $\epsilon(\bar{F}_1(t^{'}) + \bar{F}_2(t^{'}))$ for a sufficiently large t due to the negligible joint tail probability assumption. Therefore, the tail probability of $X_1 + X_2$ is bounded below as follows.

$$\begin{aligned} 1 - G(t) &= \mathbb{P}(E) \geq \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_3) \\ &= \hat{H}(t^{'}, -\delta t) + \hat{H}(-\delta t, t^{'}) - \hat{H}(t^{'}, t^{'}) \\ &\geq (1 - \frac{\epsilon}{2})(\bar{F}_1(t^{'}) + \bar{F}_2(t^{'})) - \frac{\epsilon}{2}(\bar{F}_1(t^{'}) + \bar{F}_2(t^{'})) \\ &= (1 - \epsilon)(\bar{F}_1(t^{'}) + \bar{F}_2(t^{'})). \end{aligned}$$

Similarly, in order to find the upper bound of $1 - G(t)$, we put $t^{''} = (1 - \delta)t$ with $0 < \delta < \frac{1}{2}$. Then E will be included in the set $D_1 \cup D_2 \cup D_3$ where $D_1 = \{X_1 > t^{''}\}$, $D_2 = \{X_2 > t^{''}\}$, and D_3=$\{X_1>\delta t,X_2>\delta t\}$. Then

$$\begin{aligned} 1 - G(t) &= \mathbb{P}(E) \leq \mathbb{P}(D_1) + \mathbb{P}(D_2) + \mathbb{P}(D_3) \\ &= \bar{F}_1(t^{''}) + \bar{F}_2(t^{''}) + \hat{H}(\delta t, \delta t) \\ &\leq (\bar{F}_1(t^{''}) + \bar{F}_2(t^{''})) + \epsilon(\bar{F}_1(t^{''}) + \bar{F}_2(t^{''})) \\ &= (1 + \epsilon)(\bar{F}_1(t^{''}) + \bar{F}_2(t^{''})) \end{aligned}$$

by the negligible joint tail condition and Remark 4.3. Since ϵ and δ are arbitrarily small, $1-G(t) \sim \bar{F}_1(t) + \bar{F}_2(t) = t^{-\alpha}(L_1(t) + L_2(t))$, which completes the proof. □

The above lemma asserts that the sum of two dependent random variables in the same family of regular variation $\mathcal{R}_{-\alpha}$ also belongs to the same family $\mathcal{R}_{-\alpha}$ if they satisfy the negligible joint tail condition. Lemma 4.5 can be generalized for the aggregate distribution of regularly varying

random variables $X_1, \ldots, X_n$ not necessarily independent. Beforehand we need the generalized definition of the negligible joint tail condition.

Definition 4.6 A sequence of random variables $\{X_1, X_2, \ldots\}$ is called to satisfy the negligible joint tail condition if and only if any pair $\{X_i, X_j\}$ of the sequence satisfies the negligible joint tail condition in Definition 4.2.

The following theorem is an immediate consequence of Lemma 4.5 and Definition 4.6, which provides the sufficient condition for the closure property of the distributions of dependent regularly varying random variables in (4.9).

Theorem 4.7 If $X_1, \ldots, X_n \in \mathcal{R}_{-\alpha}$ and satisfy the negligible joint tail condition in Definition 4.6, the aggregate distribution G of $X_1, \ldots, X_n$ varies regularly such that

$$1 - G(x) \sim x^{-\alpha}(L_1(x) + \cdots + L_n(x))$$

as $x \to \infty$ where $F_i(x) = x^{-\alpha}L_i(x)$ for some slowly varying function $L_i(x)$.

Proof:

It suffices to show the closure property in case $n = 3$. The general case will follow by induction. Let $S_2 = X_1 + X_2$ and $t'' = (1-\delta)t$ for $0 < \delta < 1/2$. The event $\{S_2 > t, X_3 > t\}$ is included by $\{X_1 > t'', X_3 > t\} \cup \{X_1 > \delta t, X_2 > \delta t, X_3 > t\} \cup \{X_2 > t'', X_3 > t\}$, and hence for any $\epsilon > 0$ and t sufficiently large

$$\begin{aligned}
\frac{\overline{H}_{S_2,X_3}(t,t)}{\overline{F}_{S_2}(t) + \overline{F}_3(t)} &\leq \frac{\overline{H}_{1,3}(t'',t) + \overline{H}_{1,2,3}(\delta t, \delta t, t) + \overline{H}_{2,3}(t'',t)}{\overline{F}_1(t) + \overline{F}_2(t) + \overline{F}_3(t)} \\
&\leq \frac{\overline{H}_{1,3}(t'',t)}{\overline{F}_1(t) + \overline{F}_3(t)} + \frac{\overline{H}_{1,2}(\delta t, \delta t)}{\overline{F}_1(t) + \overline{F}_2(t)} + \frac{\overline{H}_{2,3}(t'',t)}{\overline{F}_2(t) + \overline{F}_3(t)} \\
&< \epsilon
\end{aligned}$$

by Remark 4.3. Since δ, ϵ are arbitrarily small, $\{S_2, X_3\}$ satisfies the negligible joint tail condition with the joint distribution H_{S_2,X_3}. By Theorem 4.5,

$$\begin{aligned} 1-G(x) &\sim \overline{F}_{S_2}(x)+\overline{F}_{X_3}(x) \\ &\sim x^{-\alpha}\big(L_1(x)+L_2(x)+L_3(x)\big) \end{aligned}$$

which completes the proof. □

Now we prove the max-sum equivalence of dependent regularly varying random varialbes in (4.8) in the following corollary.

Corollary 4.8 If $X_1,\ldots,X_n \in \mathcal{R}_{-\alpha}$ and satisfy the negligible joint tail condition, the followings are true.

1. Max-sum equivalence: $\mathbb{P}(\max(X_1,\ldots,X_n)>x) \sim \mathbb{P}(X_1+\cdots+X_n>x)$,
2. Max-stability: $\mathbb{P}(\max(X_1,\ldots,X_n)\le x) \in \mathcal{R}_{-\alpha}$.

Proof:

Let $M_n=\max(X_1,\ldots,X_n)$ and $\hat{H}$ be the joint tail probability as in Definition 4.2. The proof is by induction on n. For $n=2$, by Lemma 4.5,

$$\begin{aligned} \mathbb{P}(M_2>x) &= \overline{F}_{X_1}(x)+\overline{F}_{X_2}(x)-\hat{H}_{X_1,X_2}(x,x) \\ &= \big(\overline{F}_{X_1}(x)+\overline{F}_{X_2}(x)\big)\Big(1-\frac{\hat{H}_{X_1,X_2}(x,x)}{\overline{F}_{X_1}(x)+\overline{F}_{X_2}(x)}\Big) \\ &\sim \overline{F}_{X_1}(x)+\overline{F}_{X_2}(x) \\ &\sim \mathbb{P}(X_1+X_2>x) \end{aligned}$$

as $x\to\infty$ and hence $\mathbb{P}(M_2\le x)\in\mathcal{R}_{-\alpha}$. Suppose the above asymptotic relation holds for $k\le n-1$, then $\{M_{n-1},X_n\}$ satisfies the negligible joint tail condition since $\mathbb{P}(M_{n-1}>x)\sim \mathbb{P}(X_1+\cdots+X_{n-1}>x)$ and

$$\begin{aligned} \mathbb{P}(M_n>x) &= \overline{F}_{M_{n-1}}(x)+\overline{F}_{X_n}(x)-\hat{H}_{M_{n-1},X_n}(x,x) \\ &= \big(\overline{F}_{M_{n-1}}(x)+\overline{F}_{X_n}(x)\big)\Big(1-\frac{\hat{H}_{M_{n-1},X_n}(x,x)}{\overline{F}_{M_{n-1}}(x)+\overline{F}_{X_n}(x)}\Big) \\ &\sim \overline{F}_{M_{n-1}}(x)+\overline{F}_{X_n}(x) \sim \overline{F}_{X_1}(x)+\cdots+\overline{F}_{X_n}(x) \\ &\sim \mathbb{P}(X_1+\cdots+X_n>x) \end{aligned}$$

and hence $\mathbb{P}(M_n \leq x) \in \mathcal{R}_{-\alpha}$, which proves both of the assertions. $\square$

4.2 Value at Risk of Regularly Varying Random Variables

In this section we first consider a simple risk process of the total aggregate loss $Y = Y_1 + \cdots + Y_d$ for a positive integer d where the regularly varying Y_i are not necessarily independent. We apply Theorem 4.7 to classify $\mathcal{R}_{-\alpha}$ into three categories, asymptotic superadditivity, asymptotic comonotonic, and asymptotic subadditivity corresponding to the extreme value index $\alpha > 0$. The individual aggregate loss $Y_i(t)_{t\geq 0}$ is also considered as a compound type generated by a general counting process $N_i(t)$ with finite means and we present the analogous result of the asymptotic properties of the value-at-risk.

4.2.1 Value at Risk of the Total Aggregate Loss

Simply speaking, the value-at-risk of a random variable at p confidence level is defined by the p-quantile of its distribution, which can be interpreted as a probable maximum loss that an institution may experience for a given period of time. Since a distribution function of a continuous random variable is generally assumed to be right-continuous, the value-at-risk can be formally defined by the generalized inverse of the distribution as follows.

Definition 4.9 Let F be the distribution function of a random variable X, then the value-at-risk of X at p level of confidence for $0 < p < 1$ is defined by

$$VaR_p(X) = \inf\{x \in \mathbb{R} | F(x) \geq p\}.$$

The subadditivity of value-at-risk doesn't hold in general . There is a partial result in the family of multivariate regularly varying distributions when the extreme value index $\alpha > 1$ (Danielsson et al., 2005). However, multivariate regularly varying distributions are different in nature from arbitrary multivariate distributions with regularly varying marginal distributions. Moreover, the extremely heavy-tailed regularly varying distributions have extreme value indices such as $0 < \alpha < 1$, in which interval of α, we don't have much knowledge of the subadditivity. In the following two

theorem we discuss the asymptotic properties of the value-at-risk of regularly varying random variables under the assumption of the negligible joint tail probability. As a result the family of the regularly varying distributions are classified into three subfamilies corresponding to the extreme value indices $\alpha > 0$.

Theorem 4.10 Let $X_1, \ldots, X_n$ be identical but not necessarily independent random variables the distributions of which vary regularly such that $F_i(x) = x^{-\alpha}L(x)$ for some slowly varying function $L(x)$ and $\alpha > 0$. Further we assume that the distribution functions of X_i are continuous almost everywhere for all $i = 1, \ldots, n$. If $X_1, \ldots, X_n$ satisfy the negligible joint tail condition, there exists a real $p_0 \in (0, 1)$ such that for all positive $p \in [p_0, 1]$,

i) (Subadditivity) For $\alpha > 1$

$$\text{VaR}_p(X_1 + \cdots + X_n) \leq \text{VaR}_p(X_1) + \cdots + \text{VaR}_p(X_n). \tag{4.13}$$

ii) (Asymptotic Comonotonicity) For $\alpha = 1$

$$\text{VaR}_p(X_1 + \cdots + X_n) \sim \text{VaR}_p(X_1) + \cdots + \text{VaR}_p(X_n) \tag{4.14}$$

where $\sim$ indicates the asymptotic equivalence as $p \to 1-$.

iii) (Superadditivity) For $0 < \alpha < 1$

$$\text{VaR}_p(X_1 + \cdots + X_n) \geq \text{VaR}_p(X_1) + \cdots + \text{VaR}_p(X_n). \tag{4.15}$$

Proof:

Let $S_n = X_1 + \cdots + X_n$ and write $v_1 = \text{VaR}_p(X_1)$ and $v_n = \text{VaR}_p(S_n)$ for a fixed positive integer n. Because we assume the distribution function is continuous almost everywhere, we may assume $\mathbb{P}(X_1 \leq v_1) = \mathbb{P}(S_n \leq v_n) = p$ for the following without loss of generality. By Theorem 4.7, whether X_i are independent or not, $1 - p = \mathbb{P}(S_n > v_n) \sim n\mathbb{P}(X_1 > v_n)$ as $p \to 1-$. Then for any given $\epsilon > 0$, there exists a $p_0 \in (0, 1)$ close enough to 1 such that for all $p \in (p_0, 1)$,

$$(1 - \epsilon)\frac{1 - p}{n} < \mathbb{P}(X_1 > v_n) < (1 + \epsilon)\frac{1 - p}{n}. \tag{4.16}$$

Similarly, since $\bar{F}_1$ is regularly varying, $\mathbb{P}(X_1 > nv_1) \sim n^{-\alpha}\mathbb{P}(X_1 > v_1)$ and

$$(1-\epsilon)n^{-\alpha}\mathbb{P}(X_1 > v_1) < \mathbb{P}(X_1 > nv_1) < (1+\epsilon)n^{-\alpha}\mathbb{P}(X_1 > v_1). \tag{4.17}$$

Subtracting (4.16) from (4.17), we have

$$(1-p)(\frac{1-\epsilon}{n^\alpha} - \frac{1+\epsilon}{n}) < \bar{F}_1(nv_1) - \bar{F}_1(v_n) < (1-p)(\frac{1+\epsilon}{n^\alpha} - \frac{1-\epsilon}{n}). \tag{4.18}$$

For $0 < \alpha < 1$, as p approach to 1, ϵ gets arbitrary small and the lower bound of (4.18) eventually becomes positive, thus there exists $p_0 \in (0,1)$ such that $0 < \bar{F}_1(nv_1) - \bar{F}_1(v_n)$ for all $p \in (p_0, 1)$. Since $\bar{F}_1(x)$ is a monotone decreasing function, we conclude $v_n \geq nv_1$ which proves the last assertion (4.15) and the other cases, (4.13) and (4.14), will follow immediately in the same manner. □

The above theorem can be used to approximate the minimum capital requirement of the total aggregate loss $Y = Y_1 + \cdots + Y_d$ for $0 < \alpha < 1$ and the upper bound of the capital requirement for $\alpha > 1$ by $d\,\text{VaR}_p(Y_1)$ for $0 < p < 1$ sufficiently close to 1.

4.2.2 Value at Risk of the Individual Aggregate Loss

For the risk modeling of individual aggregate loss $Y_i(t)$, it is more desirable to apply a continuous-time risk process with appropriate counting process as in (4.2). Suppose that the individual aggregate loss $Y_i(t)$ for each i is of compound type generated by $N_i(t)$, the number of losses up to time $t > 0$, independent of each $Y_i(t)$ such that

$$Y_i(t) = \sum_{k=1}^{N_i(t)} Y_k^{(i)}. \tag{4.19}$$

It is not clear that the risk process of compound distribution of (4.19) is subexponential or regularly varying when the random variables are identically subexponetial or regularly varying but not necessarily independent. Embrechts et al. (1997) showed that the compound distribution of $i.i.d$ subexponential random variables is again subexponential under quite general counting processes, i.e. if $F_{Y_k^{(i)}}$ is subexponetial for each k, then the aggregate distribution of the compound type of

(4.19),

$$F_{Y_i(t)}(x) = \sum_{k=0}^{\infty} q_k (F_{Y_k^{(i)}})^{k\star}(x) \quad x \geq 0 \tag{4.20}$$

is again subexponential when $q_k = \mathbb{P}(N_i(t) = k)$ and $\sum_{k=0}^{\infty} q_k(1+\epsilon)^k < \infty$ for some $\epsilon > 0$. In the following theorem, we show an analogous result of Theorem 4.10 with respect to the individual aggregate loss $Y_i(t)$ for any fixed $t > 0$.

Theorem 4.11 Let $\{X_1, X_2, \ldots\}$ be a sequence of identical but not necessarily independent random variables such that $F_i(x) = x^{-\alpha}L(x)$ for some slowly varying function $L(x)$ and $\alpha > 0$. Further we assume that the distribution functions of X_i are continuous almost everywhere for all $i = 1, 2, \ldots$. Let $N_t = N(t)$ be a counting process such that $E[N_t] < \infty$ for $t > 0$. If the sequence $\{X_1, X_2, \ldots\}$ satisfies the negligible joint tail condition, there exists a real $p_0 \in (0, 1)$ such that for all positive $p \in [p_0, 1]$,

i) (Subadditivity) For $\alpha > 1$

$$\text{VaR}_p\big(\sum_{i=1}^{N_t} X_i\big) \leq E[N_t]\text{VaR}_p(X_1). \tag{4.21}$$

ii) (Asymptotic Comonotonicity) For $\alpha = 1$

$$\text{VaR}_p\big(\sum_{i=1}^{N_t} X_i\big) \sim E[N_t]\text{VaR}_p(X_1) \tag{4.22}$$

where $\sim$ indicates the asymptotic equivalence as $p \to 1-$.

iii) (Superadditivity) For $0 < \alpha < 1$

$$\text{VaR}_p\big(\sum_{i=1}^{N_t} X_i\big) \geq E[N_t]\text{VaR}_p(X_1). \tag{4.23}$$

Proof:

We prove (4.23) only and the other cases, (4.21) and (4.22), will follow in the same manner. Suppose $0 < \alpha < 1$ and write $v_1 = \text{VaR}_p(X_1)$ and $v_{N_t} = \text{VaR}_p(\sum_{i=1}^{N_t} X_i)$. Then

$$v_{N_t} = \Big\{w \big| \mathbb{P}(\sum_{i=1}^{N_t} X_i \leq w) = p\Big\} = \Big\{w \big| \lim_{n\to\infty} \sum_{k=0}^{n} \big[\mathbb{P}(N_t = k)\mathbb{P}(\sum_{i=1}^{k} X_i \leq w)\big] = p\Big\}. \tag{4.24}$$

By Theorem 4.10, $\mathbb{P}(\sum_{i=1}^{k} X_i \leq w) \sim 1 - k\mathbb{P}(X_1 > w)$ as $p \to 1-$, and hence

$$\sum_{k=0}^{n} \big[\mathbb{P}(N_t = k)\mathbb{P}(\sum_{i=1}^{k} X_i \leq w)\big] \sim 1 - \sum_{k=0}^{n} \big[k\mathbb{P}(N_t = k)\big]\mathbb{P}(X_1 > w) \tag{4.25}$$

as and $p \to 1-$. Combining (4.24) and (4.25), we have

$$v_{N_t} \sim \Big\{w \Big| 1 - E[N_t]\mathbb{P}(X_1 > w) = p\Big\}$$

as $p \to 1-$ and $n \to \infty$. Therefore we can write

$$\mathbb{P}(X_1 > v_{N_t}) \sim (1-p)/E[N_t]$$

as $p \to 1-$. Since $\mathbb{P}(X_1 > E[N_t]v_1) \sim E[N_t]^{-\alpha}\mathbb{P}(X_1 > v_1)$, in the same manner as in the proof of Theorem 4.10, for any given $\epsilon > 0$, there exists a real $p_0 \in (0,1)$ such that

$$(1-p)(\frac{1-\epsilon}{E[N_t]^{\alpha}} - \frac{1+\epsilon}{E[N_t]}) < \bar{F}_1(E[N_t]v_1) - \bar{F}_1(v_{N_t}).$$

for all $p \in (p_0, 1)$ where the lower bound becomes positive eventually and hence $0 < \bar{F}_1(E[N_t]v_1) - \bar{F}_1(v_{N_t})$ for all $p \in (p_0, 1)$. We conclude $v_{N_t} \geq E[N_t]v_1$ which completes the proof. □

4.3 Numerical Examples

Let us consider the following total aggregate loss $Y = Y_1 + Y_2 + Y_3$, the sum of three individual aggregate losses. We assume that the marginal distributions of individual aggregate losses are identically Pareto(α, β) distributed such that

$$1 - F_{Y_i}(y) = \Big(\frac{\beta}{y+\beta}\Big)^{\alpha}$$

for $\alpha, \beta > 0$ and $y \geq 0$. It is immediate that Pareto distributions with the shape parameter $\alpha > 0$ belong to the family of regularly varying distributions $\mathcal{R}_{-\alpha}$. For $\alpha > 1$ and a positive integer $k > \alpha$, k-th moments of Pareto(α, β) are as follows,

$$E[Y^k] = \frac{\beta^k k}{(\alpha-1)\cdots(\alpha-k)}.$$

For $0 < \alpha \leq 1$, however, no finite moments exist. Pareto distribution with $0 < \alpha < 1$ is well known as an extremely heavy-tailed loss distribution in insurance and operational risk. The parameter α determines the maximum domain of attraction and it is one of the most important indices in the family of extreme value distributions. For more about the extreme value theory, see Embrechts et al. (1997).

For the dependence structure, we assume Markov property among Y_i for the computational convenience and choose the bivariate Gaussian copula with correlation coefficient ρ for each subsequent pair of Y_{i-1} and Y_i for $i = 2, 3$, defined by

$$C(u_{i-1}, u_i) = \Phi_\rho\big(\Phi^{-1}(u_{i-1}), \Phi^{-1}(u_i)\big)$$

where Φ_ρ and Φ^{-1} are the multivariate standard normal distribution and the inverse of the standard normal distribution, respectively, with $u_i = F_{Y_i}(y_i)$ for $i = 1, 2, 3$.

Fig. 4.1 and Table 4.1 are the summary of the numerical value-at-risks when $\alpha = 4.0, \beta = 1.0$ and the correlation coefficient $\rho = 0.7$ and -0.7. Comparing two graphs in Fig. 4.1, we can see clear evidence of stronger diversification effect in the second graph due to the negative dependence between subsequent losses. The negative dependence plays its role in such a way that large(or small) losses are not likely to happen subsequently, which makes the aggregate distribution more centered in the middle and less dispersed and explains the larger discrepancy between the curves of the second graph.

The second numerical example in Fig. 4.2 and Table 4.2 is similar to the previous one except different Pareto parameters $\alpha = 0.9, \beta = 0.2$. As it was mentioned above, no finite moments exist and the marginal distributions become extremely heavy-tailed. In this case, we observe that the superadditivity exists on the entire domain. The dependence effect on diversification cannot exceed the heavy-tailedness effect of the Pareto distribution with $\alpha < 1$ under Gaussian copula dependence.

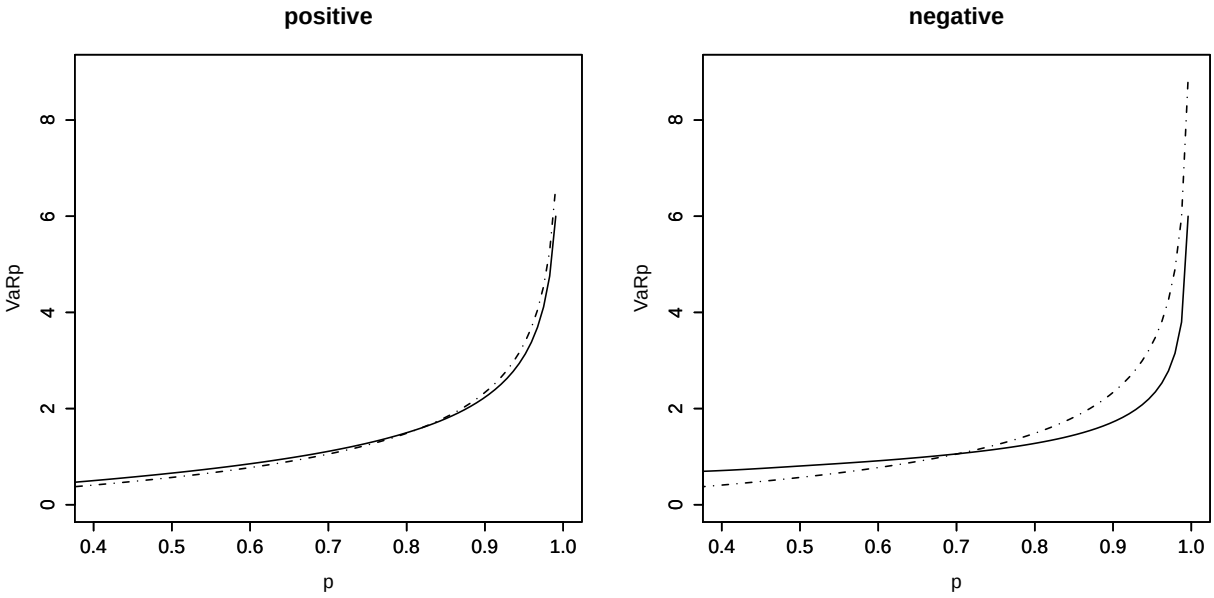


Figure 4.1 Value at risk of the total aggregate loss (solid line) vs. the sum of the value-at-risks of the individual aggregate losses (dashed line). Left: $\rho = 0.7$, $\alpha = 4.0, \beta = 1.0$. Right: $\rho = -0.7$, $\alpha = 4.0, \beta = 1.0$.

$\alpha = 4.0$	$\beta = 1.0$	$\rho = 0.7$	$\alpha = 4.0$	$\beta = 1.0$	$\rho = -0.7$
p	$VaR_p(S_3)$	$3VaR_p(Y_1)$	p	$VaR_p(S_3)$	$3VaR_p(Y_1)$
0.524	0.700	0.612	0.384	0.700	0.387
0.621	0.900	0.824	0.589	0.900	0.746
0.697	1.100	1.042	0.724	1.100	1.139
0.755	1.300	1.264	0.808	1.300	1.534
0.819	1.600	1.602	0.881	1.600	2.109
0.865	1.900	1.947	0.921	1.900	2.667
0.920	2.500	2.646	0.961	2.500	3.755

Table 4.1 Value-at-risk of the total aggregate loss vs. the sum of the value-at-risks of the individual aggregate losses with $\alpha = 4.0, \beta = 1.0$ and $\rho = \pm 0.7$.

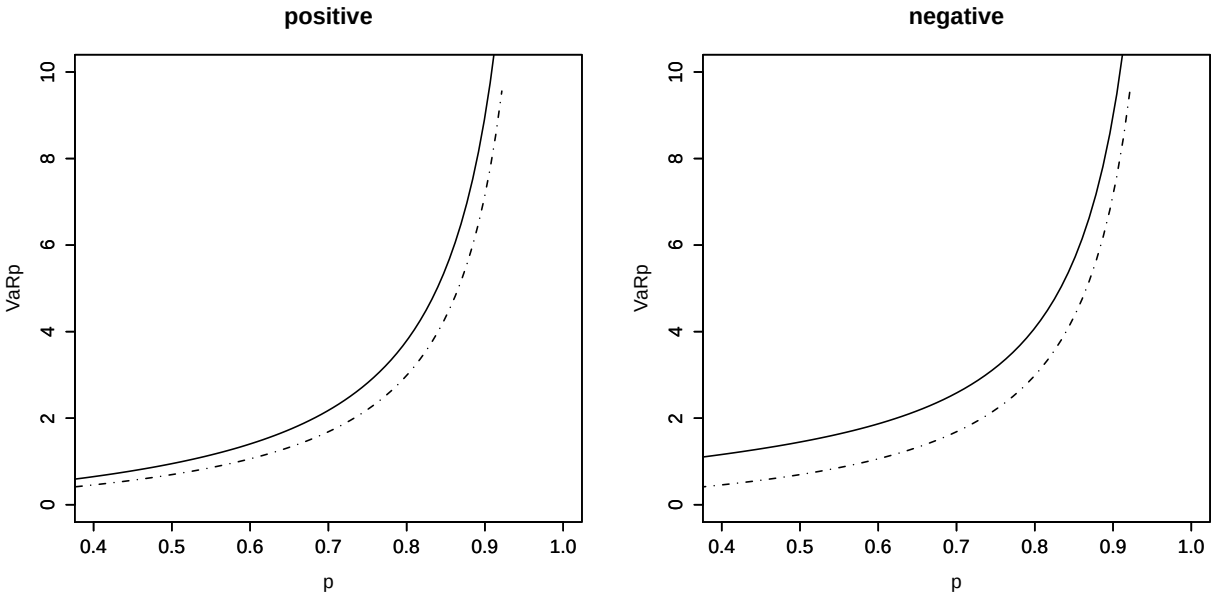


Figure 4.2 Value at risk of the total aggregate loss (solid line) vs. the sum of the value-at-risks of the individual aggregate losses (dashed line). Left: $\rho = 0.7$, $\alpha = 0.9, \beta = 0.2$. Right: $\rho = -0.7$, $\alpha = 0.9, \beta = 0.2$.

$\alpha = 0.9$	$\beta = 0.2$	$\rho = 0.7$	$\alpha = 0.9$	$\beta = 0.2$	$\rho = -0.7$
p	$VaR_p(S_3)$	$3VaR_p(Y_1)$	p	$VaR_p(S_3)$	$3VaR_p(Y_1)$
0.455	0.800	0.578	0.225	0.800	0.197
0.561	1.200	0.897	0.415	1.200	0.490
0.682	2.000	1.541	0.623	2.000	1.175
0.749	2.800	2.187	0.721	2.800	1.877
0.792	3.600	2.834	0.777	3.600	2.572
0.844	5.200	4.126	0.838	5.200	3.946
0.906	9.600	7.654	0.906	9.600	7.677

Table 4.2 Value-at-risk of the total aggregate loss vs. the sum of the value-at-risks of the individual aggregate losses with $\alpha = 0.9, \beta = 0.2$ and $\rho = \pm 0.7$.

4.4 Conclusions

Assuming the existence of the diversification of risks in practice, we have taken it for granted that the subadditivity of the value-at-risk holds. However, if risks are extremely heavy-tailed, it is essential to find the lower bound of a given risk measure. Considering dependent loss random variables with regularly varying tails, we present a sufficient condition of the joint distribution and generalize the convolution theorem of regular variations. Applying the result, we classify the family of regularly varying distributions into three subfamilies such as regularly varying distributions with asymptotic subadditivity, comonotonocity, and superadditivity which correspond to the extreme value indexes, $\alpha > 1$, $\alpha = 1$, and $0 < \alpha < 1$, respectively. In the numerical examples, we observe that heavy-tailedness of the marginal distributions has much larger influence on the value-at-risk measure than the dependence among losses above certain confidence level.

‘

DISCARD THIS PAGE

Appendix A: Extreme Value Distributions

For an insurance risk X defined as a non-negative loss random variable, a sound modeling of the risk or premium calculation has been the major goal of actuaries. Especially in property-casualty insurance, pricing large claims or catastrophe losses are gaining attention because of rapid increase of gross claim amount in recent market. Reinsurers also have sought for mathematical or empirical models of loss distributions with good fits of extremal events. Premiums should be high enough to cover the insured if claims occur so that the company stays solvent with certain confidence level even in case of catastrophe. Ruin problem in risk theory is one of the mathematical solutions to this question. For a long time extreme value theory has belonged to the standard tool kits for reliability engineers. Although it has been widely used in hydrology and climatology rather than in insurance, it has been recently noted that the extreme value theory could be one solution to the extreme events in insurance and operational risks.

A.1 Fisher-Tippet Theorem

Given a sequence of $i.i.d$ random variables $X_1, \cdots, X_n$ with common distribution F, denote the maximum of the random variables by $M_n = \max(X_1, \cdots, X_n)$. The fundamental *Fisher-Tippet theorem* classifies the possible limit laws of the maxima, M_n, as $n \to \infty$, introducing appropriate normalizing constants a_n and b_n. If there exist sequences of normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$ and a non-degenerating distribution function H such that

$$\frac{M_n - b_n}{a_n} \longrightarrow H \tag{A.1}$$

in distributions, then H belongs to one of the types as follows.

$$\begin{cases} \text{Fréchet} \quad \Phi_\alpha(x) & = & \begin{cases} 0, & \text{if } x \leq 0 \\ \exp\{-x^{-\alpha}\}, & \text{if } x > 0 \end{cases} \\ \text{Weibull} \quad \Psi_\alpha(x) & = & \begin{cases} \exp\{-(-x)^{\alpha}\}, & \text{if } x \leq 0 \\ 1, & \text{if } x > 0 \end{cases} \\ \text{Gumbel} \quad \Lambda(x) & = & \exp\{-e^{-x}\}, \quad \text{if } x \in \mathbb{R} \end{cases} \tag{A.2}$$

The distribution F of a random variable X is said to belong to the maximum domain of attraction of the extreme value distribution H and denoted by $F \in \text{MDA}(H)$ if (A.1) is satisfied.

A.2 The Generalized Extreme Value Distributions

Extreme value distributions can be written as one parameter functions, so called *Jenkinson-von Mises representation* which covers the three types of limiting distributions in (A.2). It is widely accepted and called the (standard) generalized extreme value distribution and defined by

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\}, & \text{if } \xi \neq 0 \\ \exp\{-\exp(-x)\}, & \text{if } \xi = 0 \end{cases} \tag{A.3}$$

where $1 + \xi x > 0$ and ξ is called the extreme value index or tail index. The one-parameter functional form classifies the extreme value distributions with respect to the shape parameter as follows.

Definition A.1 (Maximum Domain of Attractions) A distribution F belongs to a maximum domain of attraction of H_ξ, denoted by $F \in MDA(H_\xi)$, if and only if there exist sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq a_n x + b_n) = H_\xi(x) \tag{A.4}$$

where H_ξ is the generalized extreme value distribution with shape parameter ξ as in (A.3). The real sequences a_n and b_n are called normalizing constants for the maxima.

A.3 Approximation of the Extreme Value Distributions

The (standard) generalized extreme value distribution has a close link to the (standard) generalized Pareto distribution which is defined by

$$1 - G_\xi(x) = \begin{cases} (1+\xi x)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ e^{-x}, & \text{if } \xi = 0 \end{cases} \tag{A.5}$$

for $x \geq 0$ if $\xi \geq 0$ and $0 \leq x \leq 1/\xi$ if $\xi < 0$. Introducing the scale and location parameters, β and σ, the generalized Pareto distribution can be written as $G_{\xi;\beta,\sigma}(x)$ by simple substitution of x with $(x-\sigma)/\beta$. Let us define the distribution of excesses over threshold u by

$$F_u(x) = \mathbb{P}(X - u \leq x \,|\, X > u) = \frac{F(x+u) - F(u)}{\overline{F}(u)}.$$

The distribution of scaled excesses over the high threshold u can be approximated by the generalized Pareto distribution which has the same extreme value index ξ of the underlying distribution F due to the following *Balkema-de Haan-Pickands* theorem.

Theorem A.2 For every $\xi \in \mathbb{R}$, $F \in MDA(H_\xi)$ if and only if

$$\lim_{u \to x_F} \sup_{0 < x < x_F - u} \left| G_{\xi,\beta(u)}(x) - F_u(x) \right| = 0$$

where x_F implies the right-end point of F, and G_ξ is the generalized Pareto distribution with shape and scale parameters, ξ and some positive function $\beta(u)$ such that

$$G(x;\xi,\beta(u)) = \begin{cases} 1 - (1 + \frac{\xi x}{\beta(u)})^{-1/\xi}, & \text{if} \quad \xi \neq 0 \\ 1 - \exp(-x/\beta(u)), & \text{if} \quad \xi = 0 \end{cases} \tag{A.6}$$

for $x \geq 0$ if $\xi \geq 0$ and $0 \leq x \leq -\beta(u)/\xi$ if $\xi < 0$.

Note that β is a function of the threshold u. For more details, see Embrechts et al. (1997); Balkema and de Haan (1974); Pickands (1975). Theorem A.2 provides us a theoretical justification to use the generalized Pareto distribution for the tail approximation of the underlying loss distribution above high enough threshold.

Appendix B: Estimation of the Extreme Value Distributions

We introduce a maximum likelihood estimator of the threshold level over which the tail of loss distribution can be approximated by an appropriate extreme value distribution belonging to certain maximum domain of attraction.

B.1 Shape Parameter Estimation of the Extreme Value Distributions

Let $X_1, \ldots, X_n$ be $i.i.d.$ loss random variables whose common distribution is $F_X(x) = \mathbb{P}(X \leq x)$ for $x \geq 0$. Consider an excess of loss insurance with retention M and limit L, $M < L$. The claim amount Y covered by reinsurer given that a loss X occurs is given by

$$Y = \begin{cases} 0, & 0 \leq X < M \\ (X - M), & M \leq X < M + L \\ L, & M + L \leq X \end{cases} \tag{B.1}$$

The problem of retention level in excess of loss insurance is as important as that of pricing. They are of great relevance when we price high excess of loss layer or choose optimal retention level for insurer or reinsurer. The family of extreme value distributions proves important in the study of the limiting behavior of sample extrema, $\max(X_1, \ldots, X_n)$, as the normal distributions are essential in the study of the limiting distributions for sample sum S_n or sample mean $\overline{X}$.

B.2 Threshold Estimation of the Mixture Distribution

The proposed model by Behrens et al. (2004) assumes that the observations below a threshold u come from a certain distribution with parameters η, denoted by $F_1(x|\eta)$, where those above the threshold come from a generalized Pareto distribution $G_{\xi,\beta(u)}(x)$ which can be justified by Theorem A.2. The shape parameter ξ is determined by the maximum domain of attraction of the underlying distribution such that $F_1 \in MDA(H_\xi)$ or should be estimated together with the threshold u. The layer(conditional layer) mixture distribution $F^{(2)}$ of F_1 and $G_{\xi,\beta(u)}$ introduced in

Chapter 2 can be written as

$$F^{(2)}(x|\eta,\xi,\beta(u)) \quad = \quad \begin{cases} F_1(x|\eta) & x < u \\ F_1(u|\eta) + \overline{F}_1(u|\eta)\, G_{\xi,\beta(u)}(x-u) & x \geq u. \end{cases} \tag{B.2}$$

If we denote the parameters of our concern by $\theta = (\eta, \xi, \beta(u))$ and rearrange the sample observations $\{X_1, \ldots, X_n\}$ in increasing order by $\{X_{n,n} \leq \cdots \leq X_{1,n}\}$, the likelihood function can be written as follows.

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=k+1}^{n} f_1(x_{i,n}|\eta) \prod_{i=1}^{k} \overline{F}_1(u|\eta)\, g_{\xi,\beta(u)}(x_{i,n}) \tag{B.3}$$

where f_1 and $g_{\xi,\beta(u)}$ are the density functions of F_1 and the generalized Pareto distribution, respectively and $x_{k-1,n} < u \leq x_{k,n}$.

Behrens et al. (2004) referred to the elicitation of information within a parameterization on which experts in that field are familiar with, the basic idea of which can be found in Coles and Tawn (1996). Because analysis data in extreme values are usually sparse, information from experts in specific field can play an important role to supplement the inferential information from the data. In this manner, the authors used the experts information for the key parameters of prior distributions above the threshold. When we have such reliable information at hand, it increase the accuracy of the estimation by and large. If we had a prior distribution which can describe the true behavior of the threshold, it would be the best choice of all. However, there is no way of statistical testing that the elicitation of prior information is acceptable or not. In order to avoid assuming specific parameter values of the prior distribution, we present a Bayesian method of the threshold estimation conditioning on the number of exceedances in the following section.

B.3 Maximum Likelihood Estimation of the Threshold Conditioning on the Number of Exceedances

According to Theorem A.2, if a distribution $F \in MDA(\Phi_{1/\xi})$ for $\xi > 0$, the relation between the scaled excess over the threshold of the underlying distribution F and the generalized Pareto distribution G is as follows.

$$\lim_{u \to x_F} \mathbb{P}(\frac{X-u}{\beta(u)} > x | X > u) = (1+\xi x)^{-\frac{1}{\xi}} \tag{B.4}$$

where the function $\beta(u)$ is determined as an integrand of slowly varying function of $\overline{F}$ as follows. Note that F is regularly varying, denoted by $F \in \mathcal{R}_{-1/\xi}$, and we write $\overline{F}(x) = x^{-1/\xi}L(x)$ for some slowly varying function L. By the representation theorem for regularly varying functions (Embrechts et al., 1997; Beirlant et al., 2004), if $F \in \mathcal{R}_{-1/\xi}$, we have

$$\overline{F} = c(x) \exp\{-\int_z^x \frac{1}{\beta(t)}\} \tag{B.5}$$

where $c(t) \to c > 0$ and $\beta(t)/t \to \infty$ as $t \to \infty$. For more details, see Theorem 3.4.5 and following remarks in Embrechts et al. (1997). For example, if F is a Pareto distribution such that $1-F(x) = (1+x)^{-1/\xi}$, the scale parameter function $\beta(t) = \xi(1+t)$. Therefore the approximation of the tail of F above the threshold u is written as folows.

$$1 - F_u(x) \approx \overline{G}_{\xi,\beta(u)}(x) = (1 + \frac{\xi x}{\beta(u)})^{-1/\xi} = (1 + \frac{x}{1+u})^{-1/\xi}. \tag{B.6}$$

B.3.1 Maximum Likelihood Estimator of the Number of Exceedances

Let X_i be $i.i.d.$ loss random variables and denote the common distribution of X_i by F_1. If the sample observations, $X_1, \cdots, X_n$, are given, the likelihood function is defined as a product of the densities of the underlying distribution and the densities of the generalized Pareto distribution

conditioning on K, the number of exceedances above the threshold. The sample observation $X_{k,n}$ can be chosen as a threshold conditional on $K = k$. Therefore, conditioning on the threshold level is equivalent to conditioning on the number of exceedances.

This likelihood function also takes a different form from that of Hill's estimator. Since the latter is based on the K upper observations only, i.e. $X_{1,n}, \cdots, X_{k,n}$, it includes no information from the rest of the sample observation below $X_{k,n}$. The sample observations below the threshold can be included in the likelihood function (B.3) unless the threshold is random. Therefore, conditioning on K, we can write the likelihood function of the mixture distribution for all the sample observations $X_1, \ldots, X_n$ such that

$$L(k, \hat{\xi}_{n,k} | x_1, \cdots, x_n) = \prod_{i=k+1}^{n} f_1(x_{i,n}) \prod_{i=1}^{k} \bar{F}_1(x_{k+1,n})\, g_{\hat{\xi},\hat{\beta}(u)}(x_{i,n} - u) \tag{B.7}$$

where f_1 is the density of F_1 and $g_{\hat{\xi},\hat{\beta}(u)}$ is the density of the generalized Pareto distribution with the shape parameter estimate $\hat{\xi}_{n,k}$ by Hill's estimator. As long as Hill's estimator exists, the likelihood function is well defined and attains its maxima on the domain of K.

We assume that K, the number of exceedances above the threshold is uniformly distributed such that $K \sim$ discrete $U(0, t)$ for an integer $t > 0$ and is a function of the threshold level u, which is unknown. Using the likelihood function in (B.7) we can calculate the conditional probability function $f_{K|t}(k|t)$ and the conditional expectation $E[K|t]$. Since the functional relationship between the number of exceedances and the threshold is not one-to-one correspondence, we estimate K by $\hat{k}(t) = E[K|t]$ and approximate the corresponding threshold level $\hat{u}(t)$.

Once we are given a sample data, it is natural to assume that the domain of the uniform distribution of K is bounded by the sample size. For example, we can choose a discrete uniform distribution $U(0, t)$ where t represents for the possible maximum number of exceedances over the threshold. We can make use of the elicitation of experts prior information at this point or deduce t from the data analysis of the sample. We assume that t varies from 0 to $t_{\max}$ where $t_{\max}$ is large enough to cover realistic maximum number of exceedances. Then we can calculate the maximum likelihood estimator of the number of exceedances, $\hat{k}(t) = E[K|t]$ for each integer

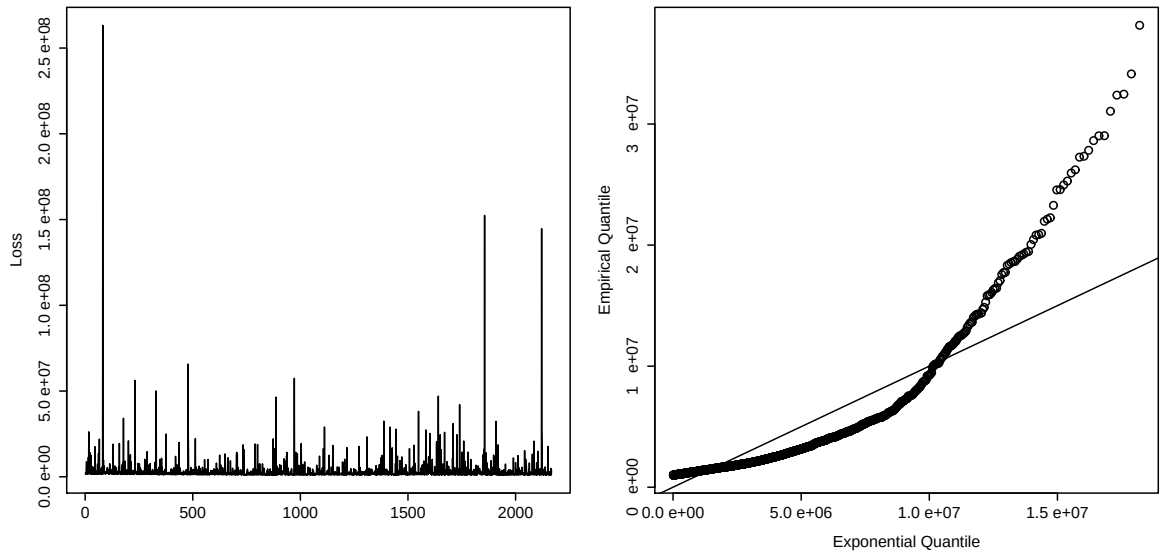


Figure B.1 Danish Fire Loss Data

$t \in [0, t_{\max}]$ by (B.7). We select an appropriate $t^\star \in (0, t_{\max})$ and estimate the threshold level u by $\hat{u}(t^\star) = X_{\hat{k}(t^\star),n}$, the sample observation corresponding to $\hat{k}(t^\star)$.

B.3.2 Numerical Example

The data used in this example is *Danish fire loss data* with losses over one million Danish Krone (R Development Core Team, 2005). The median and the mean are 1.778 and 3.385 respectively. The number of claims over 20 million Danish Krone and 10 million are respectively 36 and 109, which are 1.66% and 5.03%. The exponential quantile plot in Fig. 5.1 shows an upward trend above the straight line, which is a strong evidence of heavy tailed distribution of the data. The left graph in Fig. 5.2 is the empirical mean excess plot $\{(u, e_n(u))\}$. It follows a straight line with positive slope reasonably, which is also an indication of heavy-tailed behavior. Another useful graphical analysis is the Hill Plot. The right graph in Fig. 5.2 is the plot of $(k, \xi_{k,n})$ where $\xi_{k,n}$ is the Hill's estimator with k upper order statistics. Note that the Hill plot is stable approximately when $k > 200$.

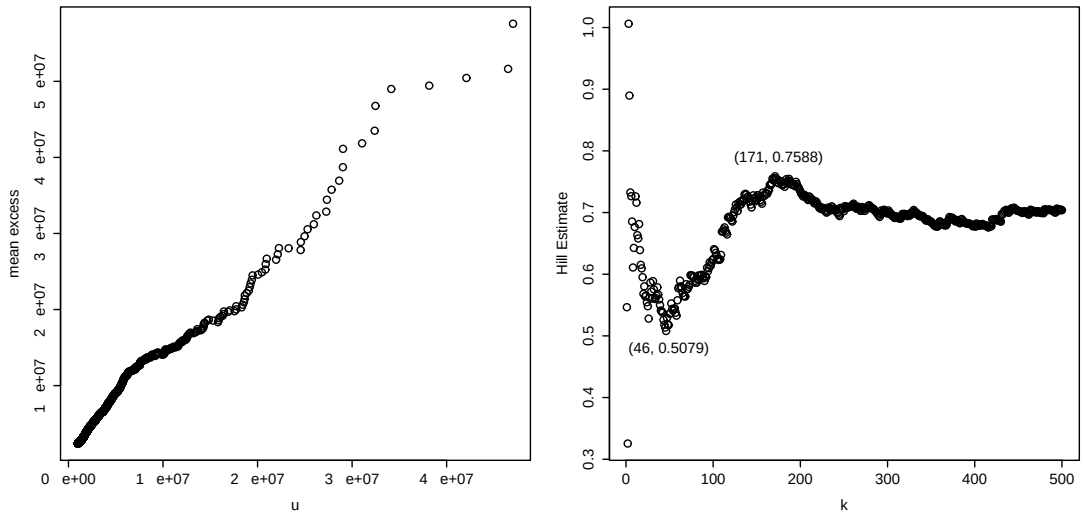


Figure B.2 Mean Excess Plot and Hill's Plot

We assume the underlying distribution F_1 is Pareto-distributed such as

$$1 - F(x) = (1 + x)^{-\frac{1}{\xi}} \tag{B.8}$$

for $\xi > 0$. Then F_X belongs to Fréchet family and $1 - F(x) \in \mathcal{R}_{-1/\xi}$. The tail of the distribution over a sufficiently high threshold can be approximated by the generalized Pareto distribution $G_{\xi,\beta(u)}$ where $\beta(u) = \xi/(1 + u)$ by (B.6).

We choose discrete uniform distributions for the prior distribution of K, the number of exceedances over the threshold, i.e. $K \sim$ discrete $U(0, t)$ for $t > 0$. For each integer value of $t \in [10, 1500]$, the conditional expectation $E[K|t]$ are calculated numerically. Fig. 5.3 is the plot $(t, E[K|t])$ for $t \in [10, 1500]$. Note that the plot shows three stable intervals. It is reasonable to select the second interval $(200 < t < 600)$ because the first $(t < 100)$ and the last $(1000 < t)$ are the results when the domains of uniform random variable K are too small or too large in terms of the possible maximum number of extremal events. Therefore we choose $E[K|t \in (200, 600)] = 177$ for $\hat{k}$ and the corresponding threshold estimate $\hat{u} = X_{177,n} = 6,234,705$ and the shape parameter

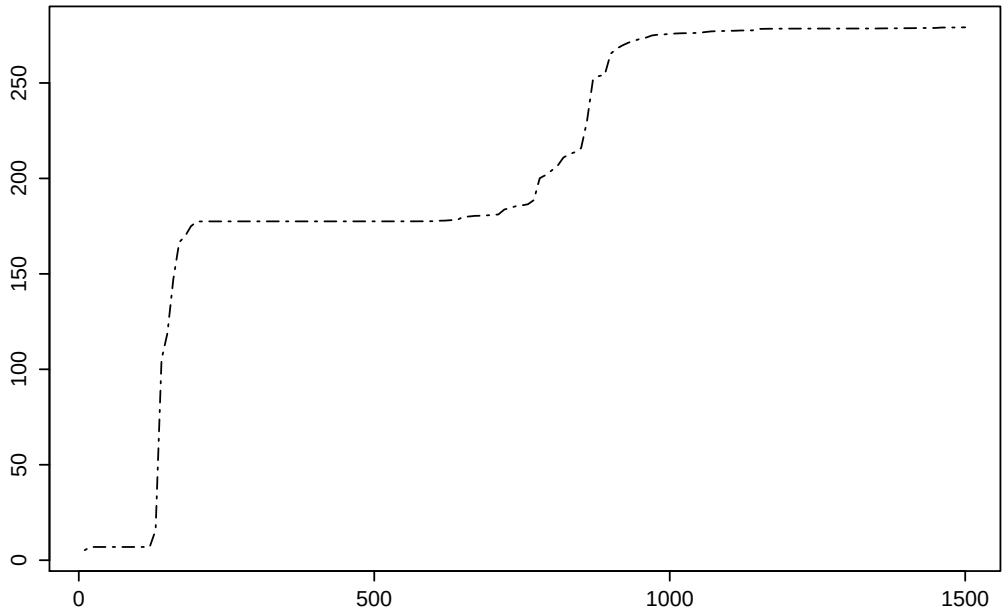


Figure B.3 Plot of the possible maximum number of exceedances, t and $\hat{k}(t)$

estimate $\hat{\xi} = \xi_{177,n} = 0.751$. The fitted distribution with the estimates $\hat{u}$ and $\hat{\xi}$ is the mixture of the underlying Pareto distribution F_1 below $\hat{u}$ and the generalized Pareto distribution $G_{\hat{\xi},\beta(\hat{\xi})}$ above $\hat{u}$.

Appendix C: Symbols and Notation

$A \backslash B$	$\{x \mid x \in A$ and $x \notin B\}$
$(\Omega, \mathcal{F}, \mathcal{P})$	probability space
$\mathbb{P}(A)$	probability of event A
$\mathbb{R}$	set of real numbers
F^{-1}	generalized inverse of distribution F
$\overline{F}$	decumulative distribution function of F
$E[X]$	expected value of random variable X
$Var(X)$	variance of random variable X
$Cov(X, Y)$	covariance of random variables X and Y
$\sigma(X)$	standard deviation of random variable X
$p(\cdot \mid \cdot)$	conditional density function
$N_t, N(t)$	counting process for $t > 0$
$\mathcal{N}(0, 1)$	standard normal distribution
Φ	cumulative standard normal distribution function
ϕ	density of standard normal distribution function
Φ_ρ	multivariate normal distribution with correlation coefficient ρ
I	unit interval $[0, 1]$
$C(u, v)$	bivariate copula
$c(u, v)$	density of bivariate copula
H_ξ	generalized extreme value distribution with shape parameter ξ
$G_{\xi;\beta,\sigma}$	generalized Pareto distribution with shape, scale, and location parameters, ξ, β, σ
$\mathcal{R}_{-\alpha}$	regularly varying distributions with extreme value index α
MDA	Maximum Domain of Attraction
$F \star G$	Convolution of F and G
VaR_p	value-at-risk measure with confidence level $0 < p < 1$
$\sim$	asymptotic equivalence
U	uniform distribution

Bibliography

E. K. Al-Hussaini and M. E. El-Adll. Asymptotic distribution of normalized maximum under finite mixture models. *Statistics & Probability Letters*, 70:109–117, 2004.

S. Alink, M. Löwe, and M.V. Wüthrich. Diversification of aggregate dependent risks. *Insurance: Mathematics & Economics*, 35:77–95, 2004.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

A.A. Balkema and L. de Haan. Residual lifetime at great age. *Annals of Probability*, 2:792–804, 1974.

Basel Committee on Banking Supervision. *International Convergence of Capital Measurement and Capital Standards*. Basel, 2004.

C. Behrens, H. Lopes, and D. Gamerman. Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4:227–244, 2004.

J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes*. John Wiley & Sons, 2004.

G. Bettina and L. Friedrich. Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics & Data Analysis*, 51(11):5247–5252, July 2007.

K. Böcker and C. Klüppelberg. Multivariate models for operational risk. *Technische Universität München, Fakultät für Mathematik, Zentrum Mathematik, Preprint*, 2006.

K. Böcker and C. Klüppelberg. Operational var: A closed form approximation. *RISK*, pages 90–93, 2005.

T. Bucar, M. Nagode, and M. Fajdiga. Reliability approximation using finite weibull mixture distributions. *Reliability Engineering & System Safety*, 84(3):241–251, June 2004.

H. Bühlmann. Experience rating and credibility. *ASTIN Bulletin*, 4:199–207, 1967.

J. Cai and Q. Tang. On max-sum equivalence and convolution closure of heavy-tailed distributions and their applications. *Journal of Applied Probability*, pages 117–130, 2004.

A. Cebrian, M. Denuit, and P. Lambert. Generalized pareto fit to the society of actuaries' large database. *North American Actuarial Journal*, 7:18–36, 2003.

V. Chavez-Demoulin, P. Embrechts, and J. Nešlehová. Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, 1:3–25, 2006a.

V. Chavez-Demoulin, P. Embrechts, and J. Nešlehová. Quantitative models for operational risk: extremes, dependence and aggregation. *Journal of Banking and Finance*, 30(10):2635–2658, 2006b.

S. Coles and J. Tawn. A bayesian analysis of extreme rainfall data. *Applied Statistics*, 45:463–478, 1996.

J. Danielsson, B. N. Jorgensen, S. Mandira, G. Samorodnitsky, and C. G. de Vries. Subadditivity re-examined: the case for value-at-risk. *London School of Economics financial markets group discussion paper series*, ISSU 549, 2005.

M. Degen, P. Embrechts, and D. D. Lambrigger. The quantitative modeling of operational risk. *preprint*, 2006.

K. Dutta and J. Perry. A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. *Federal Reserve Bank of Boston, Working Paper*, No 06-13, 2004.

P. Embrechts and C.M. Goldie. On closure and factorization properties of subexponential distributions. *Journal of Australian Mathematical Society, Series A*, 29:243–256, 1980.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, 1997.

P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: properties and pitfalls in: Risk management: Value at risk and beyond. *Cambridge University Press*, pages 176–223, 2002.

P. Embrechts, A. Hoing, and A. Juri. Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stochastics*, 7:145–167, 2003.

W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. 2, 2nd Ed.* John Wiley & Sons, New York, 1971.

R.A. Fisher and L.H.C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 24:180–190, 1928.

M. J. Frank. *Covolutions for dependent random variables: Advances in Probability Distributions with Given Marginals*. Springer, Berlin, 1991.

E.W. Frees and E.A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25, 1998.

C. Genest and J. MacKay. The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283, 1986.

C. Genest and L.-P. Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043, 1993.

J. Jang and J. Jho. Computation of the aggregate distribution of a maximal Markov sequence. 2008a. Working paper.

J. Jang and J. Jho. Asymptotic super(sub)additivity of the value-at-risk of regularly varying random variables. *Submitted to Journal of Applied Probability*, 2008b.

J. Jho and V. Kaishev. On some mixture distributions and their extreme value behavior. *Actuarial Research Paper, Cass Business School*, 185, 2008. ISBN 978-1-905752-14-0.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.

N.L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley & Sons, New York, 1972.

S. Kang and R.F. Serfozo. Extreme values of random variables with parallel-processing examples. *Journal of Applied Probability*, 36:194–210, 1999.

S.A. Klugman, H.H. Panjer, and G.E. Willmot. *Loss Models from Data to Decisions*. John Wiley & Sons, New York, 1998.

C. Klüppelberg. Risk management with extreme value theory. in: Finkenstädt, b. and rootzén, h. (eds). *Extreme Values in Finance, Telecommunication and the Environment*, Chapman and Hall / CRC:101–168, 2006.

S. Kotz and S. Nadarajah. *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London, 2000.

M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin, 1983.

J. Leslie. On the non-closure under convolution of the subexponential family. *Journal of Applied Probability*, 26:58–66, 1989.

A. McNeil. Estimation the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27:117–137, 1997.

M. Mesfioui and J. Quessy. Bounds on the value-at-risk for the sum of possibly dependent risks. *Insurance: Mathematics & Economics*, 37:135–151, 2005.

P. Mladenović. Extreme values of the sequences of independent random variables with mixed distributions. *Matematicki Vesnik*, 51-1-2:29–37, 1999.

M. Moscadelli. The modelling of operational risk: experience with the analysis of the data collected by the basel committee. temi di discussione (economic working papers). *Bank of Italy, Economic Research Department*, 2004.

R.B. Nelsen. *An Introduction to Copulas. Lecture Notes in Statistics, no.139*. Springer, New York, 1999.

P. Nurmi. Mixture models. *Working paper*, 2004.

H.H. Panjer. Recursive evaluation of a family of compound distributions. *ASTIN Bulletin*, 12: 22–26, 1981.

P. Picard and C. Lefèvre. The probability of ruin in finite time with discrete claim size distribution. *Scandinavian Actuarial Journal*, 1:58–69, 1997.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. URL `http://www.r-project.org`. ISBN 3-900051-07-0.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

B. Sundt. Linearly sufficient fun with credibility. *Insurance: Mathematics & Economics*, 10: 69–74, 1991.

M.V. Wüthrich. Asymptotic value-at-risk estimates for sums of dependent random variables. *ASTIN Bulletin*, 33(1):75–92, 2003.