School of Risk and Actuarial Studies
UNSW Business School

Pairwise versus mutual independence: visualisation, actuarial applications and central limit theorems

Guillaume Boglioni Beaulieu

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Under the supervision of:

Prof. Benjamin Avanzi
Assoc. Prof. Pierre Lafaye de Micheaux
Prof. Bernard Wong

February 2023

ORIGINALITY STATEMENT

☑ I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

COPYRIGHT STATEMENT

AUTHENTICITY STATEMENT

☑ I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate's thesis in lieu of a Chapter provided:

- The candidate contributed **greater than 50%** of the content in the publication and are the "primary author", i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

☑ The candidate has declared that **some of the work described in their thesis has been published and has been documented in the relevant Chapters with acknowledgement**.

A short statement on where this work appears in the thesis and how this work is acknowledged within chapter/s:

> Two published articles contain results that are also present in this thesis. At the beginning of the Thesis (after Acknowledgements), in a section called "Relevant Publications", I detail the contributions that I, Guillaume, made to those publications (and acknowledges the contributions made by other authors), also explaining where this content is located in this thesis.
>
> In summary, the content of the first article is located in Chapter 5, while the content of the second article is located in Chapter 6.

Candidate's Declaration

I declare that I have complied with the Thesis Examination Procedure.

ABSTRACT

Accurately capturing the dependence between risks, if it exists, is an increasingly relevant topic of actuarial research. In recent years, several authors have started to relax the traditional 'independence assumption', in a variety of actuarial settings. While it is known that 'mutual independence' between random variables is not equivalent to their 'pairwise independence', this thesis aims to provide a better understanding of the materiality of this difference. The distinction between mutual and pairwise independence matters because, in practice, dependence is often assessed via pairs only, e.g., through correlation matrices, rank-based measures of association, scatterplot matrices, heat-maps, etc. Using such pairwise methods, it is possible to miss some forms of dependence. In this thesis, we explore how material the difference between pairwise and mutual independence is, and from several angles.

We provide relevant background and motivation for this thesis in Chapter 1, then conduct a literature review in Chapter 2.

In Chapter 3, we focus on *visualising* the difference between pairwise and mutual independence. To do so, we propose a series of theoretical examples (some of them new) where random variables are pairwise independent but (mutually) dependent, in short, PIBD. We then develop new visualisation tools and use them to illustrate what PIBD variables can look like. We showcase that the dependence involved is possibly very strong. We also use our visualisation tools to identify subtle forms of dependence, which would otherwise be hard to detect.

In Chapter 4, we review common dependence models (such has elliptical distributions

and Archimedean copulas) used in actuarial science and show that they do not allow for the possibility of PIBD data. We also investigate concrete consequences of the 'non-equivalence' between pairwise and mutual independence. We establish that many results which hold for mutually independent variables do not hold under sole pairwise independent. Those include results about finite sums of random variables, extreme value theory and bootstrap methods. This part thus illustrates what can potentially 'go wrong' if one assumes mutual independence where only pairwise independence holds.

Lastly, in Chapters 5 and 6, we investigate the question of what happens for PIBD variables 'in the limit', i.e., when the sample size goes to infinity. We want to see if the 'problems' caused by dependence vanish for sufficiently large samples. This is a broad question, and we concentrate on the important classical Central Limit Theorem (CLT), for which we find that the answer is largely negative. In particular, we construct new sequences of PIBD variables (with arbitrary margins) for which a CLT does not hold. We derive explicitly the asymptotic distribution of the standardised mean of our sequences, which allows us to illustrate the extent of the 'failure' of a CLT for PIBD variables. We also propose a general methodology to construct dependent K-tuplewise independent (K an arbitrary integer) sequences of random variables with arbitrary margins. In the case $K = 3$, we use this methodology to derive explicit examples of triplewise independent sequences for which no CLT hold. Those results illustrate that mutual independence is a crucial assumption within CLTs, and that having larger samples is not always a viable solution to the problem of non-independent data.

ACKNOWLEDGEMENTS

Along the PhD journey, I met tremendous people without whom it would not have been possible to write this thesis. I can only attempt to thank them.

I thank dearly my supervisors, Prof. Benjamin Avanzi, Assoc. Prof. Pierre Lafaye de Micheaux and Prof. Bernard Wong for the advice and wisdom they have imparted me over the years.

Pierre was the first to stimulate my interest in statistics when I sat in his 'Introduction to Statistics' class at the University of Montreal, back in 2012. Inside and outside class, I had lots of questions, and he always had good answers. I learned a lot form Pierre scientifically, and I have always admired his rigour. During my PhD, he was extremely generous with his time, and gave me over the years many useful scientific and career advice. I remember fondly the long conversations had in his office about statistics and academic life. Certainly this thesis contains many ideas that either came from him, or that he helped me develop. I am very grateful he always had my best interests at heart.

Benjamin taught me how fun actuarial science can be. Sitting in his 'Risk Theory' class back in 2014 was certainly among the highlights of my bachelor degree, and I am grateful he encouraged me to do a Masters', which also prepared me to embark on this PhD journey. Along the way, he constantly made valuable comments and provided great insights I would not have seen otherwise. His career advice was always on point, and if I have chosen to try the academic career, it is in great part due to the example he set and the advice he gave. And, at times when I was down, Benjamin found the right words to pull me up and make me believe in myself. For this, I am very grateful.

Bernard gave me valuable advice all along the PhD path, and I am grateful for his many useful suggestions. His contagious enthusiasm was a source of motivation throughout. I am also very grateful he provided me with great opportunities to travel at conferences and be involved in the academic life of the School. That he confided in me the responsibility to teach a large first year course, and gave with the right advice and support to do so, was a great honour and highlight of my PhD. I thank him for trusting in my abilities, and for his great sense of fairness.

I would also like to thank my collaborator Frédéric Ouimet, whose help and suggestions contributed greatly to the quality of this thesis. I have always appreciated our frank and no-nonsense email exchanges. Hopefully we can meet in person some day.

I thank dearly the two examiners of this thesis, prof. Arthur Charpentier and prof. Jed Frees. I am very appreciative that they carefully reviewed this thesis, formulated many insightful comments and recommendations and also pointed out several useful references I had missed. They also gave me interesting ideas for future research, which I am very thankful for.

I thank the many friends I met along the PhD journey, who made the experience all the more pleasant. They also inspired me to be my best self. To Alan, Anh, Nikolay and Vincent, you are all wonderful human beings and you set the bar very high for me to follow in your footsteps. To Héloïse and Igor, our friendship has been a true blessing. I do not think I would have made it through without your support in the last rocky bits of this PhD. To Mikhaïl, what a gem of a flatmate and friend you have been; those years have past too quickly. To Gayani, you arrived more recently in my life, but were a great friend at a time I needed it the most. I will remember it.

Running has been my escape, and I thank all my training partners from "Deano's Training Group". Alas they are too numerous to name (if I tried I would be bound to forget someone). I will only mention by name the man behind this, Dean Gleeson. Thank you coach for the great support over those years, and for always putting your athletes' needs first.

I acknowledge the generous financial support I received from UNSW Sydney (via a University International Postgraduate Award and a HDR Completion Scholarship), UNSW Business School (via a Supplementary Scholarship) and the FRQNT (via a Doctoral Scholarship, B2).

Lastly, I have been incredibly fortunate in my life. To be given strong health, good education and a supporting family, as I have, is all I could have wished for. Those gifts made the writing of this thesis possible. I must especially thank my mother, whose love and optimism have no bounds. I must also thank my father, who set the best example for me to follow, in everything.

RELEVANT PUBLICATIONS

Two published articles contain results that are also present in this thesis. This section details the contributions that I, Guillaume, made to those publications (and acknowledges the contributions made by other authors), and where this content is located in this thesis.

- Avanzi et al. (2021). I am the primary author of this paper whose topic was initially suggested by one of my supervisors. Most ideas come from me, and I wrote a large part of the text in the paper. I wrote the majority of the proofs of an earlier version of the paper (found on arXiv, see arxiv:2003.01350v1), though substantial improvements were later made with the help of co-authors (especially Frédéric Ouimet), also making the results more general (in the earlier version, the main result only included continuous distributions, while the current version includes some discrete distributions). The content of this paper corresponds to Sections 5.1, 5.2, 5.3, 5.4, 5.6 and 5.B of this thesis. Some of the `R` code in Appendix A.4 is also found in the 'Supplementary Material' of the paper. Compared to the paper, small adjustments were made where necessary to better fit in this thesis in terms of narrative, notation, use of British spelling, etc.

- Boglioni Beaulieu et al. (2021). This paper stems directly from the previous, as in many ways it is a generalisation of the results from that paper (from 'pairwise' to 'triplewise' independence). I was involved in all aspects of this paper, and I contributed to the writing of all main sections. I contributed the central idea that our methodology is very general, and hence can be used to construct any number of examples (an earlier version of the paper provided only the specific examples).

I acknowledge that Frédéric Ouimet is the main contributor of Section 4 and the Appendix, while Pierre Lafaye de Micheaux is the main contributor of the simulation results from Section 5. The content of this paper corresponds to Chapter 6 of this thesis (with the exception of Section 6.4.4, 6.6 and 6.B of this thesis, which are not in the paper). Compared to the paper, Chapter 6 in this thesis contains a few more explanations. Minor adjustments were also made in terms of narrative, notation, use of British spelling, etc.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CDF	Cumulative Distribution Function
CLT	Central Limit Theorem
CR	Capital Required
DB	Diversification Benefit
i.i.d.	Independent and identically distributed
LLN	Law of Large Numbers
MSE	Mean Squared Error
PIBD	Pairwise independent but dependent
p.i.i.d.	Pairwise independent and identically distributed
PDF	Probability density function
PMF	Probability mass function
r.v.s	Random variables
t.i.i.d.	Triplewise independent and identically distributed
TVaR	Tail Value-at-Risk
VaR	Value-at-Risk

SYMBOLS AND NOTATION

$\mathbb{R}$	Set of the real numbers
$\mathbb{N}$	Set of the natural numbers
$\mathbb{P}$	Probability or probability measure
$\mathbb{E}[X]$	Expectation of the random variable X
$\mathbb{V}\mathrm{ar}[X]$	Variance of the random variable X
$\log(x)$	Natural logarithm of x
$\{X_j, j \geq 1\}$	An infinite sequence of random variables $X_1, X_2, \ldots$
$N(0,1)$	Standard Normal random variable
$\mathbb{1}_B$	Indicator function on a set B
$\mathrm{i} \coloneqq \sqrt{-1}$	Imaginary number
$\mathrm{Erf}(z) \coloneqq \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$	Error function
$\text{`}\stackrel{d}{=}\text{'}$	Equality in distribution
$\text{`}\xrightarrow{d}\text{'}$	Convergence in distribution
$\text{`}\xrightarrow{\mathbb{P}}\text{'}$	Convergence in probability
$\text{`}\xrightarrow{\text{a.s.}}\text{'}$	Almost sure convergence

CHAPTER 1

INTRODUCTION

1.1 Context

As individuals, we face constant uncertainties, big and small. An unexpected death can cripple a whole family's financial situation. A lifetime of savings can be wiped out in a single weather event. Years of future earnings can be compromised by an accident or a debilitating illness.

To diminish such uncertainty, many of us are willing to pay someone else to cover some of the *risks* we face. This possibility of a *risk transfer* is at the core of insurance companies' business model. Indeed, provided they can reliably make money out of it, insurers are happy to bear (at least some of) those risks, in exchange for a premium. Although they will lose money on some insurance policies, *on the aggregate*, they can expect to make a profit thanks to diversification benefits generated from pooling many small risks together. The unpredictability we dislike can turn into a profit for them.

It is often said that the Law of Large Numbers (LLN) offers a mathematical foundation for the insurance business (Smith and Kane, 1994; Dhaene et al., 2002b; Feng and Shimizu, 2016). As Feng and Shimizu (2016) put it

> One of the most fundamental principles for the insurance business is the pooling of funds from a large number of policyholders to pay for losses that a few policyholders incur. The mathematics behind such a business model is the law of large numbers which dictates that actual average loss would be close to the theoretical mean of loss with a large pool of homogeneous and independent risks.

A keyword in the above is 'independent'; if risks are somehow dependent, there is no guarantee that the LLN would apply and diversification is compromised. The words 'dependent' and 'independent' have here their usual probabilistic meanings. We recall formal definitions in Section 2.2.1, but an intuitive definition of independence is given as (Resnick, 1999):

> the easily envisioned property that the occurrence or non-occurrence of an event has no effect on our estimate of the probability that an independent event will or will not occur.

Hence conversely, when we speak of *dependent risks*, we mean that knowledge about one (or more) gives *some* knowledge about one (or more) other risk(s).

This thesis aims to contribute to the ongoing conversation around 'dependence' and 'independence' in actuarial science (which has been very active in recent years, to say the least). In particular, we will focus on the distinction between 'pairwise' and 'mutual' independence, which we feel is too easily overlooked. We provide much more detail in Section 1.5, but note, for example, that the ubiquitous *correlation coefficient* (the most common dependence measure) is by its very nature a *bivariate* coefficient. Since pairwise independence does not imply mutual independence, if one relies solely on correlation (or on any *bivariate* measure of dependence), one may fail to detect some forms of dependence. This can potentially have material consequences, which we will explore all along this thesis.

1.2 Diversification benefit: a definition

In the previous section, we have used the term 'diversification benefit' somewhat loosely, to mean a reduction in the overall risk a company faces and which stems from "detaining many different risks, with various probabilities of occurrence and a low probability of happening simultaneously" (Busse et al., 2014).

A more formal definition (often used in Enterprise Risk Management) is as follows. Let $\boldsymbol{X} \coloneqq (X_1, X_2, \ldots, X_d)$ be a collection of random variables which represent losses (e.g., from d business segments of an insurance company) and let 'CR' stand for Capital Required. We have a 'diversification benefit' (DB) whenever the *total* Capital Required is less than the *sum of* the 'standalone' Capitals, $\mathrm{CR}(X_i)$. We then define the DB as (see, e.g. Dacorogna, 2018, Eq.8):

$$\mathrm{DB}(\boldsymbol{X}) = 100\% - \frac{\mathrm{CR}(\sum_{i=1}^{d} X_i)}{\sum_{i=1}^{d} \mathrm{CR}(X_i)}, \tag{1.2.1}$$

where the CRs are typically computed as risk measures (VaR, TVaR, or other). Clearly, in this context the diversification benefit is function of the dependence between the losses X_i, because the distribution of the random variable $\sum_{i=1}^{d} X_i$ is function of such dependence.

Of course, certain risks insurance companies face are 'undiversifiable' (also known as *systematic risks*). For example, if those risks are extremely positively dependent (comonotonic), then increasing their number in a portfolio does not bring any substantial diversification. In Equation (1.2.1), this would correspond to the case

$$\mathrm{CR}\left(\sum_{i=1}^{d} X_i\right) \approx \sum_{i=1}^{d} \mathrm{CR}(X_i).$$

A possible example of this would be flood insurance for properties in the same flood-prone geographic location. A single flood would then trigger claims for a very large number of insurance policies, all at once. Hence, holding *more* of such insurance policies would bring very little diversification to an insurer, if any.

That said, in this thesis we will focus on the case where risks are (at least to some extent), diversifiable. For such risks, the dependence between them (if any) will impact the extent of the diversification. This is illustrated in the next section, where we give examples of possible dependencies between insurance risks (taken from the literature).

1.3 A few tales of dependence

In the fields of probability and statistics, questions of 'independence and dependence' are not new, and have given rise to a large literature (especially in the second half of the 20th century). For instance, a classical paper by Kruskal (1958) poses the question "What is meant by the degree of association or dependence between two random variables with a joint distribution?", and investigates at length three common rank-based measures of association (namely quadrant association, Kendall's tau and Spearman's rho). Other seminal articles include Rényi (1959) who proposed seven axioms that a 'good' bivariate measure of dependence should possess, and Lehmann (1966) who provided several (increasingly stronger) definitions of "positive dependence".

It is also worth noting that the 1960's was a productive period in multivariate modelling, with many authors deriving multivariate extensions of common probability distributions. Some influential articles proposed, for example, extensions of the Exponential distribution (Gumbel, 1960; Freund, 1961; Marshall and Olkin, 1966, 1967; Block and Basu, 1974), the Pareto distribution (Mardia, 1962) and the Burr distribution (Takahasi, 1965). This period also saw great innovations in the non-parametric estimation of multivariate distributions, see e.g., Loftsgaarden and Quesenberry (1965) and Cacoullos (1965).

In the same broad field of multivariate modelling, another seminal piece of work is of course the famous Sklar's theorem (Sklar, 1959), which showed that the dependence within any multivariate distribution function with continuous margins is characterised by a unique *copula* (a distribution function with Uniform$[0, 1]$ margins). Copulas have since become a very popular tool in dependence modelling, in a variety of fields (for a review of copula concepts, see Section 2.2.2).

In the actuarial literature, though, dependence has more recently become a central research topic. To quote the influential work of Embrechts et al. (2002),

> Although insurance has traditionally been built on the assumption of independence and the law of large numbers has governed the determination of premiums, the increasing complexity of insurance and reinsurance products has led recently to increased actuarial interest in the modelling of dependent risks.

When we speak of dependent risks, this can be envisioned at many levels of granularity.

For instance, it could be that different individuals are related in a way which makes the claims they each generate dependent. For instance, from Cossette et al. (2002):

> Consider a group life insurance or a group health insurance contract issued to a company for a section of its employees working in a mine, on a steel plant, in a paper mill, etc. In these cases, a single event (e.g., explosion, breakdown) influences the risks of the entire portfolio. The risks are therefore statistically dependent.

We can think of many other such situations. For instance, within a couple, "[i]t has long been documented that the death of the spouse adversely impacts the surviving spouse, accelerating the death of the latter" (Lu, 2017). That is, there is dependence between the future lifetimes of the individuals in a couple. Hence, any insurance payments contingent on those lifetimes (e.g., through life insurance or annuity contracts) would be dependent. Another example is in workers' compensation insurance, where there will likely be a link between medical costs, and 'daily allowance costs' (see, e.g. Avanzi et al., 2011).

Home insurance is another obvious example, where geographical proximity of properties can create a strong dependence, since "homes may share a common exposure to disasters (e.g., windstorm, hurricane, fire, earthquake) that may lead to possible catastrophic financial damage" (Valdez, 2014). As another example, some find reasons to believe there is dependence between individual frequencies and severities of claims, for instance Garrido et al. (2016) argue that "claim counts and amounts are often negatively associated in collision automobile insurance because drivers who file several claims per year are typically involved in minor accidents".

In addition to the individual level, dependence can also be envisioned at a higher (or more aggregate) 'macro' level. For instance in property insurance, the total daily (or weekly, or monthly) claims experienced in a given region (e.g., a state) could be related to the total claims experienced in a neighbouring region, because big storms can sweep both regions almost at the same time. This dependence can be exacerbated by worldwide weather cycles such as El Niño Southern Oscillation (ENSO). Indeed, and as noted by Boudreault et al. (2014), "[m]any authors have reported that ENSO is known to have an important influence on hurricane frequency and intensity in the Atlantic Ocean". Here, we are considering the *total amount* of claims in a region (which stem from many policies sold) to be dependent of another aggregated amount, that of a neighbouring region.

When speaking of dependence at the aggregate level, we can also imagine that whole

business segments selling different types of insurance (e.g., motor and home) might be dependent. For instance, catastrophic events can trigger several types of claims at once. Bürgi et al. (2008) describe an extreme example of this:

> The collapse of the World Trade Center towers on September 11 2001 has shown in a dramatic way how insurance products of lines of business, which had thought to be independent until then, can be triggered at once. The scenario of both towers collapsing had been assumed to be impossible [...] Several surrounding buildings have also collapsed or suffered severe damages. Insurance-wise, several property policies were triggered. Next to the damage to the buildings, there was a large amount of business interruption claims, and several life policies were triggered, while cars parked under the towers were also destroyed and triggered motor policies.

Another obvious example of a single catastrophic event inducing unexpected dependence between insurance business segments is that of the Covid-19 pandemic. As Richter and Wilson (2020) note,

> While high correlation in life and health insurance potentially covered losses is a defining feature of pandemic risk, the Covid-19 crisis has highlighted the accumulation potential also with respect to financial market losses and to property and casualty losses (in particular non-property damage business interruption and event cancelation). Combined with legal uncertainty and issues with imprecise wording this has caused an unpleasant "surprise" for the industry.

The realisation that independence is an untenable assumption in many insurance settings has given rise to a vast body of actuarial research, tackling a variety of insurance problems, e.g., claims count processes (e.g. Denuit et al., 2002; Pfeifer and Nešlehová, 2004; Avanzi et al., 2016a), dependence between severity and frequency in compound models (e.g. Peters et al., 2009; Hernández-Bastida et al., 2009; Czado et al., 2012; Garrido et al., 2016), claims reserving (e.g. Shi and Frees, 2011; De Jong, 2012; Merz et al., 2013; Abdallah et al., 2015; Avanzi et al., 2016b,d), credibility theory (e.g. Frees and Wang, 2005; Englund et al., 2008; Wen et al., 2009), ruin theory (e.g. Müller and Pflug, 2001; Bregman and Klüppelberg, 2005; Eling and Toplek, 2009; Albrecher et al., 2011), only to name a few.

This literature will be discussed in more detail in Chapter 2. For now, we proceed to a small digression on the famous correlation coefficient which (albeit often criticised) is still widely used as a measure of dependence. For instance, from Taylor (2018), "it is

correlation that is commonly used in practice as a measure of dependence for the purpose of risk margin estimation".

Remark 1.1. *Some of the examples above revealed dependence 'in space'. Another important form of dependence (which can potentially affect the diversification benefits of an insurer) is dependence 'in time'. For example, when considering a risk which evolves through time (e.g., the number of insurance claims stemming from a specific contract, or a portfolio of contracts), past experience often contains information about future experience. Furthermore, and as noted by Bermúdez et al. (2018) "the behaviour of a driver is likely to change after they have made a claim and, therefore, some kind of time dependence should be found in a panel count dataset."*

Another instance of time dependence is when the time elapsed since the last event influences the severity of the next event. For example, Boudreault et al. (2006) note that "in an earthquake risk context, one can expect that the longer is the time between two events the larger will be the claim amount for the next catastrophe" (because, perhaps, more structures will have been built in the intervening time and hence the potential for damage will be greater).

Many other factors can induce time dependence in claim processes, for example seasonality (Avanzi et al., 2016c) and inflation (Landriault et al., 2014b). In general, we have that knowledge about the number (and/or size) of claims in a given time period provides some information about the number (and/or size) of claims in a future time period. While time dependence is a broad and important topic, we note that it will not be the main focus of this thesis.

1.4 A note on Pearson's correlation

In a conversation about dependence, it is hard not to mention the notion of *correlation*. Recall that for two random variables X and Y with finite means μ_X, μ_Y and finite variances σ_X^2, σ_Y^2, Pearson's correlation is given by

$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}.$$

This correlation coefficient ρ is broadly used in risk management practice. It is, for instance, an important element in the determination of solvency capital requirements, as set by many regulatory frameworks around the world (see Section 2.3.2 for more details). Many authors (e.g. Embrechts et al., 2002; Danıelsson, 2002; Pukthuanthong and Roll, 2009; Bartram and Wang, 2015) have pointed out potential pitfalls of using correlation as a measure of dependence in finance and insurance. Here we want to briefly summarise some of those pitfalls. This will also set the scene for us to introduce the central theme of this thesis, in the next section.

One important flaw of correlation is that it detects only *linear* dependence. As the following example shows, two random variables can be strongly dependent, yet have a correlation of zero.

Example 1.1. *Let two random variables (X, Y) be defined such that*

$$X \sim \text{Exp}(1)$$
$$Y|X \sim \text{Normal}(\mu = 1, \sigma = X).$$

Then, we have that $\mathbb{E}[X] = \mathbb{E}[Y] = 1$ and $\mathbb{V}\text{ar}[X] = 1, \mathbb{V}\text{ar}[Y] = 2$. More importantly, we have that

$$\rho_{X,Y} = 0,$$

even though X and Y are dependent. Ther scatterplot of a (rank-transformed) sample of size $n = 5{,}000$ generated from this example is displayed on Figure 1.1. We can see a clear 'C-shape' pattern in the data, indicating dependence. Furthermore, if we interpret X and Y as 'risks' we can calculate the 'capital required' to cover the total 'loss' $X + Y$ using a risk measure. If we use the Value-at-Risk $\text{VaR}_{99.5\%}$ (for example), under independence

between X and Y we would have

$$\mathrm{VaR}_{99.5\%}(X+Y) = 8.351.$$

However, here the real VaR is in fact

$$\mathrm{VaR}_{99.5\%}(X+Y) = 10.405.$$

Also, the 'standalone' capitals required to cover X and Y are

$$\mathrm{VaR}_{99.5\%}(X) = 5.298, \quad \mathrm{VaR}_{99.5\%}(Y) = 6.607.$$

Recalling (1.2.1)*, we then have a diversification benefit (under independence) of*

$$1 - \frac{8.351}{5.298 + 6.607} = 29.9\%,$$

while the real diversification benefit is only

$$1 - \frac{10.405}{5.298 + 6.607} = 12.6\%.$$

That is to say, independence entails a much larger diversification benefit (more than double) than the diversification benefit under this dependence structure (even though the correlation is zero).

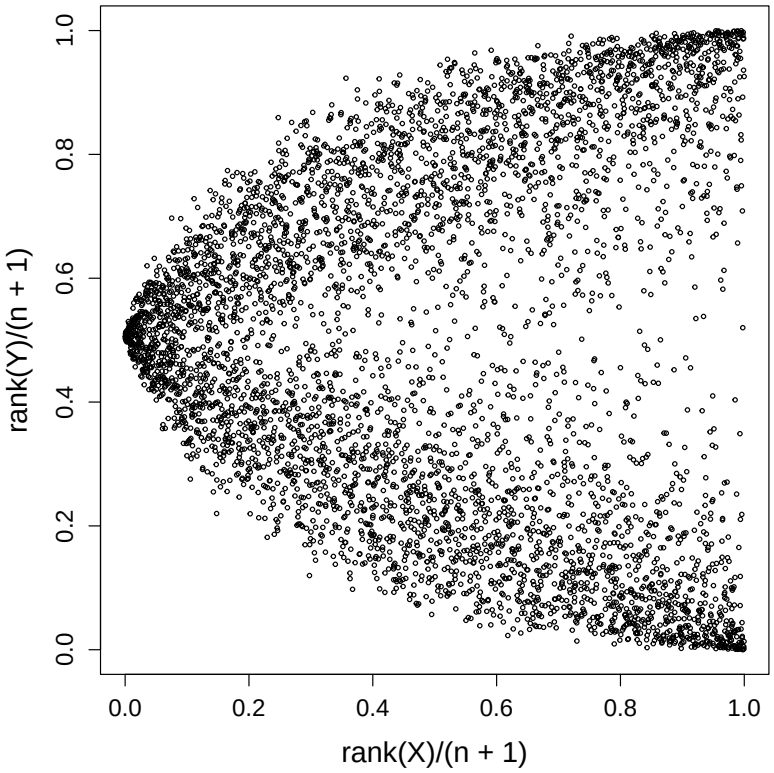

Figure 1.1: Sample of Y versus X generated under Example 1.1 (rank-transformed observations)

This example showcases that 'independence' and 'zero-correlation' are very different things. Other shortcomings of correlation include the fact it is not invariant to monotone transformations of the risks, it is defined only for risks with finite variance, and the range of its possible values depends on the marginal distributions of the risks (Embrechts et al., 2002). When estimated from data, correlation is also "very sensitive to the presence of outliers" (Abdullah, 1990). This is because its influence function is unbounded, and hence one single 'unusual' data point has the potential to dominate its value (Wilcox, 2011, chap. 9).

Lastly, we note that correlation is, after all, just a number. Hence, it cannot (in general) provide the full dependence picture. The same correlation can imply very different dependencies; Matejka and Fitzmaurice (2017) provide interesting examples of this, some of which we reproduce on Figure 1.2 (using the R package `datasauRus`). Here, all thirteen bivariate datasets (X vs Y) have almost the same correlation (always between -0.07 and -0.06), as well as almost the same means and variances for X and Y. However, all those datasets look very different. In particular, one looks like a dinosaur.

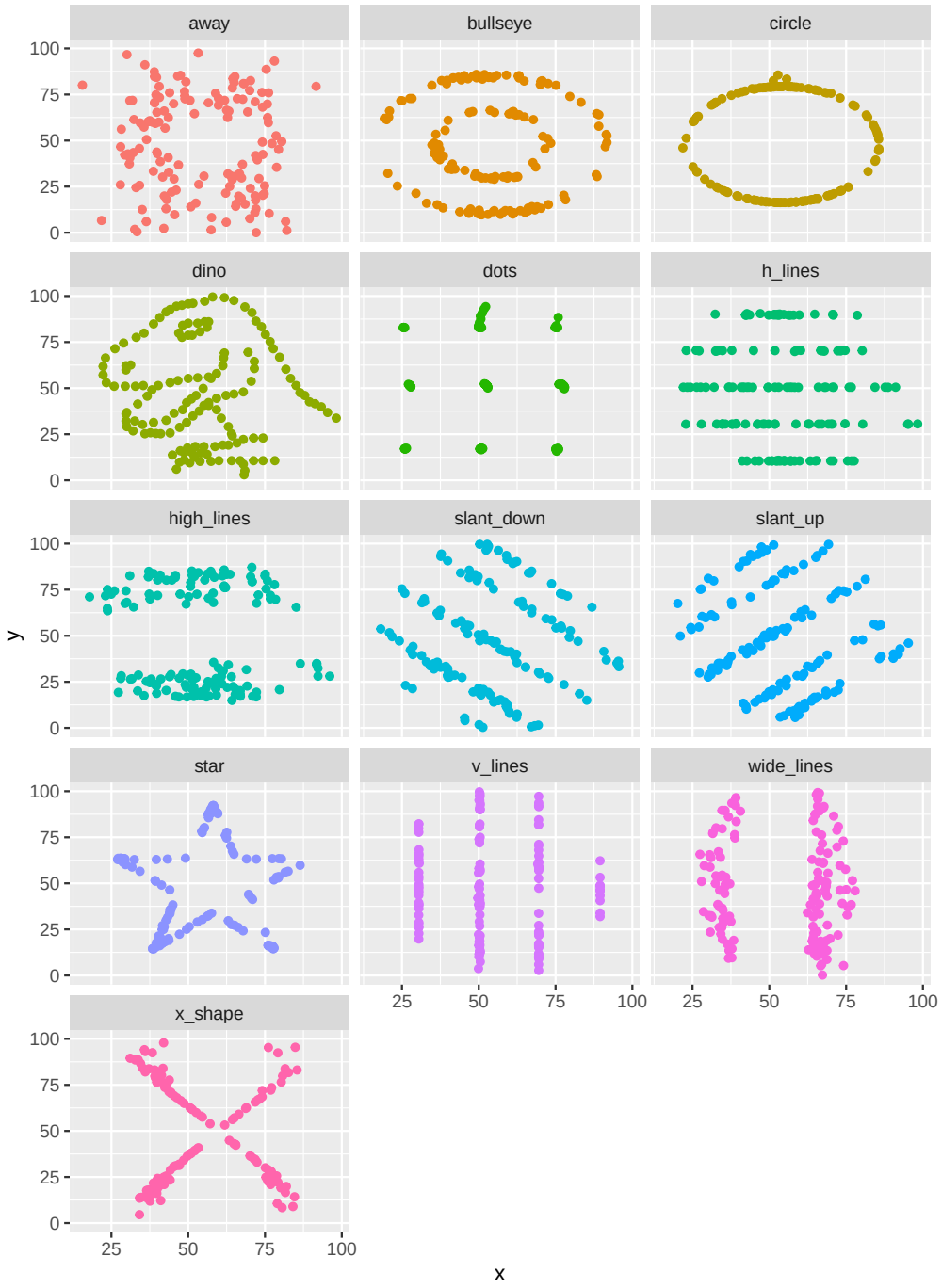

Figure 1.2: Thirteen datasets, all with (almost) the same Pearson's correlation

The bottom line here is that correlation is an imperfect measure of *bivariate* dependence. This is well known in the actuarial community. In the next section, we will see that even if one models perfectly the dependence between pairs of random variables, one can fail to capture all the dependence which exists in a set of *more than two* random variables.

Remark 1.2. *When discussing the limitations of correlation, one often makes the point that correlations can be 'spurious' or that 'correlation is not causation'. This is indeed true:*

to observe that two random variables X and Y have a strong correlation says nothing about whether X causes Y, or Y causes X, or whether another phenomenon influences both (such that the result is a linear trend between X and Y but with no direct causation). The website tylervigen.com/spurious-correlations collates interesting (and funny!) examples of this.

This point also holds about dependence more generally (not just correlation): we cannot infer from the fact that two variables are dependent (even strongly) that one causes the other. To give an actuarial example (which we take from Lu, 2017), it has been documented that the future lifetimes of two spouses are positively correlated. But can we conclude from this that the death of one spouse 'causes' an increased mortality for the other spouse, perhaps because of "psychological shock, and the subsequent change of the life style after the loss of the spouse" (Lu, 2017)? Or, is the dependence explained simply by the fact that two spouses usually share common risk factors (e.g., lifestyle, wealth)? It takes careful analysis to answer this question, and Lu (2017) finds evidence of both effects. Though this distinction between 'dependence' and 'causation' is important, it will not be a focus of this thesis. In speaking of 'dependence', we will intend any pattern between random variables (not necessarily causal ones).

1.5 Pairwise independence versus mutual independence

Upon realising that correlation can be inadequate to capture pairwise dependence, one might think sufficient to 'improve' correlation. One may want, for example, to use a robust measure of bivariate dependence such as Kendall's tau. Or, one could turn to a more sophisticated measure of dependence (capable of capturing complex non-linear dependencies), such as distance-correlation (Székely et al., 2007), Maximal Information Coefficient (Reshef et al., 2011), generalised R-squared (Wang et al., 2017) or Hellinger correlation (Geenens and Lafaye de Micheaux, 2022), only to name a few. This could be appropriate, but it could also be insufficient. Indeed, a fact about dependence which is less studied (but crucial to this thesis) is that

> Even a perfect characterisation of *pairwise* dependence can be insufficient to paint the full *multivariate* dependence picture.

For example, a vector of three random variables X_1, X_2, X_3, can have all its pairs (X_1, X_2), (X_1, X_3) and (X_2, X_3) be bivariate Gaussian, while the triplet (X_1, X_2, X_3) is *not Gaussian* (see an explicit example in Loisel, 2009). But in particular, the focus of this will be the additional fact that

> A series of risks can be perfectly **Pairwise Independent But still Dependent** (PIBD).

Hence, being confident that every pair of risks in a portfolio is independent still does not guarantee their *mutual* independence (see Section 2.2 for formal definitions of pairwise versus mutual independence). Said otherwise, *mutual independence* of a sequence of random variables $\{X_j, j \geq 1\}$ is a stricter condition than their *pairwise independence*:

i. mutual independence of $\{X_j, j \geq 1\} \implies$ pairwise independence of $\{X_j, j \geq 1\}$

ii. pairwise independence of $\{X_j, j \geq 1\} \not\Longrightarrow$ mutual independence of $\{X_j, j \geq 1\}$.

This is problematic, because independence is usually assessed by pairs, e.g., through correlation matrices (which are still vastly used for risk aggregation, see, e.g., Taylor, 2018), rank-based measures of association, scatterplot matrices (Hofert and Oldford, 2018), heatmaps, pair-copula constructions, etc. While this 'non-equivalence' of pairwise independence and mutual independence is known in the probability literature (see, e.g., Derriennic and Kłopotowski, 2000; Nelsen and Ubeda-Flores, 2012), we believe it has been largely

overlooked in the actuarial literature, and its consequences are not well known. This thesis is an effort towards filling this gap. Along it, we will see that the distinction between *pairwise* and *mutual* independence can be stark, and we will cover many situations where things can 'go wrong' in typical insurance settings if only *pairwise* (but not *mutual*) independence holds.

We note that this phenomena of PIBD variables is interesting in part because it goes against the usual ways in which we think of dependence. For example, in models of times series or spatial data, it is usually observations close to one another that are the most dependent (and dependence diminishes for observations further apart). In contrast, with PIBD data, consecutive observations (pairs) are independent, while larger groups may exhibit dependence.

As a teaser for the rest of this thesis, we present here two toy examples which illustrate the difference between 'pairwise' and 'mutual' independence.

Example 1.2. *Consider a classical formulation of the 'Individual Risk Model' (see, e.g., Klugman et al., 2012, Section 9.8), where n insurance policies can each produce, or not, a claim. The risk X_k (for the kth policy, $k = 1, \ldots, n$) is usually written as*

$$X_k = \begin{cases} B_k & \text{if } I_k = 1 \\ 0 & \text{if } I_k = 0, \end{cases}$$

where each I_k is a Bernoulli random variable, equal to 1 if a claim occurs for policy k, and where B_k is a strictly positive random variable which represents the claim amount for policy k (given a claim occurs). Typically, the B_k's are assumed independent of each other, and also independent of the I_k's, which themselves are independent of each other.

For our toy example, consider the case $n = 3$, $B_k \sim Exp(\lambda)$ and $I_k \sim Bernoulli(1/2)$, $k = 1, 2, 3$. Under the (typical) assumption that X_1, X_2, X_3 are mutually independent, the distribution of the aggregate amount of claims

$$S = X_1 + X_2 + X_3,$$

is straightforward to derive. Indeed, we have that

$$I_1 + I_2 + I_3 = \begin{cases} 0 & \textit{with probability } 1/8 \\ 1 & \textit{with probability } 3/8 \\ 2 & \textit{with probability } 3/8 \\ 3 & \textit{with probability } 1/8, \end{cases}$$

and consequently, S is a mixture of the constant 0, an Exp(λ), a Gamma($2, \lambda$) and a Gamma($3, \lambda$), with weights $1/8, 3/8, 3/8, 1/8$, respectively.

Now assume that we change just one 'detail'. We still let the B_k's be independent, and independent of the I_k's. We also let I_1, I_2, I_3 be independent by pairs (meaning that $\mathbb{P}[I_j = 1, I_k = 1] = 1/4$, $j \neq k$), but not mutually independent. Following a classical example from Bernšteĭn (1927), this will be the case if, for instance,

$$\begin{aligned} \mathbb{P}[I_1 = 1, I_2 = 1, I_3 = 1] &= 1/4 \\ \mathbb{P}[I_1 = 1, I_2 = 0, I_3 = 0] &= 1/4 \\ \mathbb{P}[I_1 = 0, I_2 = 1, I_3 = 0] &= 1/4 \\ \mathbb{P}[I_1 = 0, I_2 = 0, I_3 = 1] &= 1/4. \end{aligned} \tag{1.5.1}$$

In that case, the distribution of S is significantly different from that under mutual independence of the I's. Indeed, since here we have

$$I_1 + I_2 + I_3 = \begin{cases} 1 & \textit{with probability } 3/4 \\ 3 & \textit{with probability } 1/4, \end{cases}$$

S would be a mixture of an Exp(λ) and a Gamma($3, \lambda$), with weights $3/4$ and $1/4$, respectively. As illustrated on Figure 1.3, the PDF and CDF of S in both scenarios is very different. Most notably, the fundamental nature of S has changed. Indeed, in the fully independent case, S has a mixed distribution with $\mathbb{P}[S = 0] = 1/8$, while in the pairwise independent case S is fully continuous.

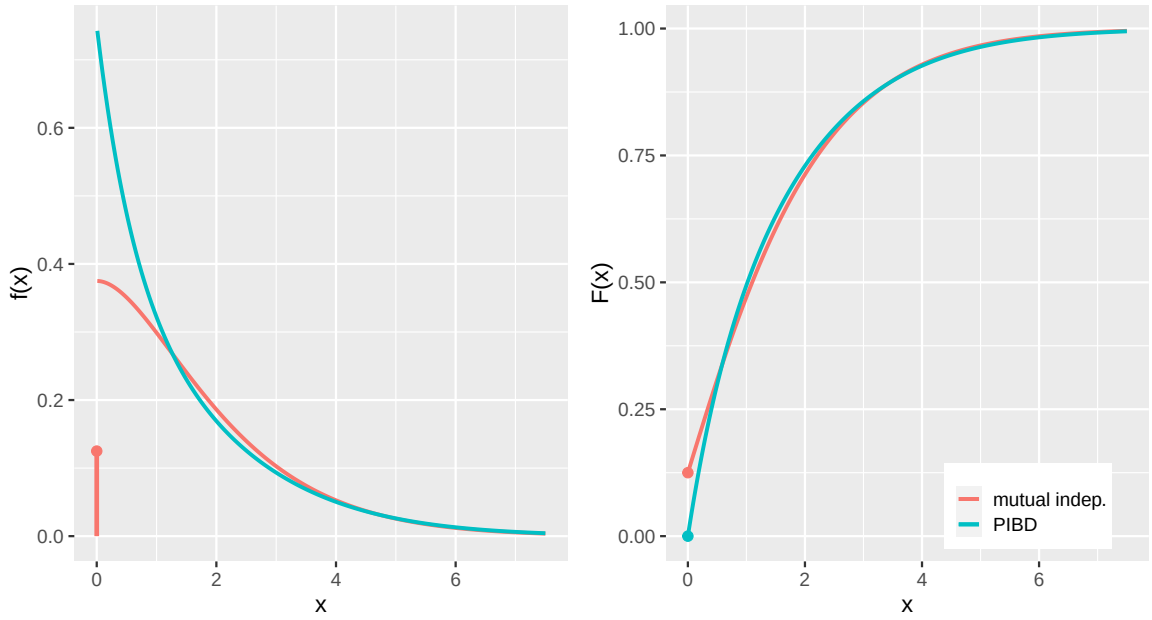

Figure 1.3: PDF (left) and CDF (right) of S as defined in Example 1.2. In the 'mutual independence' case, S is a mixed random variable, so an additional mass of $1/8$ is shown on the PDF plot.

Our second example is taken from Romano and Siegel (1986, Ex. 2.12) and features three standard Normal random variables which are PIBD.

Example 1.3. *Consider a triplet (X, Y, Z) of random variables constructed as follows:*

$$X, Y, W \overset{\text{i.i.d.}}{\sim} N(0,1)$$
$$Z = |W| \operatorname{sign}(X \cdot Y),$$

where

$$\operatorname{sign}(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ -1 & \text{if } t < 0. \end{cases}$$

This construction yields that Z is also a standard Normal random variable. Furthermore, the triplet (X, Y, Z) is pairwise independent, i.e., all pairs of random variables (X, Y), (X, Z) and (Y, Z) are independent. However, those three variables are not mutually independent. There are many ways to see this, but one that is visually striking is to look at the distribution of the sum

$$S = X + Y + Z.$$

If one were to conclude (falsely) that because the variables are pairwise independent they are

also mutually independent, then one would expect S to be a Normal random variable (with mean 0 and variance 3). However this is 'very false', as shown on Figure 1.4 where we plot the density function and cumulative distribution function (CDF) of S (obtained through simulations) against that of a Normal($\mu = 0, \sigma^2 = 3$). We note that the distribution of S is rather unusual. Indeed, it is asymmetrical (with a high peak followed by a smaller 'bump') and hence very far from a Normal.

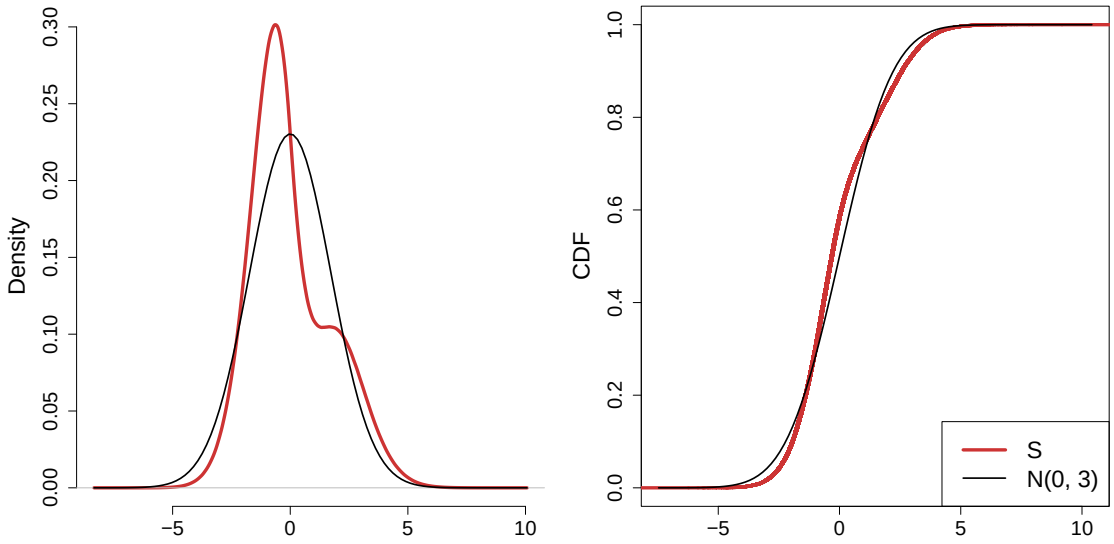

Figure 1.4: Density and CDF of S against that of a Normal($\mu = 0, \sigma^2 = 3$)

Examples 1.2 and 1.3 are toy examples, involving only three random variables (or 'risks'), and one might think that such 'odd behaviour' would not happen for a larger pool of risks. This intuition is however challenged by the fact Central Limit Theorems (CLTs) (and many other results involving large numbers of random variables) can also be violated under pairwise independence (the particular topic of CLTs will be treated in great detail in Chapters 5 and 6). That is to say, the unusual behaviour seen in those examples is not caused by the small sample size. Rather, it is caused by the fundamental difference between pairwise and mutual independence, and we will explore such difference furthermore throughout this thesis.

1.6 Statement of contributions

In previous sections, we have seen that accurately capturing the (possible) dependence between risks (in a variety of different contexts) is an increasingly relevant area of actuarial research. We have also reviewed some important flaws of the famous correlation coefficient (which has been, and still is, broadly used in risk models). Moreover, we established that the independence of all pairs of random variables in a set does not guarantee their mutual independence. This means that many typical dependence models which rely on the characterisation of dependence via pairs can potentially miss some forms of dependence. We have abbreviated as 'PIBD' the phrase 'pairwise independent but still dependent', and we will frequently use this abbreviation throughout the remainder of this thesis.

We now briefly summarise the main contributions this thesis makes, by chapter.

- Chapter 3. We provide new tools to visualise dependence, which are especially suited to PIBD data. We also present many theoretical examples (some of them new) of PIBD random variables and use our visualisation tools to better understand them.

- Chapter 4. We prove that many common dependence models do not allow for PIBD observations, and we show that many results routinely used in actuarial studies and relying on the independence assumption are not valid under sole pairwise independence. Those proofs are, to our knowledge, new.

- Chapter 5. We construct a new sequence of pairwise independent random variables (with arbitrary margins) for which a CLT does not hold. We also derive explicitly the distribution of the standardised mean of that sequence. This allows us to illustrate the extent of the 'failure' of a CLT for PIBD variables. We conduct an analysis of a parameter (which we call 'r') arising in the asymptotic distribution of the sample mean of our sequence. This analysis shows that r is a measure of tail-heaviness. We note some of the content of Chapter 5 has been published, see Avanzi et al. (2021).

- Chapter 6. We propose a general methodology to construct dependent K-tuplewise independent ($K \geq 2$ an integer) sequences of random variables with arbitrary margins. For the case $K = 3$, we use this methodology to derive new explicit examples of triplewise independent sequences for which no CLT hold. We note the content of this chapter has been published, see Boglioni Beaulieu et al. (2021).

To give further motivation for this thesis (and before arriving at our original contributions),

we will survey in Chapter 2 some important areas within actuarial research where the modelling and characterisation of dependence is essential. We will also review existing results around the difference between pairwise and mutual independence.

CHAPTER 2

CHALLENGING THE INDEPENDENCE ASSUMPTION: LITERATURE REVIEW

2.1 Overview

It is traditionally assumed in many actuarial models that risks are mutually independent of each other. However, and as motivated earlier in Chapter 1, in the last few decades many streams of actuarial literature have emerged from the realisation that independence is not always an appropriate assumption. It has also been recognised that dependencies between risks can take many complex forms and that pairwise correlations are not always sufficient to capture such dependencies. It is difficult to give an exhaustive summary of all those streams of literature, but in this section we will give an overview of some important ones. More precisely:

- In Section 2.2, we provide essential definitions about independence and dependence.
- In Section 2.3, we review the topic of risk aggregation for the purpose of setting insurance companies' capital requirements (we especially review the dependence assumptions underlying such aggregation methods). This is also a topic we will revisit in Section 4.2.2.
- In Section 2.4, we review important models and settings of actuarial science where independence has traditionally been assumed, along with recent efforts made to relax this assumption.

This literature review serves a number of purposes:

- To highlight that the concern around possible dependence is at the forefront of many new developments in actuarial science, and in a large breadth of different applications.
- To show that the way dependence is modelled is very varied and increasingly sophisticated.
- To highlight that it is often the case that the models used do not allow for the possibility of PIBD. I.e., within those models, independence by pairs is equivalent to mutual independence (which is not true in general).

In Section 2.5 we review some literature on the difference between pairwise and mutual independence (which is the central theme of this thesis), also highlighting gaps and where this thesis makes original contributions. Lastly, we briefly cover the topic of 'independence tests' in Section 2.6 and 2.A.

2.2 Definitions

The literature review (and the rest of this thesis) will make use of some concepts we prefer to define now in a separate section. The reader could skip this section and refer back to it if/when necessary.

2.2.1 Independence

We start by stating a few classical definitions about 'independence', since this is such a fundamental concept for this thesis. Those can be found in most probability textbooks. We use Resnick (1999) as our main reference.

We first note that the independence of random variables is typically defined via the independence of their generated σ-fields, which is itself defined by the independence of *events* in those σ-fields. Hence, we start by recalling the definition of independent events.

Definition 2.1 (Resnick (1999), Definition 4.1.2)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The events $A_1, \ldots, A_n \in \mathcal{F}$, $(n \geq 2)$ are independent if*

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i), \quad \text{for all finite } I \subset \{1, \ldots, n\}.$$

In words, the probability of an intersection of independent events is equal to the product of all events' probabilities. Next, we use this definition to define the independence of *classes*. If Ω is a sample space, we call a class of Ω any collection of subsets of Ω. Said otherwise, a class is any collection of events in Ω. Obviously, any σ-field of Ω is then a class of Ω.

Definition 2.2 (Resnick (1999), Definition 4.1.3)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{C}_i \subset \mathcal{F}$, $i = 1, \ldots, n$ be classes of Ω. The classes $\mathcal{C}_i$ are independent if for any choice of $A_1, \ldots, A_n$, with $A_i \in \mathcal{C}_i, i = 1, \ldots, n$, we have that the events $A_1, \ldots, A_n$ are independent (in the sense of Definition 2.1).*

Definition 2.2 defines the independence of a *finite* number of classes, and is next extended to the case of an infinitely large number of classes.

Definition 2.3 (Resnick (1999), Definition 4.1.4)**.** *Let T be an arbitrary (possibly infinitely large) index set. The classes $\mathcal{C}_t, t \in T$ are independent if for each finite I, $I \subset T$, $\{C_t, t \in I\}$ is independent (in the sense of Definition 2.2).*

We next turn to the (mutual) independence of random variables. If X is a real-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote by $\sigma(X) \subset \mathcal{F}$ the σ-field generated by X, i.e.,

$$\sigma(X) = \{[\omega \in \Omega : X(\omega) \in A], A \in \mathcal{B}(\mathbb{R})\},$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-field. Of course, $\sigma(X)$ is a class of Ω, and we are ready to state the definition of (mutually) independent random variables.

Definition 2.4 (adapted from Resnick (1999), Definition 4.2.1)**.** *Let $\{X_t, t \in T\}$ be a collection of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where T is an index set (possibly infinitely large). Those random variables are mutually independent if $\{\sigma(X_t), t \in T\}$ are independent σ-fields (in the sense of Definition 2.3).*

Definition 2.4 is somehow technical, and we next state a theorem which gives an alternative (and arguably more intuitive) characterisation of independent random variables, in terms of their cumulative distribution functions.

Theorem 2.1 (Resnick (1999), Theorem 4.2.1)**.** *A collection of random variables $\{X_t, t \in T\}$ indexed by a set T, is mutually independent if and only if for all finite $J \subset T$,*

$$F_J(x_t, t \in J) = \prod_{t \in J} \mathbb{P}[X_t \leq x_t], \quad \forall x_t \in \mathbb{R},$$

where F_J denotes the joint distribution function of the random variables $\{X_t, t \in J\}$.

In words, a *finite* collection of random variable is mutually independent if its joint distribution function is equal to the product of its marginals' distribution functions. An *infinite* collection of random variables is mutually independent if *any* finite subset of variables in this collection is made of (mutually) independent variables.

In the case of *discrete* random variables, their mutual independence can also be characterised via their probability mass functions, as we state in the next theorem.

Theorem 2.2 (Resnick (1999), Corollary 4.2.2)**.** *Let $X_1, \ldots, X_k$ be discrete random variables with a countable range $\mathcal{R}$. They are mutually independent if and only if*

$$\mathbb{P}[X_i = x_i, i = 1, \ldots, k] = \prod_{i=1}^{k} \mathbb{P}[X_i = x_i],$$

for all $x_i \in \mathcal{R}, i = 1, \ldots, k$.

Finally, we define the notion of 'pairwise independence' (Definition 2.5), which is a weaker condition than 'mutual independence'. In words, pairwise independence simply means that all *pairs* of variables in a collection are independent.

Definition 2.5. *Let $\{X_t, t \in T\}$ be a collection of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where T is an index set (possibly infinitely large). Those random variables are pairwise independent if any two of them are independent, i.e., if for any $t_1 \in T$, $t_2 \in T$, $t_1 \neq t_2$, $\sigma(X_{t_1})$ and $\sigma(X_{t_2})$ are independent σ-fields.*

Note that in the literature, 'mutual independence' is often called just 'independence'. Since our purpose is to highlight the difference between 'pairwise' and 'mutual' independence, here we have used (and will keep using) the term 'mutual independence'.

2.2.2 Copulas

Copulas have become a very popular tool in dependence modelling, and they will be mentioned often in this thesis. Hence, we find useful to place here a few fundamental definitions and theorems about copulas. We take them from McNeil et al. (2015, Chapter 7).

Definition 2.6 (Copula)**.** *A d-dimensional copula is a cumulative distribution function on $[0,1]^d$ with standard uniform marginal distributions. We often denote such a distribution function $C(\boldsymbol{u}) = C(u_1, \ldots, u_d)$.*

The famous Sklar's Theorem shows that any multivariate CDF 'contains' a copula, and also that multivariate CDFs with specific margins can be constructed from any given copula.

Theorem 2.3 (Sklar's Theorem)**.** *Let F be a joint distribution function with margins $F_1, \ldots, F_d$. Then, there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, \ldots, x_d$ in $\bar{\mathbb{R}} = [-\infty, \infty]$,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{2.2.1}$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $\operatorname{Ran} F_1 \times \operatorname{Ran} F_2 \times \ldots \times \operatorname{Ran} F_d$, where $\operatorname{Ran} F_i = F_i(\bar{\mathbb{R}})$ denotes the range of F_i. Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined in (2.2.1) is a joint distribution function with margins $F_1, \ldots, F_d$.

For a random vector $\boldsymbol{X}$ with continuous margins, it then makes sense to speak of 'the'

copula of $\boldsymbol{X}$ (since it is unique), see the next definition.

Definition 2.7 (Copula of F)**.** *If a random vector $\boldsymbol{X}$ has joint CDF, F, with continuous marginal distributions $F_1, \ldots, F_d$, then the copula of F (or $\boldsymbol{X}$) is the CDF of $(F_1(X_1), \ldots, F_d(X_d))$.*

An important concept in actuarial science is that of 'comonotonicity', as it corresponds to perfect positive dependence, see Dhaene et al. (2002b) and Dhaene et al. (2002a) for two classical articles on the topic. A copula-based definition of comonotonicity is as follows (again taken from McNeil et al., 2015).

Definition 2.8 (Comonotonicity)**.** *The random variables $X_1, \ldots, X_d$ are said to be comonotonic if they admit as copula de Fréchet upper bound, i.e., the copula*

$$C(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d).$$

The following proposition makes clearer why comonotonicty corresponds to perfect positive dependence.

Proposition 2.1. *Random variables $X_1, \ldots, X_d$ are comonotonic if and only if*

$$(X_1, \ldots, X_d) \stackrel{d}{=} (v_1(Z), \ldots, v_d(Z))$$

for some random variable Z and increasing functions $v_1, \ldots, v_d$.

Hence, if the variables $X_1, \ldots, X_d$ represent 'risks', comonotonicity means that "there is a single source of risk and the comonotonic variables move deterministically in lockstep with that risk" (McNeil et al., 2015, p. 236).

For two random variables X_1, X_2, we can also define a concept of 'perfect negative dependence', called 'countermonotonicity', as formalised in the next definition.

Definition 2.9 (Countermonotonicity)**.** *Two random variables X_1 and X_2 are countermonotonic if they have as copula the Fréchet lower bound, i.e., the copula*

$$C(u_1, u_2) = \max(u_1 + u_2 - 1, 0).$$

Again, we can better see why this corresponds to perfect negative dependence with an alternative characterisation, given in the next proposition.

Proposition 2.2. *Random variables X_1 and X_2 are countermonotonic if and only if*

$$(X_1, X_2) \stackrel{d}{=} (v_1(Z), v_2(Z))$$

for some random variable Z with v_1 increasing and v_2 decreasing, or vice versa.

That is to say, Z is the unique source of randomness (or 'risk'), and since X_1 increases with Z and X_2 decreases with Z (or vice versa), X_1 and X_2 always move in opposite directions. Lastly, note that this concept of countermonotonicity does not generalise to more than two random variables.

2.2.3 Characteristic functions

The *characteristic function* of a random variable (or random vector) is a mathematical tool which will be especially useful for us in Chapters 5 and 6, see the next two definitions.

Definition 2.10. *Given a random variable X, its characteristic function is defined as*

$$\varphi_X(t) = \mathbb{E}[\exp(\mathrm{i}tX)] = \mathbb{E}[\cos(tX)] + \mathrm{i}\mathbb{E}[\sin(tX)], \quad t \in \mathbb{R},$$

where $\mathrm{i} = \sqrt{-1}$ is the imaginary number.

If X is rather a random vector, its characteristic function is defined in an analogous way (using the inner product between vectors).

Definition 2.11. *Given a random vector $\boldsymbol{X} \coloneqq (X_1, \ldots, X_d)'$ of size d, its characteristic function is defined as*

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp\left(\mathrm{i}\boldsymbol{t}'\boldsymbol{X}\right)] = \mathbb{E}[\cos\left(\mathrm{i}\boldsymbol{t}'\boldsymbol{X}\right)] + \mathrm{i}\mathbb{E}[\sin\left(\mathrm{i}\boldsymbol{t}'\boldsymbol{X}\right)], \quad \boldsymbol{t} = (t_1, \ldots, t_d)' \in \mathbb{R}^d,$$

where $\boldsymbol{t}'\boldsymbol{X}$ represent the usual inner product between vectors $\boldsymbol{t}$ and $\boldsymbol{X}$, i.e., $\boldsymbol{t}'\boldsymbol{X} = \sum_{j=1}^{d} t_j X_j$.

2.3 Impact of dependence assumptions on Capital Required

Setting dependence (or independence) assumptions between risks is a crucial step in the establishment of regulatory capital requirements for insurance companies. This section reviews this important topic, where we will discuss relevant regulatory rules, as well as recent actuarial literature.

2.3.1 Problem formulation

In most jurisdictions, regulators require insurance companies to hold certain minimal levels of capital. As described in Tang and Valdez (2009):

> From the insurer's perspective, the purpose of capital is to provide a financial cushion for adverse situations when its insurance losses exceed or asset returns fall below the levels expected. This cushion further enhances the insurer's ability to continue paying claims and in most instances, to continue writing new business even under unfavorable financial circumstances.

To formulate this mathematically, assume $\boldsymbol{X} \coloneqq (X_1, \ldots, X_d)$ are d random variables which represent the potential losses an insurance company will incur over a fixed period of time, for d different risk categories (e.g., losses for different lines of business or different subsidiaries of the company). The total capital required at the company level is often set to be a risk measure $\rho(\cdot)$ on

$$S = X_1 + \cdots + X_d, \tag{2.3.1}$$

the aggregated loss. For an overview of this topic, one can review Dhaene et al. (2005) or McNeil et al. (2015, Chapter 8) and references therein. We note that, since the X_i are potential losses, *bigger* values of X_i equate to *worse* results.

The function $\rho(\cdot)$ is often a translation invariant *risk measure* $\rho(\cdot)$ computed on S. It maps a random risk S to a real number that represents the 'riskiness' of S. Many choices of risk measures are possible (see Dhaene et al., 2006, for a good review of the most common risk measures). The Value-at-Risk (VaR) is the most used in the insurance sector (Abbasi and Guillen, 2013; Bernard and Vanduffel, 2015), and it is also common in the banking/financial sector (Ziegel, 2016). The Value-at-Risk at level α (VaR_α) is simply

the α-quantile of the distribution (call it F_S) of the risk S, i.e.,

$$\mathrm{VaR}_\alpha(S) = \inf\{t \in \mathbb{R} : F_S(t) \geq \alpha\}. \tag{2.3.2}$$

Remark 2.1. *For insurance applications where S in an aggregate loss (hence a larger S is worse) VaR would be evaluated for values of α close to 1. For instance, Solvency II prescribes the use of $\alpha = 0.995$ (EIOPA, 2014). Then, Value-at-Risk has a simple interpretation: it is the amount of capital needed to have a probability of at least 99.5% that losses will not exceed capital (in a given year).*

Remark 2.2. *Another very common risk measure is the Tail Value-at-Risk (TVaR), which we define later, see Equation* (2.3.6)*. A good discussion on some advantages of TVaR can be found in Emmer et al. (2015).*

It is obvious that any dependence between the X's in (2.3.1) will affect the distribution of S, and hence the Capital Required $\rho(S)$. However, it is usually difficult to specify a complete model for the distribution of $X_1, \ldots, X_d$ (including both the marginals of the X's and the dependence between them), such that we can extract the distribution of S precisely. Therefore, simplified risk-aggregation rules are often used instead.

Arguably the most used rule to compute the Capital Required as function of d 'individual' Capitals Required, call them $\mathrm{CR}(X_i)$, is the so-called 'standard formula' (otherwise known as the *correlation adjusted summation* formula see, e.g., McNeil et al., 2015, p. 300). This formula reads

$$\mathrm{CR}\left(\sum_{i=1}^{d} X_i\right) = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d} \rho_{ij}\ \mathrm{CR}(X_i)\ \mathrm{CR}(X_j)}. \tag{2.3.3}$$

Where ρ_{ij} usually denotes the correlation between risks X_i, X_j $(i, j = 1, \ldots, d)$. This formula is convenient because, as long as not all correlations are equal to 1, it automatically induces a diversification benefit (in the sense of Equation 1.2.1). Furthermore, it results in a diversification benefit which is function only of the standalone capitals and the correlations between the risks.

However, in most cases formula (2.3.3) is only an approximation of the 'real' risk measure $\rho(S)$. Indeed, from Proposition 8.29 in McNeil et al. (2015), we know that this formula gives the correct 'global' risk measure of $S = \sum_{i=1}^{d} X_i$ only if the joint distribution of such risks is *elliptical*. But this 'ellipticity' assumption is quite strong, and questionable for

insurance and finance applications. This opinion is shared by many authors, e.g. McNeil et al. (2015, p.301): "[clearly] the elliptical assumption is unlikely to hold in practice", Embrechts et al. (2014): "[u]nfortunately, and this in particular in moments of stress, the world of finance may be highly non-elliptical", or Scherer and Stahl (2021): "the class of elliptical distributions is not compatible with the skewed distribution of typical insurance risks such as NatCat and credit default, to name but a few. In this respect, the class of elliptical distributions is not compatible with the empirical and epistemic (contextual) knowledge about insurance risks".

Hence, there is no guarantee that Equation (2.3.3) would yield a 'correct' required capital. Nonetheless, this formula is used in many regulatory frameworks (which, perhaps, contributes to its popularity). We next review some of those regulatory frameworks.

Remark 2.3. *The standard formula* (2.3.3) *has another obvious flaw (which, given the topic of this thesis, is of special interest to us), in that it cannot distinguish between pairwise and mutual independence. Indeed, if the risks $X_1, \dots, X_n$ are all pairwise independent, then $\rho_{ij} = 0$ for all $i \neq j$. Hence, we would have that in both the cases of mutual and pairwise independence, the formula would set the CR as*

$$\mathrm{CR} = \sqrt{\sum_{i=1}^{d} \mathrm{CR}(X_i)}.$$

Remark 2.4. *It seems that, in practice, building correlation matrices between risk categories often relies on 'informed guesswork' (see, e.g. Avanzi et al., 2018). For example, Avanzi et al. (2016c) mention that in the establishment of risk margins, the Australian industry "frequently relies" on correlations matrices found in Bateup and Reed (2001) or Collings and White (2001), which are "based largely on the judgement of a small number of actuaries".*

2.3.2 Regulatory frameworks

The importance of dependence between large categories or risks is recognised by regulators of the insurance industry around the world, e.g., EIOPA (European Insurance and Occupational Pensions Authority) in the European Union or APRA (Australian Prudential Regulation Authority) in Australia. Indeed, under many existing legislations, insurance companies must assume some level of dependence between business segments and/or risk categories in the calculation of their capital requirements, usually via an aggregation rule

of the type given by the 'standard formula' (2.3.3).

For example, in Australia the general insurance prudential standard on 'Capital Adequacy: Internal Model-based Method' (GPS113, 2019) states:

> In combining components of risk, the model must make appropriate allowance for correlation between risks, particularly correlations in the tail of distributions. A regulated institution wishing to incorporate diversification assumptions in respect of operational risk must demonstrate an adequate process for estimating dependencies (particularly for extreme losses) and must apply conservatism in its assumptions that is commensurate with the uncertainty of those assumptions.

Furthermore, the general insurance prudential standard on 'Capital Adequacy', (GPS110, 2019) gives an explicit formula for the 'aggregation benefit' between asset risk and insurance risk, which is

$$\text{Aggregation benefit} = A + I - \sqrt{A^2 + I^2 + 2 \times \text{correlation} \times A \times I} \tag{2.3.4}$$

where A is the total capital required to cover 'asset risk', (broadly speaking the combination of market and credit risk) and I is the total capital required to cover insurance risk. The correlation assumed here between A and I is set to be 20% "for all insurers except lenders mortgage insurers" and "50 per cent for lenders mortgage insurers". This formula is analogue to formula (2.3.3), only it applies to just two broad categories of risks. The implied assumption made is however the same, that correlation aptly characterises dependence between risks, at least in the sense that it determines what the diversification benefit should be.

To give another example, Canadian regulation on "Canadian property and casualty insurance companies that are not mortgage insurance companies", established by the Office of the Superintendent of Financial Institutions (OSFI) justifies a 'diversification credit' in the following manner (OSFI, 2019):

> Because losses arising across some risk categories are not perfectly correlated with each other, a company is not likely to incur the maximum possible loss at a given level of confidence from each type of risk simultaneously. Consequently, an explicit credit for diversification is permitted between the sum of credit and market risk requirements[.]

This diversification formula is exactly the same as in the Australian regulation, but with a correlation set at 50%.

Another prominent example is Solvency II (the regulatory framework of the insurance business in the European Union), which explicitly uses the standard formula (Bølviken and Guillen, 2017). In a report by the European Insurance and Occupational Pension Authority (EIOPA, 2014) about the main assumptions of Solvency II, we read:

> The underlying assumptions for the correlations in the standard formula can be summarised as follows:
>
> - The dependence between risks can be fully captured by using a linear correlation coefficient approach.
> - Due to imperfections that are identified with this aggregation formula (e.g., cases of tail dependencies and skewed distributions) the correlation parameters are chosen in such a way as to achieve the best approximation of the 99.5% VaR for the overall (aggregated) capital requirement.

Hence, we can see that Solvency II recognises the limits of blindly using the standard formula. Within Solvency II, there is also the possibility for insurance companies to deviate substantially from the standard formula and instead use an 'internal model' (see Eling and Jung, 2020, for a discussion), which nonetheless must be approved by the regulator. We note that this possibility of internal models opens the door to more 'sophisticated' dependence modelling (some common dependence models for a fixed number of risks are reviewed in Section 2.4.1).

That said, one needs to be careful, as picking the 'wrong' dependence assumption can have a substantial impact on the Capital Required of an insurance company. Tang and Valdez (2009) provide a detailed simulation study on the sensitivity of capital requirements to the choice of copula, and conclude that "the choice of copula has a dramatic effect on both the capital requirement and diversification benefit for a multi-line insurer".

Moreover, it can be difficult to choose and/or fit a specific multivariate model for $\boldsymbol{X}$. As Bernard et al. (2014) argue: "there are computational and convergence issues with statistical inference of multidimensional data, and the choice of multivariate distributions is quite limited compared to the modeling of marginal distributions". Instead, many authors have derived worst-case bounds on risk measures on S in (2.3.1) when only partial dependence information is known. This topic of research is commonly referred to as 'risk

aggregation under dependence uncertainty', and we review it in some detail in the next section.

2.3.3 Risk aggregation under dependence uncertainty

A recent stream of literature is interested in finding lower and upper bounds for risk measures (especially for Value-at-Risk) on a sum of dependent risks $S = X_1 + \cdots + X_d$ when the marginals of the X's (and/or their dependence structure) are not fully known. Deriving such bounds is a "natural way to measure model uncertainty" (Puccetti et al., 2017), since they indicate the range of possible values for the risk measure when only partial information about the risks (and their dependence) is known. Embrechts et al. (2013) proposed what they called the Rearrangement Algorithm (RA) to approximate bounds on $\mathrm{VaR}_\alpha(S)$ when the marginal distributions of the risks are known but their dependence structure is totally unspecified. Bernard et al. (2017a) note that

> So far, numerical experiments have shown that the RA presents very good accuracy. However, the gap between upper and lower VaR bounds is wide, a feature that can only be explained by the nonuse of dependence information.

This means we might obtain an upper bound on the VaR which is very high, making it unrealistic for a company to hold as much capital. That said, it may be overly pessimistic to assume *nothing at all* is known about the dependence structure of the risks. Hence, some authors have suggested to incorporate *some* information about dependence in the derivation of bounds for the VaR. For example, Bernard et al. (2017a) derive bounds in the case the variance of S is known to be below a certain level s^2. Defining the *upper* VaR, i.e.,

$$\mathrm{VaR}_\alpha^+(S) = \sup\{t \in \mathbb{R} : F_S(t) \leq \alpha\}, \tag{2.3.5}$$

as well as the Tail Value at Risk (another popular risk measure):

$$\mathrm{TVaR}_\alpha(S) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u^+(S)\mathrm{d}u, \tag{2.3.6}$$

and finally the Left Tail Value at risk,

$$\mathrm{LTVaR}_\alpha(S) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_u(S)\mathrm{d}u, \tag{2.3.7}$$

their main result (Theorem 5) is that for $\alpha \in (0,1), X_i \sim F_i$ $(i = 1, 2, \ldots, d)$ if $\mathbb{V}\text{ar}[S] \leq s^2$ then

$$\max\left(\mu - s\sqrt{\frac{1-\alpha}{\alpha}}, A\right) \leq \text{VaR}_\alpha(S) \leq \min\left(\mu + s\sqrt{\frac{\alpha}{1-\alpha}}, B\right). \tag{2.3.8}$$

(where $\mu = \mathbb{E}[S]$, $A = \sum_{i=1}^d \text{LTVaR}_\alpha(X_i)$ and $B = \sum_{i=1}^d \text{TVaR}_\alpha(X_i)$). Here, perfect knowledge of all the marginal distributions F_i, $i = 1, \ldots, d$ is assumed. Unsurprisingly, we see that as the bound s^2 on the total variance decreases (indicating we have more information about the dependence) the bounds on the VaR become tighter.

Remark 2.5. *We note here that, if the risks $X_1, \ldots, X_d$ are assumed to be pairwise independent (and with known margins), then the variance of S is known exactly. Hence, the results in Bernard et al. (2017a) can be used indirectly to obtain (some) bounds on the VaR of S for the case where the risks are PIBD (though these bounds would not be sharp).*

Many other articles investigated the question of deriving bounds on risk measures in a variety of setting where partial dependence information is included. Bernard and Vanduffel (2015) argue that in practice one often fits a multivariate model to a sample collected from the d risks, but that this model cannot be 'trusted' on the whole sample space $\mathbb{R}^d$. Hence, they propose to split $\mathbb{R}^d$ into two (disjoint) subsets: the 'trusted' region (the one for which we assume the multivariate model is correct) and the 'untrusted' region (the one for which we are not sure the multivariate model is correct). These authors obtain new lower and upper bounds on the Value-at-Risk in this setting.

Bernard et al. (2016) investigate what happens if we *do not* have perfect knowledge of the marginals, but instead we have knowledge about the first few moments of the aggregated risk S. They argue that "in practice, loss statistics may only be available at the portfolio level, or may not be rich enough to derive the marginal distributions of the risks involved". Then, they derive upper and lower bounds on the VaR (as well as TVaR), for the case where we either know with precision the first few moments of S and for the case where we have upper bounds on these moments.

Puccetti et al. (2016) provide an upper bound on the VaR of S when marginals are fixed and partial dependence information is known, in the form of positive dependence on a subset of the domain of the distribution function of the joint risk portfolio. Puccetti et al. (2017) derive lower and upper bounds on the VaR of S when marginals are known and independence is assumed between some subgroups of the risks $X_1, \ldots, X_d$.

Bernard et al. (2017b) provide risk bounds on S for both the VaR and for convex risk measures (e.g., TVaR), in the case where risks $X_1, \dots, X_d$ follow a partially specified factor model. Lux and Papapantoleon (2019) derived bounds on the VaR of S when marginals are known, along with some 'extreme value information' (i.e., the distribution of the minima and maxima of some subsets of the risks $X_1, \dots, X_d$ is assumed known), as well as with information on the copula of $\boldsymbol{X} \coloneqq (X_1, \dots, X_d)$, which is assumed known only on a subset of its domain (or lying in the vicinity of a reference copula). Bernard et al. (2020) derive bounds for the VaR, TVaR and Range Value-at-Risk (RVaR, see Section 2 of Fissler and Ziegel, 2021, for a definition) of S when S is known to be unimodal, and the mean and variance of S are also known. Chen et al. (2022) obtain bounds on the Value-at-Risk of a sum $S = X_1 + X_2$ for two random risks X_1, X_2 which have known marginals, and when in addition it is known that $X_1 \leq X_2$ (almost surely).

This growing body of literature illustrates the importance of dependence information in capital requirement calculations (and how various assumptions can have a large impact on such requirements). To our knowledge, in this area of the literature, no results specifically on pairwise independence have been obtained. Pairwise independence is perhaps reasonable in certain contexts (or at least, more reasonable than mutual independence, since it is a weaker requirement), and we will come back to this question in Section 4.2.2. For now, we review a few more actuarial models and settings where (in)dependence assumptions play a crucial role.

2.4 Actuarial models where independence is often assumed

In this section, we present important models and settings in actuarial science where it has traditionally been assumed that independence holds between certain risks. We also review some recent attempts made to relax this assumption. We stress that the list we present is not exhaustive, and serves to showcase the importance of (in)dependence assumptions in actuarial science, in general (rather than be a comprehensive summary).

2.4.1 The individual risk model

The deterministic sum of n random risks within an insurance portfolio, expressed in its most general form as

$$S = \sum_{i=1}^{n} X_i, \tag{2.4.1}$$

(where $i \in \{1, \ldots, n\}$ represent different insurance contracts) is often referred to as the 'Individual Risk Model' (IRM) (see, e.g., Klugman et al., 2012, Definition 9.2). Here $X_i, i = 1, \ldots, n$ is the total loss for contract i, and S is the total loss for the entire insurance portfolio. It is traditionally assumed that the X's are mutually independent (though not necessarily identically distributed). This independence assumption (along with knowledge of the marginal distributions) then allows one to compute the distribution of S, and hence other quantities of interest such as the Value-at-Risk of S.

Remark 2.1. *We note here that such a deterministic sum* (2.4.1) *has the same form as the random sum discussed in Section 2.3. However, the IRM usually refers to the aggregation of individual risks (at a portfolio level), not of large categories of risk as in the previous section. What we discuss here is then the possible dependence between risks at a much more granular level, with 'n' in* (2.4.1) *potentially very large. Another difference with the literature cited previously in Section 2.3.3, is that here we present papers which make use of some specific dependence models (as opposed to only using partial dependence information).*

Discussing sums of random variables (as in 2.4.1) in an insurance context, Dhaene et al. (2002b) note that "[t]he assumption of mutual independence between the components of the sum is very convenient from a computational point of view, but sometimes not realistic." Aiming at producing more realistic models, a large literature has emerged

where various assumptions are made about the dependence between the X's in (2.4.1). In this literature, the behaviour of S (particularly in its tails) is also frequently studied. We give here a few important examples.

Bäuerle and Müller (1998) propose a model where each risk X_i is an (increasing) function $g(V, G_\nu, Z_i)$ of three quantities: V is a global risk (affecting all risks), G_ν is a class-specific risk (which affects only a subset of the risks) and Z_i is an individual risk.

As seen in Example 1.2, the IRM is sometimes expressed with the random variables $X_1, \ldots, X_n$ in (2.4.1) further decomposed as

$$X_k = I_k B_k, \quad k \in \{1, \ldots, n\} \tag{2.4.2}$$

where each I_k is a Bernoulli random variable (equal to 1 if a claim occurs for policy k), and where B_k is a random variable which represents the claim amount for policy k (given a claim occurs). Again, the customary assumption is that the random variables $I_1, \ldots, I_n, B_1, \ldots, B_n$ are mutually independent, though some authors have introduced dependence between the X's via dependence between the I's. For example, in a model by Cossette et al. (2002), a single event can trigger claims for many, or even all, risks. In another model, they introduce dependence between the I's via copulas. Specifically, they use the Cook-Johnson and Gumbel copulas.

Wüthrich (2003) sets the X_i in (2.4.1) to be dependent via an Archimedean copula (also assuming they share the same continuous distribution), and then studies the tail behaviour of the sum S. The use of Archimedean copulas is motivated as follows:

> Archimedean copulas are interesting in practice because they are very easy to construct, but still we obtain a rich family of dependence structures. Usually they have only one parameter which is a great advantage when one needs to estimate parameters from data.

Those results were extended by Alink et al. (2004) (who gave more explicit expressions for the asymptotic VaR of the sum S), as well as Embrechts et al. (2009). Alink et al. (2005) also assumed an Archimedean copula for $\boldsymbol{X} \coloneqq (X_1, \ldots, X_n)$ and obtained expressions for the asymptotic TVaR of S. More results in this same setting (IRM under an Archimedean copula) were provided by Chen et al. (2012). Note that, given their importance, we will discuss Archimedean copulas in more detail in Section 4.3.2 (and, in particular, we will see that they do not allow for PIBD random variables).

Barbe et al. (2006) extended the results of Wüthrich (2003) and Alink et al. (2004) beyond the case where $\boldsymbol{X}$ has a Archimedean copula. In their setting, each X_k is regularly varying with index $-\beta < 0$, i.e.,

$$\lim_{t\to\infty} \frac{\mathbb{P}[X_k > xt]}{\mathbb{P}[X_k > t]} = \frac{1}{x^\beta}, \quad x > 0,$$

with the vector $\boldsymbol{X} \coloneqq (X_1, \ldots, X_n)$ also multivariate regularly varying of index $-\beta < 0$ (see Resnick, 2004, Theorem 1, for different characterisations of multivariate regular variation). Some of the results in Barbe et al. (2006) were further extended by Kortschak and Albrecher (2009) to the case of non-identically distributed X's dependent via an extreme value copula.

Furman and Landsman (2005) derive explicit expressions for the TVaR of S, in the specific scenario where $\boldsymbol{X}$ has a multivariate Gamma distribution (see their Definition 1 for a definition of the multivariate Gamma). Those results were later extended by Furman and Landsman (2008) to the case where $\boldsymbol{X}$ has a multivariate Tweedie distribution.

Many other results of this sort have been derived, where $\boldsymbol{X}$ is assumed dependent according to specific models, for example a Farlie-Gumbel-Morgenstern copula (Cossette et al., 2013), Sarmanov distribution (Hashorva and Ratovomirija, 2015), multivariate Pareto (Sarabia et al., 2016), phase-type distribution (Furman et al., 2021) or yet again Archimedean copula (Sarabia et al., 2018; Cossette et al., 2018).

Remark 2.6. *Part of the appeal of Archimedean copulas is that they can usually be defined in any dimension, which is essential when modelling dependence between the risks of an insurance portfolio. We also note that Archimedean copulas always induce an exchangeable dependence structure. That is, for variables having an Archimedean copula, the dependence between any subset of those variables (taken in any order) is identical. While 'exchangeability' is a less restrictive (hence, more realistic) alternative to mutual independence, it is still quite limiting in the types of dependence it allows. A flexible and popular alternative for building multivariate copulas (possibly in high dimension) is the so-called pair-copula construction (see, e.g,. Aas et al., 2009), which we will review in Section 4.3.3.*

We note that we are not aware of any actuarial papers discussing the possibility of PIBD sums of random variables (nor did we find papers discussing this possibility in other fields, outside probability and statistics). This will be a topic we will treat in Sections 4.2.1 and 4.2.2, where we will see that such sums can behave much differently than under mutual independence (this point was already illustrated in the Introduction, recall Examples 1.2 and 1.3). We will also see in Chapters 5 and 6 that, in the limit, sums of PIBD variables

are not necessarily Gaussian (as they would be under mutual independence via Central Limit Theorems).

2.4.2 The collective risk model

The so-called 'Collective Risk Model' (CRM) also plays an important role in loss modelling. It is a model for "the amount paid on all claims occurring in a fixed time period on a defined set of insurance contracts" (Klugman et al., 2012, p. 138). The total amount paid (or 'aggregate losses') incurred by an insurer is expressed as

$$S = \sum_{i=1}^{N} X_i, \tag{2.4.3}$$

where N is the (random) number of claims incurred, and for $i \in \{1, \ldots, N\}$, X_i is the ith claim payment amount. Traditionally, it was assumed under this model that:

- Conditionally on N, the random variables $X_1, \ldots X_N$ are mutually independent (and identically distributed).
- The (common) distribution of those X's does not depend on N.
- The distribution of N does not depend on the values of the X's.

Those are traditional assumptions. However, often motivated by empirical findings, many authors have recently proposed to relax those assumptions in various ways.

Czado et al. (2012) introduce dependence by using a bivariate Gaussian copula between the joint distribution of the claim count N and an individual claim size X. Krämer et al. (2013) extend this work by allowing the use of bivariate copulas other than the Gaussian. Vernic et al. (2022) recently proposed to use a bivariate Sarmanov distribution (see, e.g., Ting Lee, 1996, for details on the Sarmanov distribution) to model the dependence between N and the average claim amount $\bar{X}$.

Another approach to introduce dependence in the CRM is to use the claim count N as a covariate when modelling the claim sizes $X_1, \ldots, X_N$. For instance, Frees et al. (2011) use a mixed linear regression model where $\log(X)$ (for X a claim size) depends linearly on N (conditional on N being non-zero). A more general model is proposed in Garrido et al. (2016) who introduce dependence between N and the X's through a Generalised Linear Model (GLM) where N is one of the covariates which affects the expectation of $\bar{X}$, i.e.,

$$\mathbb{E}[\bar{X}|N, x] = g^{-1}(\boldsymbol{x}\beta + \theta N),$$

where g is a link function and $\boldsymbol{x} = (x_1, \ldots, x_p)$ are a set of covariates.

Others have proposed alternative modelling of dependence between frequency and severity, and this seems to be an active area of research, see Cossette et al. (2019), Lee and Shi (2019) Jeong and Valdez (2020), Oh et al. (2021) for some recent developments. Lastly, for results of the type discussed in Section 2.3.3 (i.e., risk aggregation under dependence uncertainty), but under the framework of the CRM, see Liu and Wang (2017).

2.4.3 Ruin theory

Ruin theory is an important topic of actuarial research. The foundational model within this field is the Cramér–Lundberg model (see, e.g., Schmidli, 2017, Section 5.1), which expresses the surplus of an insurance portfolio at time t, call it $R(t)$, as

$$R(t) = u + ct - \sum_{k=1}^{N(t)} X_k, \tag{2.4.4}$$

where u is the initial surplus, c is the (constant) premium rate, and $N(t)$ is a Poisson process (with constant rate λ) which denotes the number of insurance claims in the interval $(0, t]$. The claim sizes $\{X_k, k \geq 1\}$ are i.i.d. positive random variables, also independent of the process $N(t)$. We note that the compound sum in (2.4.4) resembles the CRM in (2.4.3), only now it is a stochastic process. We also note that the more general case where $N(t)$ is any renewal process (not necessarily Poisson) is called the Sparre Andersen model.

Within model (2.4.4), a quantity of interest is the probability of ruin, defined as

$$\psi(u) = \mathbb{P}[R(t) < 0 \text{ for some } t > 0]. \tag{2.4.5}$$

Yet again, many authors have remarked that, in this context, "the independence assumption can be too restrictive in practical applications and it is natural to look for explicit formulas for $\psi(u)$ and related quantities in the presence of dependence among the risks" (Albrecher et al., 2011). We give here a few examples of such developments.

Albrecher and Boxma (2004) introduce dependence by letting the distribution of the time between two claim occurrences depend on the previous claim size. More precisely, they let 'thresholds' $\{T_k, k \geq 1\}$ be i.i.d. random variables (also independent of the claims X_k), and then if a claim X_k is larger than threshold T_k "the time until the next claim is exponentially distributed with rate λ_1, otherwise it is exponentially distributed with rate λ_2". That is to say, the rate of claim arrivals can dependent on the size of a previous claim.

The authors then derive exact formulas for the ruin probability (2.4.5) in this setting.

Many other authors proposed models with dependence between interclaim arrival times and claim sizes. For example, Boudreault et al. (2006) let the distribution of a claim X depend on the time elapsed since the last claim occurred (call this time W), as follows:

$$f_{X|W}(x) = e^{-\beta W} f_1(x) + (1 - e^{-\beta W}) f_2(x),$$

where f_1, f_2 are two arbitrary density functions. This can be understood as follows. If we expect that, for example, a longer elapsed time until the next claim is more likely to generate a larger next claim, we could choose f_2 to be 'riskier' (e.g., with a heavier tail) than f_1.

Albrecher and Teugels (2006) introduce more general dependence between interclaim arrival time and the subsequent claim size. Indeed, they model such dependence with an arbitrary copula (and also in the more general case where the process $N(t)$ in (2.4.4) is a renewal process). They obtain asymptotic results for both finite-time and infinite-time probabilities or ruin.

We note here that such type of dependence (between interclaim time and claim size) is by nature bivariate. A different type of dependence (possibly of a multivariate nature), is that of dependence between claim sizes. For example, Albrecher et al. (2011) formulate a model where the claims sizes $\{X_k, k \geq 1\}$ are dependent as follows: for each n,

$$\mathbb{P}[X_1 > x_1, \ldots, X_n > x_n | \Theta = \theta] = \prod_{k=1}^{n} e^{-\theta x_k}, \tag{2.4.6}$$

where Θ is itself random. That is to say, given $\Theta = \theta$, the X_k's are conditionally independent and distributed as Exponential(θ). Of course, unconditionally they are not independent.

Remark 2.7. *We note that models such as* (2.4.6) *are often called 'mixture' models. They are part of a much broader class of models called 'latent variable models', which are a classical approach to induce dependence between random variables (see, e.g., Bartholomew et al., 2011, for a general reference on this topic).*

Albrecher et al. (2011) show that the above model (2.4.6) has a dependence structure equivalent to an Archimedean survival copula (see their Proposition 2.1). For this general class of models, the probability of ruin $\psi(u)$ can then be obtained explicitly. In a similar

fashion, they also propose to let the intensity λ of the Poisson process $N(t)$ be random. This then makes the arrival times of the claims (rather than their amounts) dependent.

The idea of using 'mixtures' to introduce dependence in the Cramér–Lundberg model (or variations of it) can be found in many other papers, see for example Constantinescu et al. (2011) and Dutang et al. (2013), who also apply this idea to discrete-time versions of model (2.4.4). Constantinescu et al. (2019) also introduced dependence between claim amounts via mixing, but within the 'compound binomial risk model', which is a discrete-time approximation of model (2.4.4), see, e.g., Gerber (1988).

We stress again that this review is not exhaustive, and many more authors have developed ruin theory models under some form of dependence between claim sizes and/or claim arrival times, see, for example, Badescu et al. (2009), Cheung et al. (2010), Ignatov and Kaishev (2012), Willmot and Woo (2012), Zhang et al. (2012), Woo and Cheung (2013), Chadjiconstantinidis and Vrontos (2014), Heilpern (2014), Landriault et al. (2014a), Cossette et al. (2015), Cheung and Woo (2016), Constantinescu et al. (2016), Avram et al. (2016), Eryilmaz and Gebizlioglu (2017), Cheung et al. (2018), Bladt et al. (2019).

Our bottom line is that myriads authors have worked on relaxing the independence assumption traditionally made in ruin theory, and in many different ways.

Remark 2.8. *We note here that the 'mixture' approach to dependence modelling is certainly not unique to applications to ruin theory, and is a very common approach to dependence modelling, in general. For example,...*

2.4.4 Life insurance models

The models presented in the previous three sections (2.4.1-2.4.3) are most typically used in general insurance. In this section, we survey articles treating dependence in the context of life insurance. Within a portfolio of life insurance contracts, it is often assumed that lives are independent. Or at least, this is a routine assumption one encounters in actuarial textbooks when life insurance contracts issued on multiple lives are introduced (see, e.g., Gerber, 1997, Section 8.2).

This can be reasonable on some level, because people who are not related, and sampled from a large population, would be exposed to different mortality risks. However, and as noted by Denuit et al. (2001):

> Standard actuarial theory of multiple life insurance traditionally postulates

> independence for the remaining lifetimes in order to evaluate the amount of premium relating to an insurance contract involving multiple lives. Nevertheless, this hypothesis obviously relies on computational convenience rather than realism. A fine example of possible dependence among insured persons is certainly a contract issued to a married couple.

This question of the dependence between the future lifetimes of spouses (or in general, between two given lifetimes) is only one example of possible dependence between lives. Though, it is an important one which has been studied extensively in the actuarial literature. The dependence modelling approaches used for this purpose are also very varied. They include, for example, 'extreme dependence' via Fréchet-Hoeffding bounds (Denuit and Cornet, 1999; Denuit et al., 2001), Markov models (Denuit and Cornet, 1999; Denuit et al., 2001; Spreeuw and Wang, 2008), semi-Markov models (Ji et al., 2011), correlation (Ribas et al., 2003), copulas (Frees et al., 1996; Carriere, 2000; Denuit et al., 2001; E. Shemyakin and Youn, 2006; Spreeuw, 2006; Luciano et al., 2008, 2016), bivariate Weibull models and a semi-parametric model (Sanders and Melenberg, 2016), mixed proportional hazards model with treatment effects (Lu, 2017) and extended Marshall–Olkin models (Gobbi et al., 2019). Henshaw et al. (2020) also proposed to model the joint mortality within a couple as a non-mean-reverting Cox-Ingersoll-Ross stochastic process.

Since those papers are concerned with the joint survival of two lives, the dependence implied is by nature *bivariate*. Because this thesis is concerned with multivariate dependence (and especially, the possibility of dependence being present even if independence by pairs holds), we do not review those models specifically. Instead, we present in some more detail a few *multivariate* models of lifetime dependencies.

Indeed, some authors have introduced models to better understand and capture the (possible) dependence between more than two lives. Motivation for this can be found, for example, in Alai et al. (2016a), who argue

> The study of lifetime dependence is highly important in actuarial science. A positive pattern of dependence may range from exposure to similar risk-factors among a small group of individuals (say, a couple) all the way to systematic mortality improvements experienced by a population, and hence, the link with longevity risk is noteworthy.

A good example of this might be workers' compensation for members of a team (or multiple teams) of workers in a mine. The workers are then exposed to a common risk which affects

the health and lifetime of all of them.

For the general case of n dependent lives, Dhaene and Goovaerts (1997) studied the impact of dependence between lives for a simplified life insurance portfolio. In their setting, each risk $X_1, \ldots, X_n$ has a (possibly different) two-point distribution, i.e., for $i = 1, \ldots, n$,

$$\mathbb{P}[X_i = 0] = p_i, \qquad \mathbb{P}[X_i = \alpha_i] = 1 - p_i,$$

where $\alpha_i > 0$. This represents a situation where a benefit of α_i is paid if the ith person dies during a reference period (and nothing is paid otherwise). The authors then show that, in this setting, the type of dependence "if a person dies then all persons with lower survival probabilities will die too" yields the riskiest portfolio (in the sense that it yields the largest possible stop-loss premiums).

Milevsky et al. (2006) provide a simple 'pedagogical' example of the effect of dependence on life insurance products. In their setting, they assume the individuals of a population have a *random* probability of survival p, where p has a two-point distribution. This is in contrast with the usual assumption that p is a deterministic quantity, and it induces a dependence between the lives involved. Indeed, in this setting the Bernoulli variables representing the survival of the individuals are not mutually independent. The authors then use this example to illustrate the breakdown of the LLN for a portfolio of insurance policies. That is: contrary to the independent case, the standard deviation of the average insurance payment does not go to zero as the number of policies increases (hence, the diversification benefit is compromised).

A stream of literature has also proposed specific parametric models for the joint distribution of lifetimes. For example, Alai et al. (2013) use a multivariate gamma distribution for this purpose. They also propose estimators for the parameters of their model (which work with possible truncation of the observed data) and assess the impact of such dependence on the valuation of a portfolio of annuities. In their model, individuals are pooled into M pools (each pool is made of individuals sharing common risk factors). $T_{i,j}$ then denotes the survival time of individual $i \in \{1, \ldots, N\}$ in pool $j \in \{1, \ldots, M\}$. The model for the individual lifetimes is then:

$$T_{i,j} = Y_{0,j} + Y_{i,j}, \tag{2.4.7}$$

where

- the $Y_{0,j}$ follow a gamma distribution with common shape parameter γ_0 and rate parameter α_j specific to pool $j \in \{1, \ldots, M\}$.
- the $Y_{i,j}$ follow a gamma distribution with shape parameter γ_j and rate parameter α_j, for $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, M\}$.
- the $Y_{i,j}$ are mutually independent for $i \in \{0, \ldots, N\}$ and $j \in \{1, \ldots, M\}$.

We can see that $Y_{0,j}$ induces a systematic dependence between all lives within pool j. However, within this model, any two lives from different pools are independent. Alai et al. (2016b) offer a generalisation of the previous model, where instead of a multivariate gamma distribution, a multivariate Tweedie distribution is proposed (which allows for more flexibility). The general form of their model is given as in (2.4.7), but with all random variables involved distributed as the more general Tweedie (rather than gamma) distribution. They also develop parameter estimation procedures for the Tweedie model. Alai et al. (2015) extended such estimation procedure to the case of censored observations.

Alternatively, Alai et al. (2016a) proposed to use a 'multivariate Pareto' distribution to model dependent lifetimes. We note that many multivariate extensions of the Pareto distribution have been proposed in the literature, going back at least to Mardia (1962). The version used in Alai et al. (2016a) is simple, in that it only features two parameters ($\alpha > 0$ and $\sigma > 0$). Its joint survival function (call it $\bar{F}$) is also given by a simple expression. Indeed, for $\boldsymbol{X} = (X_1, \ldots, X_n)$ a multivariate Pareto, and $\boldsymbol{x} = (x_1, \ldots, x_n) \in [0, \infty)^n$, we have

$$\bar{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \left(1 + \frac{\sum_{i=1}^{n} x_i}{\sigma}\right)^{-\alpha}.$$

Aside the fact that, in this model, risks are identically distributed, an important restriction is that the same parameter α influences both the shape of marginal distributions and the strength of the dependence. Indeed, reducing α both increases the correlation between any two risks (X_j, X_k), and makes individual risks heavier-tailed.

In closing, we note that none of the multivariate models presented in this section allow for the possibility of PIBD random variables. Said otherwise, within those models, if all variables are pairwise independent, then they are automatically mutually independent. It remains to be seen, however, exactly 'how close' those two types of independence (pairwise and mutual) are. This will be the topic of Section 2.5, where we review some results around this question. This will also help motivate the work done in the rest of this thesis.

2.4.5 Dependence in other actuarial problems

The issue of dependent risks extends much beyond the settings we covered here. In closing this section, we mention briefly two more actuarial areas where a surge of innovations around dependence considerations recently emerged.

The first one is 'loss reserving', i.e., the problem for a general insurer to estimate its total liabilities from *future claims* (claims yet to be paid, but arising from current or old policies). Here, some authors aim at incorporating dependence 'within triangles' (i.e., dependence within cells of a single triangle) but also 'across triangles' (for different business segments). The literature on this topic is fairly new and appears to still be growing, see, e.g., Shi and Frees (2011), Merz et al. (2013), Happ and Wuthrich (2013), Abdallah et al. (2015), Abdallah et al. (2016), Côté et al. (2016), Avanzi et al. (2016b), Hahn (2017), Badounas and Pitselis (2020), Nieto-Barajas and Targino (2021), Araiza Iturria et al. (2021) for recent developments.

Another area of research where the inclusion of dependence seems increasingly relevant is 'optimal reinsurance'. By reinsurance, we mean the practice by which insurance companies sometimes transfer a part of their risks to a reinsurer (in exchange for a premium). Finding an 'optimal' reinsurance strategy, i.e., one that optimises a certain criterion (expected utility of net profits, probability of ruin, etc.) is then an important topic of actuarial research (see for example the book by Albrecher et al., 2017, for a general reference). As stated by Guerra and de Moura (2021), "[i]n most research on optimal reinsurance, independence is assumed. Indeed, for many years dependence has not been considered in research on optimal risk transfer, possibly due to its complexity." Papers investigating this question under some dependence assumptions between risks are, for example, de Lourdes Centeno (2005), Cai and Wei (2012), Cheung et al. (2014), Zhang et al. (2015), Yuen et al. (2015), Ming et al. (2016), Liang and Yuen (2016), Bi et al. (2016), Han et al. (2019), Guerra and de Moura (2021).

2.5 Difference between pairwise and mutual independence

It has long been known that *pairwise* independence among random variables is a necessary *but not sufficient* condition for them to be *mutually* independent. The earliest counterexample can be attributed to Bernšteĭn (1927), followed by a few other authors, see, e.g., Geisser and Mantel (1962); Pierce and Dykstra (1969); Joffe (1974); Bretagnolle and Kłopotowski (1995); Derriennic and Kłopotowski (2000). That said, we could only find a few papers discussing explicitly 'how close' pairwise and mutual independence are, and we present here some of their findings.

Mroz et al. (2021) investigate the flexibility of the so-called 'simplifying assumption' in pair-copula constructions (also called 'vine copulas', see Section 4.3.3 for details). While this is not the main topic of their paper, they provide a result on the 'distance' between a specific 3-dimensional pairwise independent copula, call it C, and the mutual independence copula (in dimension 3), call it Π^3. This distance is found to be

$$d_\infty(C, \Pi^3) = 1/8, \tag{2.5.1}$$

where d_∞ is the metric defined by

$$d_\infty(C_1, C_2) \coloneqq \max_{\boldsymbol{u}\in[0,1]^n} |C_1(\boldsymbol{u}) - C_2(\boldsymbol{u})|,$$

for two n-dimensional copulas C_1, C_2. As a point of reference, the distance between the comonotonic copula in dimension 3 (call it M^3) and Π^3 is

$$d_\infty(M^3, \Pi^3) = \frac{2}{3\sqrt{3}} \approx 0.385. \tag{2.5.2}$$

(This is straightforward to show: we note that for any $\boldsymbol{u} = (u_1, u_2, u_3) \in [0,1]^3$,

$$|M^3(\boldsymbol{u}) - \Pi^3(\boldsymbol{u})| \leq \min(u_1, u_2, u_3) - \min(u_1, u_2, u_3)^3,$$

so we only need to maximise $u - u^3$ for $u \in [0,1]$, which by trivial calculations yields the result).

Equation (2.5.1) hence indicates that there can be a significant difference between pairwise and mutual independence, but of course it is a specific result (i.e., obtained for a specific copula, and using a specific metric).

Nelsen and Ubeda-Flores (2012) provide more general results. Consider three pairwise independent continuous random variables X_1, X_2, X_3 with copula C. Then, for any $\boldsymbol{u} = (u_1, u_2, u_3) \in [0,1]^3$, we have

$$S_3(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq T_3(\boldsymbol{u}), \tag{2.5.3}$$

where

$$\begin{aligned} S_3(\boldsymbol{u}) &= \max(u_3(u_1+u_2-1), u_2(u_1+u_3-1), u_1(u_2+u_3-1), 0), \\ T_3(\boldsymbol{u}) &= \min(u_1u_2, u_2u_3, u_1u_3, (1-u_1)(1-u_2)(1-u_3)+u_1u_2u_3). \end{aligned}$$

Here, S_3 and T_3 are quasi-copulas (see, e.g., Rodríguez-Lallena and Úbeda-Flores, 2009, for a definition of quasi-copulas) which improve the usual Fréchet-Hoeffding bounds. Indeed, denote by M^n and W^n (superscripts denote dimension) the usual upper and lower bounds (respectively) on a copula. That is, for any n-dimensional copula C and any $\boldsymbol{u} \in [0,1]^n$,

$$W^n(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M^n(\boldsymbol{u}).$$

Then, because for all $\boldsymbol{u}$

$$W^3(\boldsymbol{u}) \leq S_3(\boldsymbol{u}), \quad \text{and} \quad T_3(\boldsymbol{u}) \leq M^3(\boldsymbol{u}),$$

we see that (2.5.3) constitutes a tightening of the Fréchet-Hoeffding bounds. Nelsen and Ubeda-Flores (2012) then propose a measure to judge how significant this tightening is. Let Q be a quasi-copula. Define

$$\xi_n(Q) = \frac{\int_{[0,1]^n} Q(\boldsymbol{u})d\boldsymbol{u} - \int_{[0,1]^n} W^n(\boldsymbol{u})d\boldsymbol{u}}{\int_{[0,1]^n} M^n(\boldsymbol{u})d\boldsymbol{u} - \int_{[0,1]^n} W^n(\boldsymbol{u})d\boldsymbol{u}},$$

so that ξ_n "measures how far a given n-quasi-copula Q is from W^n". Then, because

$$\xi_3(T_3) - \xi_3(S_3) = 9/50,$$

(which is significantly smaller than $\xi_n(M^n) - \xi_n(W^n) = 1$), we see that, based on this metric ξ_n, the bounds in (2.5.3) are substantially tighter than the Fréchet-Hoeffding bounds.

For the more general n-variate case, Nelsen and Ubeda-Flores (2012) present the following result. Let $X_1, \ldots, X_n$ be continuous pairwise independent random variables with copula

C. Let $\Pi^n(\boldsymbol{u}) \coloneqq u_1 \times \cdots \times u_n$ be the (mutual) independence copula. Then,

$$\frac{\int_{[0,1]^n} |C(\boldsymbol{u}) - \Pi^n(\boldsymbol{u})| d\boldsymbol{u}}{\int_{[0,1]^n} M^n(\boldsymbol{u}) d\boldsymbol{u} - \int_{[0,1]^n} W^n(\boldsymbol{u}) d\boldsymbol{u}} \tag{2.5.4}$$

goes to 0 as $n \to \infty$. To quote directly from the authors, this means that "the normalised L_1-distance between the copula C of a vector $\boldsymbol{X}$ of pairwise independent random variables and the copula Π^n of the corresponding vector of mutually independent random variables approaches 0 as the dimension increases".

From this, it would seem that, as n increases, pairwise and mutual independence get more and more similar. However, we note from (2.5.4) that the distances between C and Π^n diminishes only relative to the maximal possible distance betwen two copulas, i.e., the denominator in (2.5.4). Using such a criteria, we also have that the distance between the mutual independence copula Π^n and W^n gets to 0 for large n. This is easy to see, as

$$\frac{\int_{[0,1]^n} |\Pi^n(\boldsymbol{u}) - W^n(\boldsymbol{u})| d\boldsymbol{u}}{\int_{[0,1]^n} M^n(\boldsymbol{u}) d\boldsymbol{u} - \int_{[0,1]^n} W^n(\boldsymbol{u}) d\boldsymbol{u}} = \xi_n(\Pi^n) = \frac{(n+1)! - 2^n}{(n! - 1)2^n}, \tag{2.5.5}$$

where the last equality comes directly from Nelsen and Ubeda-Flores (2012). We then conclude that (2.5.5) converges to 0 as $n \to \infty$. But of course, from this we cannot conclude that, for large n, Π^n and W^n are essentially the same (especially since W^n is not even a copula, only a quasi-copula).

Likewise, the convergence of (2.5.4) to 0 is insufficient to conclude whether in a specific context the difference between pairwise and mutual independence would be material. In fact, many asymptotic results valid under mutual independence do not hold under pairwise independence. A prominent example is the classical Central Limit Theorem which, in general, does not hold for PIBD sequences. Specific examples of this can be found in Romano and Siegel (1986, Example 5.45), Bradley (1989), Janson (1988) or Cuesta and Matrán (1991). This will also be the topic of Chapters 5 and 6, where a more detailed literature review on CLTs under pairwise independence (as well as new results) is provided. The novelty in our results is that we build new sequences of pairwise independent (Chapter 5) and triplewise independent (Chapter 6) random variables with *arbitrary* marginal distribution, yet a known asymptotic distribution for the standardised mean of those sequences (seen to be non-Gaussian, and heavier-tailed than a Gaussian).

2.6 Independence tests

Lastly, we should mention that one can check 'formally' for dependence in a dataset via *statistical tests of independence*. Developing such tests is an active topic of research in statistics, and one we thought relevant to mention here, albeit briefly.

The problem of testing the independence between *two* random variables (or vectors) is not new, and has been widely researched in the literature (for reviews, see Josse and Holmes, 2016; Tjøstheim et al., 2022). However, even powerful tests of bivariate independence can fail to detect the types of dependence we are concerned about in this thesis (since, for PIBD variables, there is simply *no* dependence to be found if only pairs of variables are considered). To test the independence of *more than two* random variables at a time, a few tests have also been proposed, e.g., Fan et al. (2017), Pfister et al. (2018), Jin and Matteson (2018), Chakraborty and Zhang (2019) or Böttcher et al. (2019). For the interested reader, we place in Appendix 2.A a description of some of those tests. Here, we simply provide a general description of the statistical problem at hand.

Consider d random variables $\boldsymbol{X} \coloneqq (X_1, \dots, X_d)$, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We want a procedure to test the null hypothesis

$$H_0 : X_1, \dots, X_d \text{ are mutually independent,} \tag{2.6.1}$$

against the alternative

$$H_1 : X_1, \dots, X_d \text{ are not mutually independent.}$$

We denote by $\tilde{\boldsymbol{X}} \coloneqq (\tilde{X}_1, \dots, \tilde{X}_d)$ a random vector with mutually independent components and such that for every $j \in \{1, \dots, d\}$, $\tilde{X}_j \overset{d}{=} X_j$. That is to say, under H_0, $\tilde{\boldsymbol{X}}$ and $\boldsymbol{X}$ have the same distribution. To test H_0, many existing statistical tests rely on measuring a certain notion of 'distance' between the distribution of $\boldsymbol{X}$ (call it $\mathbb{P}^{\boldsymbol{X}}$) and that of $\tilde{\boldsymbol{X}}$ (call it $\mathbb{P}^{\tilde{\boldsymbol{X}}}$). That is, those tests are based on a statistical functional of the form

$$T(\mathbb{P}^{\boldsymbol{X}}) = \text{dist}\left(\mathbb{P}^{\boldsymbol{X}}, \mathbb{P}^{\tilde{\boldsymbol{X}}}\right), \tag{2.6.2}$$

where $\text{dist}(\cdot, \cdot)$ is some notion of distance between two multivariate distributions. Provided that $\text{dist}\left(\mathbb{P}^{\boldsymbol{X}}, \mathbb{P}^{\tilde{\boldsymbol{X}}}\right) = 0$ if and only if $\mathbb{P}^{\boldsymbol{X}} = \mathbb{P}^{\tilde{\boldsymbol{X}}}$, we have that $T(\mathbb{P}^{\boldsymbol{X}}) = 0$ characterises

the mutual independence of $\boldsymbol{X}$. Many choices of functional T are possible, and for at least two reasons:

1. Many functions uniquely characterise the distribution $\mathbb{P}^{\boldsymbol{X}}$ of a random vector: its characteristic function, its distribution function, its density function (provided it exists), etc.

2. Many notions of 'distance' between two functions can be chosen.

In practice, the distribution of $\boldsymbol{X}$ is unknown, and hence the quantity T is unknowable. From an observed i.i.d. sample of size n,

$$\boldsymbol{X_1} = (X_{11}, \ldots, X_{1d}), \ldots, \boldsymbol{X_n} = (X_{n1}, \ldots, X_{nd}), \tag{2.6.3}$$

one can approximate the distributions $\mathbb{P}^{\boldsymbol{X}}$ and $\mathbb{P}^{\tilde{\boldsymbol{X}}}$ with some empirical counterparts (call those $\widehat{\mathbb{P}}^{\boldsymbol{X}}$ and $\widehat{\mathbb{P}}^{\tilde{\boldsymbol{X}}}$) and obtain an empirical version of T, denote it

$$\widehat{T}_n(\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}) = \text{dist}\left(\widehat{\mathbb{P}}^{\boldsymbol{X}}, \widehat{\mathbb{P}}^{\tilde{\boldsymbol{X}}}\right).$$

For a predetermined level α, if $\widehat{T}_n$ exceeds its H_0 $(1-\alpha)$ quantile, then mutual independence is rejected. Many different approaches to this problem are possible (e.g., different functions characterising $\mathbb{P}^{\boldsymbol{X}}$, different distances in (2.6.2), different estimators of $\widehat{\mathbb{P}}^{\boldsymbol{X}}$, etc.), and hence several different tests have been proposed in recent years, some of which are described in Appendix 2.A.

Remark 2.9. *We are not aware of any large study assessing which multivariate independence tests perform better under specific dependence scenarios. Hence, we think it is not obvious which test one should choose, in practice, to detect dependence for pairwise independent variables. We also note that statistical tests, upon rejection of independence, do not tell a story about the shape of the dependence. In this regard, adequate visualisation of the data can help. This partly motivates the next chapter of this thesis, where we will develop not statistical tests, but visualisation tools to detect, visually, dependence in a dataset. Those visualisation tools will be especially relevant to the case of PIBD variables.*

Remark 2.10. *Some statistical tests can be used to detect dependence not only between random variables, but between random vectors, see e.g., Beran et al. (2007); Heller et al. (2012); Fan et al. (2017); Bilodeau and Nangue (2017); Jin and Matteson (2018); Shi et al. (2022).*

To simplify matters, let us consider the case of two vectors, $\boldsymbol{X} \in \mathbb{R}^p$ and $\boldsymbol{Y} \in \mathbb{R}^q$. That is, $\boldsymbol{X}$ is p-dimensional and $\boldsymbol{Y}$ is q-dimensional, for p, q two positive integers, and we could write them as

$$\boldsymbol{X} = (X_1, \ldots, X_p), \qquad \boldsymbol{Y} = (Y_1, \ldots, Y_q).$$

Testing the independence between two (or more) random vector is relevant when one is not concerned about dependence within the components of the vectors, but rather about possible dependence between vectors.

As a possible actuarial application, consider the Collective Risk Model (described in Section 2.4.2), where the aggregate risk is a sum of a random number N of individual 'severities' $X_1, \ldots, X_N$. As mentioned in Section 2.4.2, the traditional assumption of independence between N and the severities $\{X_1, X_2, \ldots\}$ is not always realistic. Such an assumption could be checked with a statistical test of independence, testing whether the variable N and the vector $\boldsymbol{X} \coloneqq (X_1, \ldots, X_N)$ are dependent. We note, however, that this situation presents an additional complexity: the size N of the vector $\boldsymbol{X}$ is not deterministic. Hence, it is not clear whether usual tests apply here, or if a new test should be developed. This question is left for future research.

2.7 Summary of literature review

Along this chapter, we have seen that understanding the possible dependence between risks is a central issue of actuarial research, and in a variety of settings. In Section 2.3, we saw that it is common for insurance companies to establish capital requirements by aggregating 'standalone capitals' (e.g., from different risk categories or business segments). This is commonly done via the so-called 'standard formula' (2.3.3), which makes explicit use of the correlations between the risks. This formula is also prescribed under regulatory frameworks. A prominent example is Solvency II, which however also allows insurance companies to use 'internal models' instead of the standard approach. This opens the door to more 'sophisticated' dependence modelling.

Motivated by the fact it can be hard to fit multivariate dependence models to data, a body of literature investigates 'risk aggregation under dependence uncertainty'. That is, it tries to establish how large (or small) a risk measures $\rho(S)$ can get, for $S = X_1 + \cdots + X_n$ a sum of risks, when only partial dependence information is known. This was reviewed in Section 2.3.3.

In Section 2.4, we reviewed some important models in actuarial science where independence assumptions are traditionally made. We also saw that a large body of literature has developed with the aim to relax such independence assumptions. In particular, we reviewed the Individual Risk Model, the Collective Risk model, as well as Ruin Theory and Life Insurance problems, and saw that the approaches developed to model dependence are numerous and varied.

The fact that dependence modelling has become such an important topic in actuarial research, along with the fact that little seems to be known about 'how close' pairwise and mutual independence are (a question surveyed in Section 2.5) gives motivation for the rest of this thesis. In the next chapter, we give many more examples of PIBD random variables and develop visualisation tools to better 'see' this type of dependence. In Chapter 4, we investigate specific situations relevant to actuarial science where there is a material difference between pairwise and mutual independence. In Chapters 5, we then investigate more precisely the case of 'Central Limit Theorems' under pairwise independence. In particular, we provide new instances of infinite sequences of PIBD random variables which do not have an asymptotic Gaussian mean. In Chapter 6, we present extensions of those results to 'triplewise independent' variables, and also derive a general methodology to build

sequences of 'K-tuplewise independent' random variables (for arbitrary integer $K \geq 2$). This general methodology allows others to derive more examples, perhaps suiting other purposes.

2.A Some multivariate independence tests

In this section, we review a few recent multivariate independence tests, i.e., statistical tests for the null hypothesis as given by (2.6.1). In what follows, we let $F_{\boldsymbol{X}}(\cdot)$ be the joint CDF of $\boldsymbol{X} \coloneqq (X_1, \ldots, X_d)$, while $F_j(\cdot)$ denotes the marginal CDF of X_j, for $j \in \{1, \ldots, d\}$. For $\boldsymbol{t'} \coloneqq (t_1, \ldots, t_d)' \in \mathbb{R}^d$ we also denote by $\varphi_{\boldsymbol{X}}(\boldsymbol{t})$ and $\varphi_{\tilde{\boldsymbol{X}}}(\boldsymbol{t})$ the characteristic functions of $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, respectively, i.e.,

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}\left[\exp(\mathrm{i}\boldsymbol{t'}\boldsymbol{X})\right] \quad \text{and} \quad \varphi_{\tilde{\boldsymbol{X}}}(\boldsymbol{t}) = \prod_{j=1}^{d} \mathbb{E}\left[\exp(\mathrm{i}t_j X_j)\right]. \tag{2.A.1}$$

The Euclidean norm of a vector of real values $\boldsymbol{t} \in \mathbb{R}^d$ is denoted by $|\boldsymbol{t}|_d$. For a complex number $z \in \mathbb{C}$, we denote its squared modulus as $|z|^2$. Recall that $|z|^2 = z\bar{z}$ for $\bar{z}$ the complex conjugate of z.

Genest and Rémillard (2004) use the empirical copula process to design their tests, which yields 'simple' test statistics using only the ranks of the collected sample. Those test statistics are given by

$$T_{A,n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{n} \prod_{j \in A}^{n} \left\{ \frac{2n+1}{6n} + \frac{R_{ij}(R_{ij}-1)}{2n(n+1)} + \frac{R_{kj}(R_{kj}-1)}{2n(n+1)} - \frac{\max(R_{ij}, R_{kj})}{n+1} \right\}, \tag{2.A.2}$$

where R_{ij} denotes the rank of X_{ij} among $X_{1j}, X_{2j}, \ldots, X_{nj}$, i.e., if X_{ij} is the smallest among those observations, than $R_{ij} = 1$ and so on. Here A denotes any subset of the total d variables. Therefore (2.A.2) actually defines $2^d - d - 1$ statistics, which can be used for 'individual' tests of independence, i.e., tests of independence among any particular subset of the d variables. For instance, the test statistic $T_{\{1,2,3\},n}$ can be used to test if X_1, X_2 and X_3 are independent. Genest and Rémillard (2004) also propose a way to combine the p-values of all those statistics, which yields a more powerful test of mutual independence among all the d variables.

Ghoudi et al. (2001) proposed an independence test defined for continuous random variables. This test was generalised to random vectors (with possibly discrete components) by Beran et al. (2007). Those tests use the cumulative distribution function (CDF) as their main ingredient, and are based on the following characterisation of mutual independence. Let $\mathcal{I}_d = \{A \subset \{1, \ldots, d\} : |A| > 1\}$ where $|A|$ is the cardinality of set A. For $\boldsymbol{t} \in \mathbb{R}^d$ and

for any $A \in \mathcal{I}_d$, define

$$\mu_A(\boldsymbol{t}) = \sum_{B \subset A} (-1)^{|A \setminus B|} F_{\boldsymbol{X}}(\boldsymbol{t}_B) \prod_{j \in A \setminus B} F_j(t_j), \tag{2.A.3}$$

where a null product $\prod_{\emptyset} = 1$ and where the vector $\boldsymbol{t}_B = (t_B^{(1)}, \ldots, t_B^{(d)}) \in \mathbb{R}^d$ is defined as

$$t_B^{(j)} = \begin{cases} t_j, & \text{if } j \in B \\ \infty, \text{otherwise.} \end{cases}$$

Mutual independence between $X_1, \ldots, X_d$ then holds if and only if $\mu_A(\boldsymbol{t}) = 0$ for all $\boldsymbol{t} \in \mathbb{R}^d$ and all $A \in \mathcal{I}_d$. Because knowledge of all $\mu_A(\boldsymbol{t})$ implies knowledge of the distribution $\mathbb{P}^{\boldsymbol{X}}$, one can use the functions μ_A to measure a 'departure from independence' in the style of (2.6.2). Call $V_{n,A}(\boldsymbol{t})$ an appropriate empirical version of (2.A.3). Ghoudi et al. (2001) propose a statistic of the Cramér-von Mises type

$$T_{n,A} = \int V_{n,A}^2(\boldsymbol{t}) \mathrm{d}F_n(\boldsymbol{t}),$$

which aggregates the 'evidence against H_0' over all possible values of $\boldsymbol{t}$ (but only for the variables X_j's with $j \in A$). Because for some sets $A \in \mathcal{I}_d$ it may be that $\mu_A(\boldsymbol{t}) = 0$ even if H_0 is false, a test of mutual independence must consider all sets A at the same time. Ghoudi et al. (2001) propose global statistics of the form

$$\sum_A T_{n,A} \quad \text{or} \quad \max_A \{T_{n,A}\},$$

while Beran et al. (2007) combine the p-values from individual tests (one for each set A) in the manner of Fisher (see, e.g. Elston, 1991).

Fan et al. (2017), Bilodeau and Nangue (2017), as well as Jin and Matteson (2018) use a definition of T in (2.6.2) based on characteristic functions, i.e.,

$$T(\mathbb{P}^{\boldsymbol{X}}) = \int_{\mathbb{R}^d} \left| \varphi_{\boldsymbol{X}}(\boldsymbol{t}) - \varphi_{\tilde{\boldsymbol{X}}}(\boldsymbol{t}) \right|^2 w(\boldsymbol{t}) d\boldsymbol{t}, \tag{2.A.4}$$

where $w(\boldsymbol{t})$ is a certain weight function (for which the integral exists). Fan et al. (2017, Section 5) suggest five choices of weight function, all of the form:

$$w(\boldsymbol{t}) = \prod_{j=1}^{d} v(t_j),$$

for v a non-negative, continuous and symmetric function (i.e., $v(t) = v(-t)$ for all $t \in \mathbb{R}$). Jin and Matteson (2018) suggest a different weight function (implemented in the R package `EDMeasure`), defined as

$$w(\boldsymbol{t}) = \left(K_d \, |\boldsymbol{t}|_d^{d+1}\right)^{-1},$$

where $K_d = \pi^{(d+1)/2}/\Gamma[(d+1)/2)]$ and Γ is the gamma function. The derivation of appropriate empirical versions of $T(\mathbb{P}^{\boldsymbol{X}})$ in (2.A.4) (and corresponding approximation of its null distribution in order to compute p-values) is not straightforward and details are deferred to the original paper.

Pfister et al. (2018) propose a *multivariate* independence test based on a functional they call the d-variable Hilbert-Schmidt independence criterion (dHSIC), which conforms to the general form (2.6.2). It is a generalisation of the bivariate ($d = 2$) independence test proposed in (Gretton et al., 2007), and is based on the *mean embedding* of probability distributions into *reproducing kernel Hilbert spaces* (RKHSs). In order to present this test, it is necessary to first outline the theory of RKHSs. We largely follow the presentation in Pfister et al. (2018) (with perhaps small changes in notation for consistency). We start with the definition of a positive semidefinite kernel[1].

Definition 2.12. *Given a set $\mathcal{X}$, a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive semidefinite kernel if for any set of points $(x_1, \ldots, x_n) \in \mathcal{X}^n$ the $n \times n$ matrix whose entry (i, j) equals $k(x_i, x_j)$ is positive semidefinite, i.e., if*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(x_i, x_j) \geq 0,$$

for any $c_1, \ldots, c_n \in \mathbb{R}$.

Denote by $\mathcal{F}(\mathcal{X})$ the space of functions from $\mathcal{X}$ to $\mathbb{R}$. A RKHS on $\mathcal{X}$ is a subclass of $\mathcal{F}(\mathcal{X})$, which, in particular, is a Hilbert space (i.e., a complete vector space equipped with an inner product, see, e.g., Kreyszig, 1978, Chapter 3) with some useful properties for the purpose of independence testing. We next define RKHSs.

Definition 2.13. *Let $\mathcal{X}$ be a set and let $\mathcal{H} \subseteq \mathcal{F}(\mathcal{X})$ be a Hilbert space with inner product denoted by $\langle f, g \rangle_{\mathcal{H}}$ for $f, g \in \mathcal{H}$. Then, $\mathcal{H}$ is called a RKHS if there is a kernel k on $\mathcal{X}$ satisfying:*

a. $\forall\, x \in \mathcal{X} : k(x, \cdot) \in \mathcal{H}$, and

[1] We use the term positive *semidefinite* kernel to stick to the convention in Pfister et al. (2018), but several authors would use the term positive *definite* kernel instead.

b. $\forall f \in \mathcal{H}, \forall x \in \mathcal{X} : \langle f, k(x,\cdot)\rangle_{\mathcal{H}} = f(x)$. This is often called the 'reproducing property', and we call k a reproducing kernel of $\mathcal{H}$.

Importantly, a RKHS uniquely determines a positive semidefinite kernel k, and vice versa (see Muandet et al., 2016, Theorem 2.5).

The dHSIC test is based on embedding multivariate probability functions ('complicated objects') into a RKHS (which consists of 'simpler objects'). This is useful because, once we are in a RKHS, computations are easier. For instance, via the reproducing property the computation of an inner product reduces to a simple function evaluation. Denoting

$$\mathcal{M}(\mathcal{X}) \coloneqq \{\mu | \mu \text{ is a finite Borel measure on } \mathcal{X}\},$$

we next define a *mean embedding* function (which can be applied to any Borel measure).

Definition 2.14. *Let $\mathcal{X}$ be a separable metric space, k a continuous bounded positive semidefinite kernel and $\mathcal{H}$ the RKHS with reproducing kernel k. The mean embedding (associated with k) $\Pi : \mathcal{M}(\mathcal{X}) \to \mathcal{H}$ is a function defined as*

$$\Pi(\mu) \coloneqq \int_{\mathcal{X}} k(x,\cdot)\mu(dx). \tag{2.A.5}$$

Remark 2.2. *In the general case where $\mathcal{X}$ is any separable metric space, the integral in* (2.A.5) *should be interpreted as a Bochner integral. However, for our purposes we can see it as the usual Lebesgue integral on a set $\mathcal{X} \subseteq \mathbb{R}$.*

Remark 2.3. *$\Pi(\mu)$ is a function in the space $\mathcal{H}$, hence it can be seen as a 'simpler' object than a probability measure (and also, arguably, easier to estimate).*

We next want to embed the probability distributions $\mathbb{P}^{\boldsymbol{X}}$ and $\mathbb{P}^{\tilde{\boldsymbol{X}}}$ into an appropriate RKHS and, in that space, check if the embedded elements are equal, as in (2.6.2). We consider the following setting.

For $j \in (1,\dots,d)$, $X_j : \Omega \to \mathcal{X}^j$ where $\mathcal{X}^j$ is a separable metric space (in our case, think simply of $\mathbb{R}$). Further, let $\boldsymbol{\mathcal{X}} = \mathcal{X}^1 \times \cdots \times X^j$ be the product space. For $j \in (1,\dots,d)$, let $k^j : \mathcal{X}^j \times \mathcal{X}^j \to \mathbb{R}$ be continuous, bounded and positive semidefinite kernels. Denote by $\mathcal{H}^j$ their corresponding RKHS. Define $\boldsymbol{\mathcal{H}} = \mathcal{H}^1 \otimes \cdots \otimes \mathcal{H}^d$ as the projective tensor product of the RKHSs $\mathcal{H}^j$'s and $\boldsymbol{k} = k^1 \otimes \cdots \otimes k^d$ as the tensor product of the kernels k^j's (for details on tensor products see Berlinet and Thomas-Agnan, 2004, Section 4.6). Further, assume that $\boldsymbol{k}$ is characteristic (see Muandet et al., 2016, Definition 3.2). Then,

let $\Pi : \mathcal{M}(\boldsymbol{\mathcal{X}}) \to \boldsymbol{\mathcal{H}}$ be the mean embedding function associated with kernel $\boldsymbol{k}$ as in (2.A.5).

In this setting, we then have that $\boldsymbol{\mathcal{H}}$ is a RKHS with reproducing kernel $\boldsymbol{k}$, and that Π is injective. We can now define the dHSIC statistical functional

$$\text{dHSIC}(\mathbb{P}^{\boldsymbol{X}}) \coloneqq ||\Pi(\mathbb{P}^{\tilde{\boldsymbol{X}}}) - \Pi(\mathbb{P}^{\boldsymbol{X}})||^2_{\boldsymbol{\mathcal{H}}}, \tag{2.A.6}$$

where $||\cdot||_{\mathcal{H}}$ is the norm induced by the inner product on the space $\boldsymbol{\mathcal{H}}$. Equation (2.A.6) is then seen to be a distance between $\mathbb{P}^{\boldsymbol{X}}$ and $\mathbb{P}^{\tilde{\boldsymbol{X}}}$ after they have been embedded in the RKHS $\boldsymbol{\mathcal{H}}$. Importantly, since Π is injective we have (Pfister et al., 2018, Proposition 1):

$$\text{dHSIC}(\mathbb{P}^{\boldsymbol{X}}) = 0 \quad \iff \quad \mathbb{P}^{\boldsymbol{X}} = \mathbb{P}^{\tilde{\boldsymbol{X}}}.$$

An empirical version of dHSIC (and associated independence test) is then developed in Pfister et al. (2018), and we refer the reader to that paper for the details (see Section 2.3 and Section 3). We note that three options are proposed to conduct the test and derive p-values: 'permutation', 'bootstrap' and 'Gamma approximation', and the former is the default option in the R package `dHSIC`.

Remark 2.4. *The dHSIC independence test (and, in general, methods based on RKHSs) is very general, as it is defined for random elements taking values in separable metric spaces $\mathcal{X}$ (which are generalizations of the usual Euclidean spaces). In this thesis we deal with real random variables, so it does not hurt to see sets $\mathcal{X}$ as simply $\mathbb{R}$. However, one should keep in mind that an important strength of kernel methods is that they apply more generally to any structured data, which also explains their popularity in machine learning. As an example, Gretton et al. (2007) apply the bivarite version of dHSIC to test the independence between English texts and their French translation.*

CHAPTER 3

UNDERSTANDING PIBD VARIABLES WITH EXAMPLES AND NEW VISUALISATION TOOLS

3.1 Introduction

The notion that random variables can be *pairwise independent but still dependent* (PIBD) is somewhat abstract. It is often mentioned as a comment in probability textbooks, with perhaps a toy example to prove the point, but not discussed further. Given the ubiquitous importance of dependence (or independence) assumptions in actuarial science, the broad goal of this chapter is to make this notion more concrete. In particular, we want to illustrate as best as we can what PIBD data may look like. We do this following two main steps:

1. we provide a series of theoretical examples (with accompanying visualisations) where three random variables are PIBD. This serves to illustrate what types of dependence patterns are possible if pairs of observations are independent. In particular, we will see that many different patterns are possible, and that the underlying dependence can possibly be very strong. This is done in Section 3.2;

2. we develop visualisation tools to better 'see' this possible type of dependence in our data. This is done in Section 3.3, where we also apply such tools to synthetic data stemming from the examples of Section 3.2.

Remark 3.1. *In the subsequent Chapters 4, 5 and 6 of this thesis, we will provide more theoretical results about the difference between mutual and 'pairwise only' independence. That said, we thought important, as a first step, to establish what this type of dependence can look like (especially since we have not encountered such visualisation elsewhere in the literature we surveyed).*

3.2 Examples of PIBD random variables

From a technical definition of 'pairwise independence' like Definition 2.5, it can be difficult to grasp what PIBD data can look like. In this section, we address this issue by presenting a series of PIBD examples (with visualisations). We focus on the case of three Uniform$[0,1]$ random variables (U_1, U_2, U_3), a sample of which can be represented on a 3D scatterplot. The use of Uniforms is a natural choice, as it is widely recognised that the study of dependence should be divorced from the specific margins of the random variables involved. Furthermore, we know that a sample from three mutually independent Uniform$[0,1]$ random variables should fill the $[0,1]^3$ cube uniformly. This provides an easy benchmark to visually detect dependence: any significant deviation from a 'constant density of points' indicates some form of dependence. We also note that the use of Uniforms is done without loss of generality, since a continuous random variable X can always be transformed to a Uniform via the probability integral transform, i.e.,

$$F(X) \sim \text{Uniform}(0,1)$$

where $F(\cdot)$ is the CDF of X.

In Section 3.2.1, we present our 'main examples', while in Section 3.2.2 we establish that an infinity of new examples can be obtained by simply *mixing* existing ones.

3.2.1 Main examples

Example 3.1. *Our first example is taken from Driscoll (1978). Let U_1 and U_2 be independent Uniform$[0,1]$ r.v.s, and let*

$$U_3 = (U_1 + U_2) \bmod 1 = \begin{cases} U_1 + U_2 & \text{if } U_1 + U_2 < 1 \\ U_1 + U_2 - 1 & \text{if } U_1 + U_2 \geq 1. \end{cases} \tag{3.2.1}$$

Then, U_3 is also a Uniform$[0,1]$. Furthermore, U_1 is independent of U_3, and U_2 is independent of U_3. Scatterplots based on $n = 2{,}000$ observations confirm this pairwise independence; see Figure 3.1. However, the triplet of variables (U_1, U_2, U_3) is strongly dependent: knowing two of them is sufficient to deduce the third precisely. This dependence can be seen clearly on Figure 3.2, which displays the same sample, but on a 3D scatterplot. We notice an extreme form of dependence: all points lie on a subset of null volume, i.e., on

two parallel planes.

In addition to the 3D scatterplot, we provide a 2D scatterplot of U_1 versus U_2 (see Figure 3.6) where the value of U_3 is represented by a colour (whose scale is on the right of the figure). Likewise, we present on Figure 3.3 a 2D scatterplot of U_1 versus U_3 (with U_2 represented as a colour). We omit to present U_2 versus U_3, since the pattern is identical to that of U_1 versus U_3.

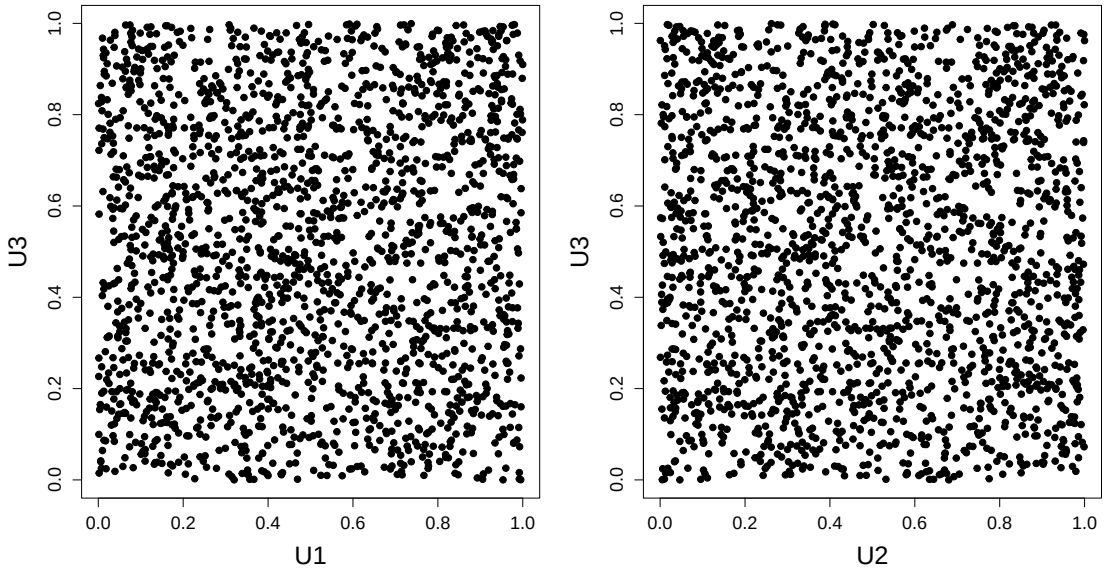

Figure 3.1: 2D scatterplots (U_3 vs U_1 on the left, U_3 vs U_2 on the right) of a sample generated from (3.2.1)

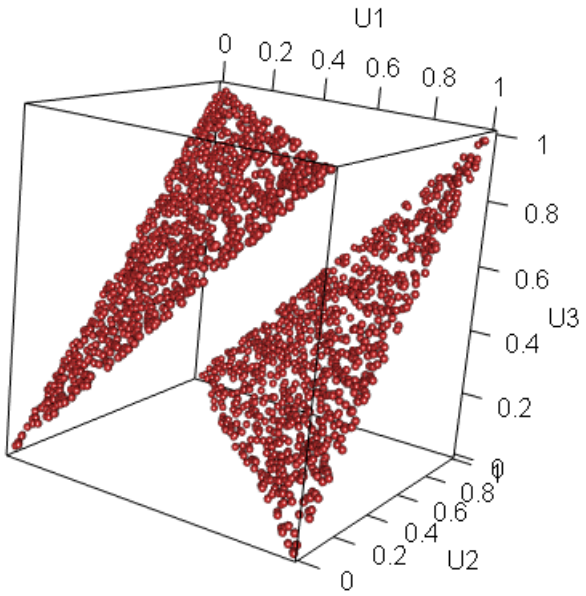

Figure 3.2: 3D scatterplots of a sample (U_1, U_2, U_3) generated from (3.2.1)

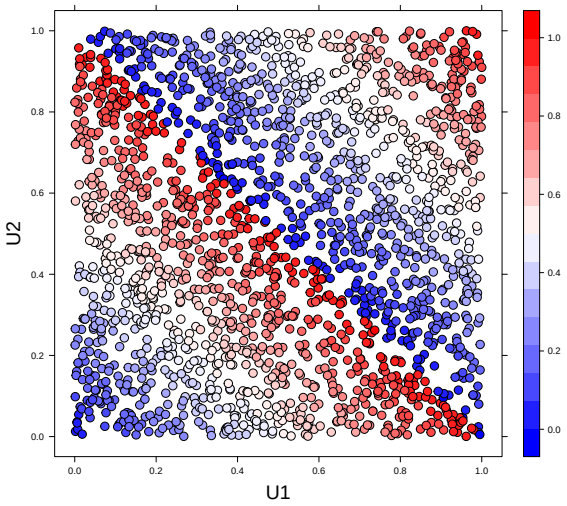

Figure 3.3: 2D scatterplot (U_1 vs U_2) of a sample generated from (3.2.1). The colour of each sample point represents the value of the third variable (U_3).

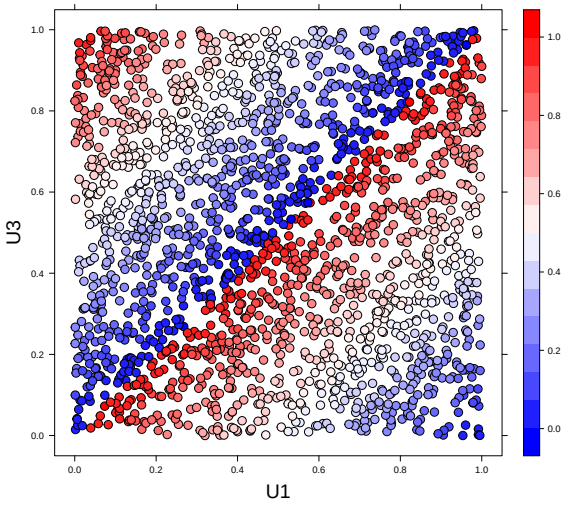

Figure 3.4: 2D scatterplot (U_1 vs U_3) of a sample generated from (3.2.1). The colour of each sample point represents the value of the third variable (U_2).

Example 3.2. *In a second example, consider three Bernoulli(1/2) r.v.s I_1, I_2, I_3 defined as in our Example 1.2 from the Introduction; see Equation* (1.5.1)*. Further, let V_1, V_2, V_3 be three mutually independent Uniform[0,1] r.v.s. Then, create three new r.v.s U_1, U_2, U_3 as*

$$U_1 := \frac{I_1 + V_1}{2}, \qquad U_2 := \frac{I_2 + V_2}{2}, \qquad U_3 := \frac{I_3 + V_3}{2}. \tag{3.2.2}$$

This yields that U_1, U_2, U_3 are themselves Uniform[0, 1], and still pairwise independent (but not mutually independent). We first visualise the dependence between U_1, U_2, U_3 on

Figure 3.5 which features a 3D scatterplot of a sample of size $n = 2{,}000$ *generated from* (3.2.2). *As in the previous example, we also provide a 2D scatterplot of* U_1 *versus* U_2 *(see Figure 3.6) where the value of* U_3 *is represented by a colour (whose scale is on the right of the figure). This highlights the pattern in the data: values of* U_3 *bigger than their mean (1/2) occur exclusively in the first and third quadrant. We note that this type of dependence is perhaps possible in some real-life situations, as it corresponds to a scenario where a variable tends to be high when two other variables concord (i.e., are either high together, or low together). Lastly, we note that an equivalent formulation of this example (expressed in terms of joint density) is found in Nelsen (1996, Example 6).*

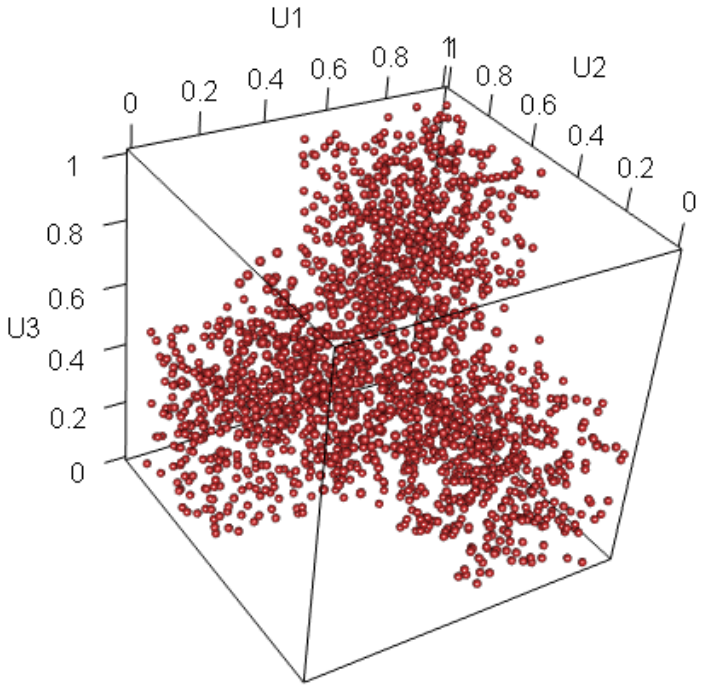

Figure 3.5: 3D scatterplots of a sample (U_1, U_2, U_3) generated from (3.2.2)

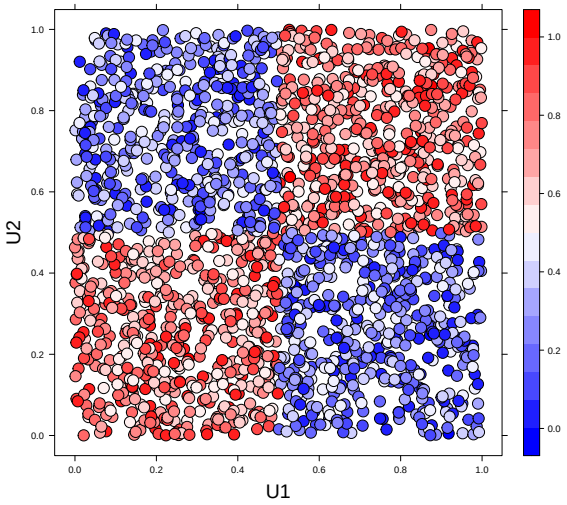

Figure 3.6: 2D scatterplot (U_1 vs U_2) of a sample generated from (3.2.2). The colour of each sample point represents the value of the third variable (U_3).

Example 3.3. *In a third example, we let points* (U_1, U_2, U_3) *be uniformly distributed on the four faces of a tetrahedron whose vertices are (1,0,0), (0,1,0), (0,0,1), (1,1,1). One way to obtain this pattern is to independently generate* $U_1 \sim$ *Uniform*$[0,1]$ *and* $U_2 \sim$ *Uniform*$[0,1]$. *Then,*

- with 50% probability, let

$$U_3 = \begin{cases} U_1 + U_2 - 1 & \text{if } U_1 + U_2 \geq 1, \\ 1 - (U_1 + U_2) & \text{if } U_1 + U_2 < 1. \end{cases} \tag{3.2.3}$$

- with 50% probability, let

$$U_3 = \begin{cases} 1 - (U_1 - U_2) & \text{if } U_1 - U_2 \geq 0, \\ 1 + (U_1 - U_2) & \text{if } U_1 - U_2 < 0. \end{cases} \tag{3.2.4}$$

It is then not hard to show that U_3 *will be a Uniform[0,1] and that the triplet* (U_1, U_2, U_3) *will be pairwise independent (for a proof see Proposition 3.2 in Appendix 3.A). However, much in the spirit of the previous examples, there is a strong dependence between those variables, since all points occupy only a fraction of the available space, as illustrated on Figures 3.7 (3D view) and 3.8 (2D view). We note that the idea of creating three PIBD variables by uniformly placing them on a tetrahedron was mentioned in Nelsen (1996, Example 1), though a proof of pairwise independence was not provided.*

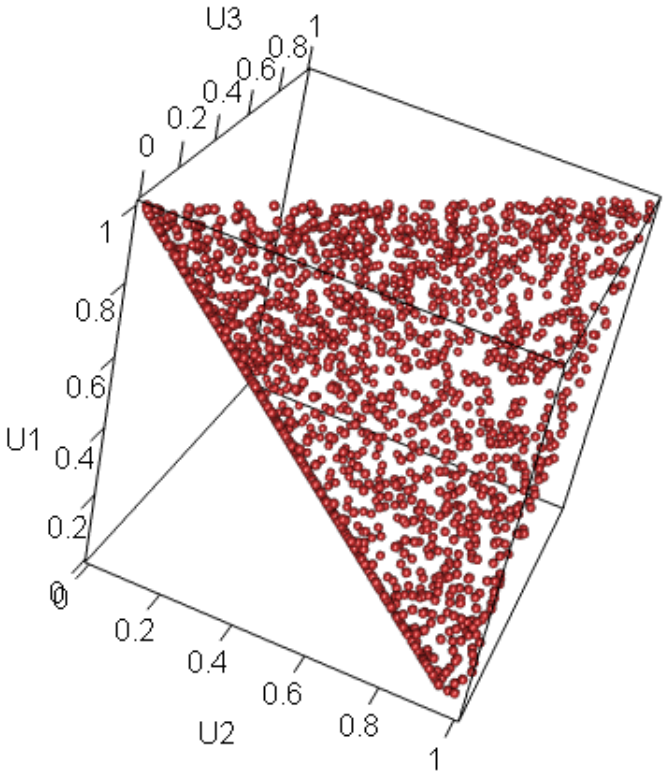

Figure 3.7: 3D scatterplot of a sample (U_1, U_2, U_3) generated from Example 3.3

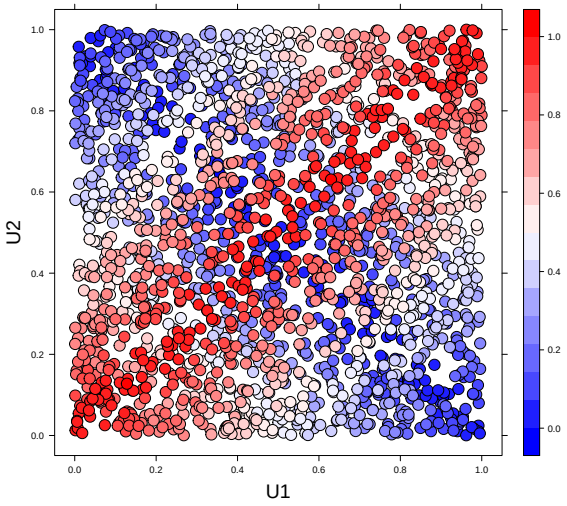

Figure 3.8: 2D scatterplot (U_1 vs U_2) of a sample generated from Example 3.3. The colour of each sample point represents the value of the third variable (U_3).

Example 3.4. *A modification of an example from Janson (1988) yields another example of PIBD variables. As before, we let* U_1 *and* U_2 *be two independent* $U[0,1]$ *random variables.*

Then, we define U_3 *as*

$$U_3 = F\left(\frac{\cos\left(2\pi(U_1+U_2)\right)+1}{2}\right),$$

where $F(\cdot)$ *is the CDF of a Beta*$(\alpha = 1/2, \beta = 1/2)$ *random variable. This yields that* U_3 *is also a* $U[0,1]$ *and that the triplet* (U_1, U_2, U_3) *is PIBD (for a proof, see Proposition 3.3 in Appendix 3.A). Yet again, although pairwise independent, those variables* (U_1, U_2, U_3) *are strongly dependent: as seen on Figure 3.9, all points of a sample generated from this example sit on a tri-dimensional 'W'. Figures 3.10 and 3.11 further provide the 2D scatterplots of the generated sample (as before, the third variable is represented as a colour).*

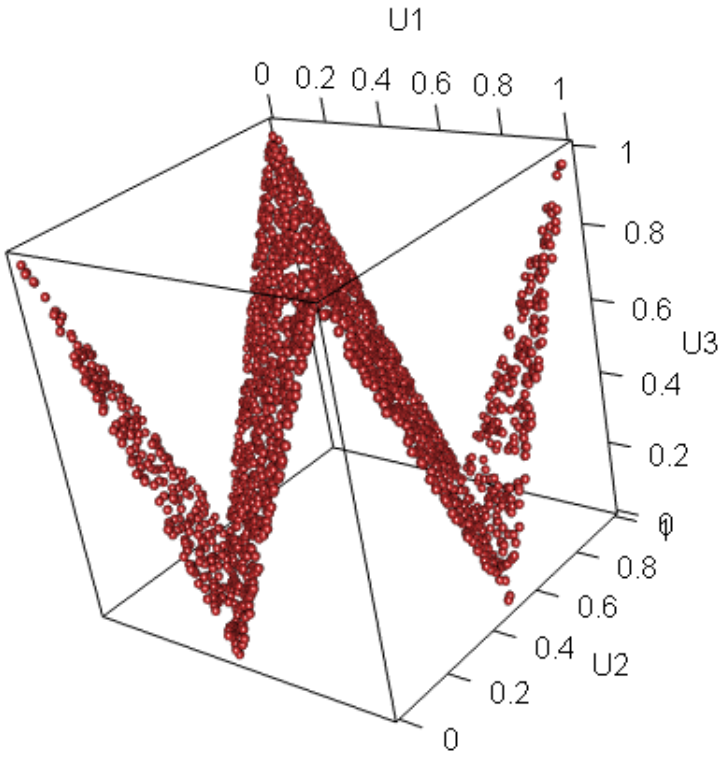

Figure 3.9: 3D scatterplot of a sample (U_1, U_2, U_3) generated from Example 3.4

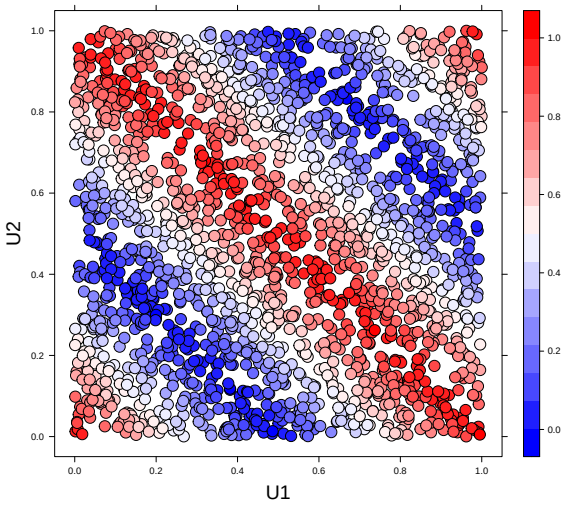

Figure 3.10: 2D scatterplot (U_1 vs U_2) of a sample generated from Example (3.4). The colour of each sample point represents the value of the third variable (U_3).

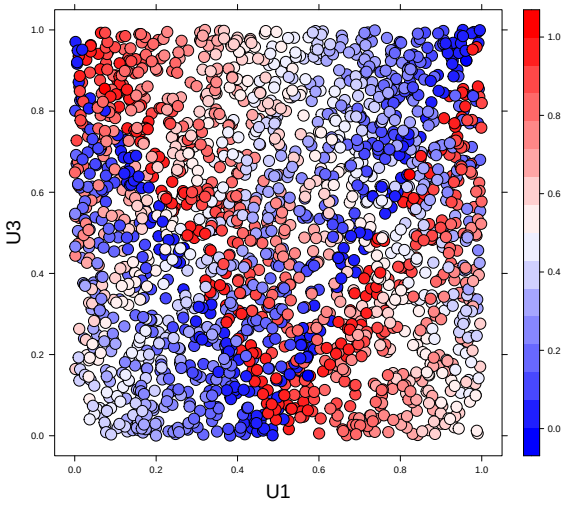

Figure 3.11: 2D scatterplot (U_1 vs U_3) of a sample generated from Example (3.4). The colour of each sample point represents the value of the third variable (U_2).

Example 3.5. *Durante et al. (2014) provide explicitly a tridimensional copula that is PIBD, which we use as a fifth example. This copula C is given by*

$$C(u_1, u_2, u_3) = u_1 u_2 u_3 \left(1 + \alpha(1-u_1)(1-u_2)(1-u_3)\right), \tag{3.2.5}$$

where $\alpha \in [-1, 1], \alpha \neq 0$. Let U_1, U_2, U_3 be three $U[0,1]$ random variable with copula C. The dependence between those variables is quite 'mild' (at least compared to that featured in Examples 3.1 to 3.4). To see this, consider the copula density associated with copula

$C(u_1, u_2, u_3)$, *i.e.*

$$c(u_1, u_2, u_3) = \frac{\partial C(u_1, u_2, u_3)}{\partial u_1 u_2 u_3} = 1 + \alpha(1 - 2u_1)(1 - 2u_2)(1 - 2u_3),$$

from which we have that

$$1 - |\alpha| \leq c(u_1, u_2, u_3) \leq 1 + |\alpha|. \tag{3.2.6}$$

Then, from (3.2.6)*, there are no regions of the cube* $[0,1]^3$ *where the concentration of points is 'drastically' higher than what it would be under mutual independence (recall the independence copula density is* $c^{\perp}(u_1, u_2, u_3) = 1$ *for all* $(u_1, u_2, u_3) \in [0,1]^3$*). Likewise, there are no regions of positive volume where the density* $c(u_1, u_2, u_3)$ *is 0, and therefore all regions of the* $[0,1]^3$ *are 'possible' (unlike in Examples 3.1 to 3.4). As a specific example, consider the case* $\alpha = 1$*. Then, for values* $(u_1, u_2, u_3) \approx (0, 0, 0)$*,* $c(u_1, u_2, u_3) \approx 2$*, and hence the region of the unit cube around the* $(0, 0, 0)$ *corner has a higher density than under mutual independence. On the other hand, for values* $(u_1, u_2, u_3) \approx (1, 1, 1)$*,* $c(u_1, u_2, u_3) \approx 0$*, hence the region around the corner* $(1, 1, 1)$ *has fewer points than expected under mutual independence.*

To try and 'see' what this dependence looks like, we generate a sample of $n = 2{,}000$ *observations from copula* C*, with* $\alpha = 1$ *(details on the generation procedure can be found in Appendix 3.B). On Figure 3.12, the 3D scatterplot of this sample is presented, while Figure 3.13 shows the corresponding 2D scatterplot (*U_1 *versus* U_2*, with* U_3 *as a colour). The dependence pattern is subtle, and quite difficult to see. By looking closely we may note that certain corners of the unit cube have a smaller concentration of points than other corners. However, this dependence is far weaker than in previous examples, and barely visible on both Figures 3.12 and 3.13. That is to say, pairwise independent observations may be dependent in a subtle way which is hard to detect. In Section 3.3, we will develop new visualisation tools which allow to see much more clearly such subtle dependence.*

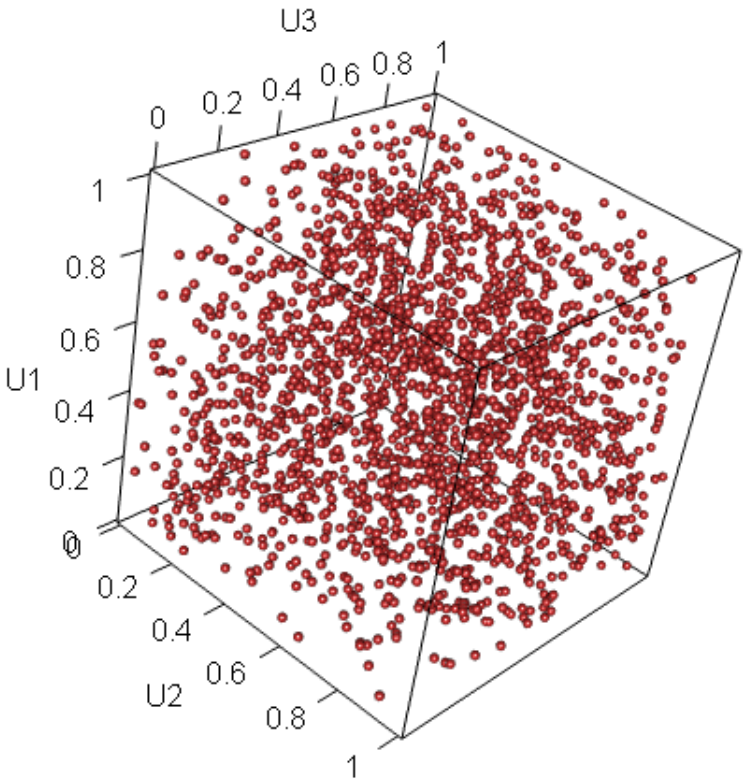

Figure 3.12: Sample generated from the PIBD copula (3.2.5), with $\alpha = 1$

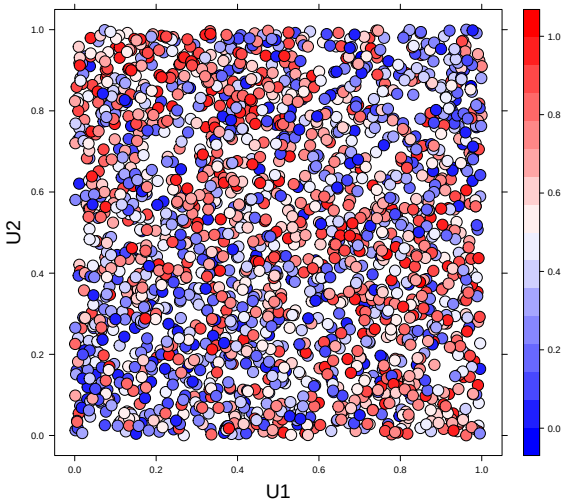

Figure 3.13: 2D scatterplot of random variables U_1 vs. U_2 generated according to Example 3.5 with $\alpha = 1$. The colour of each sample point represents the value of the third variable (U_3).

3.2.2 Mixtures of pairwise independent random variables

Examples 3.1, 3.2, 3.3, 3.4 and 3.5 are only specific instances of PIBD variables. One could call Examples 3.1, 3.2, 3.3 and 3.4 'extreme', but this is precisely our point: pairwise independence need not be 'close' to mutual independence. One could think that those are just a few 'pathological' cases, among infinitely many possible dependence structures, and hence that it is unlikely, in practice, to encounter pairwise independent but dependent data. We do not think this is so. Indeed, by *mixing* any 'pairwise independent but dependent' structures, one obtains a new dependence structure which is again pairwise independent (but not mutually independent). Since there are infinitely many of such possible mixtures, there are infinitely many 'pairwise but not mutually independent' dependence structures. Those are not all 'extreme' or 'pathological', and some are very close (and can be made arbitrarily close to) mutual independence. This argument is formalised in Proposition 3.1.

Proposition 3.1. *Let $F_1, \ldots, F_d$ be univariate CDFs, with $d \geq 2$. For $\ell = 1, 2, \ldots, m$, let $G_\ell(x_1, \ldots, x_d)$ be d-variate CDFs whose marginal CDFs are $F_1, \ldots, F_d$, and whose components are pairwise independent. Further, let H be any mixture of those m distributions, i.e.,*

$$H(x_1, \ldots, x_d) = \sum_{\ell=1}^{m} a_\ell G_\ell(x_1, \ldots, x_d),$$

with $\sum_{\ell=1}^{m} a_\ell = 1$ and $0 < a_\ell$ for all $\ell = 1, \ldots, m$. Then, H is also a distribution function whose components are pairwise independent.

Proof. Let $(X_1, \ldots, X_d)$ be a vector of random variables with distribution H, and pick any two variables $X_i, X_j, i \neq j$ from this vector. For any $\ell = 1, \ldots, m$, denote by $G_{\ell,ij}$ the bivariate distribution function of the i^{th} and j^{th} components of G_ℓ. Let $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$. We have

$$\begin{aligned}
\mathbb{P}[X_i \leq x_1, X_j \leq x_2] &= \sum_{\ell=1}^{m} a_\ell G_{\ell,ij}(x_1, x_2) \\
&= \sum_{\ell=1}^{m} a_\ell F_{X_i}(x_1) F_{X_j}(x_2) \\
&= F_{X_i}(x_1) F_{X_j}(x_2).
\end{aligned}$$

Since the choice of the pair (X_i, X_j) was arbitrary, the proof is complete. □

The fact that Proposition 3.1 applies to mixtures of G_ℓ's ($\ell = 1, \ldots, m$) whose margins

are identical (across different ℓ's) is not a restriction to our argument. Indeed, what characterises the dependence of any random vector is its copula (see Sklar's Theorem 2.3), and Proposition 3.1 applies to copulas (as stated in the following Corollary).

Corollary 3.1. *Let a copula $C(u_1, \ldots, u_d)$ be created by mixing any number (say m) of other copulas (call them $C_1, \ldots, C_m$) whose components are pairwise independent. That is,*

$$C(u_1, \ldots, u_d) = \sum_{\ell=1}^{m} a_\ell C_\ell(u_1, \ldots, u_d),$$

with $\sum_{\ell=1}^{m} a_\ell = 1$, $0 < a_\ell$ for all $\ell = 1, \ldots, m$. Then, copula C is also pairwise independent.

Proof. Copulas are multivariate CDFs whose marginal components are all Uniform[0,1]. Hence, Proposition 3.1 applies to copulas. □

Remark 3.2. *In Proposition 3.1, a case of special interest is that of a mixture between mutual independence and any pairwise independent (but dependent) structure. This automatically yields another pairwise independent (but dependent) structure. Because the weight given to the mutually independent part of such a mixture can be made arbitrarily close to 1, one can create a dependence structure that is arbitrarily close to mutual independence. In other words, pairwise independent random variables can be dependent in a very subtle way, making such dependence hard to detect.*

Example 3.6. *Using the idea from this section, we generate a sample of size $n = 2{,}000$ under a mixture of two dependence structures (in the sense of Proposition 3.1), where the 'weights' are:*

- 90% on the dependence structure of Example 3.1
- 10% on mutual independence.

We obtain a sample as displayed on Figure 3.14 (3D scatterplot) and 3.15 (2D scatterplot of U_1 versus U_2). We could see this as having injected some 'noise' to what is otherwise a very strong relationship between three variables.

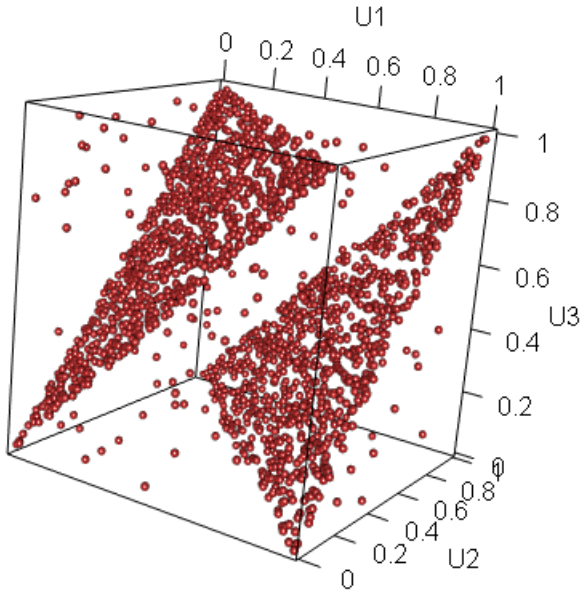

Figure 3.14: 3D scatterplots of a sample (U_1, U_2, U_3) generated from (3.6)

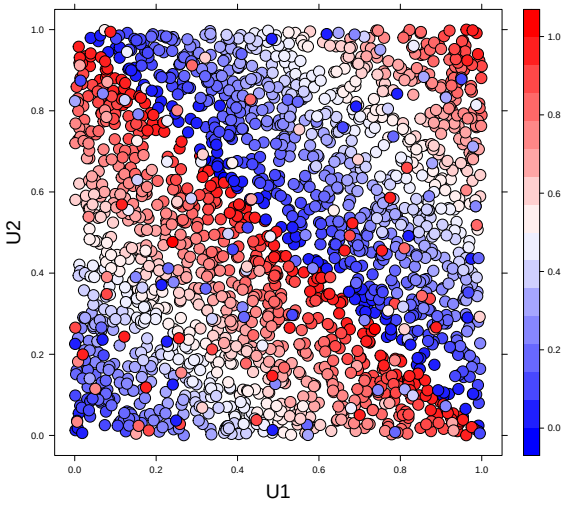

Figure 3.15: 2D scatterplot (U_1 vs U_2) of a sample generated from (3.6). The colour of each sample point represents the value of the third variable (U_3).

3.3 Dependence visualisation tools

In this section, we develop dependence visualisation tools which are especially well suited to detect dependence between variables that are pairwise independent. We focus on visualising dependence between three random variables (Sections 3.3.1 and 3.3.2), but we also introduce a tool to visualise dependence between four variables (Section 3.3.3).

We note that the existing literature on dependence visualisation is primarily devoted to visualising dependence between *two* numerical random variables, call them generically X and Y, with joint distribution F. For a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ stemming from F, the simplest and most used method to visualise if the variables X and Y are independent is the scatterplot. Another common method is to plot the contours of the (estimated) copula density of (X, Y), see Chapter 3 in Czado (2019) or Section 3 in Yan (2007) for examples. Note that the estimation of copula densities is a non-trivial task, see Charpentier et al. (2007) for a review of different approaches. Another visualisation tool for bivariate dependence is the heatmap, where on a two-dimensional grid a colour-code is used to highlight the 'concentration' of points in specific areas. Though, with a heatmap, one must make a choice of what exactly the colour represents. Typically, it represents the copula itself or copula density, but other suggestions have been made in the literature, e.g. the 'local correlation' (Tjøstheim and Hufthammer, 2013) or the 'quantile dependence function', which is a "normalized difference between the underlying copula $C(u, v)$ and the independence copula $\Pi(u, v) = uv$" (Ćmiel and Ledwina, 2020).

Of course, such visualisation tools are of no use for PIBD data, since by definition for such data there is no dependence to be seen when considering only two variables at a time.

The visualisation of dependence between three or more variables is a harder task, and fewer visualisation tools are available for that purpose. A relatively straightforward method is to use 3D contour plots of the copula density, as done for example in Killiches et al. (2017). We believe the methods we develop in Sections 3.3.1 and 3.3.2 are an alternative to contour plots (advantages of our method are discussed in Remark 3.4). Another visual method to detect dependence in dimension three (or more) is the multivariate version of the 'Kendall plot' (see Section 6 of Genest and Boies, 2003). This tool adapts the concept of a QQ plot to the detection of dependence, though the authors themselves highlighted that their tool is not well suited to visualise dependence in the case of PIBD data. Another approach to visualise the dependence between three variables is to plot a bivariate dependence measure

between two of them (e.g., Kendall's tau), as function of the value taken by the third one (see Gijbels et al., 2011). Of course, such visualisation is impacted by *which* bivariate measure is chosen.

Lastly, we want to mention that there is a well established literature on the problem of visualising the effects of predictor variables on a response variable in 'black box' learning models. For this purpose, a popular tool is the so-called 'partial dependence plot' (PDP) due to Friedman (2001), which plots the (average) change in the the predicted outcome of a learning model when one (or perhaps two) predictor(s) changes. This method was extended by Goldstein et al. (2015) to what they call 'individual conditional expectation plots' (showing multiple individual paths rather than the 'average' one, which may reveal dependence patterns a PDP cannot). While those methods provide a way to visualise the dependence between a predicted response and one (or two) predictors, it is of an 'explanatory' nature in a context where an appropriate model has already been fit to the data. In this section, our purpose is different, as we are interested in visualising the 'pure dependence', before any predictive model has been fit to data. Hence, we would qualify our tools as 'exploratory' in nature.

3.3.1 2D visualisation

For the purpose of visualising dependence between two or more numerical variables, it is standard to display the data on a [0,1] scale, as it makes it easier to spot any dependence pattern. It is well known that, for a continuous random variable X with CDF $F_X(\cdot)$, $F_X(X)$ has a Uniform[0,1] distribution. Hence, for two random variables X_1, X_2 (with CDFs F_1, F_2, respectively), if it was possible one would want to display a sample of the pair

$$F_1(X_1),\ F_2(X_2)$$

on a scatterplot. In practice with real data, we do not know the margins F_1, F_2. However, they can be approximated by their empirical counterparts, $\widehat{F}_j$. The empirical CDF is typically computed as (see, e.g., Gibbons and Chakraborti, 2003, p.37):

$$\widehat{F}_j(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_{ij} \leq t\}}, \tag{3.3.1}$$

where $X_{1j}, \ldots, X_{nj}$ is a random sample from variable X_j (and $\mathbb{1}_A$ denotes the indicator function on the set A). We then have that, for any X_k having distribution F_j,

$$\widehat{F}_j(X_k) \tag{3.3.2}$$

has an approximately Uniform$[0, 1]$ distribution. It is only *approximately* Uniform$[0, 1]$, because, strictly speaking, the distribution of $\widehat{F}_j(X_k)$ is discrete (it is often called a *pseudo uniform* distribution).

Now, if some variables in a dataset are pairwise independent, standard 2D scatterplots are of no use to reveal potential dependence patterns between them (as was the case, for example, on Figure 3.1). This is also true of more sophisticated 2D visualisation tools like those relying on the 'local Gaussian correlation' (Tjøstheim and Hufthammer, 2013) or the 'quantile dependence function' (Ćmiel and Ledwina, 2020). This is because, between any two of PIBD variables, there is simply no dependence to see.

However, and as seen already in the examples of the previous section, a simple idea to detect possible dependence between *three* variables using a 2D scatterplot is to add a third variable to the plot, represented as a colour (recall Figures 3.3, 3.4, 3.6, 3.8, 3.10, 3.11, 3.13, 3.15). Indeed, if any particular colour appears more (or less) frequently in certain areas of the 2D plot (compared to how the colours would appear if they were determined by pure chance), then some dependence is present. This is because under mutual independence we expect the colours to be completely randomly spread on the $[0, 1]^2$ range of the plot.

To go a bit further, when investigating the possible dependence between three variables, we propose to present the data on a 3×3 matrix of plots, where:

- on the upper part of the 3×3 matrix, we display the three colour-coded 2D scatterplots of: $\widehat{F}_1(X_1)$ versus $\widehat{F}_2(X_2)$, $\widehat{F}_1(X_1)$ versus $\widehat{F}_3(X_3)$ and $\widehat{F}_2(X_2)$ versus $\widehat{F}_3(X_3)$.
- On the lower part of the matrix, we display the same plots but in black-and-white, so that any 'purely pairwise' pattern can be spotted, if any.
- On the diagonal of the matrix, we display the histograms of the three variables. This provides information on the marginal distributions of the variables. This information would otherwise be lost, since we transformed the data to pseudo uniforms to create the scatterplots.

We present on Figure 3.16 this idea applied to a sample of size $n = 2{,}000$ stemming from

the dependence structure of Example 3.1, only now with variables having non-uniform margins. This means that the data presented on the scatterplots has been transformed via Equation 3.3.2. We do the same for a sample stemming from the dependence structure of Example 3.2 and present the result on Figure 3.17. In both cases, we notice at a glance the strong dependence patterns on the scatterplots of the upper part of the matrix. We also see that the dependence structure of Example 3.2 is such that the variables are exchangeable within this structure. In Example 3.1, they are not.

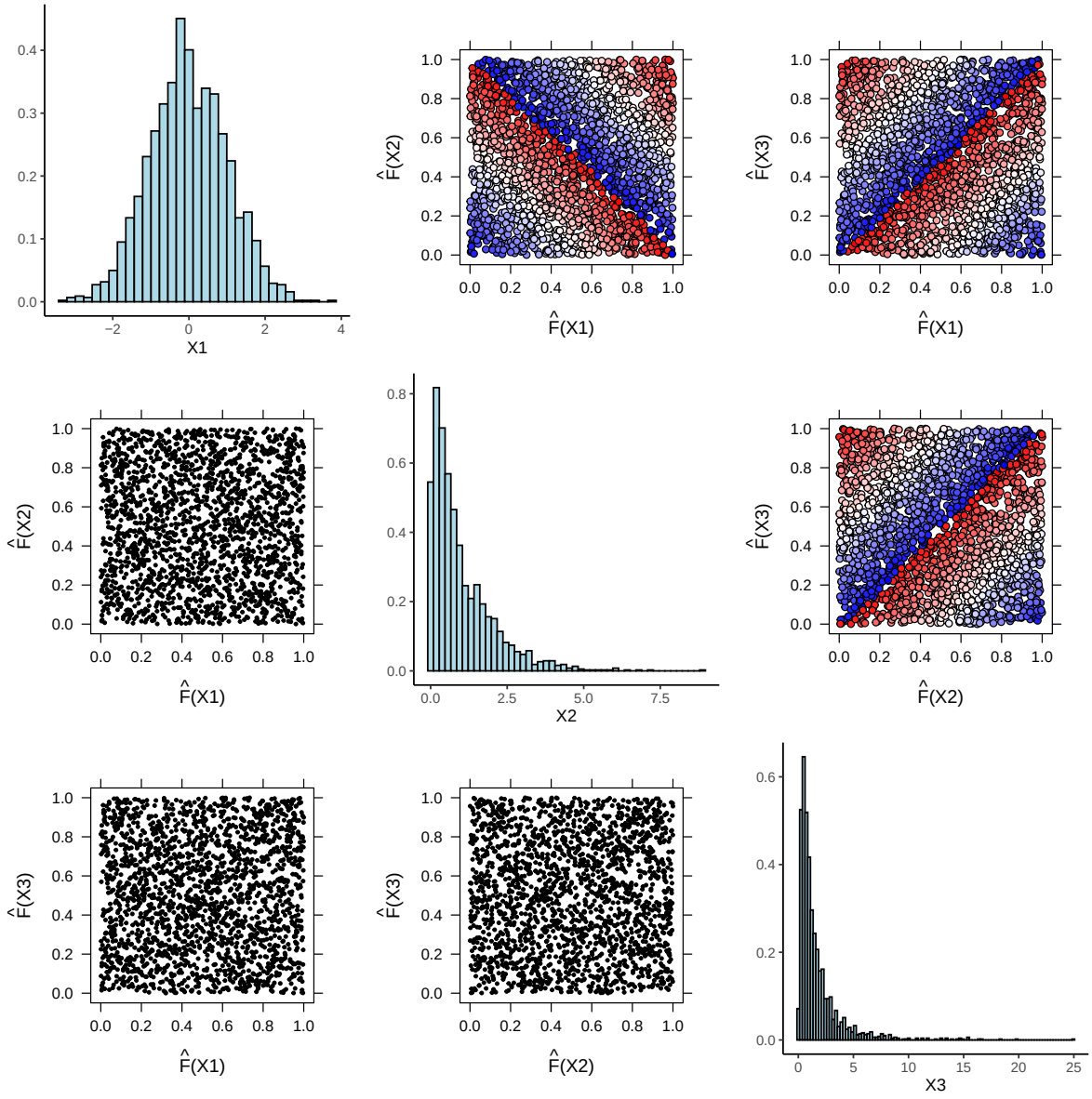

Figure 3.16: 2D scatterplots of a sample of $(\widehat{F}_1(X_1), \widehat{F}_2(X_2), \widehat{F}_3(X_3))$ generated from Example 3.1 but with non-uniform margins (histograms of every variable are shown on the diagonal). On the coloured scatterplots, the third variable is represented as a colour, with convention 0 = `blue`, 1 = `red`.

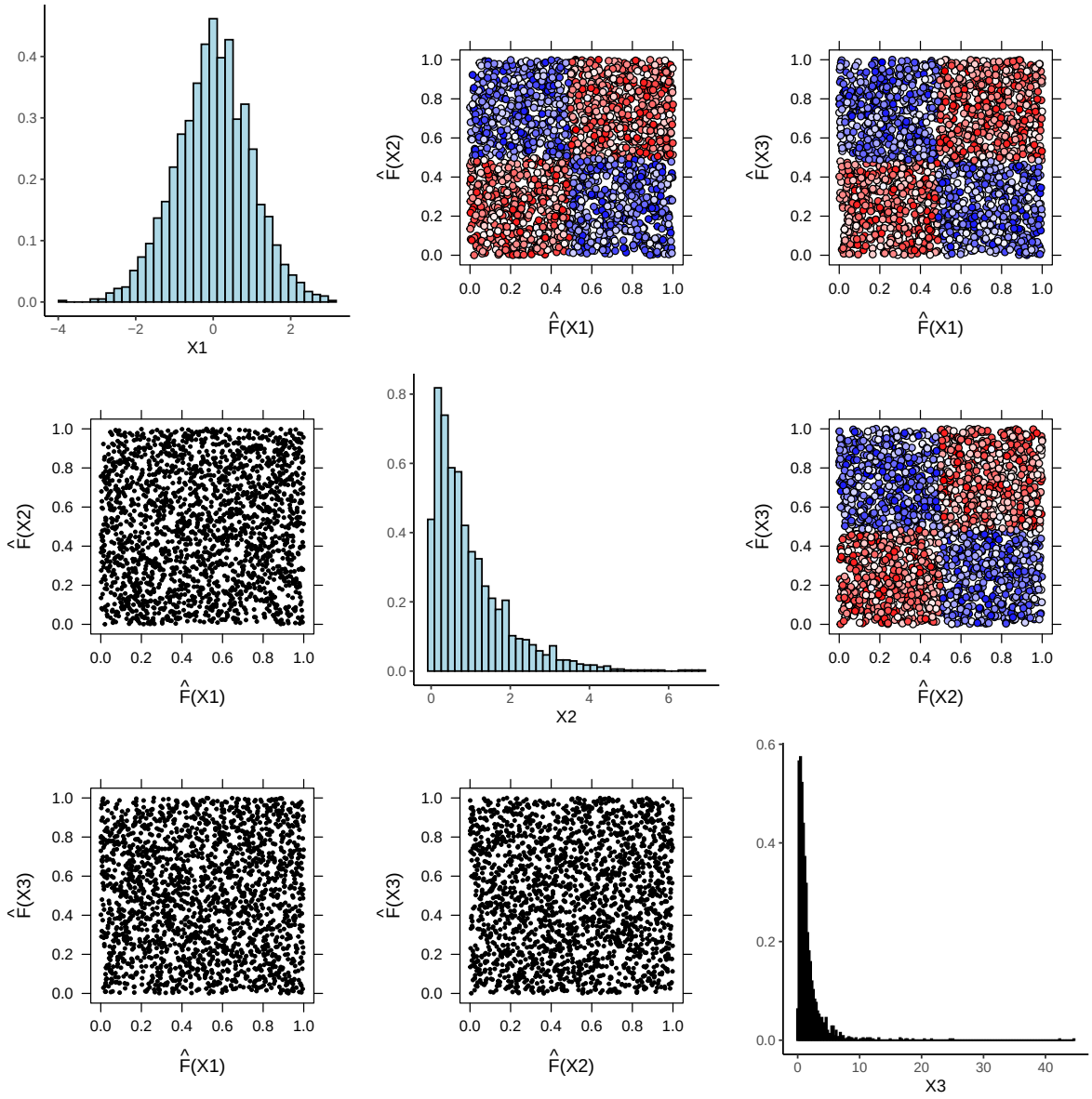

Figure 3.17: 2D scatterplots of a sample of $(\widehat{F}_1(X_1), \widehat{F}_2(X_2), \widehat{F}_3(X_3))$ generated from Example 3.2 but with non-uniform margins (histograms of every variable are shown on the diagonal). On the coloured scatterplots, the third variable is represented as a colour, with convention 0 = `blue`, 1 = `red`.

3.3.2 3D visualisation

To detect multivariate dependence, scatterplots in 2D have their limits. For example, on Figure 3.13 it was quite hard to distinguish a clear dependence pattern, even though the data was generated under a 'non independent' copula (see 3.2.5).

Hence, for data which appears pairwise independent (or close to), we also propose to use colour-coded 3D scatterplots. This can help assess whether some 'triplewise dependence' is present. For this purpose, assume a sample of size n stems from a random

vector with uniform margins, call it $U \coloneqq (U_1, U_2, U_3) \sim C(u_1, u_2, u_3)$, where $C(\cdot, \cdot, \cdot)$ is a three-dimensional copula. Note that non-uniform continuous variables can always be transformed to pseudo uniforms via Equation (3.3.2).

To visualise the type of dependence associated with copula $C(\cdot, \cdot, \cdot)$, we can display this data as n points on a 3D scatterplot, as we have done in Section 3.2 (recall Figures 3.2, 3.5, 3.7, 3.9 and 3.12). This can be enough to spot a strong dependence pattern. However, for more subtle forms of dependence (such as that of Example 3.5), simple 3D scatterplots might not be enough to detect a pattern.

Hence, to better visualise possible dependence, we propose in this section to colour each point of a 3D scatterplot according to how 'concentrated' other points are around it. For instance, low density points could be 'blue', average density points could be 'grey' and high-density points could be 'red'. A significant departure from a 'constant and average concentration' (every point is grey) would then be spotted easily, indicating some dependence. That said, quantitatively, it is not obvious how this colour-code should be designed. The purpose of the next sections is to develop an original methodology to do this.

3.3.2.1 Concentration index and colour-coding

For any point $u \coloneqq (u_1, u_2, u_3) \in [0, 1]^3$, we must choose a 'concentration index' $h(u)$ and then map it to a series of colours. What we call $h(u)$ is a quantitative measure of how concentrated the points are around u. This concentration index will take:

$$\begin{aligned} \text{a minimum value:} \quad & h(\cdot) = a \\ \text{a baseline value:} \quad & h(\cdot) = b \\ \text{a maximum value:} \quad & h(\cdot) = c. \end{aligned} \tag{3.3.3}$$

Note that what we call the 'baseline value' is the one which corresponds to the mutual independence case. Next, any scale of colours (e.g., blue-to-grey-to-red) can be represented by numerical values in the interval $[0, 1]$, such that:

- `colour1` (e.g., 'blue') has numerical value '0'. A point u with the smallest possible concentration ($h(u) = a$) would have this colour.
- `colour2` (e.g., 'grey') has numerical value '1/2'. A point u with a concentration corresponding to mutual independence ($h(u) = b$) would have this colour.
- `colour3` (e.g., 'red') has numerical value '1'. A point u with the largest possible

concentration ($h(u) = c$) would have this colour.

One may want to simply let the 'concentration index' $h(u)$ be the copula density $\frac{\partial^3 C(u_1,u_2,u_3)}{\partial u_1 \partial u_2 \partial u_3}$. However, this is impractical for at least three reasons:

- The copula density does not always exist at every point.
- Even if the copula density exists everywhere, it is not universally bounded (so we cannot choose a universal constant c).
- The copula density is hard to estimate empirically (especially past the bivariate case), because the bounded support $[0,1]^3$ causes boundary bias problems, see, e.g., Geenens et al. (2017).

Hence, in what follows we instead let $h(u)$ be the *probability a random point $U \sim C$ is close to u*. By being 'close', we mean the probability an observation $U \sim C$ is inside a ball of radius r centered at u. Denoting a closed ball of center u and radius r by $B[u,r]$, we define

$$h^*(u,r) \coloneqq \mathbb{P}[U \in B[u;r]] = \mathbb{P}[|U-u| \leq r], \tag{3.3.4}$$

where $|\cdot|$ denotes Euclidean distance. How to choose r (and how 'small' it should be) is a delicate matter, discussed in the next section (Section 3.3.2.3).

Importantly, in the case where C is the independence copula, we want our index to be constant for all $u \in [0,1]^3$. Hence, $h^*(u,r)$ as defined in (3.3.4) cannot be used 'as is' and needs to be adjusted. This is because for a point u close to one or multiple faces of the unit cube, the small ball $B[u,r]$ will be partly outside the unit cube.

Denote by $B^*[u,r]$ the intersection of a ball $B[u,r]$ with the unit cube $[0,1]^3$ and by $V^*(u,r)$ its volume. Let $v = \frac{4\pi r^3}{3}$ be the volume of a ball of radius r and further define

$$\delta(u,r) = \frac{V^*(u,r)}{v}, \tag{3.3.5}$$

i.e., $\delta(u,r)$ is the proportion (volume-wise) of the ball of radius r centered at u which is comprised within the unit cube $[0,1]^3$. Note that deriving closed-formed equations for the value of $\delta(u,r)$ (for any $u \in [0,1]^3$ and any $r < 1/2$) is not a trivial task, but has been done in Freireich et al. (2010). Next, we define our 'adjusted' index, call it $h(u,r)$, as

$$h(u,r) = \frac{h^*(u,r)}{\delta(u,r)}, \tag{3.3.6}$$

and it is then the case that $h(u,r) = v$ is constant under the independence copula (for any u). We then let b in (3.3.3) be

$$b = v.$$

Now, contrary to the copula density, the (adjusted) probability $h(u,r)$ is always defined, and bounded. Note that the lower bound '0' is strict, i.e., it would be possible that, for some copula C and some point u, $h(u,r) = 0$. Hence, in (3.3.3), it is reasonable to set

$$a = 0.$$

For the upper bound 'c', we let

$$c = 2r/\sqrt{3}. \tag{3.3.7}$$

Indeed, this is the maximum value of $h^*(u,r)$ under the comonotonic copula. Because the comonotonic copula represents the situation of "perfectly positively dependent" variables (McNeil et al., 2015, p.226), we believe it is natural to use it as the benchmark for the 'maximal concentration'.

Remark 3.3. *We think it is the case that $c = 2r/\sqrt{3}$ is in fact the supremum of $h^*(u,r)$ over the set of all possible 3-copulas. That is, we conjecture that for any $r < 1/2$,*

$$\sup_{\substack{C \in \mathcal{C}_3 \\ u \in [0,1]^3}} \mathbb{P}[|U - u| \leq r] = 2r/\sqrt{3},$$

where $\mathcal{C}_3$ is the set of all 3-copulas. Even though we were unable to prove this statement, our methodology does not crucially depend on it. Indeed, in the hypothetical case of a point with higher concentration $h(u,r)$ than that under the comonotonic copula, our methodology would simply assign the 'maximum colour' to that point. Furthermore, we note that an 'elementary bound' is simply $2r$, since

$$\mathbb{P}[|U - u| \leq r] \leq \mathbb{P}[|U_1 - u_1| \leq r] \leq 2r,$$

which is not too far from our proposed 'tighter bound' $c = 2r/\sqrt{3}$.

Lastly, we need to assign colours to values of h. Assume a scale of colours with numerical

values $y \in [0,1]$. We need a function f which maps h to y, and such that

$$\begin{aligned} f(a) &= 0 \\ f(b) &= 1/2 \\ f(c) &= 1. \end{aligned} \tag{3.3.8}$$

A straightforward choice is the power function:

$$f(h) = \left(\frac{h-a}{c-a}\right)^{\beta}, \tag{3.3.9}$$

where, to satisfy (3.3.8), we set

$$\beta = -\frac{\log(2)}{\log\left(\frac{b-a}{c-a}\right)}.$$

Note that for $a = 0, b = v$ and $c = 2r/\sqrt{3}$, we obtain

$$\beta = -\frac{\log(2)}{\log(2\pi r^2/\sqrt{3})}.$$

Remark 3.4. *We note that our approach for visualising 3D dependence is an alternative to using 3D contour surfaces, as done for instance in Killiches et al. (2017). In this paper, the authors visualise the dependence between three variables by displaying the contour surfaces of the corresponding trivariate density (but where all univariate marginals are set to standard Gaussians). The authors state that "[t]his is done because on the uniform scale copula densities would be difficult to interpret and hardly comparable with each other". That may be true, but this presents the downside that such plots do not display 'pure dependence', but also the effect of the Gaussian margins (which can complicate interpretation). In comparison, our method completely removes the effect of the marginals when examining dependence.*

3.3.2.2 Empirical estimation of $h(\cdot,\cdot)$

The next step is to estimate $h^*(u,r)$, and hence $h(u,r)$, empirically. To that end, assume a sample

$$U_1, \ldots, U_n,$$

where $U_i := (U_{i1}, U_{i2}, U_{i3})$ for $i = 1, \ldots, n$, has been obtained from the three-dimensional copula C. For any point U_j, a 'rough' estimator of $h^*(U_j, r)$, call it $\widehat{h}^*(U_j, r)$, is simply

the empirical proportion of points U_i's falling within the ball $B[U_j, r]$, i.e.

$$\widehat{h}^*(U_j, r) = \frac{\sum_{i=1, i\neq j}^{n} \mathbb{1}_{U_i \in B[U_j, r]}}{n-1}, \tag{3.3.10}$$

where $\mathbb{1}_A$ is the indicator function on the set A. Then, the 'adjusted' estimator is given simply by:

$$\widehat{h}(U_j, r) = \frac{\widehat{h}^*(U_j, r)}{\delta(U_j, r)},$$

where $\delta(\cdot, \cdot)$ is the adjustment defined previously in (3.3.5). However, this estimator is not very smooth, and therefore we introduce a smoother version. For any given point, we let this smoother estimator be a weighted average of the values of $\widehat{h}(\cdot, \cdot)$ for points around it (i.e., within a radius r of it), where the weights are inversely proportional to the distances from that point. That is, let U_j be an arbitrary point in the sample, and let n_j be the number of points within a radius r of U_j (including the point U_j itself, so that $n_j \geq 1$). Then, call

$$h_1, \ldots, h_{n_j}$$

the values of $\widehat{h}(\cdot, \cdot)$ for those points, when ordered from closest to farthest from U_j (so that $h_1 = \widehat{h}(U_j, r)$). Further, denote the Euclidean distances between U_j and those points (still ordered from closest to farthest) by

$$d_1 = 0, d_2, \ldots, d_{n_j}.$$

We define weights as the inverse of those distances, i.e.

$$w_k = 1/d_k,$$

with the exception that

$$w_1 = \begin{cases} \frac{1}{d_2} & \text{if } n_j \geq 2 \\ 1 & \text{otherwise} \end{cases}$$

(that is, to avoid an infinite weight assigned to the point j itself, we assign to it the same weight as that of its closest neighbour). Then, we define our final estimate of $h(U_j, r)$ as:

$$\widetilde{h}(U_j, r) = \frac{\sum_{k=1}^{n_j} w_k h_k}{\sum_{k=1}^{n_j} w_k}. \tag{3.3.11}$$

Remark 3.5. *Such an 'inverse-distance weighting' method is often used for interpolating*

spatial data in geostatistics, see, e.g., Webster and Oliver (2007, Section 3.1.4), although here we are not interpolating, but simply computing a weighted average.

3.3.2.3 Choice of the radius r

The next step is to choose the radius r in (3.3.4). In theory, if the copula C was known, we could pick r arbitrarily small. However, since in practice we must estimate $h(\cdot,\cdot)$ empirically, r is constrained by how much data is available (and we henceforth denote it r_n, for n the sample size). It seems reasonable that we should choose r_n such that:

1. r_n is decreasing in n. This will produce a colour-coding which becomes more 'precise' as n increases (i.e., reflecting the concentration of points closer and closer to any point u).

2. nr_n^3 is increasing in n. This way, the expected number of points inside any small ball $B[\cdot, r_n]$ will increase as the sample size increases (making the estimator $\widetilde{h}(\cdot,\cdot)$ less variable as the sample size increases).

The above two requirements still leave a lot of leeway to choose r_n, since for any $0 < q < 1$ and some constant $\gamma > 0$,

$$r_n^3 = \frac{\gamma}{n^q}$$

satisfies them both. In what follows, we propose a scheme to calibrate r_n, which is motivated by two heuristic criteria:

1. It seems natural to calibrate r_n under the 'null assumption':

$$H_0 : C \text{ is the independence copula,}$$

   because this is the assumption we want to 'reject' (if it is false).

2. Since we are primarily interested in data visualisation, we want to choose r_n such that, under H_0, the colour-coding of data points (as described in Section 3.3.2.1) yields 'mostly colourless' points (this is because, under H_0, we expect a constant concentration of points everywhere).

Recall that every colour on our scale is represented by a numerical value in the interval $[0, 1]$, and that we have mapped the concentration index $h(\cdot,\cdot)$ to those colors via the function

$$f(h) = \left(\frac{h-a}{c-a}\right)^{\beta},$$

see (3.3.8) and (3.3.9). Under H_0 (independence), the target value of $f(\widetilde{h}(u,r))$ is 1/2 for any point u and any r. Hence, to achieve the second goal above, we propose to use the mean squared error

$$\mathrm{MSE}(f(\tilde{h})) = \mathbb{E}\left[\left(f(\tilde{h}) - \frac{1}{2}\right)^2\right] \tag{3.3.12}$$

as a criteria to judge how good the estimator $f(\tilde{h})$ is (as a function of r). As is well known, the MSE of an estimator is equal to its variance plus the square of its bias. Here, it is important to minimise both the bias (we want the 'average colour' to be 1/2) and the variance (from one point to the next, we do not want too much variation in the colours).

Minimising the MSE in (3.3.12), the r_n we obtain tends to be quite large (around $r = 0.25$, even for large n). However, the shape of the MSE as function of n tends to be always similar (regardless of n), and as displayed on Figure 3.18 (for the case $n = 500$, and based on 3,000 simulations). That is, the MSE drops quickly at first, but soon the rate of decrease slows down substantially. In this example (for $n = 500$), while the minimum is attained around $r \approx 0.27$, the benefit of increasing r past $r \approx 0.20$ is not very significant. Through simulations ran for $n = 200$ up to $n = 3{,}000$ (detailed in Appendix 3.C), we have found that setting:

$$r = 0.61n^{-0.18}, \tag{3.3.13}$$

gives a satisfying compromise between having a low MSE and having a r_n which decreases as n increases. In the next section, we showcase our colour-coding methodology (including the choice of r_n with formula 3.3.13) on the various examples of Section 3.2.

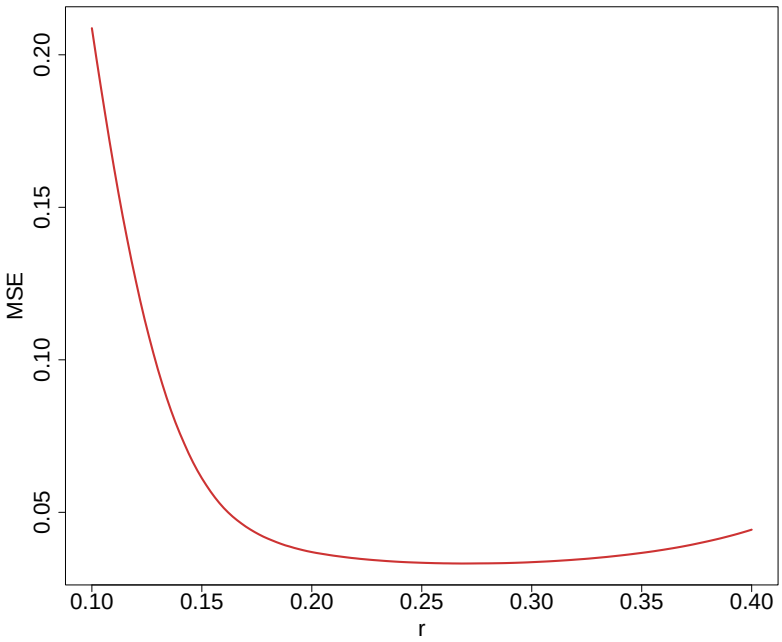

Figure 3.18: MSE of $f(\widetilde{h})$ as function of r, for $n = 500$

3.3.2.4 Examples using the estimator $\widetilde{h}(\cdot, \cdot)$

Here we apply the estimator (3.3.11) to samples of size $n = 1{,}000$ stemming from the various examples of Section 3.2. In accordance with (3.3.13), we use a radius $r = 0.176$. As explained in Section 3.3.2.1, every value $\widetilde{h}(\cdot, \cdot)$ is mapped to a number between 0 and 1 via the function $f(\cdot)$ in (3.3.9). Every number is then assigned a colour, and we use a colour code such that

- `0 = blue`
- `1/2 = whitesmoke`
- `1 = red`.

Hence, if all points $U_1, \ldots U_n$ of the sample are pale (either grayish, slightly blue or slightly red), this corresponds to the independence copula. In addition to the choice of the three colours above, we must also choose the three constants a, b, c as in (3.3.3). As explained in Section 3.3.2.1, a natural choice is to set

$$a = 0, \qquad b = 3\pi r^3/4, \qquad c = 2r/\sqrt{3}.$$

We call this choice the 'absolute' colour scale, because the constants a, b, c are then independent of the specific sample at hand. However, if one wishes to highlight more sharply zones of higher than average concentration (lots of red points) or lower than average

concentration (lots of blue points), one can also set

$$a = \min_{1\le j\le n} \widetilde{h}(U_j, r), \qquad b = 3\pi r^3/4, \qquad c = \max_{1\le j\le n} \widetilde{h}(U_j, r)$$

(provided that $a < b < c$). We call this the 'relative' colour scale. Note it will always yield that a point is of the brightest red possible, and another one is of the brightest blue possible.

On Figure 3.19 is the 'mutual independence' case, where colours appear fairly randomly, and are quite pale (as expected). Of course, random variations still create some areas of more or less dense points, hence the points are not totally colourless.

On Figure 3.20 (absolute scale) and then Figure 3.21 (relative scale) is a sample from Example 3.2. We see that most points are red, which is expected since in this example only half of the unit cube is occupied by points (which are otherwise uniformly distributed).

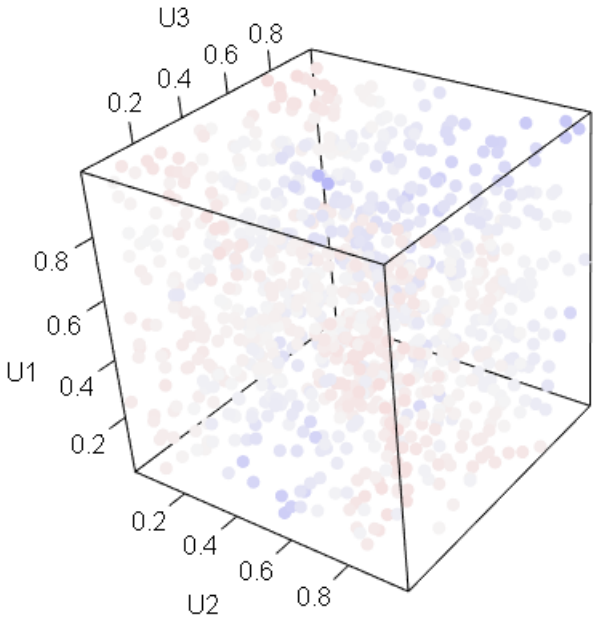

Figure 3.19: Mutually independent sample of $U_1, U_2, U_3 \sim$ Uniform[0,1]

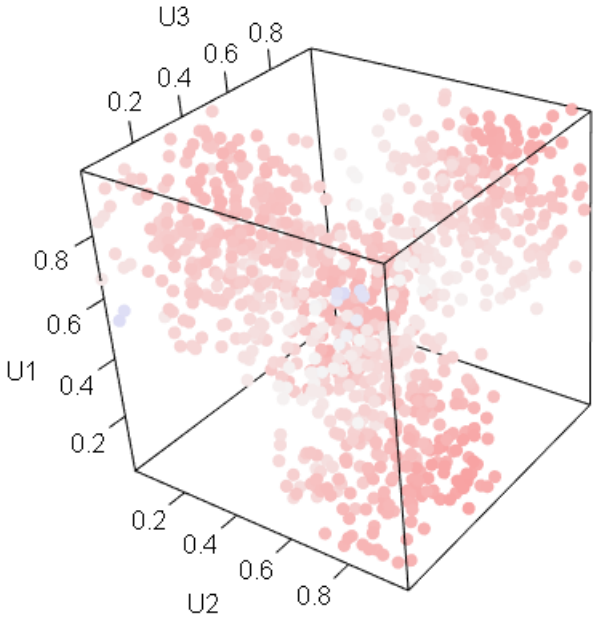

Figure 3.20: Sample from Example 3.2, absolute scale

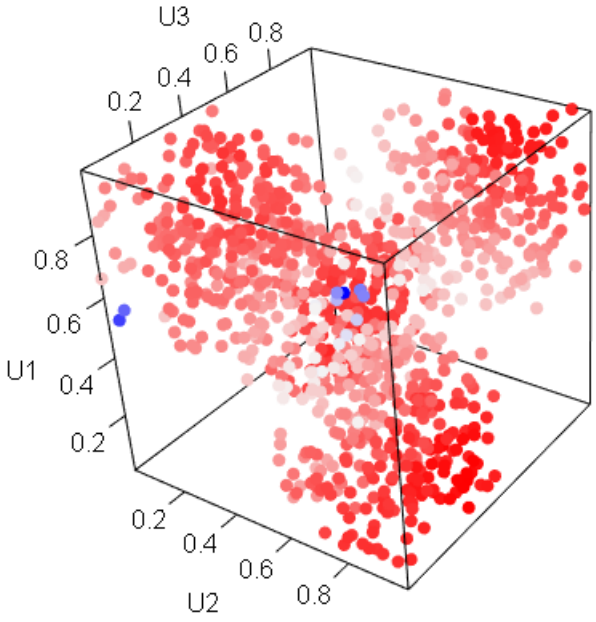

Figure 3.21: Sample from Example 3.2, relative scale

On Figure 3.22, 3.23 and 3.24 are samples from Examples 3.1, 3.3 and 3.4, respectively (all on the absolute scale). Here again we note mostly red points due to the high density of points in those examples. We do not present plots on the 'relative scale', since the dependence is already very obvious using the absolute scale.

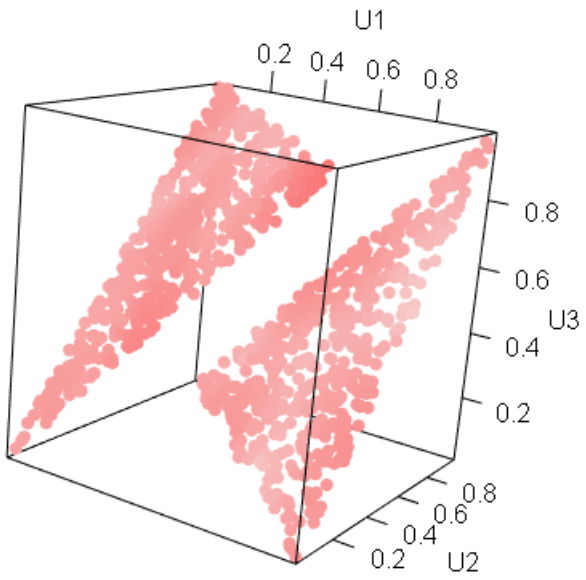

Figure 3.22: Sample from Example 3.1, absolute scale

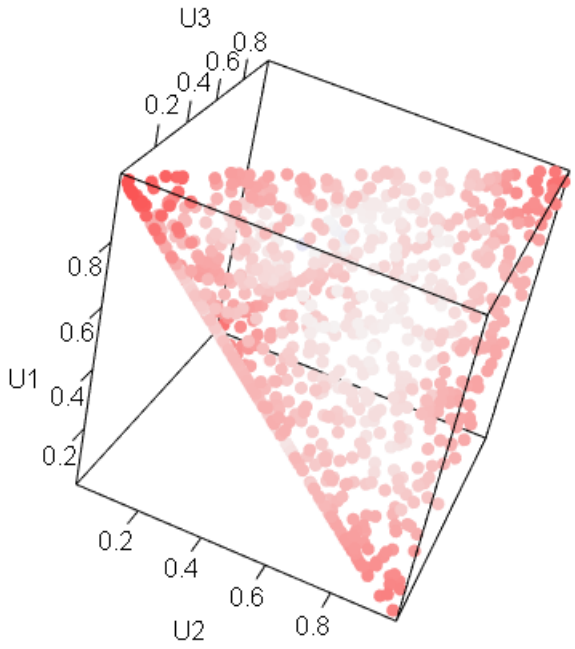

Figure 3.23: Sample from Example 3.3, absolute scale

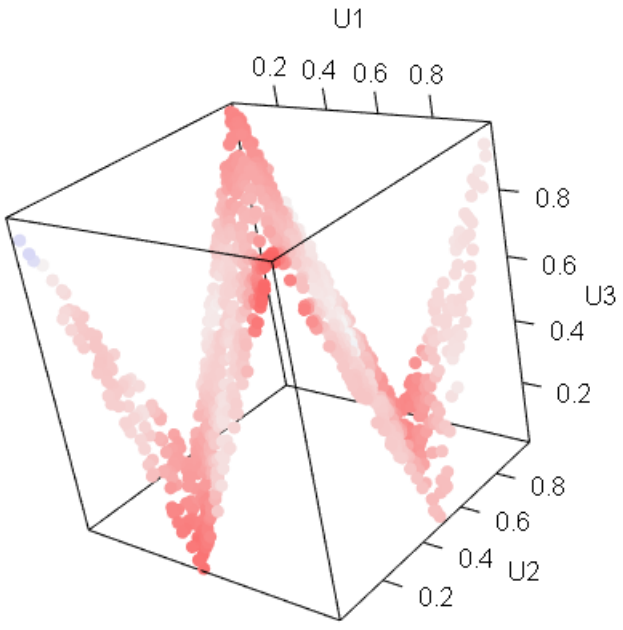

Figure 3.24: Sample from Example 3.4, absolute scale

For a sample from Example 3.5, the dependence is more subtle. On the 'absolute scale' (see Figure 3.25), we get some idea that some corners have a higher concentration of points than others. To see this a bit more clearly, we present on Figure 3.26 the same plot but where only the 'above average' density points are coloured. Lastly, on the 'relative scale' the dependence become more apparent, see Figure 3.27. Here, it is clearer that some corners of the unit cube have a higher concentration, while others have lower concentration of points. Though, as we have said before, this is a much weaker type of dependence than previous examples.

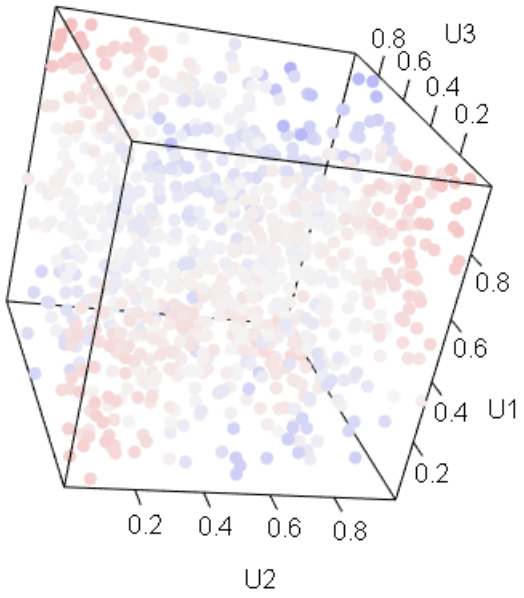

Figure 3.25: Sample from Example 3.5, absolute scale

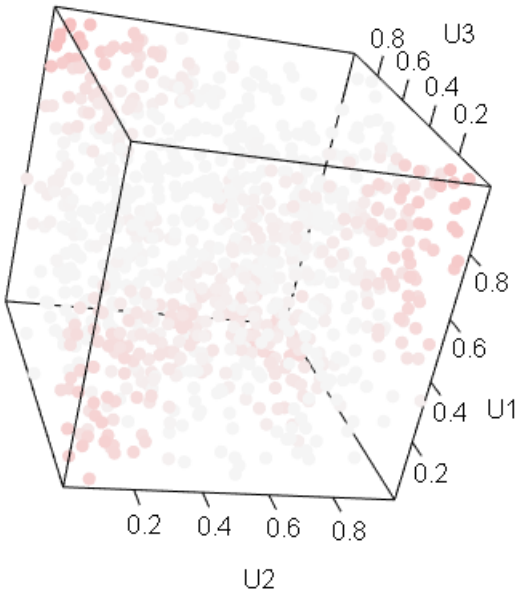

Figure 3.26: Sample from Example 3.5, blue removed

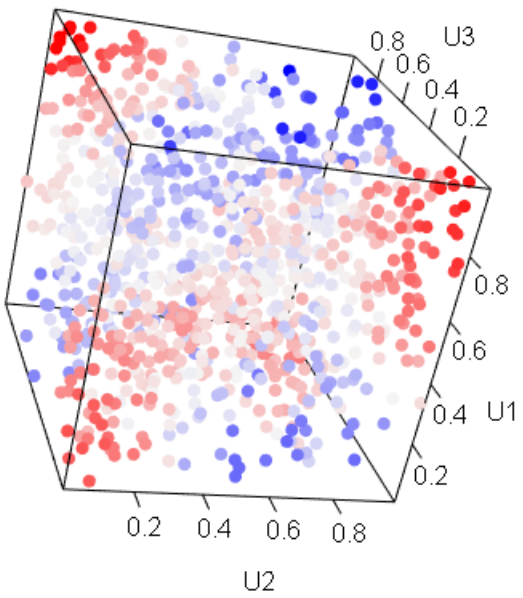

Figure 3.27: Sample from Example 3.5, relative scale

3.3.2.5 Alternative method using a 'fixed grid'

Lastly, we present an alternative method of visualising dependence between three random variables. Instead of having one 'coloured sphere' per data point, we can place spheres at regular intervals on a $(g \times g \times g)$ grid. Everything else is then the same: the colours of those spheres represent the concentration of points around them, via the estimation method of Section 3.3.2.2. This is especially helpful for large sample sizes, where displaying every point on a regular 3D scatterplot can yield cluttered figures. Here, we find that smaller values of 'r' yield better results. In the examples below, for samples of size 5,000 we used $r = 0.103$, which comes from Equation 3.C.1 (see Appendix 3.C for details on how this value was derived). In addition, here we also make the size of a sphere centered at U_i proportional to the value of the estimated $\widetilde{h}(U_i, r)$. This way, areas of low density are easier to identify.

In what follows, we use grids of size $8 \times 8 \times 8$. Figure 3.28 displays the results for a mutually independent sample. This methods work better for more 'extreme' types of dependence, and we present on Figures 3.29, 3.30 and 3.31 the results for samples coming from Examples 3.1, 3.2 and 3.3, respectively.

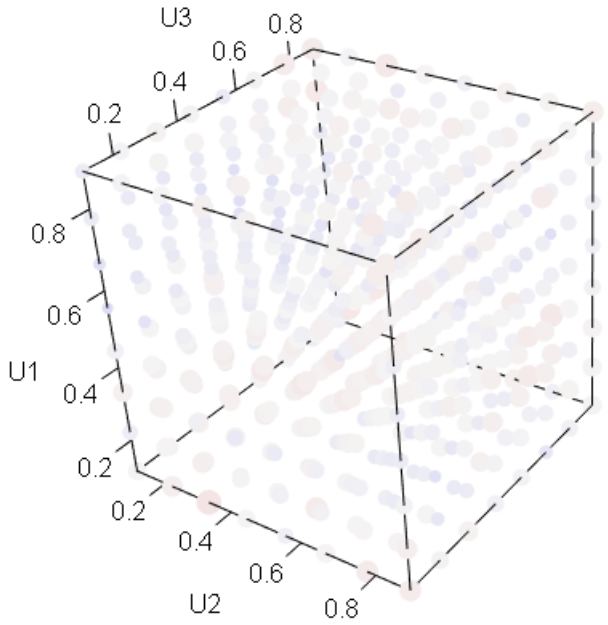

Figure 3.28: Mutually independent sample of $U_1, U_2, U_3 \sim$ Uniform[0,1], fixed grid method and absolute scale

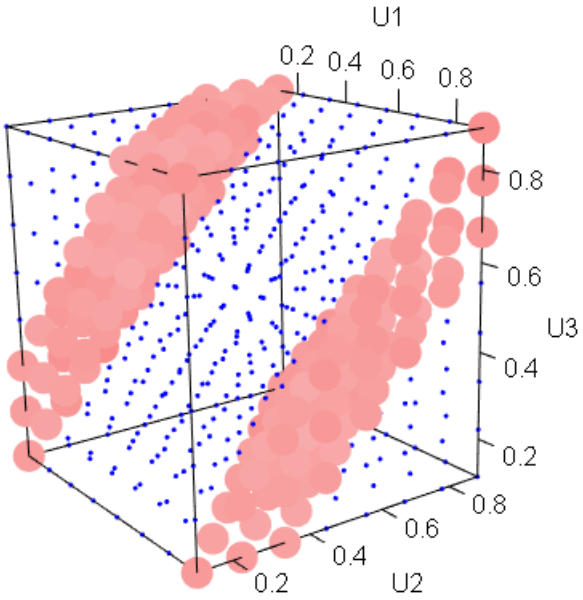

Figure 3.29: Sample from Example 3.1, fixed grid method and absolute scale

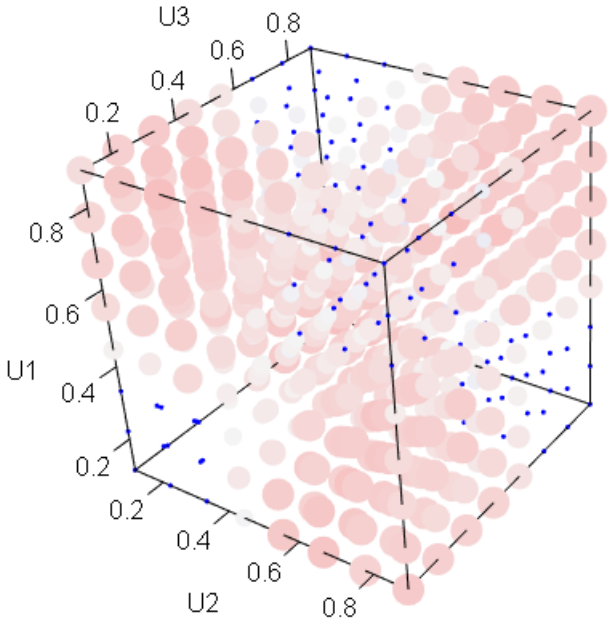

Figure 3.30: Sample from Example 3.2, fixed grid method and absolute scale

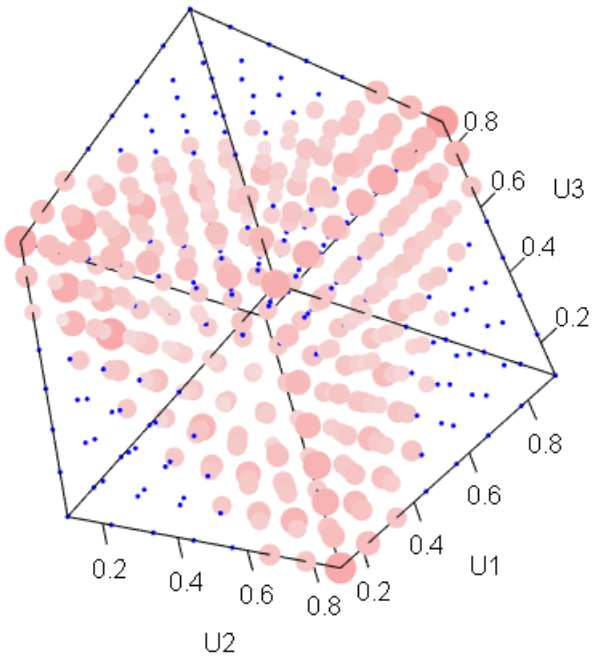

Figure 3.31: Sample from Example 3.3, fixed grid method and absolute scale

3.3.3 4D visualisation

Although this chapter is concerned mainly with the distinction between mutual and pairwise independence, one should note that we can also define a more general notion of 'K-tuplewise independence', for an arbitrary integer $K \geq 2$. We say a random sequence $\{X_j, j \geq 1\}$ is K-tuplewise independent if for every choice of K distinct integers $j_1, \ldots, j_K$, the random variables $X_{j_1}, \ldots, X_{j_K}$ are mutually independent. Pairwise independence is then simply the case $K = 2$. Let us give a specific example for $K = 3$, i.e., 'triplewise independence'.

Example 3.7. *Let M_1, M_2, M_3, M_4 be mutually independent Bernoulli(1/2) random variables. Define*

$$X_1 = \mathbb{1}_{\{M_1 = M_3\}},$$
$$X_2 = \mathbb{1}_{\{M_1 = M_4\}},$$
$$X_3 = \mathbb{1}_{\{M_2 = M_3\}},$$
$$X_4 = \mathbb{1}_{\{M_2 = M_4\}}.$$

Then, X_1, X_2, X_3, X_4 are triplewise independent, but not mutually independent (the proof of this is not hard, and we omit it because this example is a special case of the more general sequences we will introduce in Chapter 6, Section 6.2). Further, X_1, X_2, X_3, X_4 are all Bernoulli(1/2) random variables, and we can easily create from them four new variables with Uniform[0,1] margins (call them U_1, U_2, U_3, U_4) which are still triplewise independent (but not mutually independent). Indeed, simply set

$$U_i = \begin{cases} Uniform\left[0, \frac{1}{2}\right] & \text{if } X_i = 0 \\ Uniform\left[\frac{1}{2}, 1\right] & \text{if } X_i = 1, \end{cases} \tag{3.3.14}$$

for $i = 1, 2, 3, 4$.

We then wish to visualise the dependence between the variables U_1, U_2, U_3, U_4 from Example 3.7. Of course, to do so we can generate a sample from those variables, but any 'monochrome' 3D scatterplot (say of U_1 vs. U_2 vs. U_3) would not show any dependence pattern (since those variables are, by construction, triplewise independent).

However, on such a scatterplot we can colour-code the fourth variable (say U_4), and then check for any pattern of colours (which would indicate some dependence). This is totally analogous to the approach we used in Section 3.3.1 to detect triplewise *dependence* on a 2D scatterplot (for pairwise independent variables). Note that, contrary to the previous Section 3.3.2, here the colour of a point *does not* represent the concentration of points around it. Rather, it simply represents the *value* of the fourth variable.

Figure 3.32 illustrates such a 3D scatterplot for a random sample (of size 3,000) generated from Example 3.7, with the fourth variable U_4 colour-coded such that:

- $U_4 = 0 \implies$ `blue`
- $U_4 = 1/2 \implies$ `white`
- $U_4 = 1 \implies$ `red`.

On Figure 3.32 we note a strong pattern: 'blue points' only appear in four out of eight 'sub-cubes', while 'red points' only appear in the other four 'sub-cubes'.

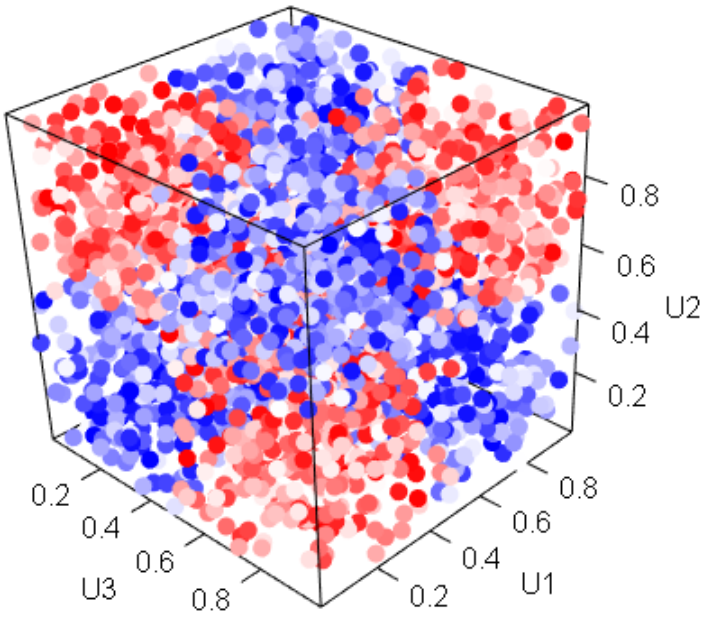

Figure 3.32: Scatterplot of a sample of U_1, U_2, U_3 from Example 3.7, where U_4 is represented by a colour on a blue-to-white-to-red scale

3.4 Conclusion

In this chapter, our goal was to visualise what dependence under 'pairwise independence' can look like. Of course, this type of dependence is not 'one specific thing'. Hence, to ask the question 'what does PIBD data look like' is a bit like asking 'what does dependent data look like'; it is not possible to give an exhaustive list of all *possible* cases. Nonetheless, we have provided in this chapter many possible examples, and we saw that this type of dependence can, at least in principle, be very strong. We also saw that by using the simple idea of 'mixing', we can create an infinite number (and continuum) of such dependence structures (where the dependence can be more or less subtle).

In addition, we provided original visualisation tools to better 'see' this dependence. Those tools can be used to explore datasets and detect possibly 'subtle' forms of dependence. In particular, we developed a colour-coding methodology to highlight which areas of a 3D scatterplot have a higher (or lower) concentration of points (where the benchmark is a constant concentration, as under mutual independence). We saw this can help detect forms of dependence otherwise hard to see, as that of Example 3.5.

After 'seeing' that there can be a significant difference between mutual and pairwise independence, we are interested to find how 'material' this difference can be, especially in common actuarial settings. Is it 'mostly okay' to assume independence in a situation where only pairwise independence holds? Or do common tools and theorems which rely on the independence assumption fail badly for PIBD variables? We investigate this question in the next chapters.

3.A Proofs

Proposition 3.2. *Let U_1, U_2, U_3 be random variables defined as in Example 3.3. Then, $U_3 \sim$Uniform$[0,1]$, and the triplet (U_1, U_2, U_3) is pairwise independent.*

Proof. Let $u_1, u_3 \in [0,1]$. Assuming that U_3 is generated under (3.2.3), by conditioning we obtain:

$$\begin{aligned}
\mathbb{P}[U_1 \le u_1, U_3 \le u_3] &= \frac{1}{2}\left(\mathbb{P}[U_1 \le u_1, U_3 \le u_3 | U_1 + U_2 \ge 1] + \mathbb{P}[U_1 \le u_1, U_3 \le u_3 | U_1 + U_2 < 1]\right) \\
&= \mathbb{P}[U_1 \le u_1, U_1 + U_2 - 1 \le u_3, U_1 + U_2 \ge 1] \\
&+ \mathbb{P}[U_1 \le u_1, 1 - (U_1 + U_2) \le u_3, U_1 + U_2 < 1] \\
&= \mathbb{P}[U_1 \le u_1, 1 - U_1 \le U_2 \le 1 + u_3 - U_1] \\
&+ \mathbb{P}[U_1 \le u_1, 1 - u_3 - U_1 \le U_2 < 1 - U_1] \\
&= \begin{cases} u_1^2/2 & \text{if } u_1 \le u_3 \\ u_1 u_3 - u_3^2/2 & \text{if } u_1 > u_3 \end{cases} + \begin{cases} u_1 u_3 & \text{if } u_1 \le 1 - u_3 \\ u_3 - u_3^2/2 - (1-u_1)^2/2 & \text{if } u_1 > 1 - u_3. \end{cases}
\end{aligned}$$

Likewise, assuming that U_3 is generated under (3.2.4),

$$\begin{aligned}
\mathbb{P}[U_1 \le u_1, U_3 \le u_3] &= \frac{1}{2}\left(\mathbb{P}[U_1 \le u_1, U_3 \le u_3 | U_1 \ge U_2] + \mathbb{P}[U_1 \le u_1, U_3 \le u_3 | U_1 < U_2]\right) \\
&= \mathbb{P}[U_1 \le u_1, 1 - U_1 + U_2 \le u_3, U_1 \ge U_2] \\
&+ \mathbb{P}[U_1 \le u_1, 1 + U_1 - U_2 \le u_3, U_1 < U_2] \\
&= \mathbb{P}[U_1 \le u_1, U_2 \le u_3 - 1 + U_1] \\
&+ \mathbb{P}[U_1 \le u_1, 1 - u_3 + U_1 \le U_2] \\
&= \begin{cases} 0 & \text{if } u_1 \le 1 - u_3 \\ (u_1 + u_3 - 1)^2/2 & \text{if } u_1 > 1 - u_3 \end{cases} + \begin{cases} u_3^2/2 - (u_3 - u_1)^2/2 & \text{if } u_1 \le u_3 \\ u_3^2/2 & \text{if } u_1 > u_3. \end{cases}
\end{aligned}$$

Then, by conditioning on whether U_3 is generated from (3.2.3) or (3.2.4) (each option having a 50% probability), we get:

$$2\mathbb{P}[U_1 \le u_1, U_3 \le u_3] = \begin{cases} u_1^2/2 + u_3^2/2 - (u_3 - u_1)^2/2 & \text{if } u_1 \le u_3 \\ u_1 u_3 - u_3^2/2 + u_3^2/2 & \text{if } u_1 > u_3 \end{cases}$$

$$
\begin{aligned}
&+ \begin{cases} u_1 u_3 & \text{if } u_1 \leq 1 - u_3 \\ u_3 - u_3^2/2 - (1-u_1)^2/2 + (u_1 + u_3 - 1)^2/2 & \text{if } u_1 > 1 - u_3 \end{cases} \\
&= \begin{cases} u_1 u_3 & \text{if } u_1 \leq u_3 \\ u_1 u_3 & \text{if } u_1 > u_3 \end{cases} + \begin{cases} u_1 u_3 & \text{if } u_1 \leq 1 - u_3 \\ u_1 u_3 & \text{if } u_1 > 1 - u_3 \end{cases} \\
&= 2u_1 u_3.
\end{aligned}
$$

This shows that U_3 is a Uniform$[0,1]$ and that U_1 and U_3 are independent. Noting that U_1 and U_2 are interchangeable in this example, we have that U_2 and U_3 are also independent. □

Proposition 3.3. *Let U_1, U_2, U_3 be random variables defined as in Example 3.4. Then,*

a. U_3 has a Uniform$[0,1]$ distribution.

b. The triplet (U_1, U_2, U_3) is pairwise independent.

Proof. a. First, note that

$$\cos\left(2\pi(U_1 + U_2)\right) = \cos\left(2\pi(U_1 + U_2) \bmod 2\pi\right).$$

Now, from Example 3.1 we know that $(U_1+U_2) \bmod 1$ has a Uniform[0,1] distribution. It follows that the random variable

$$U \coloneqq 2\pi(U_1 + U_2) \bmod 2\pi$$

has a Uniform[0, 2π] distribution. Simple calculations[1] then yield that the density function $f(\cdot)$ of $\cos(U)$ is given by

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}} \quad \text{for } -1 < x < 1.$$

This yields that the random variable $(\cos(U)+1)/2$ has a density function given by:

$$2 \cdot f(2u-1) = \frac{u^{-1/2}(1-u)^{-1/2}}{\pi} \quad \text{for } 0 < u < 1,$$

which corresponds to the density of a Beta$(\alpha = 1/2, \beta = 1/2)$ random variable.

[1] See details here: https://stats.stackexchange.com/questions/309400/pdf-of-cosine-of-a-uniform-random-variable.

b. It suffices to show that U_1 and U_3 are independent (because U_1 and U_2 are interchangeable in the construction). The proof relies largely on a result from Janson (1988): if Z_1 and Z_2 are two independent complex-valued random variables, both uniformly distributed on the unit circle, then the sequence defined as

$$X_j = Z_1^{j-1} Z_2, \quad j = 1, 2, \ldots$$

is pairwise independent. In particular, we note that Z_2 is independent of $Z_1 Z_2$. Since Z_1 and Z_2 are interchangeable in this construction, we also have that Z_1 is independent of $Z_1 Z_2$. We remark that setting

$$Z_1 \coloneqq \exp(2\pi \mathrm{i} U_1), \quad Z_2 \coloneqq \exp(2\pi \mathrm{i} U_2),$$

(where i is the unit imaginary number) makes Z_1 and Z_2 uniformly distributed on the unit circle. It is then the case that Z_1 is independent of

$$Z_3 \coloneqq Z_1 Z_2 = \exp(2\pi \mathrm{i}(U_1 + U_2)).$$

It follows that the angle of Z_1 (in its polar coordinates representation), which is simply $2\pi U_1$, is independent of the real part of Z_3, which is $\Re(Z_3) = \cos(2\pi(U_1 + U_2))$. By noting that U_3 is a (Borel-measurable) function of $\Re(Z_3)$, U_1 is also independent of U_3. This completes the proof.

□

3.B Generating variables from Example 3.5

In this short section, we explain how to generate a triplet (U_1, U_2, U_3) as in Example 3.5. We first need the conditional copula $C_{U_3|U_1,U_2}(u_1, u_2, u_3)$, which is given by

$$C_{U_3|U_1,U_2}(u_1, u_2, u_3) = \frac{\partial^2 C(u_1, u_2, u_3)}{\partial u_1 \partial u_2} = u_3 \left(1 + \alpha(1 - 2u_1)(1 - 2u_2)(1 - u_3)\right).$$

For given (u_1, u_2), $C_{U_3|U_1,U_2}(u_1, u_2, u_3)$ is a continuous univariate CDF, and we can find its inverse, call it $C^{-1}(t|u_1, u_2)$. Simple calculations yield

$$C^{-1}(t|u_1, u_2) = \frac{1 + a(u_1, u_2) - \sqrt{(1 + a(u_1, u_2))^2 - 4a(u_1, u_2)t}}{2a(u_1, u_2)},$$

where $a(u_1, u_2) = \alpha(1 - 2u_1)(1 - 2u_2)$. Then, to generate Uniform$[0, 1]$ random variables having copula C, one need only apply a two-steps protocol (see, e.g., Cherubini et al., 2004, Section 6.3).

1. Simulate three mutually independent Uniform$[0, 1]$ random variables, call them U_1, U_2 and T.
2. Set $U_3 = C^{-1}(T|U_1, U_2)$.

This results in (U_1, U_2, U_3) having copula C.

3.C Calibrating the size of r

For sample sizes $n = 200, 300, 400, 500, 1000$ we simulated 2,000 samples of mutually independent Uniform[0,1] random variables (U_1, U_2, U_3). In addition, for sample sizes $n = 2000, 3000$, we simulated 1,000 mutually independent samples. Then, we applied to each sample the methodology described in Sections 3.3.2.1 and 3.3.2.2, and for various values of the radius r. We then computed the MSE in (3.3.12) (for a given n, averaged over all samples of size n), as function of r. The curves obtained (of MSE versus r, displayed on Figures 3.33 to 3.39) all have a similar shape: the MSE decreases with r, sharply at first and then less sharply.

Note that our goal is not only to minimise the MSE, but also to pick a r_n that decreases as n increases. Hence, heuristically, we want to pick a r_n which is 'not too big', but 'big enough' such that the improvement in MSE, if we were to increase r even more, would not be substantial. Of course, what 'substantial' means is arbitrary. We find that using the following rule:

> 1% Rule: 'fix r such that increasing r by a further 0.002 generates a decrease in MSE of $\sim 1\%$.'

yields a r_n that decreases smoothly as n increases, and following a power function ($r_n = \theta n^p$ for some constants θ, p), see Figure 3.40. The fit to a power function is also very good (with a R^2 of 0.9916). Rounded to two decimal places, we have, as expressed before in (3.3.13):

$$r = 0.61 n^{-0.18}.$$

While this rule is arbitrary, one can change it to, for example, a '2% Rule', yielding smaller r_n. Doing so, we obtain another curve that decreases smoothly as a power function, also depicted on Figure 3.40 ($R^2 = 0.997$), and whose equation is:

$$r = 0.69 n^{-0.22}. \tag{3.C.1}$$

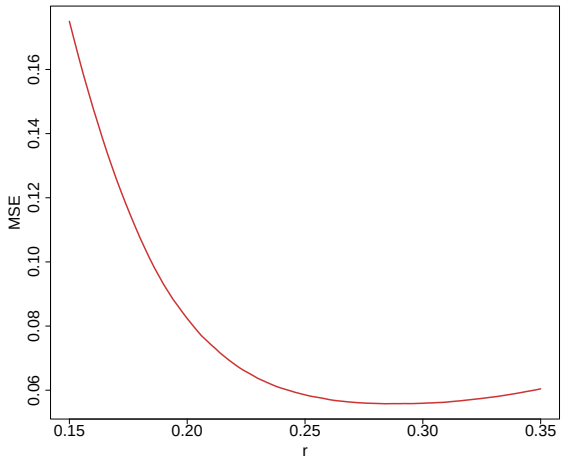

Figure 3.33: MSE of $f(\tilde{h})$ as function of r, for $n = 200$

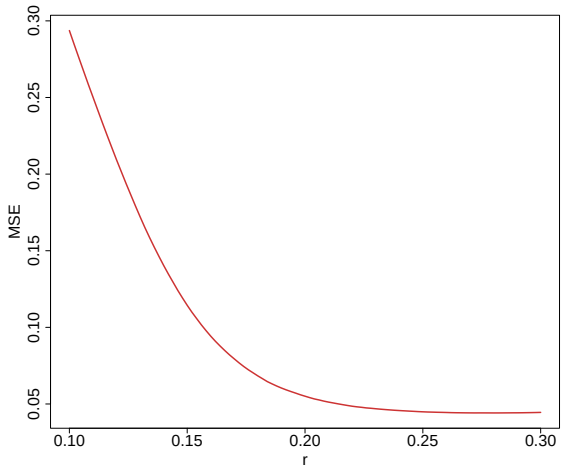

Figure 3.34: MSE of $f(\tilde{h})$ as function of r, for $n = 300$

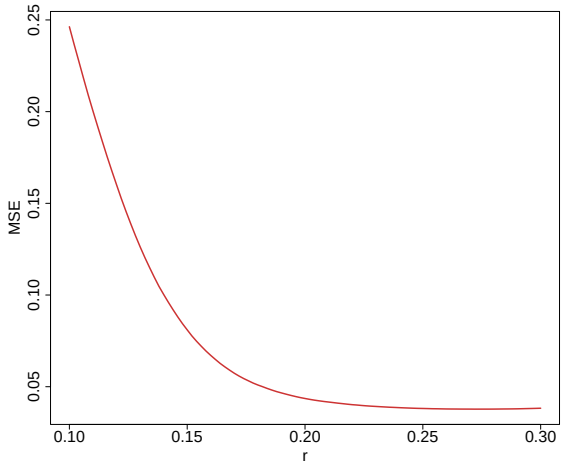

Figure 3.35: MSE of $f(\tilde{h})$ as function of r, for $n = 400$

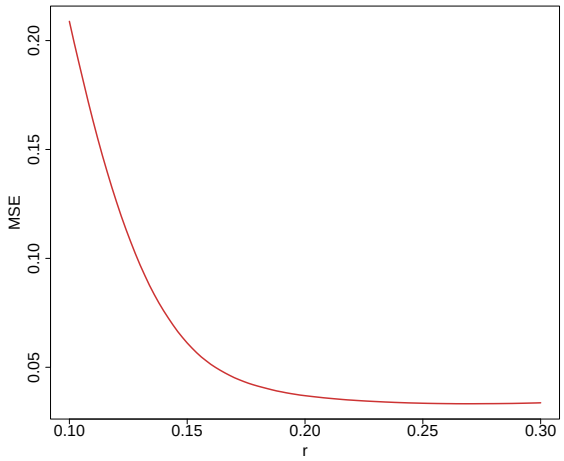

Figure 3.36: MSE of $f(\tilde{h})$ as function of r, for $n = 500$

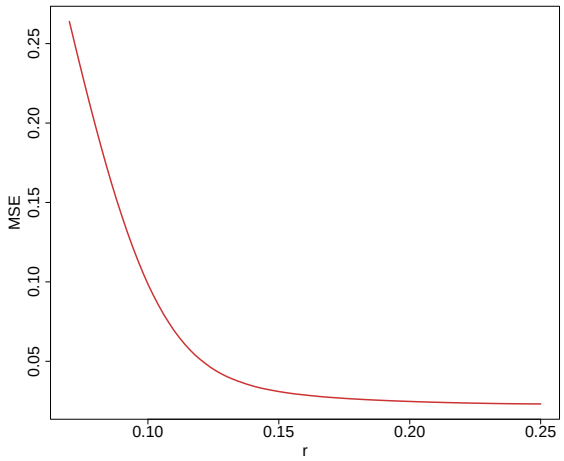

Figure 3.37: MSE of $f(\tilde{h})$ as function of r, for $n = 1{,}000$

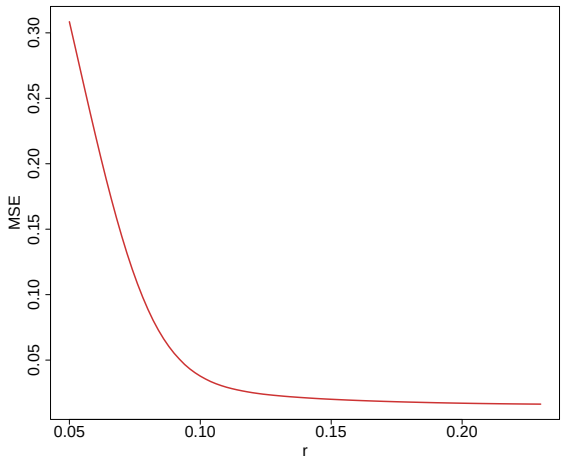

Figure 3.38: MSE of $f(\tilde{h})$ as function of r, for $n = 2{,}000$

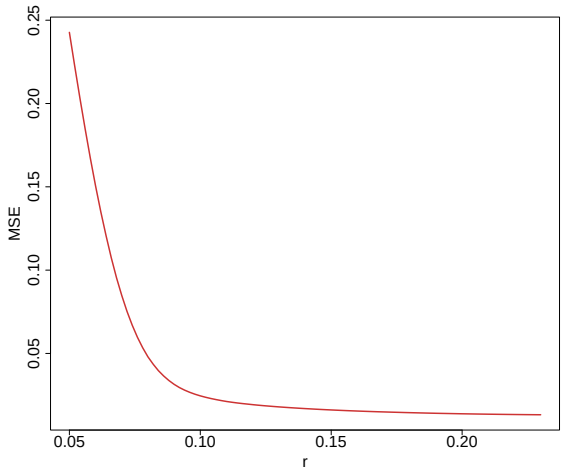

Figure 3.39: MSE of $f(\tilde{h})$ as function of r, for $n = 3{,}000$

Figure 3.40: r as a function of n when using the '1%' and '2%' rules

CHAPTER 4

WHAT CAN 'GO WRONG' UNDER PAIRWISE INDEPENDENCE

4.1 Introduction

In Chapter 2, we saw that independence is a very important assumption in many actuarial settings (and also one that many authors, in recognition that dependence does appear in real data, have started to relax by using increasingly sophisticated dependence models). Whenever mutual independence is assumed (in a model or theorem) it is relevant to know whether pairwise independence is 'enough' for the model or theorem to be valid. This is because pairwise independence is easier to justify, as it is a less stringent requirement (compared to mutual independence). It can also be sufficient for some fundamental results to hold. For example, even if it is almost always stated for mutually independent random variables, the strong Law of Large Numbers is in fact valid under sole pairwise independence (see, e.g., Etemadi, 1981).

That said, we saw in Chapter 3 that pairwise independent variables can still be strongly dependent. Hence, it is certainly not obvious that any given result assuming mutual independence would hold for PIBD variables. In Section 4.2, we review many results (important in actuarial science) valid under mutual independence and which 'fail' for PIBD variables. This serves to highlight that there is a substantial difference between 'mutual' and 'pairwise' independence. This difference matters, since modellers often stop their dependence checks at the pairwise level (perhaps because of the historical dominance of Gaussian-based models, for which zero-correlation of all pairs implies mutual independence). In Section 4.3, we establish that many dependence models (not only the multivariate Gaussian) popular in actuarial science cannot capture this difference, because they do not allow for the possibility of PIBD variables. That is, within those models, pairwise independence implies mutual independence (which of course is not true in general).

4.2 Important results which do not hold for PIBD variables

In Chapter 3, we provided examples which show, visually, that pairwise independence can be starkly different from mutual independence. Here, we present some 'consequences' of that difference. That is, we give examples of what can 'go wrong' if one assumes mutual independence where only pairwise independence holds. This question is very broad, so it is hard to cover it theoretically in an exhaustive way. Hence, we concentrate on a few situations that are relevant to the actuarial field (with no pretence that our list is exhaustive).

4.2.1 Sums of random variables

For a series of risks $X_1, \ldots, X_n$ (whose marginal distributions are known) an important problem in actuarial science is that of understanding the behaviour of their sum

$$S = X_1 + \cdots + X_n, \tag{4.2.1}$$

and under various assumptions on the dependence between risks. If the dependence structure between the X's is fully known, then the distribution of S can be derived (either analytically or numerically). If we only have *partial* information on the dependence, then there is uncertainty about the distribution of S (this is a topic we surveyed in some detail in Sections 2.3.3).

Under mutual independence, one can often deduce the distribution of the sum S in (4.2.1) as that of a 'simple' distribution. For instance, the sum of independent Normal r.v.s is again Normal, the sum of independent Exponential r.v.s is Gamma, the sum of independent Poisson r.v.s is again Poisson, etc. Those commonly used results need not hold under pairwise independence, as the following Example 4.1 demonstrates.

Example 4.1. *Let $m \geq 3$ be an integer and let $M_1, \ldots, M_m$ be a sequence of i.i.d. r.v.s with Bernoulli(1/2) distribution. For all pairs (M_i, M_j), $1 \leq i < j \leq m$, define a r.v. $D_{i,j}$ as*

$$D_{i,j} = \begin{cases} 1, & \text{if } M_i = M_j, \\ 0, & \text{otherwise.} \end{cases}$$

The $D_{i,j}$ are then also Bernoulli(1/2). For convenience, we refer to these $n = \binom{m}{2}$ random

variables $D_{1,2}, D_{1,3}, \ldots, D_{1,m}, D_{2,3}, D_{2,4}, \ldots, D_{m-1,m}$ *simply as*

$$D_1, \ldots, D_n.$$

From the sequence $D_1, \ldots, D_n$*, we can construct a pairwise independent sequence* $X_1, \ldots, X_n$ *with an arbitrary distribution* F*. The only restriction imposed on* F *is that, for a random variable* $W \sim F$*, the median of* W *(call it* $\widetilde{w}$*) must be such that*

$$\mathbb{P}[W \leq \widetilde{w}] = \mathbb{P}[W > \widetilde{w}] = 1/2.$$

Define U *and* V *to be the truncated versions of* $W \sim F$*, respectively from above its median and from below its median:*

$$U \stackrel{d}{=} W \mid \{W \leq \widetilde{w}\}, \qquad V \stackrel{d}{=} W \mid \{W > \widetilde{w}\},$$

Then, consider n *independent copies of* U*, and independently* n *independent copies of* V*:*

$$U_1, \ldots, U_n, \overset{\text{i.i.d.}}{\sim} F_U, \qquad V_1, \ldots, V_n \overset{\text{i.i.d.}}{\sim} F_V.$$

Finally, for $k = 1, \ldots, n$*, construct*

$$X_k = \begin{cases} U_k, & \text{if } D_k = 0, \\ V_k, & \text{if } D_k = 1. \end{cases}$$

By conditioning on D_k*, one can check that* $X_k \sim F$*. It is also the case that the* X*'s are pairwise independent, but not mutually independent (we defer the proof of this to Chapter 5, where this construction is made more general).*

For illustration purposes, let us now fix F *to be Poisson(*$\log(2)$*), which satisfies the restriction*

$$\mathbb{P}[X \leq \widetilde{w}] = \mathbb{P}[X \leq 0] = 1/2,$$

and let us focus on the behaviour of

$$S = X_1 + \cdots + X_n.$$

For any given m, we can derive the probability mass function (PMF) of S as

$$p_S(s) = \frac{\log(2)^s (1/2)^m}{s!} \sum_{k=0}^{m} \mathbb{1}_{\{p(k)\leq s\}} \binom{m}{k} \sum_{j=0}^{p(k)} \binom{p(k)}{j} (-1)^j \left(p(k) - j\right)^s, \qquad s = 0, 1, \ldots, \tag{4.2.2}$$

where $p(k) = (2k^2 + m^2 - 2km - m)/2$.

Remark 4.1. *The sequence defined in Example 4.1 appears somewhat 'complicated', but we note it is not easy to build PIBD sequences of large size (especially with arbitrary marginal distribution). This sequence will also be used (and made more general) in Chapter 5, see Section 5.2. Example 4.1 is an opportunity to first introduce this sequence to the reader in a simpler form (which, for now, suffices to make our points).*

Now, if the X_k's from Example 4.1 were mutually independent, then the sum S would be distributed as a Poisson($n\log(2)$). But $p_S(s)$ in (4.2.2) is never equal to the PMF of a Poisson distribution, and regardless of the m chosen. This is easy to see if one notes, for example, that for any m,

$$p_S(0) = 0,$$

which of course is not the case for a Poisson. But just how different from a Poisson is this S? We illustrate this on Figure 4.1, for various choices of sample size n. We see a marked difference between the two distributions (and in particular, the support of both distributions is different).

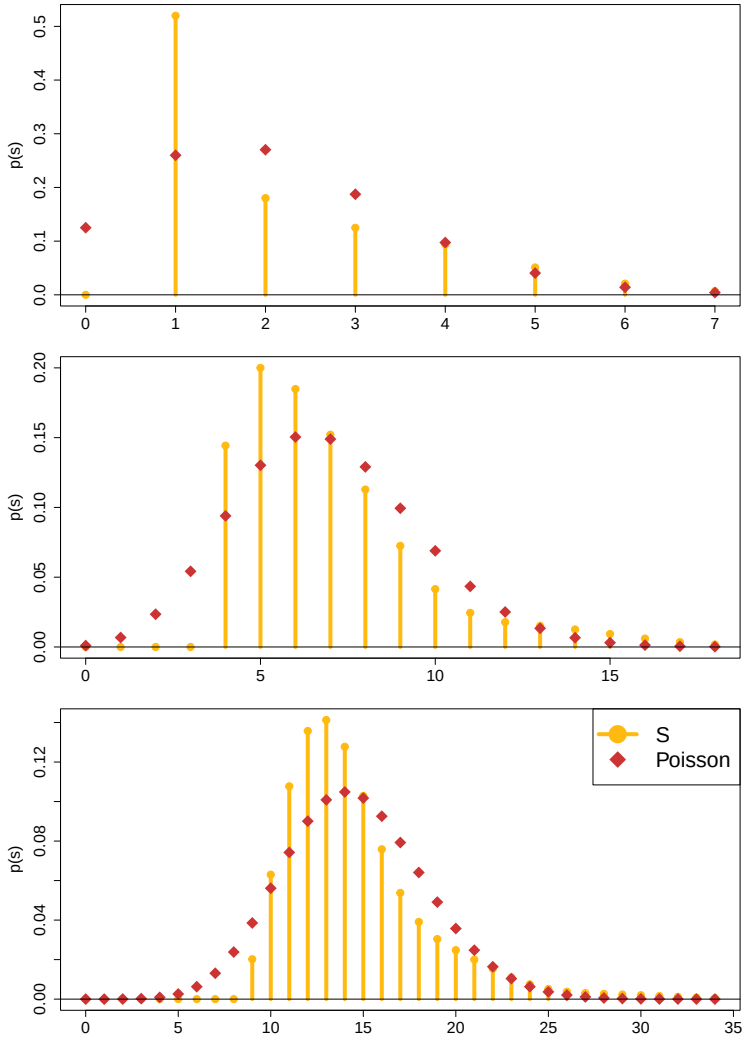

Figure 4.1: $p_S(s)$ as in (4.2.2), compared to the PMF of a Poisson$(n\log(2))$, for $n = 3$ (top), $n = 10$ (middle), $n = 21$ (bottom)

Remark 4.2. *In Example 4.1, one could ask what happens if we let 'n' increase even more. Would the distribution be closer to a Normal for a large sample size? We expect this under mutual independence, but it is not the case here. We will see in Chapter 5 that the standardised mean of that sequence does not converge to a Normal as the sample size increases. As a teaser, for Z_n the standardised version of S, i.e.,*

$$Z_n := \frac{S - \mathbb{E}[X]n}{\sqrt{n\mathbb{V}\mathrm{ar}[X]}},$$

we will show that Z_n converges in distribution to a random variable Z which can be written as

$$Z := \sqrt{1-\ln(2)}W + \sqrt{\ln(2)}\chi, \tag{4.2.3}$$

where W is a standard Normal, χ is a standardised $\chi^2(1)$, and W is independent of χ. We present on Figure 4.2 the density (left) and CDF (right) of Z, with the standard Normal density and CDF for reference. We see that Z is right-skewed (and with a heavier right tail).

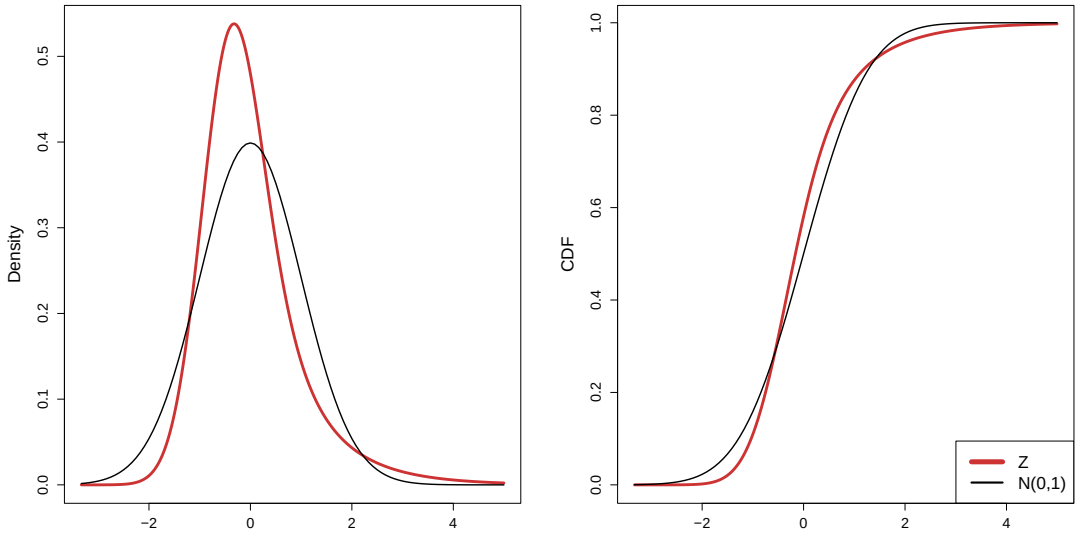

Figure 4.2: Density (left) and CDF (right) of the distribution of Z as defined in (4.2.3), compared to that of a N(0,1)

Example 4.1 (and Remark 4.2) show that a sum of PIBD variables can behave very differently than what is expected under mutual independence, the main differences being the asymmetry of the distribution of S. For PIBD variables, it is also possible for the sum S to become *more* heavy-tailed (at least, in the sense that the kurtosis increases), as the sample size *increases* (which is the opposite of what one expects under mutual independence). This is highlighted in the following example (which is a generalisation of Example 3.1).

Example 4.2. *Let ξ and η be two independent Uniform[0,1] random variables. For $j = 1, \ldots, n$, let*

$$U_j = (\eta + j\xi) \bmod 1.$$

Then, all random variables $U_1, \ldots, U_n$ are uniformly distributed U[0,1), and pairwise independent. This example stems from Example 6 in Janson (1988), though in their example the U_j have a U(-1/2, 1/2] distribution. To avoid any ambiguity, we provide a short proof that those variables are pairwise independent (see Proposition 4.3 in in Appendix 4.A).

Furthermore, the kurtosis of the sum

$$S_n = U_1 + \cdots + U_n \tag{4.2.4}$$

appears to increase linearly with n*. A proof of this has remained elusive, but via simulations it appears undeniable: for values* $n = 10^k$ *(*$k = 1, 2, 3, 4, 5$*) we generated* 3×10^6 *samples and then computed from those the empirical kurtosis of* S_n*. The results are displayed on Figure 4.3 (presented on a log-log scale). Here, increasing the sample size* n *has the opposite effect one would expect: it increases the kurtosis of the sum* S_n*. Because the kurtosis is often used as a measure of tail-heaviness, and more tail-heavy distributions are usually considered 'riskier' (e.g., when they model losses of an insurance portfolio) this result is rather surprising: increasing the size of a portfolio of PIBD risks can make the portfolio riskier (in the sense that it is more leptokurtic).*

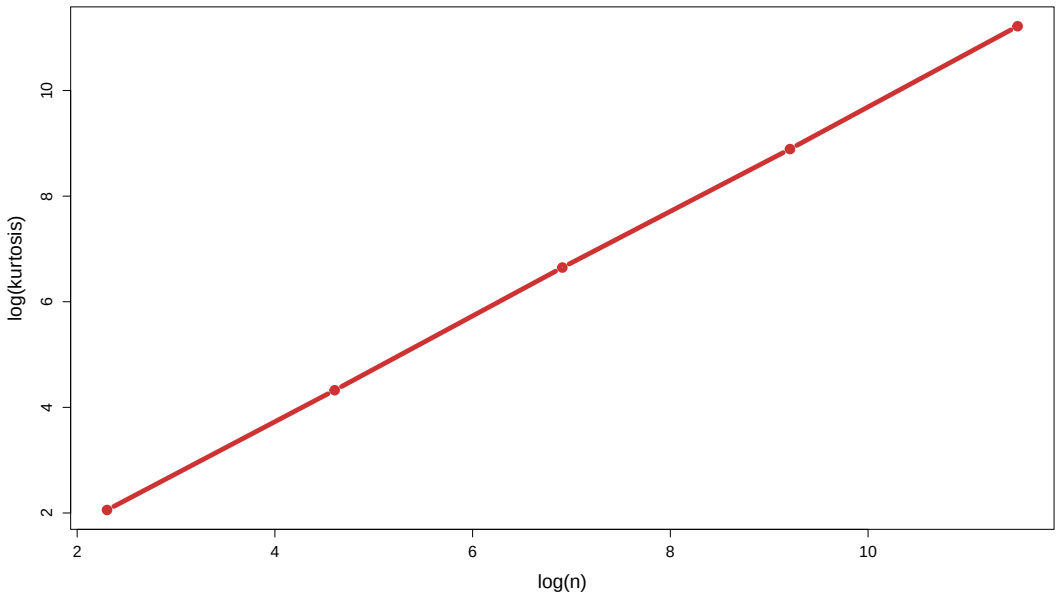

Figure 4.3: Empirical kurtosis of the sum S_n in (4.2.4), for increasing values of n

4.2.2 Risk measures on aggregate losses

As seen in Section 2.3.1, one key reason why the behaviour of a sum

$$S = X_1 + \cdots + X_n, \tag{4.2.5}$$

for $X_1, \ldots, X_n$ as series of 'risks' (typically, potential losses) is so important in actuarial science is that the capital required (CR) of an insurance company is often determined as a *risk measure* $\rho(\cdot)$ computed on S. The function $\rho(\cdot)$ is often a translation-invariant *risk*

measure $\rho(\cdot)$ such as the Value-at-Risk (VaR) or the Tail-Value-at-Risk (TVaR) computed on S. Alternatively, the required capital can be set to be the *mean adjusted* version of $\rho(\cdot)$, i.e.,

$$\text{CR} \coloneqq \rho(S - \mathbb{E}[S]) = \rho(S) - \mathbb{E}[S], \tag{4.2.6}$$

which we then interpret as the capital needed to cover unexpected losses. Of course, the assumed dependence between the risks $X_1, \ldots, X_n$ influences such a required capital, since it influences the distribution of S.

In this section, we investigate how the CR given by (4.2.6) can vary depending on whether the risks $X_1, \ldots, X_n$ are mutually independent, as compared to only *pairwise* independent. To this end, we use the examples presented in Section 3.2. We will limit our analysis to the case $n = 3$, and to the risk measures 'VaR' and 'TVaR' (which are arguably the most commonly used in insurance).

We assume that the X's are identically distributed and we compute the two risk measures (for various levels) of the sum

$$S = X_1 + X_2 + X_3,$$

and for different dependence structures, i.e., those given by Examples 3.1, 3.2, 3.3, 3.4 and 3.5 (setting the parameter $\alpha = 1$ and then $\alpha = -1$). We also vary the margins of the X_j's so that they are either

- Uniform,
- Normal,
- Exponential,
- Log-normal.

To make the comparison between different margins 'fair', we fix their parameters such that we always have $\mathbb{E}[X_j] = 1$ and $\mathbb{V}\text{ar}[X_j] = 1$, for all margins. Since the distribution of S is unknown (under the various PIBD examples), we use Monte Carlo simulations (with number of repetitions $B = 10^7$) to compute the VaR and TVaR of S under the different dependence structures. That is, we simulate a large number of times ($B = 10^7$) the variables

$$X_1, X_2, X_3,$$

under our various dependence scenarios. From those B simulations, we obtain a sample

of size B of the variable $S = X_1 + X_2 + X_3$, from which we estimate empirically both the VaR and TVaR of S[1]. We then compute the ratio:

$$\frac{\text{CR under a given PIBD example}}{\text{CR under mutual independence}}. \tag{4.2.7}$$

We report the results in Tables 4.1 (VaR) and 4.2 (TVaR). Note those are estimated values (rounded to two decimal places) and the width of a 95% bootstrap confidence interval around them is, at most, ± 0.005.

From the results about the VaR, we extract the following findings:

- The VaR for PIBD variables can differ considerably from that under mutual independence: the lowest ratio observed is 47% and the highest is 124% (both attained for Example 3.2 and Uniform margins).
- There is no general trend about which way the VaR moves towards: for different dependence structures (and/or different levels), the VaR is sometimes greater, sometimes smaller than under mutual independence. This is not too surprising, since 'pairwise independence' is not something specific (as seen before, it allows for many different dependencies).
- Within a given example (i.e., a given dependence structure), the ratio (4.2.7) varies substantially across different choices of margins. This is also not too surprising, since the quantiles of S are highly impacted by the marginal distribution chosen (not only the dependence). On first thought, it may seem surprising that the heavier-tailed Log-normal distribution *does not* produce the most 'extreme' discrepancies (quite the opposite: the ratios furthest from 1 are obtained for the short-tailed Uniform and Normal margins). However, we note that a heavier tailed distribution produces larger high quantiles under both dependence and independence. Hence, if both the numerator and denominator of (4.2.7) increase, their ratio can decrease. Said otherwise, when margins have a bigger impact on the VaR, the impact of the dependence can be comparatively less pronounced.
- We find interesting that Example 3.5 (which featured a much more 'subtle' form of dependence compared to the other examples) produces ratios that can still be substantially different from 1. This means that fairly subtle dependence (at least,

[1] We note this way to compute the VaR and TVaR is rather 'brute force'. This is sufficient for our purposes, though it would also be possible to derive the theoretical distribution of S under each scenario, and hence obtain the exact values of the risk measures.

dependence which is hard to detect with the naked eye, recall Figures 3.12 and 3.13) can still have a sizeable impact on Capital Required.

For the results about TVaR (Table 4.2), the same general observations hold, though the ratios are generally less drastically far from 1: the lowest ratio is 78% (attained for Example 3.4 and Uniform margins) and the highest is 117% (attained for Example 3.3 and Uniform and Normal margins).

Marginal	**Level**	**Ex. 3.1**	**Ex. 3.2**	**Ex. 3.3**	**Ex. 3.4**	**Ex. 3.5** ($\alpha = 1$)	**Ex.3.5** ($\alpha = -1$)
Uniform	70%	0.97	0.47	0.85	1.16	1.06	0.92
	90%	0.92	1.24	1.17	0.71	0.95	1.06
	95%	1.04	1.17	1.18	0.66	0.94	1.07
	99%	1.10	1.07	1.15	0.97	0.93	1.05
	99.5%	1.09	1.05	1.12	1.01	0.93	1.04
Normal	70%	0.95	0.72	0.94	1.11	1.05	0.94
	90%	0.96	1.16	1.10	0.81	0.97	1.03
	95%	0.99	1.14	1.12	0.81	0.95	1.05
	99%	1.07	1.10	1.15	0.94	0.93	1.05
	99.5%	1.09	1.08	1.16	0.99	0.93	1.05
Exponential	70%	0.91	0.75	0.76	1.05	1.09	0.91
	90%	0.97	1.05	1.00	0.93	1.00	1.00
	95%	0.99	1.06	1.04	0.94	0.98	1.02
	99%	1.03	1.06	1.09	0.98	0.97	1.03
	99.5%	1.05	1.06	1.11	0.99	0.96	1.03
Log-normal	70%	0.91	0.76	0.76	1.05	1.10	0.90
	90%	0.98	1.03	0.99	0.95	1.00	1.00
	95%	0.99	1.04	1.02	0.96	0.99	1.01
	99%	1.02	1.03	1.05	0.99	0.98	1.02
	99.5%	1.02	1.02	1.05	1.00	0.98	1.01

Table 4.1: Ratios of Capital Required (using VaR) under different PIBD structures (compared to the mutual independence case)

Marginal	**Level**	**Ex. 3.1**	**Ex. 3.2**	**Ex. 3.3**	**Ex. 3.4**	**Ex. 3.5** ($\alpha = 1$)	**Ex.3.5** ($\alpha = -1$)
Uniform	70%	0.97	1.13	1.12	0.83	0.97	1.03
	90%	1.04	1.15	1.17	0.78	0.94	1.06
	95%	1.08	1.10	1.16	0.87	0.93	1.05
	99%	1.09	1.05	1.12	1.01	0.93	1.04
	99.5%	1.08	1.04	1.10	1.03	0.94	1.03
Normal	70%	0.98	1.09	1.09	0.88	0.98	1.02
	90%	1.01	1.13	1.13	0.86	0.95	1.05
	95%	1.04	1.11	1.14	0.90	0.94	1.05
	99%	1.10	1.08	1.16	1.00	0.93	1.05
	99.5%	1.11	1.07	1.17	1.03	0.93	1.05
Exponential	70%	0.98	1.02	1.00	0.95	1.00	1.00
	90%	1.01	1.06	1.06	0.96	0.98	1.02
	95%	1.03	1.06	1.08	0.97	0.97	1.03
	99%	1.06	1.06	1.12	1.00	0.96	1.04
	99.5%	1.08	1.06	1.14	1.02	0.96	1.04
Log-normal	70%	0.99	1.01	1.00	0.97	1.00	1.00
	90%	1.00	1.03	1.03	0.98	0.99	1.01
	95%	1.01	1.03	1.04	0.98	0.99	1.01
	99%	1.03	1.02	1.05	1.00	0.99	1.02
	99.5%	1.03	1.02	1.05	1.01	0.99	1.01

Table 4.2: Ratios of Capital Required (using TVaR) under different PIBD structures (compared to the mutual independence case)

An alternative way to assess the impact of different dependence scenarios on Capital Required is by computing the Diversification Benefit ('DB', recall Equation 1.2.1), i.e.,

$$\mathrm{DB}(\boldsymbol{X}) = 100\% - \frac{\mathrm{CR}(\sum_{i=1}^{d} X_i)}{\sum_{i=1}^{d} \mathrm{CR}(X_i)},$$

under those different dependence scenarios. For both VaR (Table 4.3) and TVaR (Table 4.4), we report the DB in each of our dependence scenarios. In both tables, the column labelled 'Ind.' corresponds to mutual independence. Note those results are based on the same Monte Carlo simulations we performed to obtain the values in Tables 4.1 and 4.2.

We observe that the Diversification Benefit under PIBD is sometimes smaller and sometimes larger than under mutual independence. This is in line with the previous results from Tables 4.1 and 4.2, where we established that VaR and TVaR under PIBD are sometimes smaller and sometimes larger than under mutual independence.

Here, we simply note that the difference between the DB under mutual independence and pairwise independence can be significant. For example, for the VaR results (Table 4.3), looking at the results for Log-normal margins at level 70%, while under independence the

DB is -45% (meaning that pooling risks increases Capital Required), for the dependence of Example 3.2, it is substantially higher, at -10%. However, for Example 3.5 ($\alpha = 1$), the DB is even more negative, at -58%.

Looking at the TVaR results (Table 4.4), the largest differences (in absolute value) are observed for Uniform margins. Indeed, at the level 90%, under independence the DB is 36%. It is substantially higher for Example 3.4, at 50%, and substantially lower for Example 3.2, at 25%.

Marginal	**Level**	**Ind.**	**Ex.3.1**	**Ex.3.2**	**Ex.3.3**	**Ex.3.4**	**Ex.3.5** ($\alpha = 1$)	**Ex.3.5** ($\alpha = -1$)
Uniform	70%	54%	56%	78%	61%	47%	51%	58%
	90%	45%	50%	32%	36%	61%	48%	42%
	95%	38%	36%	28%	27%	59%	42%	34%
	99%	25%	17%	19%	14%	27%	30%	21%
	99.5%	20%	12%	16%	10%	19%	25%	17%
Normal	70%	42%	45%	58%	46%	36%	39%	46%
	90%	42%	45%	33%	37%	53%	44%	40%
	95%	42%	43%	34%	35%	53%	45%	39%
	99%	42%	38%	37%	34%	46%	46%	39%
	99.5%	42%	37%	37%	33%	43%	46%	39%
Expon.	70%	-1%	9%	25%	23%	-6%	-10%	9%
	90%	41%	42%	38%	41%	45%	41%	41%
	95%	45%	45%	42%	43%	48%	46%	44%
	99%	50%	48%	47%	45%	51%	52%	49%
	99.5%	51%	49%	49%	46%	52%	53%	50%
Log-norm	70%	-45%	-31%	-10%	-10%	-52%	-58%	-30%
	90%	34%	35%	32%	34%	37%	34%	34%
	95%	40%	40%	38%	39%	42%	41%	39%
	99%	48%	47%	46%	45%	48%	48%	47%
	99.5%	50%	49%	49%	47%	50%	51%	49%

Table 4.3: Diversification Benefit (using VaR) under different dependence scenarios

Marginal	Level	Ind.	Ex.3.1	Ex.3.2	Ex.3.3	Ex.3.4	Ex.3.5 ($\alpha = 1$)	Ex.3.5 ($\alpha = -1$)
Uniform	70%	44%	46%	37%	37%	53%	46%	42%
	90%	36%	33%	26%	25%	50%	40%	32%
	95%	30%	24%	23%	19%	39%	35%	26%
	99%	19%	12%	15%	9%	18%	24%	16%
	99.5%	15%	8%	12%	7%	13%	20%	13%
Normal	70%	42%	43%	37%	37%	49%	44%	41%
	90%	42%	41%	35%	35%	50%	45%	39%
	95%	42%	40%	36%	34%	48%	46%	39%
	99%	42%	37%	38%	33%	42%	46%	39%
	99.5%	42%	36%	38%	33%	41%	47%	39%
Expon.	70%	41%	42%	40%	41%	44%	41%	41%
	90%	47%	46%	44%	44%	49%	48%	46%
	95%	49%	47%	46%	45%	50%	50%	47%
	99%	52%	49%	49%	46%	52%	54%	50%
	99.5%	53%	49%	50%	47%	52%	55%	51%
Log-norm	70%	37%	38%	37%	38%	39%	37%	38%
	90%	44%	43%	42%	42%	45%	44%	43%
	95%	46%	46%	45%	44%	47%	47%	46%
	99%	51%	50%	50%	49%	51%	52%	50%
	99.5%	53%	51%	52%	50%	52%	53%	52%

Table 4.4: Diversification Benefit (using TVaR) under different dependence scenarios

Remark 4.3. *Of course, the 'pairwise independence scenarios' chosen here are not the only ones possible, and we cannot infer that other PIBD structures would generate similar results. Another limitation of our analysis is that we investigated only the case of three risks, X_1, X_2, X_3. Nonetheless, this section serves to highlight that there can be an important difference between the required capital (and diversification benefit) under an assumption of mutual independence, compared to an assumption of pairwise independence.*

4.2.3 Extreme value theory

As Embrechts et al. (1999) put it, "extreme value theory plays an important methodological role within risk management for insurance, reinsurance, and finance". Many classical results in this theory rely on the mutual independence between a series of risks, and we recall here what is arguably the most central result of classical extreme value theory (EVT), namely the Fisher-Tippett Theorem (see, e.g., Embrechts et al., 1997, Theorem 3.2.3). We then give an example to illustrate this theorem does not hold for merely *pairwise independent* variables.

Theorem 4.1. *Let $\{X_n\}$ be a sequence of i.i.d. random variables, and denote their maximum by $M_n \coloneqq \max(X_1, \ldots, X_n)$. If there exists norming constants $c_n > 0, d_n \in \mathbb{R}$ and*

some non-degenerate distribution function H such that

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H, \tag{4.2.8}$$

then H belongs to the type of one of the following three distribution functions:

- Fréchet:
$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp[-x^{-\alpha}], & x > 0 \end{cases} \qquad \alpha > 0,$$

- Weibull:
$$\Psi_\alpha(x) = \begin{cases} \exp[-(-x)^\alpha)], & x \leq 0 \\ 1, & x > 0 \end{cases} \qquad \alpha > 0,$$

- Gumbel:
$$\Lambda(x) = \exp[-e^{-x}], \quad x \in \mathbb{R}.$$

We can see this theorem as an analogue of a CLT for standardised maximum (as opposed to standardised sums). We now provide a counterexample to the theorem in the case of PIBD r.v.s. This example is interesting because:

- the standardised maximum appears to converge to a distribution which is not what is 'predicted' by the theorem (for i.i.d. variables).
- the expected value of the standardised maximum appears to go to $-\infty$ as n increases.

We note that the sequence in this example comes from Janson (1988, see Remark 2), but our application to EVT is original.

Example 4.3. *Let ξ, η be independent Uniform[0,1] random variables, and define a sequence of random variables $\{X_j, j \geq 1\}$ as*

$$X_j = \frac{\cos(2\pi(\eta + j\xi)) + 1}{2}, \quad j = 1, 2, \ldots$$

We then have that the random variables $\{X_j,\ j \geq 1\}$ are pairwise independent (though not mutually independent).

We note that the random variables $\{X_j,\ j \geq 1\}$ in Example 4.3 are identically distributed with Beta$(a = 1/2, b = 1/2)$ distribution. This follows directly from the proof of part a) of

Proposition 3.3, together with the proof of Proposition 4.3. Then, if we set the constants

$$c_n^{-1} = \left(\frac{2n}{\pi}\right)^2, \qquad d_n = 1$$

(which are the values required for convergence under mutual independence, see Proposition 4.4 in Appendix 4.A), we have that

$$H_n \coloneqq c_n^{-1}(M_n - d_n)$$

appears to converge to a non-degenerate distribution which is *not* the Weibull($\alpha = 1/2$), i.e., the distribution 'predicted' by the Fisher-Tippett Theorem for Beta distributed i.i.d. random variables. We say 'appears' because we unfortunately do not have a proof of this statement, but our simulation results are, we believe, convincing. They are presented on Figure 4.4, which displays the histogram and empirical CDF of $-\log(-H_n)$ for $n =$ 100,000 (and based on 3×10^6 simulated samples), compared to those of $-\log(-W)$ for $W \sim$ Weibull(1/2). We used the log transformation for better visualisation, since the distribution of H_n itself is extremely heavy-tailed. We notice that the distribution of H_n is not degenerate, and markedly different from the Weibull; especially, it has a much heavier left tail.

Furthermore, the first moment of H_n appears to decrease linearly as n increase, and hence it appears to tend to $-\infty$ for $n \to \infty$. This is illustrated on Figure 4.5, which shows the empirical means of H_n, for n varying from $n = 10$ to $n =$ 100,000 (results are presented on a log-log scale).

Remark 4.4. *It is not paradoxical that H_n appears to converge to a fixed distribution, while $\mathbb{E}[H_n]$ goes to $-\infty$. This is certainly theoretically possible. For example, for $F(x)$ a non-degenerate CDF (with finite expectation) and F_n ($n = 1, 2, \ldots$) a sequence of CDFs defined as:*

$$F_n(x) = \frac{n-1}{n}F(x) + \frac{1}{n}\mathbb{1}_{[n^2,\infty)}(x),$$

(i.e., F_n is a mixture of F and the constant n^2), we have that $F_n(\cdot)$ converges pointwise to $F(\cdot)$, but the expectation of a random variable $X \sim F_n$ goes to infinity for $n \to \infty$.

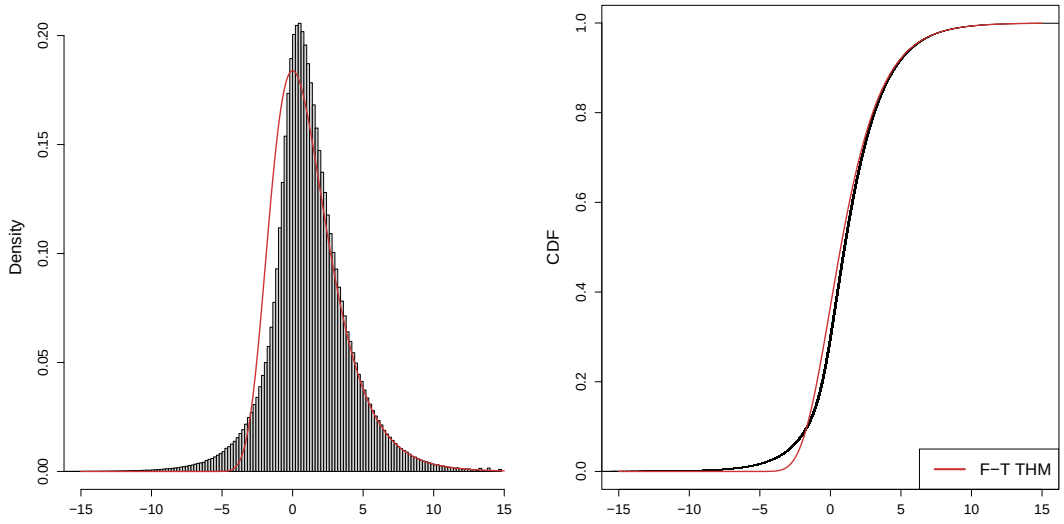

Figure 4.4: Density (left) and CDF (right) of $-\log(-H_n)$ as compared to those 'predicted' by the Fisher-Tippett Theorem (red line), for $n = 100{,}000$

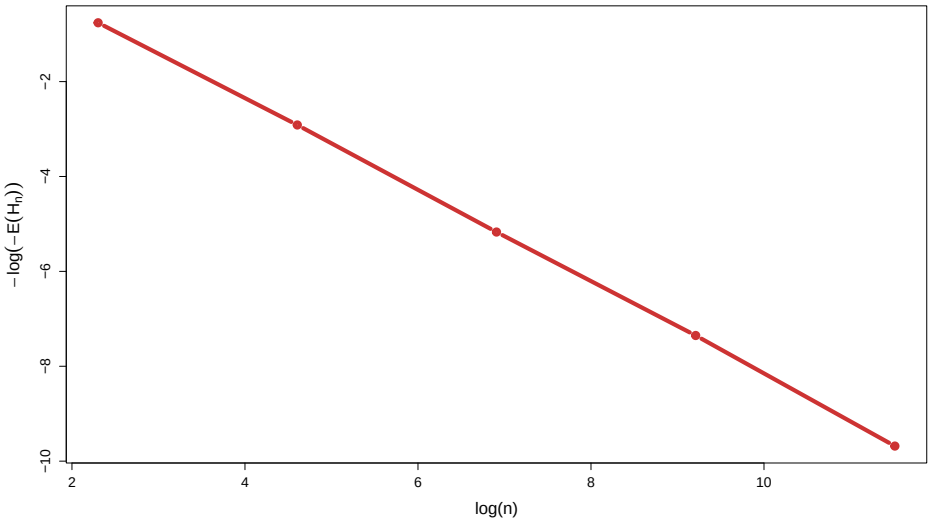

Figure 4.5: Empirical mean of H_n (log-transformed) for increasing values of n

4.2.4 Bootstrap

The bootstrap is a powerful statistical technique which can be used to estimate distributions, standard errors and confidence intervals for quantities of interest within a model, but without resorting to strong distributional assumptions (for a classical reference on the bootstrap, see Efron and Tibshirani, 1994). Bootstrapping has been used in actuarial science for many purposes, including prediction errors in claims reserving (England and Verrall, 1999), obtaining the predictive distribution of outstanding loss liabilities within the chain-ladder model (Peters et al., 2010) or measuring uncertainty of mortality projections (D'Amato et al., 2012).

A crucial assumption of the bootstrap is that observations within a collected sample are mutually independent. In this section, we showcase that the bootstrap can fail drastically if a sample is only pairwise independent. For illustrative purposes, we consider the following simple problem.

Problem 4.1.

Let $\boldsymbol{X} \coloneqq (X_1, \ldots, X_n)$ *be discrete data, with* $X_j \in \{0, 1, \ldots\}$ *for* $j = 1, \ldots, n$*. We want to estimate*

$$q \coloneqq \mathbb{P}[X_j = 0],$$

and obtain a confidence interval for q[2]*. The point estimator of* q *is given by*

$$\widehat{q} = \frac{\#\text{ of 0's in } \boldsymbol{X}}{n}, \tag{4.2.9}$$

but this is of course just a number. For simplicity, assume we want to derive a one-sided confidence interval of the form:

$$[L, 1].$$

We can use the bootstrap to obtain such a confidence interval, i.e., we can apply the following procedure.

Procedure 4.1.

i. We fix the level of our confidence interval to $1 - \alpha$. Here, we use $\alpha = 10\%$ (this is arbitrary).

ii. For $i = 1, 2, \ldots, B$ (B a 'large number', henceforth set to $B = 5000$) we generate bootstrapped samples of $\boldsymbol{X}$ (of the same original size n). Call them $\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_B^*$.

iii. For each of these bootstrapped samples we compute a new estimator of q, i.e.,

$$\widehat{q}_i^* = \frac{\#\text{ of 0's in } \boldsymbol{X}_i^*}{n}, \qquad i = 1, 2, \ldots, B.$$

iv. This yields a bootstrapped sample of the statistic of interest, $\widehat{q}$. The bootstrap

[2] An actuarial example fitting this setting could be if the X_j's represent the remaining whole years of life for individuals in an insurance portfolio (all with similar mortality). Then, q would represent the probability a given individual dies within the next year.

confidence interval (BCI) for q, at level α, is then given simply by setting

$$L = \widehat{F}^{-1}_{q^*}(\alpha),$$

where $\widehat{F}^{-1}_{q^*}(\alpha)$ is the empirical α-quantile of the bootstrapped values $\widehat{q}^*_1, \ldots, \widehat{q}^*_B$.

We follow the above procedure under two different data-generating processes, namely:

1. The X_j's are i.i.d. Poisson(log(2)).
2. The X_j's are generated as in Example 4.1, hence they are Poisson(log(2)) and PIBD.

Results for Scenario 1: Here, the bootstrap is supposed to work (since the i.i.d. assumption is verified), and we quickly check that it is the case.

We first let $n = 45$. Figure 4.6 shows the histogram of one bootstrapped sample $\widehat{q}^*_1, \ldots, \widehat{q}^*_B$. For this specific simulation, the bootstrapped confidence interval does contain the true value $q = 1/2$. We also note that the empirical bootstrap distribution looks approximately Normal (which is expected).

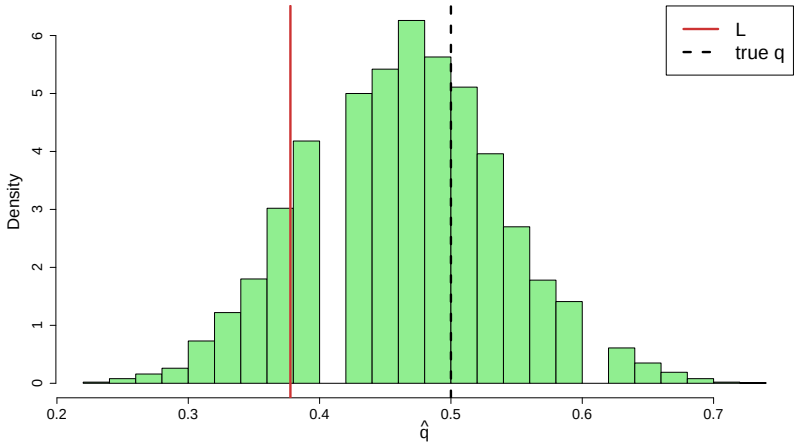

Figure 4.6: Histogram of a bootstrapped sample of empirical proportions $\widehat{q}^*$ from a i.i.d. Poisson(log(2)), with sample size $n = 45$

Figure 4.6 alone does not tell us much, as it represents only one simulation. A different sample would have yielded a different BCI interval. Hence, we repeat the procedure 10,000 times, and we count the number of times the 'true' value $q = 1/2$ falls *outside* the BCI. We obtain a proportion of

$$\widehat{\alpha} = 0.118.$$

This is relatively close to the expected level $\alpha = 0.1$. If we repeat the same procedure for

a larger sample size $n = 990$, we obtain

$$\widehat{\alpha} = 0.095,$$

which is closer to the theoretical level, indicating that the bootstrap improves as the sample size increases (which is expected).

Results for Scenario 2: Before repeating the bootstrap simulations for a PIBD sample, let us first look at what the 'true' distribution of $\widehat{q}$ looks like in that case. We generate 10,000 samples of size $n = 45$ (for PIBD variables as in Example 4.1) and compute $\widehat{q}$ for each of them, as to obtain a sample

$$\widehat{q}_1, \ldots, \widehat{q}_{10000}.$$

The histogram of those 10,000 values is presented on Figure 4.7. We see that the distribution of $\widehat{q}$ is odd, and radically different than in the i.i.d. case. It is highly skewed to the left, and only a few values are allowed.

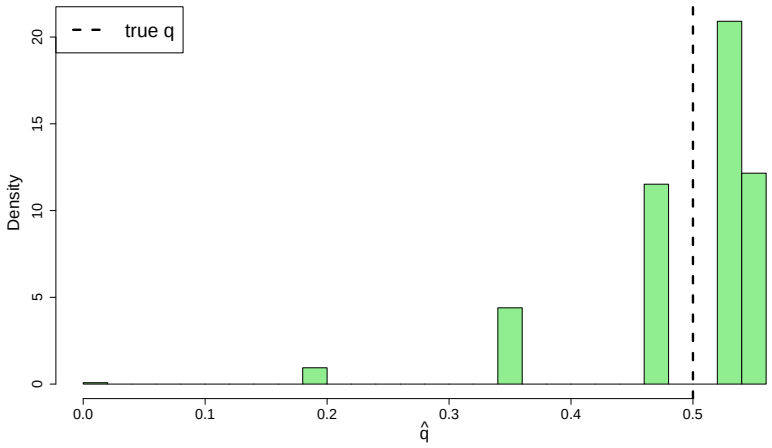

Figure 4.7: Histogram of empirical proportions $\widehat{q}$, obtained from PIBD Poisson(log(2)) variables (sample size $n = 45$)

Next, we follow the bootstrap procedure, again with $n = 45$. For a given PIBD sample, we obtain the bootstrapped sample $\widehat{q}_1^*, \ldots, \widehat{q}_B^*$, whose histogram is shown on Figure 4.8.

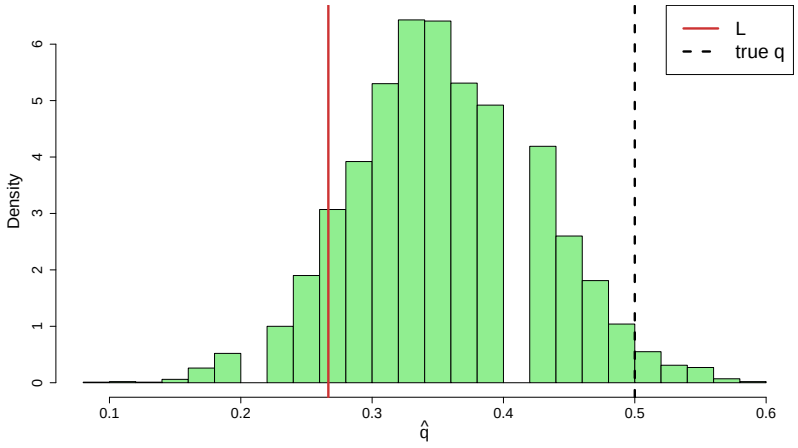

Figure 4.8: Histogram of bootstrapped empirical proportions $\widehat{q}^*$, obtained from PIBD Poisson(log(2)) variables (sample of size $n = 45$)

Perhaps unsurprisingly, Figure 4.8 reveals that the bootstrap procedure is incapable of reproducing the 'real' distribution of $\widehat{q}$ displayed on Figure 4.7. We note that the true value of q is contained in the BCI. This, again, does not tell us much since a different original sample would have yielded different results. We repeat the bootstrap procedure 10,000 times, and we record the proportion of times the true value $1/2$ is *outside* the BCI. This gives us the empirical level of the BCI. We obtain the value

$$\widehat{\alpha} = 0.00.$$

This is totally 'off', as we expect a proportion around 0.10. A sample size of $n = 45$ is relatively small, hence we repeat the procedure for $n = 990$. We obtain again an empirical level of 0.00. This means that the (bootstrapped) distribution of $\widehat{q}^*$ is unable to reproduce the 'true' distribution of $\widehat{q}$, and hence the BCI we obtain are meaningless.

Remark 4.5. *It is not surprising that the bootstrap fails here, as what causes the 'odd' distribution of $\widehat{q}$ shown on Figure 4.7 is the dependence that exists between the observations, which is not reproduced when bootstrapping the sample. Indeed, even if every pair (X_i, X_j) of observations is independent, many triplets (X_i, X_j, X_k) are strongly dependent. When randomly re-sampling the observations, those triplets are not reproduced in the bootstrapped samples, hence the 'pattern of dependence' is broken and the bootstrapped distribution obtained for $\widehat{q}$ is totally different from its true distribution.*

In closing this section, we note that many other important theorems and techniques which rely on mutual independence do not hold under sole pairwise independence (the list we have presented here is far from exhaustive). For instance, in Appendix 4.B, we provide

an additional example which shows that the Zero-One Law of Kolmogorov, as well as the Law of the Iterated Logarithm, can 'fail' for a sequence of PIBD random variables.

4.3 Popular dependence models which do not allow for PIBD variables

In the previous section, we established that there is a material difference between pairwise and mutual independence, in the sense that many results and techniques which rely on mutual independence are not valid under sole pairwise independence. In this section, we show that this difference is not captured by some of the most common dependence models used in actuarial science. That is, we show that those models do not allow for the possibility of PIBD variables. Said otherwise, within those models, pairwise independence implies mutual independence.

4.3.1 Elliptical distributions

We first recall the definition of *elliptical distributions* (the most common examples of which are the multivariate Normal and multivariate t distributions) which form a "class of distributions that has increased in popularity, both in finance and insurance" (Xiao and Valdez, 2015). A classical reference on spherical and elliptical distributions is Fang et al. (1990). In our treatment, we use the notation and definitions found in McNeil et al. (2015, Section 6.3). To define *elliptical* distributions, we must first define *spherical* distributions.

Definition 4.1. *A random vector* $\boldsymbol{X} \coloneqq (X_1, \ldots, X_d)'$ *has a spherical distribution if, for every orthogonal map* $U \in \mathbb{R}^{d\times d}$ *(i.e., maps such that* $UU' = U'U = I_d$*, where* I_d *is the identity matrix),*

$$U\boldsymbol{X} \overset{d}{=} \boldsymbol{X},$$

where '$\overset{d}{=}$' denotes 'equality in distribution'. An important result about spherical distributions is given in the next theorem (see McNeil et al., 2015, Theorem 6.18).

Theorem 4.2. *A random vector* $\boldsymbol{X}$ *has a spherical distribution if and only if there exists a function* ψ *of a scalar variable such that, for all* $\boldsymbol{t} \in \mathbb{R}^d$*,*

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}\left[\exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{X})\right] = \psi(\boldsymbol{t}'\boldsymbol{t})$$

The function ψ is called the *characteristic generator* of the spherical distribution, and we use the notation $\boldsymbol{X} \sim S_d(\psi)$. We can now introduce the definition of an *elliptical* distribution.

Definition 4.2. *A random vector* $\boldsymbol{X} \coloneqq (X_1, \ldots, X_d)'$ *has an elliptical distribution if*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + A\boldsymbol{Y},$$

where $\boldsymbol{Y} \sim S_k(\psi)$, $A \in \mathbb{R}^{d\times k}$ *is a matrix of constants, and* $\boldsymbol{\mu} \in \mathbb{R}^d$ *is a vector of constants.*

We can now state our result (which we believe, in such generality, is new).

Proposition 4.1. *Let* $\boldsymbol{X} \coloneqq (X_1, \ldots, X_d)'$ *be a random vector having an elliptical distribution. If all pairs of variables* $(X_j, X_k), j \neq k$ *of this vector are independent, then the variables* $X_1, \ldots, X_d$ *are mutually independent.*

Proof. Let $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Y}$ be the representation of $\boldsymbol{X}$ as an affine transformation of a spherical random vector $\boldsymbol{Y}$ (in the sense of Definition 4.2). The characteristic function of $\boldsymbol{X}$ can be written as

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu})\psi(\boldsymbol{t}'\Sigma\boldsymbol{t}), \tag{4.3.1}$$

where $\boldsymbol{t} \in \mathbb{R}^d$, $\Sigma = AA'$ and ψ is the generator of $\boldsymbol{Y}$ (McNeil et al., 2015, p. 200). Therefore, we have, for any $j, k \in \{1, \ldots, d\}$, $j \neq k$,

$$\begin{aligned}\varphi_{X_j}(t_j) &= \exp(\mathrm{i}t_j\mu_j)\psi(t_j^2\sigma_{jj}),\\ \varphi_{X_j,X_k}(t_j, t_k) &= \exp(\mathrm{i}t_j\mu_j)\exp(\mathrm{i}t_k\mu_k)\psi(t_j^2\sigma_{jj} + t_k^2\sigma_{kk} + 2t_jt_k\sigma_{jk}),\end{aligned}$$

where $t_j, t_k \in \mathbb{R}$ and $\sigma_{a,b}$ denotes the element in the a^{th} row and b^{th} column of Σ. Because we are under the assumption of pairwise independence, we must also have that, for any $t_j, t_k \in \mathbb{R}$,

$$\varphi_{X_j,X_k}(t_j, t_k) = \exp(\mathrm{i}t_j\mu_j)\psi(t_j^2\sigma_{jj})\exp(\mathrm{i}t_k\mu_k)\psi(t_k^2\sigma_{kk}),$$

and hence that

$$\psi(t_j^2\sigma_{jj})\psi(t_k^2\sigma_{kk}) = \psi(t_j^2\sigma_{jj} + t_k^2\sigma_{kk} + 2t_jt_k\sigma_{jk}). \tag{4.3.2}$$

Equation (4.3.2) is true for all $t_j, t_k \in \mathbb{R}$. Hence, if we let t_j be arbitrary and then set

$$t_k = -2t_j\sigma_{jk}/\sigma_{kk} \implies t_k^2\sigma_{kk} = -2t_jt_k\sigma_{jk},$$

it follows that

$$\psi(t_j^2\sigma_{jj})\psi(t_k^2\sigma_{kk}) = \psi(t_j^2\sigma_{jj}) \implies \psi(t_k^2\sigma_{kk}) = 1.$$

Then, using $t_k^2 = 4t_j^2\sigma_{jk}^2/\sigma_{kk}^2$, we have

$$\psi(4t_j^2\sigma_{jk}^2/\sigma_{kk}) = 1. \tag{4.3.3}$$

Since (4.3.3) is valid for all $t_j \in \mathbb{R}$, we deduce that $\sigma_{jk} = 0$ (note that ψ cannot be a constant function as this does not yield a valid characteristic function). In fact, since the choices of j, k were arbitrary, it also follows that $\sigma_{jk} = 0$ for all j, k such that $j \neq k$. Next, from 4.3.2 we have that

$$\psi(t_j^2\sigma_{jj})\psi(t_k^2\sigma_{kk}) = \psi(t_j^2\sigma_{jj} + t_k^2\sigma_{kk}),$$

and the only continuous function with this property is the power function[3], i.e.,

$$\psi(x) = a^x,$$

for some constants $a \in \mathbb{R}$ (ψ is continuous since it is a characteristic function). But then it follows that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu})\psi(\boldsymbol{t}'\boldsymbol{t}) = \prod_{\ell=1}^{d}\exp(\mathrm{i}t_\ell\mu_\ell)\psi(t_\ell^2\sigma_{\ell\ell}) = \prod_{\ell=1}^{d}\varphi_{X_\ell}(t_\ell), \tag{4.3.4}$$

which implies the mutual independence of all the X's.

□

4.3.2 Archimedean Copulas

Another very common class of dependence models is that of Archimedean copulas, "which enjoy considerable popularity in a number of practical applications" (McNeil and Nešlehová, 2009). In particular, we saw in Section 2.4.1 that many authors have used such copulas to model the dependence between risks within the Individual Risk Model.

We first recall the definition of Archimedean copulas, as stated for instance in McNeil and

[3]See for example on MathStackExchange: https://math.stackexchange.com/questions/1548249

Nešlehová (2009, see Definition 2.2).

Definition 4.3. *A nonincreasing and continuous function* $\psi : [0,\infty) \to [0,1]$ *which satisfies the conditions* $\psi(0) = 1$ *and* $\lim_{x\to\infty} \psi(x) = 0$ *and is strictly decreasing on* $[0, \inf\{x : \psi(x) = 0\})$ *is called an Archimedean generator. A d-dimensional copula* C *is called Archimedean if it permits the representation*

$$C(u_1,\dots,u_d) = \psi\left(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\right), \quad u_1,\dots,u_d \in [0,1]$$

for some Archimedean generator ψ *and its inverse* $\psi^{-1} : (0,1] \to [0,\infty)$ *where, by convention,* $\psi(\infty) = 0$ *and* $\psi^{-1}(0) = \inf\{u : \psi(u) = 0\}$.

Now, it is also the case that, for Archimedean copulas, pairwise independence of all components implies their mutual independence. This is formalised in the following proposition.

Proposition 4.2. *Let* $\boldsymbol{U} \coloneqq (U_1,\dots,U_d)$ *be a random vector of uniforms with Archimedean copula* C*. If this copula is such that pairs* $(U_j, U_k), j \neq k$ *are independent, then this copula is the independence copula* $C(u_1,\dots,u_d) = u_1 \times \cdots \times u_d$.

Proof. First, pairwise independence implies that

$$\psi\left(\psi^{-1}(u_1) + \psi^{-1}(u_2)\right) = u_1 \cdot u_2, \quad \text{for all } u_1, u_2 \in [0,1]. \tag{4.3.5}$$

Then, note that while for any $u \in [0,1]$,

$$\psi(\psi^{-1}(u)) = u, \tag{4.3.6}$$

for the 'other way around' we have

$$\psi^{-1}(\psi(x)) = \min\{x, \psi^{-1}(0)\}, \quad \text{for } x \in [0,\infty). \tag{4.3.7}$$

Now, let $(u_1,u_2,u_3) \in [0,1]^3$, and note that, necessarily, $u_2u_3 \in [0,1]$. From (4.3.5) and (4.3.7) we get

$$\begin{aligned}
\psi^{-1}(u_1u_2u_3) &= \psi^{-1}\left(\psi\left(\psi^{-1}(u_1) + \psi^{-1}(u_2u_3)\right)\right) \\
&= \min\{\psi^{-1}(u_1) + \psi^{-1}(u_2u_3), \psi^{-1}(0)\} \\
&= \min\left\{\psi^{-1}(u_1) + \min\{\psi^{-1}(u_2) + \psi^{-1}(u_3), \psi^{-1}(0)\}, \psi^{-1}(0)\right\} \\
&= \min\left\{\psi^{-1}(u_1) + \psi^{-1}(u_2) + \psi^{-1}(u_3), \psi^{-1}(0)\right\}.
\end{aligned}$$

By repeating the argument, we obtain that

$$\psi^{-1}\left(u_1 \times \cdots \times u_d\right) = \min\left\{\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d), \psi^{-1}(0)\right\}.$$

Next, using (4.3.6) we have

$$\begin{aligned} u_1 \times \cdots \times u_d &= \psi\left(\min\left\{\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d), \psi^{-1}(0)\right\}\right) \\ &= \begin{cases} \psi\left(\psi^{-1}(0)\right) = 0 & \text{if } \psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d) \geq \psi^{-1}(0) \\ \psi\left(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\right) & \text{otherwise.} \end{cases} \end{aligned}$$

Because $u_1 \times \cdots \times u_d = 0$ if and only if there is a $k \in \{1, \ldots, d\}$ such that $u_k = 0$, we have

$$u_1 \times \cdots \times u_d = \begin{cases} 0 & \text{if } u_k = 0 \text{ for some } k \in \{1, \ldots, d\}, \\ C(u_1, \ldots, u_d) & \text{otherwise.} \end{cases}$$

This concludes the proof, since of course $C(u_1, \ldots, u_{k-1}, 0, u_{k+1}, \ldots, u_d) = 0$. □

Remark 4.6. *It is also the case that Archimedean survival copulas are such that pairwise independence of all components implies mutual independence. Indeed, let* $\boldsymbol{U} \coloneqq (U_1, \ldots, U_d)$ *be a random vector of uniforms with Archimedean survival copula* $\widehat{C}$*, meaning that*

$$\begin{aligned} \mathbb{P}[U_1 > u_1, \ldots, U_d > u_d] &= \widehat{C}(1 - u_1, \ldots, 1 - u_d) \\ &= \psi\left(\psi^{-1}(1 - u_1) + \cdots + \psi^{-1}(1 - u_d)\right), \end{aligned}$$

for some Archimedean generator ψ *(in the sense of Definition 4.3). Then, rerunning the proof of Proposition 4.2 (with* C *replaced by* $\widehat{C}$ *and each* u_k *replaced by* $1 - u_k$*, for* $k = 1, \ldots, d$*) yields the result.*

4.3.3 Pair-copula constructions under the 'simplifying assumption'

Pair-copula constructions (PCCs), also known as 'vine copulas', have recently gained widespread popularity to model complex and/or high dimensional dependence, and in a variety of fields. Two seminal papers developing PCCs are Bedford and Cooke (2002) and Aas et al. (2009). As summarised by Aas (2016):

> A PCC is a multivariate copula that is constructed from a set of bivariate ones,

so-called pair-copulae. More specifically, the copula density is decomposed into a product of pair-copula densities. All of these bivariate copulae may be selected completely freely as the resulting structure is guaranteed to be a valid copula. Hence, PCCs are highly flexible and able to characterise a wide range of complex dependencies.

We give a simple example of a PCC in three dimensions (as given in Acar et al., 2012, see their Introduction).

Example 4.4. *Let U_1, U_2, U_3 be three Uniform[0,1] random variables. Assuming their joint density (which is a copula density) exists, it can be decomposed as*

$$c(u_1, u_2, u_3) = c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{13|2}(u_{1|2}, u_{3|2}; u_2), \tag{4.3.8}$$

where:

- $u_k \in [0, 1]$ for $k = 1, 2, 3$
- c_{12} is the copula density of the pair (U_1, U_2), and c_{23} the copula density of the pair (U_2, U_3)
- $c_{13|2}$ is the conditional copula density of the pair (U_1, U_3), given $U_2 = u_2$
- $u_{k|2} = \mathbb{P}[U_k \leq u_k | U_2 = u_2]$ for $k = 1, 3$.

We note that a decomposition such as (4.3.8) is not unique, as we could have also chosen U_1 or U_3 as the 'conditioning variable'.

While in principle any copula density (when it exists) can be expressed as that of a PCC (including, for instance, the 'unusual' dependence from Examples 3.2 and 3.5 in Chapter 3), in practice the so-called *simplifying assumption* is often made. This assumption states "that the copulas corresponding to conditional distributions are constant irrespective of the values of variables that they are conditioned on" (Stöber et al., 2013). In (4.3.8), this would correspond to the conditional copula density $c_{13|2}$ not depending on the third argument u_2, which we then write

$$c_{13|2}(\cdot, \cdot; u_2) = c_{13|2}(\cdot, \cdot).$$

The appropriateness of this simplifying assumption has been the subject of much debate, see, e.g., Haff et al. (2010), Acar et al. (2012), Killiches et al. (2017), Spanhel and Kurz

(2019), or Mroz et al. (2021) for discussions. Here, we do not not weight in on this question. We simply remark that this 'simplifying assumption' restricts the possible dependence structures allowed by PCCs. In particular, under this assumption PCCs do not allow the possibility of PIBD variables. This is easy to see via Example 4.4. Indeed, if the variables U_1, U_2, U_3 are pairwise independent, then (4.3.8) becomes

$$c(u_1, u_2, u_3) = c_{13|2}(u_{1|2}, u_{3|2}) = c_{13}(u_1, u_2) = 1,$$

i.e., U_1, U_2, U_3 are mutually independent.

Remark 4.7. *The fact that under the 'simplifying assumption' pairwise independence implies mutual independence is not only true for three-variate copulas (as in Example 4.4). It is true more generally for any so-called 'regular' PCC, and we note that regular PCCs are a very general form of PCC, of which the common 'canonical' and 'D-vines' PCC are special cases (see, e.g., Aas et al., 2009, Section 2.1).*

To see this, we note that all 'building blocks' of a regular PCC are bivariate conditional copulas. That is, for $\boldsymbol{U} \coloneqq (U_1, \ldots, U_d)$ a vector of random uniforms, the general form of the copula density of $\boldsymbol{U}$ for any regular PCC is given as (see Czado, 2010, Equation 9):

$$c(u_1, \ldots, u_d) = \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)|D(e)} \left(u_{j(e)|D(e)}, u_{k(e)|D(e)} \right) \tag{4.3.9}$$

where

- $\boldsymbol{u} \coloneqq (u_1, \ldots, u_d) \in [0,1]^d$,
- E_1, E_2, E_{d-1} are the sets of edges in the 'tree' making up the PCC (see more details in Section 2.2 of Czado, 2010),
- $j(e), k(e) \in \{1, 2, \ldots, d\}$,
- $D(e) \subset \{1, \ldots, d\}$,
- $c_{j(e),k(e)|D(e)}$ is the bivariate copula density of $U_{j(e)}, U_{k(e)}$ given $\boldsymbol{U}_{D(e)} = \boldsymbol{u}_{D(e)}$, where $\boldsymbol{U}_{D(e)}$ is the sub random vector of $\boldsymbol{U}$ containing variables with indices $D(e)$, and $\boldsymbol{u}_{D(e)}$ is the subvector of $\boldsymbol{u}$ containing variables with indices $D(e)$,
- $u_{j(e)|D(e)} = \mathbb{P}\left[U_{j(e)} \leq u_{j(e)} | \boldsymbol{U}_{D(e)} = \boldsymbol{u}_{D(e)}\right]$.

While the conditioning in any copula $c_{j(e),k(e)|D(e)}$ can be done on a large number of variables (i.e., the sets $D(e)$ can contain many variables), the simplifying assumption (that

any bivariate conditional copula does not dependent on variables in the set $D(e)$) implies that

$$c_{j(e),k(e)|D(e)} = c_{j(e),k(e)} = 1$$

for pairwise independent variables. Hence, in that case the copula density in (4.3.9) *reduces to*

$$c(u_1, \ldots, u_d) = 1,$$

i.e., the mutual independence copula.

Remark 4.8. *In writing this section, our intention was not to discredit the models we have mentioned. Rather, we wanted to point out a 'potential limitation' of those models which we believe has not been highlighted before. To use such models is to implicitly assume that pairwise independence is equivalent to mutual independence. We are not saying this is always problematic. However, and as we have seen, there are potential dangers in assuming such an equivalence.*

4.4 Conclusion

We have seen in this chapter that pairwise independence can be a poor substitute to mutual independence. Indeed, we saw that many results commonly used in actuarial science are not valid for merely 'pairwise independent' observations (i.e., full 'mutual independence' is required for them to hold). Some of those included results about the distribution of sums of random variables (see Section 4.2.1 and 4.2.2), the Fisher-Tippett theorem from extreme value theory (Section 4.2.3) and the bootstrap technique (Section 4.2.4). Of course, our list is far from exhaustive (see Appendix 4.B for an additional example of PIBD variables for which important theorems 'fail'). On the other hand, we saw that many popular dependence models do not allow for the possibility of PIBD variables (Section 4.3). That is to say, within those models pairwise independence implies mutual independence (which can be seen as a limitation of those models).

We note that, in this chapter and the previous, we have dealt with finite numbers of random variables $X_1, \ldots, X_n$, and we did not establish asymptotic results for $n \to \infty$. As a next step, it is interesting to wonder what can happen for infinitely large sequences of PIBD random variables. Is it the case that the impact of the dependence gets weaker for a large enough sample size? As can been intuited already from Examples 4.2 and 4.3, we will see that the answer is largely negative. While the current chapter gave an overview of many different topics, the next two chapters will investigate more generally the case of Central Limit Theorems (CLTs) for PIBD variables. We will see that a PIBD sequence need not verify a CLT, and in particular the novelty in our results will be to provide sequences with arbitrary marginal distributions, and for which we will obtain explicitly the (non-Normal) asymptotic distribution of the standardised sample mean.

4.A Proofs

Proposition 4.3. *Let $\{X_j, 1 \leq j \leq d\}$ be random variables defined as in Example 4.2. Then, all X_j's are pairwise independent and identically distributed with Uniform$[0,1)$ distribution.*

Proof. From Janson (1988, see Remark 2), we know that random variables defined a

$$Y_j \coloneqq \exp(2\pi \mathrm{i}(\eta + j\xi)) \qquad \text{for } j = 1, 2, \ldots$$

have a uniform distribution on the (complex) unit circle, and are pairwise independent. It follows that the angles (often called 'arguments') of any of those Y_j (in their polar coordinates representation), which are simply given by

$$2\pi(\eta + j\xi) \bmod 2\pi,$$

are uniformly distributed on [0, 2π), and also pairwise independent. Hence, we must have that

$$X_j = (\eta + j\xi) \bmod 1$$

are uniformly distributed on [0,1), and pairwise independent. □

Proposition 4.4. *Let $\{X_j, j \geq 1\}$ be i.i.d. Beta$(a = 1/2, b = 1/2)$ random variables, and let $M_n \coloneqq \max(X_1, \ldots, X_n)$. With norming constants*

$$c_n^{-1} = \left(\frac{2n}{\pi}\right)^2, \qquad d_n = 1,$$

we have that $c_n^{-1}(M_n - d_n) \xrightarrow{d} \Psi_{1/2}$, i.e., a Weibull$(\alpha = 1/2)$.

Proof. Call $\bar{F}(x)$ the survival function of a Beta$(a = 1/2, b = 1/2)$ random variable. From Example 3.3.17 in Embrechts et al. (1997), we have that $\bar{F}(1 - 1/x)$ is regularly varying (with index $-b = -1/2$), and also that

$$\bar{F}(1 - 1/x) \sim \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b+1)}(1-x)^b = \frac{2}{\pi}(1-x)^{1/2},$$

for $x \uparrow 1$. Hence, by Example 3.3.16 in Embrechts et al. (1997), $F \in \mathrm{MDA}(\Psi_{1/2})$, and the

result holds with norming constants

$$c_n = (n \cdot 2/\pi)^{-2}, \qquad d_n = 1.$$

□

4.B Other theorems which 'fail' for PIBD variables

We present an additional example (taken from Cuesta and Matrán, 1991) which shows that the Zero-One Law of Kolmogorov and the Law of Iterated Logarithms (two central results of probability theory) also 'fail' for PIBD random variables.

Example 4.5. *Let p be a prime number and let $\mathcal{P} \coloneqq \{0, 1, \ldots, p-1\}$. Let Y_0 and $\{Z_{n \cdot p}, n = 0, 1, \ldots\}$ be mutually independent random variables with uniform distribution on the set $\mathcal{P}$. Define*

$$Z_{0+k} = Z_0 \oplus k \cdot Y_0, \quad k = 0, 1 \ldots, p-1,$$
$$Z_{n \cdot p+k} = Z_{n \cdot p} \oplus k \cdot Y_0, \quad k = 0, 1 \ldots, p-1, \quad n = 1, 2, \ldots,$$

where $\oplus$ means 'addition modulo p'. The sequence $\{Z_0, Z_1, \ldots\}$ hence defined is then pairwise independent[4], and it does not satisfy the ZOL (nor the Law of Iterated Logarithms, nor a CLT).

We next recall the ZOL, for which we first need the definition of a tail-σ-field.

Definition 4.4 (Resnick (1999), p.107)**.** *Given $\{X_j, j \geq 1\}$ a sequence of random variables, let*

$$\mathcal{F}_n \coloneqq \sigma(X_{n+1}, X_{n+2}, \ldots), \quad n = 1, 2, \ldots.$$

The tail-σ field of $\{X_j, j \geq 1\}$, denoted $\mathcal{T}$, is the σ-field defined as

$$\mathcal{T} \coloneqq \bigcap_{n=1}^{\infty} \mathcal{F}_n.$$

From Definition 4.4, we see that the events contained in $\mathcal{T}$ are those that do not depend on any finite number of random variables in the sequence $\{X_j\}$. Rather, they depend on the 'tail' of the sequence. For instance, let Ω be the sample space on which the X's are defined, and let $S_n = X_1 + \cdots + X_n$. The event

$$\left\{\omega \in \Omega : \lim_{n \to \infty} \frac{S_n(\omega)}{n} = 0\right\}$$

belongs to $\mathcal{T}$. We now state the Kolmogorov Zero-One Law, as Theorem 4.1.

[4]Note that the sequence $\{Z_0, Z_1, \ldots\}$ is not stationary, and that Cuesta and Matrán (1991) define a further pairwise independent sequence which is stationary. However, the sequence $\{Z_0, Z_1, \ldots\}$ is sufficient for our purposes and we stick to it for simplicity.

Theorem 4.1 (Resnick (1999), Theorem 4.5.3)**.** *Let $\{X_j, j \geq 1\}$ be a sequence of independent random variables, with tail-σ-field $\mathcal{T}$. Then, for any $A \in \mathcal{T}$, $\mathbb{P}[A] = 0$ or 1.*

In Theorem 4.1, the assumption of mutual independence cannot be lessened to pairwise independence, and Example 4.5 provides a good illustration of this. Indeed, for this sequence $\{Z_j, j \geq 0\}$ consider the event

$$A \coloneqq \{Z_{n \cdot p} = Z_{n \cdot p+1} \text{ i.o.}\},$$

where 'i.o.' stands for infinitely often. Then, A is in the tail-σ-field of $\{Z_j, j \geq 0\}$. Next, note that, conditionally on the event $\{Y_0 \neq 0\}$, we have

$$Z_{n \cdot p} \neq Z_{n \cdot p+1}, \quad \forall n,$$

so that $\mathbb{P}[A|Y_0 \neq 0] = 0$. On the other hand, conditionally on the event $\{Y_0 = 0\}$,

$$Z_{n \cdot p} = Z_{n \cdot p+1}, \quad \forall n,$$

so that $\mathbb{P}[A|Y_0 = 0] = 1$. Unconditionally, we then have that $\mathbb{P}[A] = \mathbb{P}[Y_0 = 0] = 1/p$. This probability being neither 0 nor 1, the pairwise independent sequence $\{Z_j, j \geq 0\}$ does not respect the ZOL.

Remark 4.9. *This failure of the ZOL for a pairwise independent sequence provides useful insight on the difference between mutual and pairwise independence. Indeed, consider first what the ZOL says about a sequence of mutually independent random variables. It tells us that any event relating to the tail of that sequence is either improbable (it has a probability of 0) or certain (it has a probability of 1). This is a strong statement. For instance, in the infinite coin-tossing of a fair coin (with independent trials), we are certain that the frequency of 'Heads' will converge to 1/2. We are also certain that a sequence of 'heads only' has a probability of 0, and that any specific pattern of finite length such has 'Heads-Tails-Head' will appear infinitely often. Etc.*

On the other hand, for a sequence which is only pairwise independent, we have no such guarantees. Indeed, a given tail event can have a probability strictly between 0 and 1, meaning that this event 'might or might not happen'. Hence, in a sense, we know far less about that sequence than we know about a mutually independent sequence.

Next, we recall as Theorem 4.2 the 'classical' version of the LIL (noting that many varia-

tions exist, see Gut (2013, Chapter 8) for an overview).

Theorem 4.2 (Gut (2013), Chapter 8, Theorem 1.2)**.** *Let $\{X_j, j \geq 1\}$ be i.i.d. random variables with mean 0 and finite variance σ^2, and set $S_n = \sum_{k=1}^{n} X_k, n \geq 1$. Then*

$$\limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log\log n}} = +1 \quad a.s.,$$

while

$$\liminf_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log\log n}} = -1 \quad a.s.$$

Remark 4.10. *This theorem sits 'in between' the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). Indeed, consider the Strong LLN, which states that*

$$\bar{X}_n = \frac{S_n}{n} \xrightarrow{a.s.} 0,$$

where '$\xrightarrow{a.s.}$' denotes almost sure convergence. Qualitatively, this means that the denominator n grows to be 'very large' compared to the numerator S_n, in that, for large n, it 'flattens out' all the random variations in S_n. On the other hand, the CLT says that if we use the much smaller denominator $\sqrt{n}$, we obtain

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

In words, it means $\sqrt{n}$ is 'small enough' so that $S_n/\sqrt{n}$ does not get 'flattened out' and stays a non-degenerate random variable, which 'visits infinity' infinitely often. Indeed, as a consequence of the CLT and of Kolmogorov's Zero-One Law (see Theorem 4.1), we have that, almost surely,

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = +\infty \quad \text{and} \quad \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty.$$

Hence, the LIL provides a 'balance' between the LLN and the CLT, since the denominator $\sigma\sqrt{2n\log\log n}$ is small enough so that $S_n/(\sigma\sqrt{2n\log\log n})$ still fluctuates around 0, but big enough so that, almost surely, it stays between two fixed bounds -1 and $+1$.

Now, the LIL is not necessarily valid for PIBD random variables. To see this, from the sequence $\{Z_j, j \geq 0\}$ in Example 4.5, define a new sequence $\{X_j, j \geq 0\}$ as its zero-mean

version, i.e.,

$$X_j = Z_j - \frac{p-1}{2}, \qquad \text{for } j = 0, 1, \ldots$$

Then, consider the sequence $Z_0, Z_1, \ldots$ but conditionally on the event $\{Y_0 = 0\}$. This sequence can be written as

$$\underbrace{Z_0, \ldots, Z_0}_{p \text{ times}}, \underbrace{Z_p, \ldots, Z_p}_{p \text{ times}}, \underbrace{Z_{2p}, \ldots, Z_{2p}}_{p \text{ times}}, \ldots$$

that is to say, the same random variables is repeated p times in the sequence. Therefore, we have, for $n = k \cdot p$ (with k an integer),

$$S_n = \sum_{j=1}^{n} X_j \overset{d}{=} p \sum_{j=1}^{k} Z_j,$$

from which it follows that

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0, p),$$

where σ^2 denotes the variance of $X_0, X_1, \ldots$. From there, it is not hard to show that, conditionally on the event $\{Y_0 = 0\}$,

$$\begin{aligned} \limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log\log n}} &= \sqrt{p} \quad a.s., \\ \liminf_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log\log n}} &= -\sqrt{p} \quad a.s.. \end{aligned} \tag{4.B.1}$$

Since the event $\{Y_0 = 0\}$ has positive probability (for any p), Equation (4.B.1) is also valid unconditionally, and hence the LIL is not verified.

CHAPTER 5

A FAILURE OF CENTRAL LIMIT THEOREMS FOR PIBD RANDOM VARIABLES

5.1 Introduction

In Chapter 3, we saw through many examples (and accompanying visualisations) that pairwise independent variables can still be strongly dependent. Then, in Chapter 4, we reviewed many results relying on mutual independence which do not hold under sole pairwise independence. In this chapter, we cover in greater detail one such result, i.e., the classical CLT (which is one of the most fundamental results in statistics). This theorem states that the standardised sample mean of a sequence of n *mutually independent* and identically distributed random variables with finite second moment converges in distribution to a standard Gaussian as n goes to infinity. For practitioners of statistics, knowing the distribution of a sample mean is crucially important to, for instance, build confidence intervals and conduct statistical tests. Consequently, the classical CLT is also very important in actuarial science. As Dhaene et al. (2002b) explain,

> [i]nsurance is based on the fact that by increasing the number of insured risks, which are assumed to be mutually independent and identically distributed, the average risk gets more and more predictable because of the Law of Large Numbers. This is because a loss on one policy might be compensated by more favorable results on others. The other well-known fundamental law of statistics, the Central Limit Theorem, states that under the assumption of mutual independence, the aggregate claims of the portfolio will be approximately normally distributed, provided the number of insured risks is large enough. Assuming independence is very convenient since the mathematics for dependent risks are less tractable, and also because, in general, the statistics gathered by the insurer only give information about the marginal distributions of the risks, not about their joint distribution, i.e., the way these risks are interrelated.

As we have illustrated before, mutual independence is a strong assumption and it is relevant to understand what happens when it is not met. In this Chapter, we highlight just how crucial this assumption is to the classical CLT. We do so by constructing explicitly a sequence of *pairwise* independent and identically distributed (p.i.i.d.) random variables (r.v.s) whose common margin F can be chosen arbitrarily (under very mild conditions) and for which the (standardised) sample mean is *not* asymptotically Gaussian. We give a closed-form expression for the limiting distribution of this sample mean. It is, to the best of our knowledge, the first example of this kind for which the asymptotic distribution of

the sample mean is explicitly given, and known to be skewed and heavier tailed than a Gaussian distribution, for *any* choice of margin. Our sequence thus illustrates nicely why *mutual* independence is such a crucial assumption for the (classical) CLT to hold. It also allows us to quantify how far away from the Gaussian distribution one can get under the less restrictive assumption of *pairwise* independence.

'A' CLT for a sequence $\{X_j, j \geq 1\}$ of r.v.s. is a result establishing the convergence in distribution, under some conditions, of a normalised sum $(\sum_{j=1}^{n} X_j - a_n)/b_n$ to the standard Gaussian distribution. The established terminology is to refer (somewhat abusively) to 'the' CLT when the X_j's are *independent*, even though various sets of more or less restrictive conditions on $\{X_j, j \geq 1\}$ exist under which 'a' CLT holds. Moreover, most textbooks in mathematical statistics or introductory statistics only focus on the case of *independent* variables, even if we know (at least since Hoeffding (1948)) that a CLT can possibly hold for dependent variables. As Stoyanov (2013) puts it in his book: counterexamples can serve to "demonstrate the range of validity of the CLT and examine the importance of the conditions under which the CLT holds". He and others (e.g., Bagui et al. (2013)) give several counterexamples to CLTs for independent X_j's. But, surprisingly, counterexamples for *dependent* sequences are scarce in the literature. This lack of counterexamples might explain the commonly held belief—and often unverified assumption—among users of statistics that a large sample size is sufficient to ensure approximate normality of the sample mean, even though it is not. In many fields where applied statistics are used, articles typically discuss what the right sample size should be in order to confidently use 'the' CLT, but those articles do not also address the fundamental issue of *mutual* independence as a crucial assumption; examples can be found in biology (Fay and Gerow, 2013), medicine (Altman and Bland, 1995; Ghasemi and Zahediasl, 2012; Cundill and Alexander, 2015), psychology (Anderson, 2010), engineering (Huberts et al., 2018), or economics (Kresojević and Gajić, 2019). In fact, after an extensive search, we have not found any article published in those fields that contains an *explicit* discussion of this crucial independence assumption.

Now, recall that the classical (or basic) CLT is stated for a sequence $\{X_j, j \geq 1\}$ of i.i.d. random variables with mean μ and standard deviation $0 < \sigma < \infty$ as follows:

$$S_n := \frac{\sum_{j=1}^{n} X_j - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} Z, \quad \text{as } n \to \infty, \tag{5.1.1}$$

where the random variable Z has a standard Gaussian distribution, noted thereafter

$N(0,1)$, and '$\xrightarrow{d}$' denotes convergence in distribution. The first 'i' in the acronym i.i.d. stands for 'independent', which itself stands for '*mutually* independent', while the last 'i.d.' stands for *identically distributed*.

As we have seen through several examples in this thesis, *pairwise* independence among random variables is a necessary *but not sufficient* condition for them to be *mutually* independent. However, from examples of 'finite size' alone it can be hard to understand how bad of a substitute to mutual independence pairwise independence is. One way to study this question is to consider those fundamental (asymptotic) theorems of mathematical statistics that rely on the former assumption; do they 'fail' under the weaker assumption of *pairwise* independence? A definite answer to that question is beyond the scope of this work, as it depends on which theorem is considered. As we noted before, the Law of Large Numbers, even if almost always stated for mutually independent r.v.s, does hold under pairwise independence. The same goes for the second Borel-Cantelli lemma, usually stated for mutually independent events but valid for pairwise independent events as well; see Erdős and Rényi (1959). The CLT (for i.d. r.v.s), however, does 'fail' under pairwise independence. Since it is arguably the most crucial result in all of statistics, and since pairwise independence and uncorrelatedness are concepts vastly used by practitioners, we will focus on this case from now on.

Révész and Wschebor (1965) were the first to provide a pairwise independent sequence for which S_n does not converge in distribution to a $N(0,1)$. For their sequence, which is binary (i.e., two-state), S_n converges to a standardised χ^2_1 distribution. Romano and Siegel (1986, Example 5.45) provide a two-state, and Bradley (1989) a three-state, pairwise independent sequence for which S_n converges in probability to 0. Janson (1988) provides a broader class of pairwise independent counterexamples, most defined with X_j's having a continuous margin and for which S_n converges in probability to 0. The author also constructs a pairwise independent sequence of $N(0,1)$ r.v.s for which S_n converges to the random variable $S = R{\cdot}Z$, with R a r.v. whose distribution can be arbitrarily chosen among those with support $[0,1]$, and Z a $N(0,1)$ r.v. independent of R. The r.v. S can be seen as 'better behaved' than a $N(0,1)$, in the sense that it is symmetric with a variance smaller than 1 (regardless of the choice of R). Cuesta and Matrán (1991, Section 2.3) construct a sequence $\{X_j, j \geq 1\}$ of r.v.s taking values uniformly on the integers $\{0,1,\ldots,p-1\}$, with p a prime number, for which S_n is 'worse behaved' than a $N(0,1)$. Indeed, their S_n converges in distribution to a mixture (with weights $(p-1)/p$ and $1/p$ respectively) of the constant 0 and of a centered Gaussian r.v. with variance p. This distribution is symmetric

but it has heavier tails than that of a $N(0,1)$.

Other authors go beyond *pairwise* independence and study CLTs under 'K-tuplewise independence', for $K \geq 3$. A random sequence $\{X_j, j \geq 1\}$ is said to be K-tuplewise independent if for every choice of K distinct integers $j_1, \ldots, j_K$, the random variables $X_{j_1}, \ldots, X_{j_K}$ are mutually independent. Kantorovitz (2007) provides an example of a triplewise independent two-state sequence for which S_n converges to a 'misbehaved' distribution —that of $Z_1 \cdot Z_2$, where Z_1 and Z_2 are independent $N(0,1)$. Pruss (1998) presents a sequence of K-tuplewise independent random variables $\{X_j, j \geq 1\}$ taking values in $\{-1, 1\}$ for which the asymptotic distribution of S_n is never Gaussian, for whichever choice of K. Bradley and Pruss (2009) extend this construction to a *strictly stationary* sequence of K-tuplewise independent r.v.s whose margin is uniform on the interval $[-\sqrt{3}, \sqrt{3}]$. Weakley (2013) further extends this construction by allowing the X_j's to have any symmetrical distribution (with finite variance). Takeuchi (2019) showed that even if K grows linearly with the sample size n, a CLT need not be valid.

In the body of research discussed above, a non-degenerate and explicit limiting distribution for S_n is obtained only for very specific choices of margin for the X_j's. In this paper, we allow this margin to be almost any non-degenerate distribution, yet we still obtain explicitly the limiting distribution of S_n. This distribution depends on the choice of the margin, but it is always skewed and heavier tailed than a Gaussian. By the generality of our construction (the class of marginals allowed is very broad), and the fact we explicitly find the asymptotic distribution of the standardised mean, this work raises new awareness on the dangers of using a CLT on a sample that is only *pairwise* independent.

The rest of this chapter is organised as follows. In Section 5.2, we construct our pairwise independent sequence $\{X_j, j \geq 1\}$. In Section 5.3, we derive explicitly the asymptotic distribution of the standardised mean of that sequence. In Section 5.4, we study key properties of such a distribution. In Section 5.5, we analyse a parameter which arises naturally in the asymptotic distribution of the sample mean of our sequence, and explain how this parameter reflects the *tail-heaviness* of the marginal distribution of the sequence. In Section 5.6, we conclude.

5.2 Construction of the Gaussian-Chi-squared pairwise independent sequence

In this section, we build a sequence $\{X_j, j \geq 1\}$ of p.i.i.d. (but not mutually independent) r.v.s for which a CLT does *not* hold. We show in Section 5.4 that the asymptotic distribution of the (standardised) sample mean of this sequence can be conveniently written as that of the sum of a Gaussian r.v. and of an independent scaled Chi-squared r.v. Importantly, the r.v.s forming this sequence have a common (but arbitrary) marginal distribution F satisfying the following condition:

Condition 5.1. *For any r.v. $W \sim F$, the variance $\mathbb{V}\mathrm{ar}(W)$ is finite and there exists a Borel set A for which $\mathbb{P}(W \in A) = \ell^{-1}$, for some integer $\ell \geq 2$, and $\mathbb{E}[W|W \in A] \neq \mathbb{E}[W|W \in A^c]$.*

As long as the variance is finite, the restriction on F includes *all* distributions with an absolutely continuous part on some interval. It also includes *almost all* discrete distributions with at least one weight of the form ℓ^{-1}; see Remark 5.2. Also, note that, for a given F, many choices for A (with possibly different values of ℓ) could be available, depending on F. Note that we will explain in Remark 5.1 why our construction requires the existence of such a set A (with $\mathbb{P}(W \in A) = \ell^{-1}$). For specific examples of distributions satisfying Condition 5.1, see Example 5.1 and all examples of Section 5.B.

We begin our construction of th, see Example e sequence $\{X_j, j \geq 1\}$ by letting F be a distribution satisfying Condition 5.1. For a r.v. $W \sim F$, let A be any Borel set such that

$$\mathbb{P}(W \in A) = \ell^{-1}, \quad \text{for some integer } \ell \geq 2. \tag{5.2.1}$$

Then, for an integer $m \geq 2$, let $M_1, \ldots, M_m$ be a sequence of i.i.d. r.v.s with discrete distribution on the set $\{1, 2, \ldots, \ell\}$ and defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $i = 1, 2, \ldots, \ell$, let

$$p_i := \mathbb{P}(M_j = i) = \ell^{-1}, \quad \text{for } j = 1, 2, \ldots, m, \tag{5.2.2}$$

i.e., the M's are discrete uniforms on the set $\{1, \ldots, \ell\}$. For all pairs (M_i, M_j), $1 \leq i <$

$j \leq m$, define a r.v. $D_{i,j}$ as

$$D_{i,j} = \begin{cases} 1, & \text{if } M_i = M_j, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2.3}$$

The $D_{i,j}$ are p.i.i.d (but not mutually independent) with $\mathbb{P}(D_{i,j} = 1) = \ell^{-1}$; see Remark 5.1. For convenience, we refer to these $n = \binom{m}{2}$ random variables

$$D_{1,2}, D_{1,3}, \ldots, D_{1,m}, D_{2,3}, D_{2,4}, \ldots, D_{m-1,m}$$

simply as

$$D_1, \ldots, D_n, \tag{5.2.4}$$

where for $1 \leq i < j \leq m$, $D_{k(i,j)} \coloneqq D_{i,j}$ with $k(i,j) = [i(2m-1) - i^2]/2 + j - m$. Note that when $\ell = 2$ and $p_1 = p_2 = 1/2$, the M_j's are (shifted) Bernoulli(1/2) r.v.s, and the sequence (5.2.4) is equivalent to a pairwise independent sequence first mentioned in Geisser and Mantel (1962) and for which we already know that a CLT does not hold, see Révész and Wschebor (1965).

From the sequence $D_1, \ldots, D_n$, we now construct a new pairwise independent sequence $X_1, \ldots, X_n$ such that $X_k \sim F$ for all $k = 1, \ldots, n$. Define U and V to be the truncated versions of W, respectively off and on the set A:

$$U \stackrel{d}{=} W \mid \{W \in A^c\}, \qquad V \stackrel{d}{=} W \mid \{W \in A\}, \tag{5.2.5}$$

and denote

$$\mu_U \coloneqq \mathbb{E}[U], \qquad \mu_V \coloneqq \mathbb{E}[V]. \tag{5.2.6}$$

Then, consider n independent copies of U, and independently n independent copies of V:

$$U_1, \ldots, U_n, \overset{\text{i.i.d.}}{\sim} F_U, \qquad V_1, \ldots, V_n \overset{\text{i.i.d.}}{\sim} F_V, \tag{5.2.7}$$

both defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Finally, for $\omega \in \Omega$ and for $k = 1, \ldots, n$, construct

$$X_k(\omega) = \begin{cases} U_k(\omega), & \text{if } D_k(\omega) = 0, \\ V_k(\omega), & \text{if } D_k(\omega) = 1. \end{cases} \tag{5.2.8}$$

By conditioning on D_k, one can check that

$$F_{X_k}(x) = (1 - \ell^{-1})F_{U_k}(x) + \ell^{-1}F_{V_k}(x) = F(x). \tag{5.2.9}$$

In the next section, we will derive the asymptotic distribution of the sample mean of those X's, and see that it is *not* Gaussian.

The fact we do not have a CLT for the sequence in (5.2.4) can be explained heuristically as follows. Within the sequence $D_1, \ldots, D_n$, there can be a 'very high' proportion of 1's. This occurs if the sequence $M_1, \ldots, M_m$ contains a large proportion of equal variables. However, by definition of $D_{i,j}$, in order to have a very large proportion of 0's among the D's, one would require a large proportion of pairs (M_i, M_j), $1 \leq i < j \leq m$, to be such that $M_i \neq M_j$. This is impossible, since *all* the possible pairs are used to form the sequence of D's. This very asymmetrical situation makes the asymptotic distribution of the standardised sample mean of the D's highly skewed to the right.

Remark 5.1. *In Condition 5.1, the restriction $\mathbb{P}(W \in A) = \ell^{-1}$ for some integer ℓ may seem arbitrary. Likewise, in* (5.2.2) *the choice $p_i = \ell^{-1}$ for $i = 1, \ldots, \ell$ may also seem arbitrary. We establish here that none of these choices are arbitrary. Indeed, assume first that the only restriction on $p_1, p_2, \ldots, p_\ell \in (0, 1)$ is that*

$$\begin{aligned} &(1) \ : \ p_1 + p_2 + \cdots + p_\ell = 1, \\ &(2) \ : \ p_1^2 + p_2^2 + \cdots + p_\ell^2 = w, \\ &(3) \ : \ p_1^3 + p_2^3 + \cdots + p_\ell^3 = w^2, \end{aligned} \tag{5.2.10}$$

for some $w \in (0, 1)$. Condition (1) *is necessary for the distribution* (5.2.2) *to be well-defined, and conditions* (2) *and* (3) *are rewritings of*

$$\mathbb{P}(D_{i,j} = 1) = w, \quad 1 \leq i < j \leq m, \tag{5.2.11}$$

and

$$\mathbb{P}(D_{i,j} = 1, D_{j,k} = 1) = \mathbb{P}(D_{i,j} = 1)\mathbb{P}(D_{j,k} = 1), \quad 1 \leq i < j < k \leq m, \tag{5.2.12}$$

which are sufficient to guarantee that the D's are identically distributed and pairwise independent. Now, the solution $p_i = \ell^{-1}$ to (5.2.10) *is unique. Indeed, by squaring condi-*

tion (2) in (5.2.10) *then applying the Cauchy-Schwarz inequality, one gets*

$$w^2 = \Big(\sum_{i=1}^{\ell} p_i^{3/2} p_i^{1/2}\Big)^2 \leq \sum_{i=1}^{\ell} p_i^3 \sum_{i=1}^{\ell} p_i = \sum_{i=1}^{\ell} p_i^3 \tag{5.2.13}$$

where the last equality comes from condition (1) in (5.2.10)*. Then, condition* (3) *requires that we have the equality in* (5.2.13)*, and this happens if and only if* $p_i^{3/2} = \lambda p_i^{1/2}$ *for all* $i \in \{1, \ldots, \ell\}$ *and for some* $\lambda \in \mathbb{R}$*. In turn, this implies* $p_i = \lambda = \ell^{-1}$ *because of* (1) *and since* $p_i > 0$*, which then implies* $w = \ell^{-1}$ *by* (2)*. This reasoning shows that we cannot extend our method to an arbitrary* $\mathbb{P}(W \in A) \in (0, 1)$ *in* (5.2.1)*.*

Remark 5.2. *There is no easy characterisation of all discrete distributions with finite variance such that* $\mathbb{P}(W \in A) = \ell^{-1}$ *for some Borel set* A *and some integer* $\ell \geq 2$*, but for which the last part of Condition 5.1 is not satisfied. However, the proportion of such distributions can be expected to be very small. As a simple but convincing example, consider the set of discrete distributions on three values* $-\infty < x < y < z < \infty$ *with weights* $p_x, p_y, p_z \in (0, 1)$*. The variance is finite and say one of the three p's has the form* ℓ^{-1} *for some integer* ℓ*. The only way that* $\mathbb{E}[W|A] = \mathbb{E}[W|A^c]$ *is satisfied is by having* A *contain* y *and only* y *so that we must have* $p_y = \ell^{-1}$*,* $p_x = p$ *and* $p_z = (1 - p - \ell^{-1})$ *for some parameter* $p \in (0, 1)$*, and* $xp_x + zp_z = y\,(1 - \ell^{-1})$*. In other words, once* ℓ *is fixed, there is only freedom in the choice of* x*,* z *and* p*. If we remove the restriction* $\mathbb{E}[W|A] = \mathbb{E}[W|A^c]$ *(i.e.,* $xp_x + zp_z = y\,(1 - \ell^{-1})$*), it gives us at least one more dimension of freedom in the selection of* x, y, z, p_x, p_y, p_z*. Hence, in this case, the proportion is actually '0'. An analogous argument can be made for other discrete distributions of this kind. The restriction* $\mathbb{E}[W|A] = \mathbb{E}[W|A^c]$ *will always remove a dimension of freedom in the choice of the range of values or the weights.*

Remark 5.3. *In our construction, because the sample size is* $n = m(m-1)/2$*, it can only take specific values (*$n = 1, 3, 6, 10, \ldots$*). We do not think this is a limitation, as it is easy to build a similar sequence of arbitrary size with the same properties (pairwise independence, arbitrariness of the margins, non-Gaussian asymptotic distribution for the mean). More details are given in Appendix 5.A.*

5.3 Main result

We now state our main result.

Theorem 5.1. *Let $X_1, \ldots, X_n$ be random variables defined as in* (5.2.8) *and denote their mean and variance by μ and σ^2, respectively. Then,*

(a) *$X_1, \ldots, X_n$ are pairwise independent;*

(b) As $m \to \infty$ (and hence as $n \to \infty$), the standardised sample mean $S_n := \big(\sum_{k=1}^n X_k - \mu n\big)/\sigma\sqrt{n}$ converges in distribution to a random variable

$$S := \sqrt{1 - r^2}Z + r\,\chi, \tag{5.3.1}$$

where $Z \sim N(0,1)$, χ is independently distributed as a standardised $\chi^2_{\ell-1}$ and $r := \sqrt{\ell^{-1}(1-\ell^{-1})}(\mu_V - \mu_U)/\sigma$ with μ_U, μ_V defined in (5.2.6).

Remark 5.4. *Interestingly, since a standardised chi-squared distribution converges to a standard Gaussian as its degree of freedom tends to infinity, we see that $S \xrightarrow{d} N(0,1)$ as $\ell \to \infty$.*

Remark 5.5. *When removing the restriction $\mathbb{E}[W|A] \neq \mathbb{E}[W|A^c]$ in Condition 5.1, the case $r = 0$ (i.e., $\mu_U = \mu_V$) is possible, so our construction also provides a new instance of a pairwise independent (but not mutually independent) sequence for which a CLT does hold.*

Proof of Theorem 5.1. Proving (a) is straightforward. Simple calculations show that $D_1, \ldots, D_n$ are pairwise independent; recall (5.2.12). Now, for any $k, k' \in \{1, 2, \ldots, n\}$ with $k \neq k'$, the r.v.s

$$D_k, U_k, V_k, D_{k'}, U_{k'}, V_{k'}$$

are mutually independent and one can write $X_k = g(D_k, U_k, V_k)$ and $X_{k'} = g(D_{k'}, U_{k'}, V_{k'})$, for g a Borel-measurable function. Since X_k and $X_{k'}$ are integrable, the result follows from Pollard (2002, Section 4.1, Corollary 2).

The proof of (b) is more involved. We prove (5.3.1) by obtaining the limit of the characteristic function of S_n, and then by invoking Lévy's continuity theorem. Namely, we show

that, for all $t \in \mathbb{R}$,

$$\begin{aligned}\varphi_{S_n}(t) &\underset{m\to\infty}{\longrightarrow} \varphi_{\sqrt{1-r^2}Z}(t)\cdot\varphi_{r\,\chi}(t)\\ &= \exp\left(-\frac{1}{2}(1-r^2)t^2 - \mathrm{i}tr\sqrt{(\ell-1)/2}\right)\left(1-\mathrm{i}tr\sqrt{2/(\ell-1)}\right)^{-(\ell-1)/2}. \quad (5.3.2)\end{aligned}$$

First, let us define by $N_i = N_i(m)$ the number of M_j's equal to i ($i = 1, 2, \ldots, \ell$) within the sample $\{M_j;\ j = 1, \ldots, m\}$. Then, $\boldsymbol{N} := (N_1, \ldots, N_\ell) \sim \text{Multinomial}(m, (\ell^{-1}, \ldots, \ell^{-1}))$. Importantly, if $\boldsymbol{N}$ is known, then the number $p(\boldsymbol{N})$ of 1's in the sequence $\{D_j, j = 1, \ldots, n\}$ can be deduced as

$$\begin{aligned}p(\boldsymbol{N}) &= \sum_{i=1}^{\ell-1}\binom{N_i}{2}\mathbb{1}_{\{N_i\geq 2\}} + \binom{m-\sum_{i=1}^{\ell-1}N_i}{2}\mathbb{1}_{\{m-\sum_{i=1}^{\ell-1}N_i\geq 2\}}\\ &= \sum_{i=1}^{\ell-1}\frac{N_i(N_i-1)}{2} + \frac{(m-\sum_{i=1}^{\ell-1}N_i)(m-\sum_{i=1}^{\ell-1}N_i-1)}{2}\\ &= \frac{1}{2}\sum_{i=1}^{\ell-1}N_i^2 + \frac{1}{2}\sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1}N_iN_{i'} - m\sum_{i=1}^{\ell-1}N_i + \frac{m(m-1)}{2}\\ &= \frac{1}{2}\sum_{i=1}^{\ell-1}(N_i-m\ell^{-1})^2 + \frac{1}{2}\sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1}(N_i-m\ell^{-1})(N_{i'}-m\ell^{-1}) - \frac{\ell(\ell-1)}{2}m^2\ell^{-2} + \frac{m(m-1)}{2}\\ &= \frac{m\ell^{-1}}{2}\sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1}\Big(\frac{1}{\ell^{-1}}\mathbb{1}_{\{i=i'\}} + \frac{1}{\ell^{-1}}\Big)\frac{\big(N_i-m\ell^{-1}\big)}{\sqrt{m}}\frac{\big(N_{i'}-m\ell^{-1}\big)}{\sqrt{m}} - \frac{m}{2}(1-\ell^{-1}) + n\ell^{-1},\end{aligned} \quad (5.3.3)$$

where $\mathbb{1}_B$ denotes the indicator function on the set B. The covariances of a

$$\text{Multinomial}(m, (p_1, p_2, \ldots, p_\ell))$$

distribution are well known to be $m\Sigma$ where $\Sigma_{i,i'} = p_i\mathbb{1}_{\{i=i'\}} - p_ip_{i'}$, for $1 \leq i, i' \leq \ell-1$, and it is also known that $(\Sigma^{-1})_{i,i'} = p_i^{-1}\mathbb{1}_{\{i=i'\}} + p_\ell^{-1},\ 1 \leq i, i' \leq \ell-1$; see (Tanabe and Sagae, 1992, eq. 21). Therefore, with $p_i = \ell^{-1}$ for all i, we see from (5.3.3) that

$$\begin{aligned}\frac{p(\boldsymbol{N})-n\ell^{-1}}{\sqrt{n\ell^{-1}(1-\ell^{-1})}} &= \sqrt{\frac{m}{m-1}}\left[\frac{\sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1}(\Sigma^{-1})_{i,i'}\frac{(N_i-m\ell^{-1})}{\sqrt{m}}\frac{(N_{i'}-m\ell^{-1})}{\sqrt{m}}}{\sqrt{2(\ell-1)}} - \sqrt{\frac{\ell-1}{2}}\right]\\ &\xrightarrow{d} \frac{\xi-(\ell-1)}{\sqrt{2(\ell-1)}}, \quad \text{where } \xi \sim \chi^2_{\ell-1}. \quad (5.3.4)\end{aligned}$$

Now, let

$$\widetilde{U}_k := \frac{\sigma_U}{\sigma}\cdot\frac{U_k-\mu_U}{\sigma_U} \quad \text{and} \quad \widetilde{V}_k := \frac{\sigma_V}{\sigma}\cdot\frac{V_k-\mu_V}{\sigma_V}, \quad (5.3.5)$$

then we can write

$$S_n = \frac{1}{\sqrt{n}}\left(r\,\frac{\big(p(\boldsymbol{N}) - n\ell^{-1}\big)}{\sqrt{\ell^{-1}(1-\ell^{-1})}} + \sum_{\substack{k=1\\D_k=0}}^{n} \widetilde{U}_k + \sum_{\substack{k=1\\D_k=1}}^{n} \widetilde{V}_k \right), \tag{5.3.6}$$

since, from (5.2.9), we know that

$$\mu = (1-\ell^{-1})\mu_U + \ell^{-1}\mu_V. \tag{5.3.7}$$

With the notation $t_n := t/\sqrt{n}$, the mutual independence between the U_k's, the V_k's and $\boldsymbol{M} := \{M_j\}_{j=1}^m$ yields, for all $t \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}\big[\exp(\mathrm{i}tS_n)|\boldsymbol{M}\big] &= \exp\left(\mathrm{i}tr\,\frac{(p(\boldsymbol{N}) - n\ell^{-1})}{\sqrt{n\ell^{-1}(1-\ell^{-1})}}\right) \prod_{\substack{k=1\\D_k=0}}^{n} \mathbb{E}[\exp(\mathrm{i}t_n\widetilde{U}_k)|\boldsymbol{M}] \prod_{\substack{k=1\\D_k=1}}^{n} \mathbb{E}[\exp(\mathrm{i}t_n\widetilde{V}_k)|\boldsymbol{M}] \\
&= \exp\left(\mathrm{i}tr\,\frac{(p(\boldsymbol{N}) - n\ell^{-1})}{\sqrt{n\ell^{-1}(1-\ell^{-1})}}\right) [\varphi_{\widetilde{U}}(t_n)]^{n(1-\ell^{-1})} [\varphi_{\widetilde{V}}(t_n)]^{n\ell^{-1}} \left[\frac{\varphi_{\widetilde{V}}(t_n)}{\varphi_{\widetilde{U}}(t_n)}\right]^{p(\boldsymbol{N})-n\ell^{-1}} \\
&= \exp\left(\mathrm{i}tr\,\frac{(p(\boldsymbol{N}) - n\ell^{-1})}{\sqrt{n\ell^{-1}(1-\ell^{-1})}}\right) \cdot [\varphi_{\widetilde{U}}(t_n)]^{n(1-\ell^{-1})} [\varphi_{\widetilde{V}}(t_n)]^{n\ell^{-1}} \\
&\quad \cdot \left[\frac{[\varphi_{\widetilde{V}}(t_n)]^n \cdot \exp\left(\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2\right)}{[\varphi_{\widetilde{U}}(t_n)]^n \cdot \exp\left(\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2\right)}\right]^{\frac{p(\boldsymbol{N})-n\ell^{-1}}{n}} \cdot \left[\frac{\exp\left(-\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2\right)}{\exp\left(-\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2\right)}\right]^{\frac{p(\boldsymbol{N})-n\ell^{-1}}{n}} .
\end{aligned} \tag{5.3.8}$$

(The reader should note that, for n large enough, the manipulations of exponents in the second and third equality above are valid because the highest powers of the complex numbers involved have their principal argument converging to 0. This stems from the fact that $0 \le p(\boldsymbol{N}) \le n$, and the quantities $[\varphi_{\widetilde{U}}(t_n)]^n$ and $[\varphi_{\widetilde{V}}(t_n)]^n$ both converge to real exponentials as $n \to \infty$, by the classical CLT.) We now evaluate the four terms on the right-hand side of (5.3.8). For the first term in (5.3.8), the continuous mapping theorem and (5.3.4) yield

$$\exp\left(\mathrm{i}tr\,\frac{(p(\boldsymbol{N}) - n\ell^{-1})}{\sqrt{n\ell^{-1}(1-\ell^{-1})}}\right) \xrightarrow{d} \exp\left(\mathrm{i}tr\frac{\xi-(\ell-1)}{\sqrt{2(\ell-1)}}\right), \quad \text{as } m \to \infty. \tag{5.3.9}$$

For the second term in (5.3.8), the classical CLT yields

$$[\varphi_{\widetilde{U}}(t_n)]^{n(1-\ell^{-1})}[\varphi_{\widetilde{V}}(t_n)]^{n\ell^{-1}} \underset{m\to\infty}{\longrightarrow} \exp\left(-\frac{1}{2}\cdot(1-\ell^{-1})\frac{\sigma_U^2}{\sigma^2}t^2\right)\exp\left(-\frac{1}{2}\cdot\ell^{-1}\frac{\sigma_V^2}{\sigma^2}t^2\right)$$
$$= \exp\left(-\frac{1}{2}(1-r^2)t^2\right), \tag{5.3.10}$$

where in the last equality we used that, from (5.2.9), we have

$$\sigma^2 = \mathbb{E}[X^2] - \mu^2 = (1-\ell^{-1})\sigma_U^2 + \ell^{-1}\sigma_V^2 + \ell^{-1}(1-\ell^{-1})(\mu_U-\mu_V)^2. \tag{5.3.11}$$

For the third term in (5.3.8), the quantity inside the bracket converges to 1 by the classical CLT. Hence, the elementary bound

$$|\exp(z)-1| \le |z| + \sum_{j=2}^{\infty}\frac{|z|^j}{2} \le |z| + \frac{|z|^2}{2(1-|z|)} \le \frac{1+\ell^{-1}}{2\ell^{-1}}|z|, \quad \text{for all } |z|\le 1-\ell^{-1}, \tag{5.3.12}$$

and the fact that $\left|\frac{p(\boldsymbol{N})-n\ell^{-1}}{n}\right| \le 1-\ell^{-1}$ yield, as $m\to\infty$,

$$\left|\left[\frac{[\varphi_{\widetilde{V}}(t_n)]^n\cdot\exp\left(\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2\right)}{[\varphi_{\widetilde{U}}(t_n)]^n\cdot\exp\left(\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2\right)}\right]^{\frac{p(\boldsymbol{N})-n\ell^{-1}}{n}} - 1\right| \le \frac{1-\ell^{-2}}{2\ell^{-1}}\left|\mathrm{Log}\left[\frac{[\varphi_{\widetilde{V}}(t_n)]^n\cdot\exp\left(\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2\right)}{[\varphi_{\widetilde{U}}(t_n)]^n\cdot\exp\left(\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2\right)}\right]\right|$$
$$\longrightarrow 0. \tag{5.3.13}$$

For the fourth term in (5.3.8), the continuous mapping theorem and $\frac{p(\boldsymbol{N})-n\ell^{-1}}{n} \xrightarrow{\mathbb{P}} 0$ (recall (5.3.4)) yield

$$\left[\frac{\exp\left(-\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2\right)}{\exp\left(-\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2\right)}\right]^{\frac{p(\boldsymbol{N})-n\ell^{-1}}{n}} \xrightarrow{\mathbb{P}} 1, \quad \text{as } m\to\infty. \tag{5.3.14}$$

By combining (5.3.9), (5.3.10), (5.3.13) and (5.3.14), Slutsky's theorem implies, for all $t\in\mathbb{R}$,

$$\mathbb{E}\big[\exp(\mathrm{i}tS_n)|\boldsymbol{M}\big] \xrightarrow{d} \exp\left(\mathrm{i}tr\frac{\xi-(\ell-1)}{\sqrt{2(\ell-1)}}\right)\exp\left(-\frac{1}{2}(1-r^2)t^2\right), \quad \text{as } m\to\infty. \tag{5.3.15}$$

Since the sequence $\{|\mathbb{E}[\exp(\mathrm{i}tS_n)|\boldsymbol{M}]|\}_{m\in\mathbb{N}}$ is uniformly integrable (it is bounded by 1),

Theorem 25.12 in Billingsley (1995) shows that we also have the mean convergence

$$\mathbb{E}\big[\mathbb{E}\big[\exp(\mathrm{i}tS_n)|\boldsymbol{M}\big]\big] \longrightarrow \mathbb{E}\left[\exp\left(\mathrm{i}tr\frac{\xi-(\ell-1)}{\sqrt{2(\ell-1)}}\right)\right]\exp\left(-\frac{1}{2}(1-r^2)t^2\right), \quad \text{as } m\to\infty, \tag{5.3.16}$$

which proves (5.3.2). The conclusion follows. □

5.4 Properties of S

Recall that F denotes the marginal distribution of the r.v.s $X_1, \ldots, X_n$ in (5.2.8). Theorem 5.1 states that S_n, the standardised sample mean of these r.v.s, converges in distribution to a r.v. S whose characteristic function is given by (5.3.2). When the choice $\ell \geq 2$ is fixed, the distribution of S has only one parameter, r, defined as

$$r = \frac{\sqrt{\ell^{-1}(1-\ell^{-1})}(\mu_V - \mu_U)}{\sigma} = \frac{\mu_V - \mu}{\sigma\sqrt{\ell - 1}}, \tag{5.4.1}$$

where the second equality stems from (5.3.7). Hence, r depends on the margin F (through the quantities A, μ_U, μ_V and σ). The behavior of S with respect to F (via r) is now studied. From (5.3.11), we see that

$$r^2 = 1 - \frac{(1-\ell^{-1})\sigma_U^2 + \ell^{-1}\sigma_V^2}{\sigma^2}, \quad \text{and thus} \quad 0 \leq r^2 \leq 1. \tag{5.4.2}$$

Example 5.1 (and several other examples in the Appendix 5.B) show how the critical points $r^2 = 0, 1$ can be achieved or approached when F is discrete or absolutely continuous. See also Appendix A.4 for `R` computing codes to generate observations from these examples. (Note that these examples could serve as scenarios of dependence to compare, via Monte-Carlo simulations, various tests of independence; see, e.g., Hušková and Meintanis (2008).)

Example 5.1 (r arbitrarily close to 0 when F is absolutely continuous)**.** *Let $\ell = 2$, $A = [1, \infty)$ and let $W \sim f$ where f is the density of a Log-normal$(0, \beta)$ distribution. Note that* $\text{median}(W) = 1$, $\mathbb{E}[W] = \exp(\beta/2)$, *and* $\mathbb{V}\text{ar}[W] = [\exp(\beta) - 1]\exp(\beta)$. *Furthermore,*

$$\mu_V = \int_1^\infty 2xf(x)dx = \sqrt{\frac{2}{\pi\beta}} \int_1^\infty \exp\left(-\frac{(\log x)^2}{2\beta}\right) \mathrm{d}x = \exp(\beta/2)\left[1 + \text{Erf}\big(\sqrt{\beta/2}\big)\right], \tag{5.4.3}$$

where the integral on the second line was solved with `Mathematica`*. Hence, from* (5.4.1)*,*

$$r = \frac{\mu_V - \mathbb{E}[W]}{\sqrt{\mathbb{V}\text{ar}[W]}} = \frac{\exp(\beta/2)\text{Erf}\big(\sqrt{\beta/2}\big)}{\sqrt{[\exp(\beta) - 1]\exp(\beta)}} = \frac{\text{Erf}\big(\sqrt{\beta/2}\big)}{\sqrt{\exp(\beta) - 1}}, \tag{5.4.4}$$

and it is straightforward to see that $r \to 0$ as $\beta \to \infty$.

Next, recall that the characteristic function on the right-hand side of (5.3.2) is that of

$$S = \sqrt{1 - r^2}Z + r\,\chi, \tag{5.4.5}$$

where the r.v.s $Z \sim N(0,1)$ and $\chi \sim [\chi^2_{\ell-1} - (\ell-1)]/\sqrt{2(\ell-1)}$ are independent. This makes it clear that, when $\ell \geq 2$ is fixed, r completely determines the shape of S; the closer r gets to 0, the closer the distribution of S is to a standard Gaussian, while the closer r gets to ± 1, the closer the distribution of S is to a standardised $\pm\chi^2_{\ell-1}$. This shift from a Gaussian distribution towards a $\chi^2_{\ell-1}$ distribution is represented graphically in Figure 5.1 (where $\ell = 2$ and r varies). On the other hand, regardless of r, if ℓ increases then S gets closer to a $N(0,1)$, as illustrated in Figure 5.2 (where $r = 0.9$ and ℓ varies). These figures illustrate clearly that pairwise independence might be a very poor substitute to mutual independence as an assumption in CLTs.

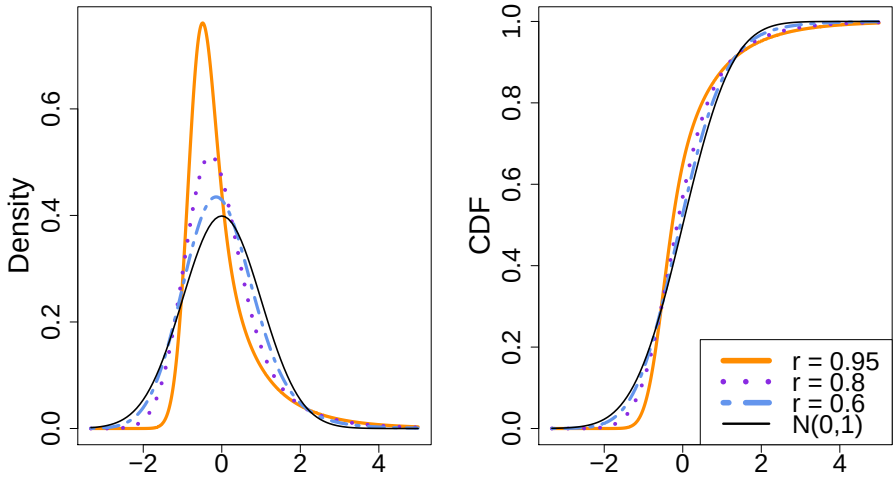

Figure 5.1: Density (left) and CDF (right) of S for fixed $\ell = 2$ and varying r ($r = 0.6, 0.8, 0.95$), compared to those of a $N(0,1)$. This illustrates that CLTs can 'fail' substantially under pairwise independence.

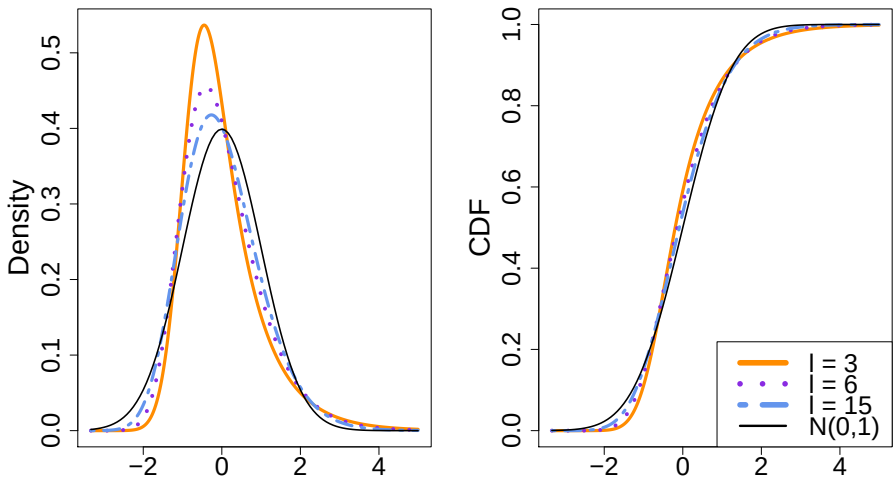

Figure 5.2: Density (left) and CDF (right) of S for fixed $r = 0.9$ and varying ℓ ($\ell = 3, 6, 15$), compared to those of a $N(0,1)$. This illustrates that S converges to a $N(0,1)$ as ℓ grows.

In terms of moments, simple calculations with `Mathematica` yield that

$$\mathbb{E}[S] = 0, \quad \mathbb{E}[S^2] = 1, \quad \mathbb{E}[S^3] = \sqrt{\frac{8}{\ell - 1}}\, r^3 \quad \text{and} \quad \mathbb{E}[S^4] = 3 + \frac{12}{\ell - 1} r^4, \tag{5.4.6}$$

so that upper bounds on the skewness and kurtosis of S are $\sqrt{8/(\ell-1)}$ and $3+12/(\ell-1)$, respectively. The limiting r.v. S can therefore be much more skewed and heavy-tailed than the standard Gaussian distribution, which is also confirmed by Figure 5.1.

Lastly, let us comment on the r parameter, and explain why a r close to 1 yields a more 'drastic' departure from normality. First, recall that a CLT does not apply to the sequence of pairwise independent r.v.s $D_1, \ldots, D_n$ given in (5.2.4) because the proportion of 1's in that sequence can be very large, whereas the proportion of 0's can never be large. Consequently, the distribution of the asymptotic sample mean of this sequence is asymmetrical (skewed to the right). When we 'assign' an arbitrary margin to the D's in order to create our sequence $\{X_j, j \geq 1\}$, we can attenuate (to a certain degree) this asymmetry. Consider for example the case $\ell = 2$ and $A = [\widetilde{w}, \infty)$, where $\widetilde{w}$ denotes the median of an absolutely continuous distribution F. In this case, the X's, as opposed to the D's, take a continuous range of values, and hence X's 'above the median' can be quite close to their mean (whereas the D's are all either 'much bigger' or 'much smaller' than their mean). The parameter $r = (\mu_V - \mu)/\sigma$ measures to what extent this 'attenuation of asymmetry' happens. Indeed, if r is close to 0, the X's observations above the median are not too far away from the mean (on average). This implies that, even if the proportion of observations above the median is huge, it will not overly boost the overall mean of the sample, and the distribution of this mean will not be overly asymmetrical.

To give a concrete example (again with $\ell = 2, A = [\widetilde{w}, \infty)$), let $X \sim$ Log-normal(α, β). In that case, simple calculations (see Example 5.1 for details) yield

$$r = \frac{\mathrm{Erf}\big(\sqrt{\beta/2}\big)}{\sqrt{\exp(\beta) - 1}},$$

a decreasing function of β. On the other hand, it is well known that the kurtosis of X is an increasing function of β. So, increasing β makes X heavier tailed, while giving a lower value of r. Hence, at least for the Log-normal, a *heavier tail* implies a *less drastic* failure of the CLT for the sequence $X_1, \ldots, X_n$.

Interestingly, we find that the same pattern (low r implies tail-heaviness) is true for a number of common distributions. In the next section, we explain this trend in more detail,

and compare r to other possible measures of tail-heaviness. We note this next Section 5.5 is a small digression away from the main topic of this thesis (pairwise independence). Nonetheless, we find relevant to investigate the link between r and tail-heaviness, as tail-heaviness is an important topic in actuarial science.

5.5 The r coefficient as a measure of tail-heaviness

5.5.1 Overview

The concept of *tail-heaviness* is ubiquitous in actuarial science. Many insurance loss data are found to be tail-heavy (see, e.g. Resnick, 2007; Ibragimov and Prokhorov, 2017), and "accurately fitting the tail" is of great importance in risk modelling (Ahn et al., 2012). For definitions and relevant literature review on tail-heaviness, see Appendix 5.D.

In this section, we study a special case of the parameter r (recall 5.4.1), obtained if in Condition 5.1 we assume the distribution F from is continuous, with choices

$$\ell = 2, \quad A = (\text{median}(F), \infty).$$

In that case, we have:

$$r = \frac{\mu_V - \mu}{\sigma}$$

(and this is the definition of 'r' we assume for the whole of Section 5.5). We argue this r is a possible measure of tail-heaviness, and we investigate its behaviour. Because it is bounded (i.e., $0 < r < 1$), has an easy interpretation and exists for all finite-variance continuous distributions, we think it is an interesting alternative to the traditional kurtosis coefficient κ (which, albeit commonly used as a measure of tail-heaviness, has many shortcomings). We note that this section is a small 'detour', away from the main topic of this thesis, but we thought interesting to understand better what this r represents, especially given its apparent link to tail-heaviness.

In Section 5.5.2, we give a few alternative definitions of r and explain why it is a possible measure of tail-heaviness. In Section 5.5.3, we survey where r has appeared in the literature before. In Section 5.5.4 we list the value of r for many common distributions (as function of the parameters of those distributions), and in Section 5.5.5 we compare r to other measures of tail-heaviness, also commenting on the results.

Throughout, we consider only continuous random variables. For X a continuous random variable, we let μ_X, m_X and σ_X denote respectively the mean, median and standard deviation of X (provided they exist).

5.5.2 Interpreting r as a measure of tail-heaviness

Let X be a continuous random variable with (finite) variance σ_X^2, mean μ_X and median m_X. We first note that r can be expressed in many ways. As seen before, we have

$$r = \frac{\mathbb{E}[X|X > m_X] - \mu_X}{\sigma_X}. \tag{5.5.1}$$

But also, because $\mu_X = (\mathbb{E}[X|X > m_X] + \mathbb{E}[X|X \leq m_X])/2$, we have

$$r = \frac{\mathbb{E}[X|X > m_X] - \mathbb{E}[X|X \leq m_X]}{2\sigma_X}.$$

Furthermore, noting that

$$\begin{aligned}
\mathbb{E}\left[|X - m_X|\right] &= \frac{1}{2}\mathbb{E}\left[X - m_X|X > m_X\right] + \frac{1}{2}\mathbb{E}\left[-(X - m_X)|X \leq m_X\right] \\
&= \frac{\mathbb{E}\left[X|X > m_X\right] - \mathbb{E}\left[X|X \leq m_X\right]}{2} \\
&= r\sigma_X,
\end{aligned}$$

we can write r as

$$r = \frac{\mathbb{E}\left[|X - m_X|\right]}{\sigma_X}, \tag{5.5.2}$$

where the numerator $\mathbb{E}\left[|X - m_X|\right]$ is called the *average absolute deviation from the median* (MAAD). Expressed as in (5.5.2), r is then seen to be the ratio of two different measures of spread: one which is 'more robust' (MAAD) and one which is 'less robust' (the standard deviation). Note that, in the actuarial literature, it has been suggested that MAAD "is better suited to determine the safety loading for insurance premiums than standard deviation" (Denneberg, 1990).

Because MAAD is more robust, we expect distributions with heavy tails (loosely speaking, distributions having some significant probability of extreme values) to have a much bigger standard deviation than MAAD, and hence to have a low r. For reference, note that the Uniform, Normal and Exponential distributions have values for r of (approximately): 0.866, 0.798 and 0.693, respectively. For distributions like the Log-normal, Gamma, Weibull, Pareto and Fréchet, r can be made arbitrarily close to 0 by varying the shape parameter; see Table 5.1 in Section 5.5.4 for values of r corresponding to common distributions.

We note that r has some convenient properties:

- It is a ratio whose numerator is always smaller than its denominator, hence $0 < r < 1$. This eases its interpretation (e.g., $r = 0.5$ means that the standard deviation of X is twice as large as its MAAD).
- It is defined for random variables with finite variance (hence, it is defined more generally than the kurtosis coefficient).
- It is invariant to shifting and scaling, see Proposition 5.1.

Proposition 5.1. *Let X be a continuous random variable with finite variance, and let $Z = aX + b$, for $a > 0$ and $b \in \mathbb{R}$. Denote r_X and r_Z the r coefficients corresponding to X and Z, respectively. Then, $r_X = r_Z$.*

Proof. First, we note that

$$\mu_Z = a\mu_X + b, \quad m_Z = am_X + b, \quad \sigma_Z = a\sigma_X.$$

We also have that

$$\begin{aligned}\mathbb{E}[Z|Z > m_Z] &= \mathbb{E}[Z - b|Z > am_X + b] + b \\ &= a\mathbb{E}[X|X > m_X] + b.\end{aligned}$$

Hence, using (5.5.1),

$$r_Z = \frac{a\mathbb{E}[X|X > m_X] + b - \mu_Z}{\sigma_Z} = \frac{\mathbb{E}[X|X > m_X] - \mu_X}{\sigma_X} = r_X.$$

□

Using Proposition 5.1, we can better interpret what r represents and how it relates to tail-heaviness. As before, let X be a continuous random variable with finite variance, and $Z = (X - \mu_X)/\sigma_X$ its standardised version. Then,

$$r_X = r_Z = \mathbb{E}[Z|Z > m_Z].$$

That is, for any continuous random variable X (with standardised version Z), r is simply the *mean of Z given it exceeds its median.* If Z is very heavy tailed, it implies mass 'far away' from the center of the distribution. But to 'compensate' the large mass away

from the center (while keeping a zero-mean and unit-variance) there must also be a large concentration of mass around the center of the distribution (close to the median). Hence, the mean of values above the median, i.e., r, is small. As a specific (but representative) illustration of this phenomenon, consider the following example.

Example 5.2. *Let $X \sim$ Log-normal(μ, β), and let $Z = (X - \mu_X)/\sigma_X$. In this case, we have*

$$r = \frac{\text{Erf}\left(\sqrt{\beta/2}\right)}{\sqrt{\exp(\beta) - 1}}$$

(see Appendix 5.C for details), where Erf() is the error function. It is well known that the tail-heaviness of X increases if β increases, while r is a decreasing function of β (hence, a smaller r corresponds to a heavier-tailed X). To illustrate this, consider Figure 5.3, which displays the density of Z for different values of β (and hence for different values of r). For $r = 0.75$, the distribution is not too far from a standard Normal (although with a positive skew). As β gets bigger, mass is pushed to the right tail. But in order for the mean to stay at 0 and the variance to stay at 1, this also means that more mass must accumulate close to the median. Hence, the mean of observations above the median, i.e., r, gets smaller.

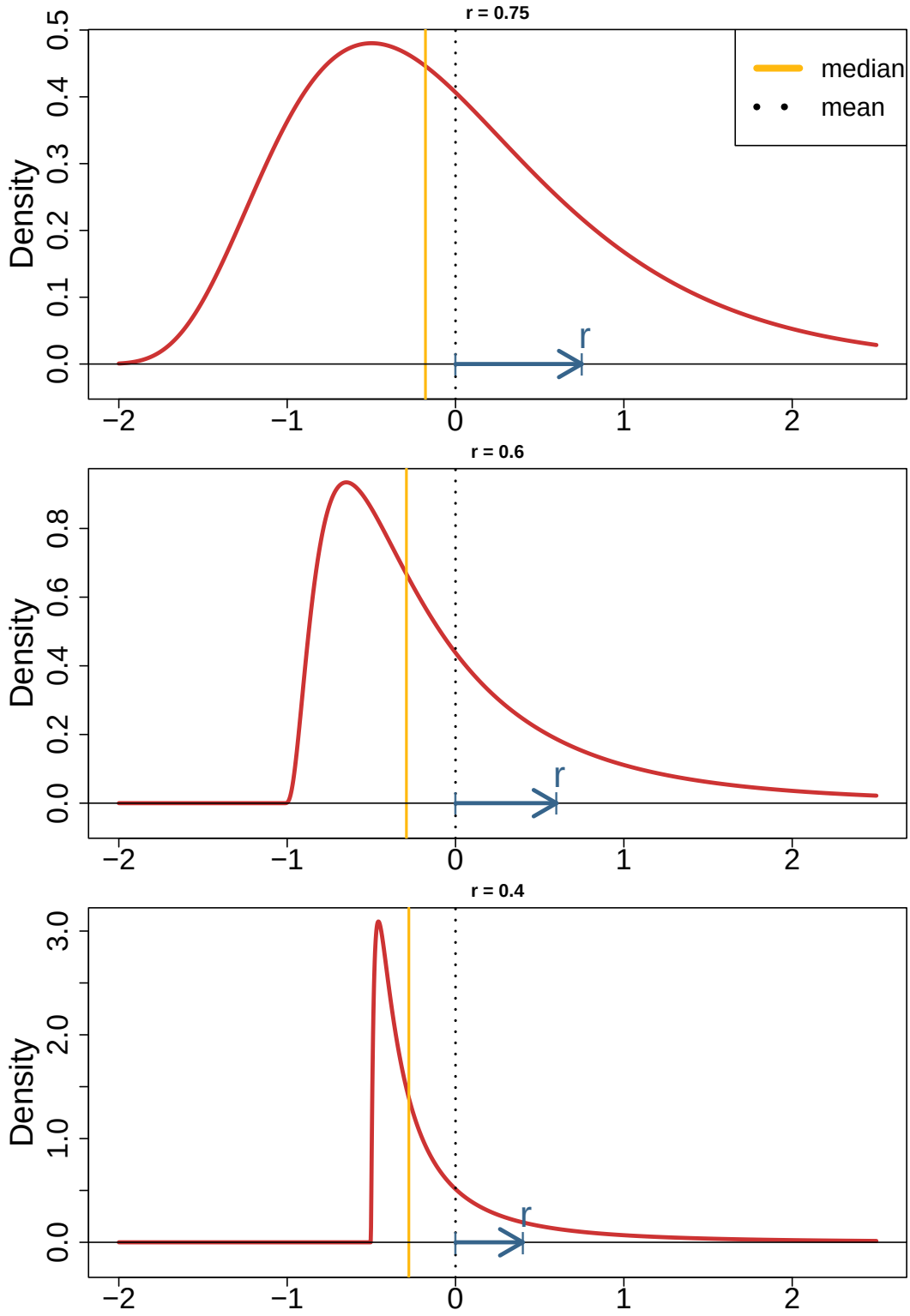

Figure 5.3: Density of the standardised Log-normal distribution for $r = 0.75$ (top), $r = 0.6$ (center), $r = 0.4$ (bottom)

5.5.3 r in the existing literature

For any continuous distribution F_θ, where θ is a set of parameters, r can be derived and expressed as a function of θ. As such — a theoretical quantity which depends on a distribution's parameters — we have not encountered r in the literature. This is perhaps surprising, given how simple a measure it is.

However, as an *empirical quantity*, an equivalent of r has been suggested before. In particular, we are not the first to notice that an empirical r (or a function of it) provides information about the tail-heaviness of a distribution. Indeed, a few statistical tests have been proposed, where an empirical version of r is used to assess the goodness-of-fit to certain distributions (in particular, the Normal distribution).

Hogg (1972) used this approach. To test if a sample $X_1, \ldots, X_n$, comes from the Normal distribution, he proposed the test statistic

$$\frac{s}{\sum_{i=1}^{n} |X_i - m|/n}, \tag{5.5.3}$$

where s is the sample standard deviation, and m the sample median. Equation (5.5.3) is easily seen to be an empirical version of $1/r$. Hogg (1972) proposed this statistic to reject normality specifically in the case where the alternative hypothesis is the heavier-tailed Laplace distribution. Smith (1975) later found that this statistic is a good choice, more generally, to detect tails that are heavier than that of the Normal distribution (and especially in the case of a small sample size). Gel et al. (2007) also used the statistic (5.5.3) as the basis for a new test of normality, and conducted a power analysis. These authors concluded that their test had improved power "when the data come from distributions that are at least as heavy-tailed as a double-exponential or the t-distribution with 3 degrees of freedom". In another paper, Gel and Gastwirth (2008) proposed an adaptation of the classical Jarque-Bera test of normality, where MAAD was used (instead of the standard deviation). Lastly, González-Estrada and Villaseñor (2016) used (5.5.3) as well, but to test the goodness-of-fit to the Laplace distribution.

From those papers, it is clear that an empirical r has been recognised already to be a valid way to test if a distribution has heavier tails than a Normal. However, none of those papers defined a 'population version' of their test statistic. Here, we suggest that r is a meaningful theoretical measure of tail-heaviness. As it appears r has not been computed for known distributions, in the next section we derive it for many distributions commonly

used in actuarial science. This yields simple expressions, which are seen to be function of the shape parameter of the distributions involved. This also allows us to compare r to other measures of tail-heaviness, see Section 5.5.5.

5.5.4 Values of r for common distributions

In this section, we present the value of the r coefficient (5.5.2) for many commonly used distributions (as a function of the parameters of those distribution). Table 5.1 reports the values, and we defer the derivations to Appendix 5.C. As we can see, r only depends on the shape parameter of those distributions (if there is a shape parameter).

Distribution	**Support**	**Density $f(x)$**	**r**
Uniform(a,b)	$x \in [a,b]$	$\frac{1}{b-a}$	$\frac{\sqrt{3}}{2}$
Normal(μ,σ^2)	$x \in \mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\sqrt{\frac{2}{\pi}}$
Exponential(λ)	$x \in [0,\infty)$	$\lambda e^{-\lambda x}$	$\log(2)$
Student(ν)	$x \in \mathbb{R}$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$2\sqrt{\frac{\nu-2}{\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu-1)}, \quad \nu > 2$
Log-normal(μ,β)	$x \in (0,\infty)$	$\frac{1}{x\sqrt{2\pi\beta}} \exp\left(-\frac{(\log x-\mu)^2}{2\beta}\right)$	$\frac{\mathrm{Erf}\left(\sqrt{\beta/2}\right)}{\sqrt{e^{\beta}-1}}$
Pareto (α,λ)	$x \in [0,\infty)$	$\frac{\alpha}{\lambda}\left[1+\frac{x}{\lambda}\right]^{-(\alpha+1)}$	$\sqrt{\alpha(\alpha-2)}\left(2^{\frac{1}{\alpha}}-1\right), \quad \alpha > 2$
Gamma(α,β)	$x \in [0,\infty)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\frac{2(m_X)^\alpha e^{-m_X}}{\Gamma(\alpha)\sqrt{\alpha}}$
Weibull(λ,k)	$x \in [0,\infty)$	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$	$\frac{2\Gamma(1+1/k,\log(2))-\Gamma(1+1/k)}{\sqrt{\Gamma(1+2/k)-(\Gamma(1+1/k))^2}}$
Fréchet(α)	$x \in (0,\infty)$	$x^{-1-\alpha} e^{-x^{-\alpha}}$	$\frac{\Gamma(1-1/\alpha)-2\Gamma(1-1/\alpha,\log(2))}{\sqrt{\Gamma(1-2/\alpha)-(\Gamma(1-1/\alpha))^2}}, \quad \alpha > 2$

Table 5.1: Values of the r coefficient for common distributions

Note that in Table 5.1, for the Gamma, m_X represents the median of a random variable $X \sim \text{Gamma}(\alpha, \beta = 1)$, and it does not have a closed-form expression.

5.5.5 Comparisons to other measures of tail-heaviness

In this section, we compare the coefficient r with two other measures of tail-heaviness. In particular, we assess whether r 'agrees' with those measures for distributions commonly used in actuarial science. For illustration purposes, in what follows we compare

$$r^* = 1 - r,$$

to the other measures. We use r^* instead of r to ease interpretation: a *bigger* r^* will mean a *heavier* tail. We compare r^* to the very common kurtosis coefficient κ but also to a more robust measure proposed by Brys et al. (2006), denoted RQW (with specific choice $q = 0.875$, see (5.D.3) in Appendix 5.D for details).

For all distributions considered (Log-normal, Weibull, Gamma, Pareto, Fréchet, Student), the three measures depend only on one parameter, call it generically θ. This facilitates comparisons: for two measures (e.g., r^* and κ) we can plot the pair $(r^*(\theta), \kappa(\theta))$ for a large range of values of θ. We can do this for all distributions, and then compare the curves obtained. If for two distributions the curves are similar, then it means the two measures assess similarly the tail-heaviness of those distributions.

A first (perhaps unsurprising) observation to make is that the kurtosis κ 'disagrees' strongly with the robust measure RQW, as seen on Figure 5.4. We see that, for a fixed value of RQW, the value of the kurtosis varies wildly across distributions. In particular, the Log-normal has a much higher kurtosis than the other two. Those large discrepancies are not surprising, since κ is moment-based, while RQW is quantile-based.

On Figure 5.5 we display κ against r^*. Here as well, κ disagrees with r^*, although to a lesser extent. Indeed, while the kurtosis of the log-Normal is again much higher (for fixed r^*) than it is for the other two distributions, it is not completely 'off the chart' as we saw on Figure 5.4. Note that in Figures 5.4 to 5.5, we did not show the Pareto, Fréchet and Student distributions since their kurtosis is not defined for a large subset of the parameters used in the comparisons.

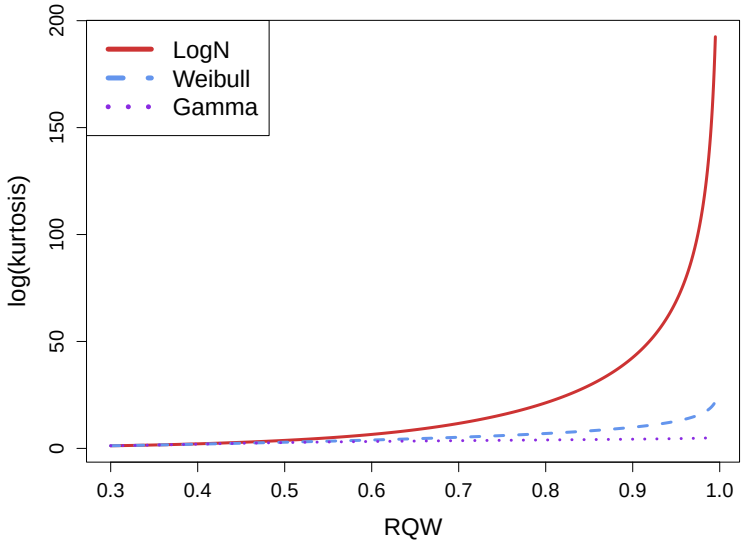

Figure 5.4: Log(kurtosis) against RQW($q = 0.875$) for common distributions

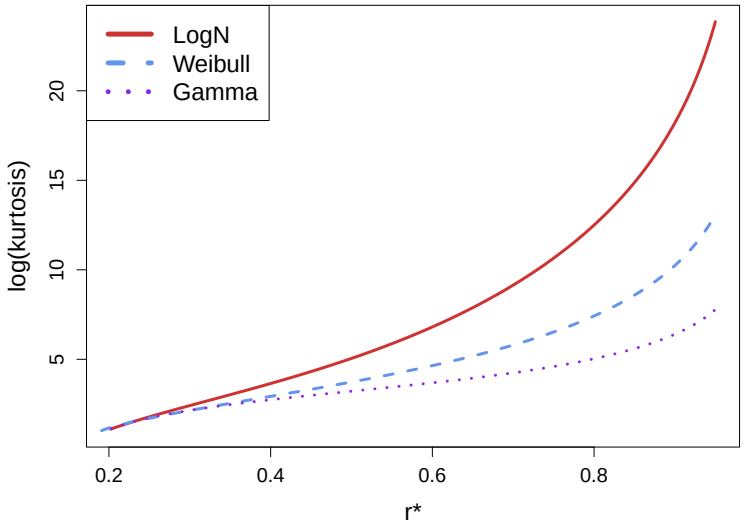

Figure 5.5: Log(kurtosis) against r^* for common distributions

We next now compare r^* to RQW (Figure 5.6), for six distributions. Figure 5.6 reveals that for low or moderate tail-heaviness ($r^* \leq 0.4$), RQW agrees with r^* for all distributions except the Student (but it should be noted that the Student is the only symmetric distribution considered). For the Pareto and the Fréchet, r^* and RQW agree to a large extent (although RQW never gets to its maximal level, which would be achieved for values of those distributions' parameters for which r^* is not defined). Interestingly, for the Gamma distribution, we see that RQW reaches its maximum when $r^* \approx 0.8$, meaning that there is a large range of possible parameter values where RQW is not able to capture the increase in tail-heaviness (having already reached its maximum).

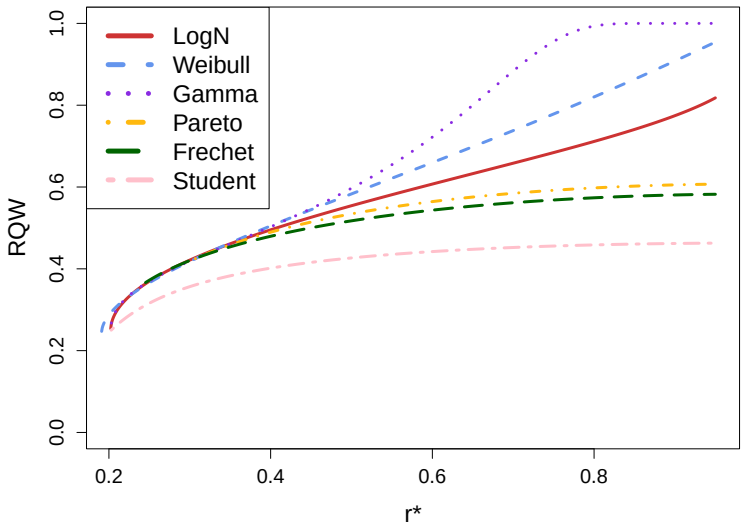

Figure 5.6: RQW($q = 0.875$) against r^* for common distributions

Overall, r^* always increases when the other two measures increase. Hence, it is clear that r^* does measure tail-heaviness, in some way, though it 'disagrees' to varying degrees with the other measures. We can perhaps see r^* as a 'middle ground' between on one side the kurtosis κ (which is moment-based) and on the other side RQW (which is quantile-based).

5.6 Conclusion

We constructed a pairwise independent sequence of identically distributed r.v.s $\{X_j, j \geq 1\}$ having *any* distribution that satisfies Condition 5.1, and for which no CLT holds. We obtained the asymptotic distribution of the standardised sample mean S_n of our sequence, and found it to be always 'worse behaved' than a Gaussian. Furthermore, the extent of this departure from normality depends on the initial common margin of the X_j's. Said otherwise, the same *dependence structure* yields fairly different behaviours in the asymptotic mean (depending on which margin is chosen). This is in contradiction with any CLT under which, regardless of the margin, S_n always converges to a Gaussian.

A corollary of our main result is that there exists a sequence of pairwise independent Gaussian r.v.s (and hence uncorrelated) for which the limiting distribution of S_n is substantially 'worse behaved' than a Gaussian, being asymmetric and heavier tailed. To our knowledge, no other such example exists in the literature. Given the widespread use of CLTs, even in standard parametric statistical techniques such as tests and confidence intervals for means and variances (see, e.g., Coeurjolly et al., 2009), this constitutes a serious warning to practitioners of statistics who may think that, to invoke 'the' CLT, all one needs is for the original random variables $\{X_j, j \geq 1\}$ to be approximately Gaussian or to have a large enough sample size. Mutual independence *is* a crucial assumption that should not be forgotten, nor misunderstood.

Lastly, we analysed in some detail the meaning of 'r', a parameter which arises in the asymptotic distribution of the sample mean of our sequence (see 5.3.1). In the special case of a continuous distribution F (and choices $\ell = 2$ and $A = (\text{median}(F), \infty)$ in Condition 5.1), we saw that r is a measure of the tail-heaviness of F.

In closing this chapter, we note that some authors have studied CLTs under K-tuplewise independence (for $K > 2$); see, e.g., Pruss (1998); Bradley and Pruss (2009); Bradley (2010); Weakley (2013); Takeuchi (2019). We find interesting to generalise our construction in that direction, and this will be the topic of the next chapter. For the case $K = 3$ ('triplewise independence'), we will provide two distinct sequences for which a CLT does not hold. Our methodology can also be used to build K-tuplewise independent sequences (for arbitrary K), though for $K \geq 4$, a CLT appears to always hold in our construction.

5.A Remark on the 'non-arbitrariness' of the sample size n

In our construction of the sequence $X_1, \ldots, X_n$ (see 5.2.8), the sample size n can only take a specific set of values (call this set $\mathbb{N}^*$), i.e.,

$$\mathbb{N}^* = \bigcup_{k=2}^{\infty} \left\{ \frac{k(k-1)}{2} \right\} = \{1, 3, 6, 10, 15, \ldots\}. \tag{5.A.1}$$

While one could see this as a 'limitation', we do not think it is, because one can easily build a sample of *arbitrary* size n which preserves the main properties of our example, i.e.,

- it is pairwise independent
- it has arbitrary margins
- its standardised mean has an asymptotic non-Normal distribution with characteristic function given by (5.3.2).

To build such a sample is simple. Let $n \in \mathbb{N}$ be arbitrary, and define $n_1(n)$ to be the biggest integer such that $n_1(n) \leq n$ and $n_1(n) \in \mathbb{N}^*$. Likewise, define $n_2(n) = n - n_1(n)$. To ease notation, let us denote $n_1(n)$ and $n_2(n)$ by simply n_1 and n_2. It is straightforward that

$$\lim_{n\to\infty} \frac{n_1}{n} = 1.$$

Then, for any n we can define a sample $X_1, X_2, \ldots, X_{n_1}$ exactly as in Section 5.2. Furthermore, we can define another sample $X_{n_1+1}, \ldots, X_n$, independent of the first one, and where all X's are i.i.d.[1] The resulting 'total' sample of size n, i.e.,

$$X_1, X_2, \ldots, X_{n_1}, X_{n_1+1}, \ldots, X_n$$

is then pairwise independent. Furthermore, consider the standardised mean of this sample, i.e.,

$$S_n = \frac{1}{\sigma\sqrt{n}} \left(\sum_{j=1}^{n} X_j - \mu n \right),$$

[1] In the case that $n = n_1$, this sample is empty.

and rewrite it as

$$\begin{aligned} S_n &= \frac{1}{\sigma\sqrt{n}}\left(\sum_{j=1}^{n_1} X_j + \sum_{j=n_1+1}^{n} X_j - \mu(n_1+n_2)\right) \\ &= \frac{\sum\limits_{j=1}^{n_1} X_j - \mu n_1}{\sigma\sqrt{n_1+n_2}} + \frac{\sum\limits_{j=n_1+1}^{n} X_j - \mu n_2}{\sigma\sqrt{n}} \\ &= \frac{\sum_{j=1}^{n_1} X_j - \mu n_1}{\sigma\sqrt{n_1}\sqrt{1+\frac{n_2}{n_1}}} + \frac{\sum_{j=n_1+1}^{n} X_j - \mu n_2}{\sigma\sqrt{n}}. \end{aligned} \tag{5.A.2}$$

Since $\sqrt{1+\frac{n_2}{n_1}}$ converges to 1, the first sum in (5.A.2) converges in distribution to a random variable with characteristic function given by (5.3.2). This is because the sample $X_1, \ldots, X_{n_1}$ is defined exactly as the original sample described in Section 5.2. Furthermore, the second sum in (5.A.2) converges in distribution to 0. Indeed, for any n, the expectation of that sum is 0, while its variance is given by n_2/n, which converges to 0. By Slutsy's theorem, we then have that S_n converges in distribution to a random variable with characteristic function given by (5.3.2).

5.B Additional examples

We provide additional examples of distributions F (discrete or absolutely continuous) and their associated r parameter. Those examples serve to showcase the range of possible values for r, and in particular we provide examples such that $r = 0$ or $r^2 = 1$.

Example 5.3 ($r = 0$ when F is discrete)**.** *For any integer $\ell \geq 2$, take*

$$A = \{-1, 1\}, \quad \mathbb{P}(W = -1) = \mathbb{P}(W = 1) = \tfrac{\ell^{-1}}{2} \quad \textit{and} \quad \mathbb{P}(W = -2) = \mathbb{P}(W = 2) = \tfrac{1-\ell^{-1}}{2}, \tag{5.B.1}$$

since it implies $\mu_U = \mu_V = 0$.

Example 5.4 ($r = 0$ when F is absolutely continuous)**.** *For any integer $\ell \geq 2$, take*

$$A = [-\ell^{-1}, \ell^{-1}] \quad \textit{and} \quad W \sim \textit{Uniform}[-1, 1], \tag{5.B.2}$$

since again it implies $\mu_U = \mu_V = 0$.

Example 5.5 ($r^2 = 1$ when F is discrete)**.** *Let $\ell \geq 2$ be any integer. To get $r = 1$, take*

$$A = \{1\}, \quad \mathbb{P}(W = 1) = \ell^{-1} \quad \textit{and} \quad \mathbb{P}(W = -1) = 1 - \ell^{-1}, \tag{5.B.3}$$

since this means $\sigma_U = \sigma_V = 0$ and $\mu_V > \mu_U$. By symmetry, taking $A^c = \{1\}$ instead yields $r = -1$.

Example 5.6 (r arbitrarily close to 1 when F is absolutely continuous)**.** *Let f_1 be the density function of a N($-\ell^{-1}, \sigma^2$), f_2 the density function of a N($1 - \ell^{-1}, \sigma^2$), and f their mixture: $f(x) = (1 - \ell^{-1})f_1(x) + \ell^{-1}f_2(x)$. Then, for $W \sim f$, we have $\mathbb{E}[W] = 0$ and $\mathbb{V}\mathrm{ar}[W] = \mathbb{E}[W^2] = \sigma^2 + \ell^{-1}(1 - \ell^{-1})$. Assuming that $0 < \sigma \leq \frac{1}{2}\ell^{-1}$, a straightforward Gaussian tail estimate on f_1 shows that there exists $w_\ell \in (-\ell^{-1}, 1 - \ell^{-1})$ such that $\mathbb{P}(W \in [w_\ell, \infty)) = \ell^{-1}$. If we take $A = [w_\ell, \infty)$, then we have*

$$\begin{aligned}
\mu_V &= \ell \int_{w_\ell}^{\infty} x\left((1 - \ell^{-1})f_1(x) + \ell^{-1}f_2(x)\right) \mathrm{d}x \\
&= (\ell - 1) \int_{w_\ell}^{\infty} \Big(\frac{x + \ell^{-1}}{\sigma} - \frac{\ell^{-1}}{\sigma}\Big) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big(\frac{x + \ell^{-1}}{\sigma}\Big)^2\right) \mathrm{d}x \\
&\quad + \int_{w_\ell}^{\infty} \Big(\frac{x - (1 - \ell^{-1})}{\sigma} + \frac{1 - \ell^{-1}}{\sigma}\Big) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big(\frac{x - (1 - \ell^{-1})}{\sigma}\Big)^2\right) \mathrm{d}x \\
&= (\ell - 1)\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big(\frac{w_\ell + \ell^{-1}}{\sigma}\Big)^2\right) - (1 - \ell^{-1})\Psi\Big(\frac{w_\ell + \ell^{-1}}{\sigma}\Big)
\end{aligned}$$

$$+\frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\Big(\frac{w_\ell-(1-\ell^{-1})}{\sigma}\Big)^2\right)+(1-\ell^{-1})\Psi\Big(\frac{w_\ell-(1-\ell^{-1})}{\sigma}\Big), \qquad (5.B.4)$$

where $\Psi(z)$ is the survival function of the standard Gaussian. Therefore, from (5.4.1)*, we have $r\to 1$ as $\sigma\to 0$:*

$$r=\frac{\mu_V-\mathbb{E}[W]}{\sqrt{\mathbb{V}\mathrm{ar}(W)}\sqrt{\ell-1}}\underset{\sigma\to 0}{\longrightarrow}\frac{0-0+0+(1-\ell^{-1})-0}{\sqrt{0^2+\ell^{-1}(1-\ell^{-1})}\sqrt{\ell-1}}=1. \qquad (5.B.5)$$

Example 5.7 (F is a $N(\mu,\sigma^2)$)**.** *Let $\ell=2$, choose $A=[\mu,\infty)$ and let Z be a $N(0,1)$ r.v. Then,*

$$\mu_V=\mu+\sigma\,\mathbb{E}[Z|Z>0]=\mu+\sigma\,\mathbb{E}[|Z|]=\mu+\sigma\sqrt{\frac{2}{\pi}}. \qquad (5.B.6)$$

It follows that $r=\sqrt{2/\pi}\approx 0.8$ (irrespective of μ and σ). Note that this corresponds to the purple dotted curve on Figure 5.1. Hence this case provides a nice illustration of how 'badly' the CLT can fail for pairwise independent Gaussian variables.

5.C Derivations of r for common distributions

In this section, we derive the value of r as given in (5.5.1), for all distributions presented in Table 5.1.

Uniform(a, b)

Let $X \sim \text{Uniform}(-\sqrt{3}, \sqrt{3})$. Then, $\mu_X = m_X = 0, \sigma_X = 1$, so that $r = \mathbb{E}[X|X > 0] = \frac{\sqrt{3}}{2}$.

Normal(μ, σ^2)

Let $X \sim \text{Normal}(0, 1)$, so that $r = \mathbb{E}[X|X > 0]$. We note that the random variable

$$|X| \stackrel{d}{=} X|X > 0,$$

has a half-normal distribution, whose mean is known to be $\sqrt{2/\pi}$.

Exponential(λ)

Let $X \sim \text{Exponential}(\lambda)$. Then, $\mu_X = 1/\lambda$, $m_X = \log(2)/\lambda$ and $\sigma_X = 1/\lambda$. Hence,

$$r = \frac{\mathbb{E}[X|X > \log(2)/\lambda] - 1/\lambda}{1/\lambda} = \log(2).$$

Student(ν) for $\nu > 2$

Let $X \sim \text{Student}(\nu)$. Then, $\mu_X = m_X = 0$ and $\sigma_X^2 = \nu/(\nu - 2)$. We note that the random variable

$$|X| \stackrel{d}{=} X|X > 0,$$

has a folded-t distribution, whose mean is

$$\mathbb{E}\left[|X|\right] = 2\sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\,(\nu-1)},$$

see Psarakis and Panaretoes (1990, Theorem 3.1). Hence,

$$r = \frac{\mathbb{E}\left[|X|\right]}{\sqrt{\nu/(\nu-2)}} = 2\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu-1)}.$$

Log-normal(μ, β)

If $Y \sim \text{Log-normal}(\mu, \beta)$, then for any $a > 0$, $aY \sim \text{Log-normal}(\mu + \log(a), \beta)$. Because of Proposition 5.1, this means that the parameter μ does not affect the value of r. Hence, to ease calculations (and without loss of generality) we let $X \sim \text{Log-normal}(0, \beta)$. This

corresponds exactly to Example 5.1, from which we have

$$r = \frac{\text{Erf}\left(\sqrt{\beta/2}\right)}{\sqrt{e^{\beta}-1}}.$$

Pareto(α, λ) for $\alpha > 2$

Let $X \sim \text{Pareto}(\alpha, \lambda)$. Then,

$$\begin{aligned}\mathbb{E}[X|X > m_X] &= \int_{m_X}^{\infty} 2x f_X(x) dx \\ &= \frac{2\alpha}{\lambda} \int_{m_X}^{\infty} x \left(1 + \frac{x}{\lambda}\right)^{-(\alpha+1)} dx.\end{aligned}$$

Noting that $m_X = \lambda\left(2^{1/\alpha} - 1\right)$ and with the change of variable $y = x/\lambda$ we obtain:

$$\begin{aligned}\mathbb{E}[X|X > m_X] &= 2\alpha\lambda \int_{2^{1/\alpha}-1}^{\infty} y\,(1+y)^{-(\alpha+1)}\, dy \\ &= \alpha\lambda \left(\frac{2^{1/\alpha}}{\alpha - 1} - \frac{1}{\alpha}\right),\end{aligned}$$

where the last equality was obtained using `Mathematica`. Therefore, we have

$$\begin{aligned}r &= \frac{\mathbb{E}[X|X > m_X] - \mu_X}{\sigma_X} \\ &= \frac{\alpha\lambda\left(\frac{2^{1/\alpha}}{\alpha-1} - \frac{1}{\alpha}\right) - \frac{\lambda}{\alpha-1}}{\sqrt{\frac{\lambda^2\alpha}{(\alpha-1)^2(\alpha-2)}}} \\ &= \left(\frac{\alpha 2^{1/\alpha} - (\alpha-1) - 1}{\alpha - 1}\right)(\alpha-1)\sqrt{\frac{\alpha-2}{\alpha}} \\ &= \sqrt{\alpha(\alpha-2)}\left(2^{1/\alpha} - 1\right).\end{aligned}$$

Gamma(α, β)

Since β is a rate parameter (hence it does not affect the value of r), we can without loss of generality let $X \sim \text{Gamma}(\alpha, 1)$. As before, we denote m_X the median of X (specifically for $\beta = 1$). Note that unfortunately there is no closed-form expression for m_X. We have

$$\begin{aligned}\mathbb{E}[X|X > m_X] &= \int_{m_X}^{\infty} 2x f_X(x) dx \\ &= \frac{2}{\Gamma(\alpha)} \int_{m_X}^{\infty} x^{\alpha} e^{-x} dx \\ &= \frac{2\Gamma(\alpha+1, m_X)}{\Gamma(\alpha)}\end{aligned}$$

Using the property of the incomplete gamma function: $\Gamma(\alpha+1,x) = \alpha\Gamma(\alpha,x) + x^\alpha e^{-x}$, we simplify this to

$$\mathbb{E}[X|X > m_X] = \frac{2\left(\alpha\Gamma(\alpha,m_X) + (m_X)^\alpha e^{-m_X}\right)}{\Gamma(\alpha)},$$

and noting that the survival function $S(x)$ of a Gamma$(\alpha,1)$ is given by $\Gamma(\alpha,x)/\Gamma(\alpha)$, this further simplifies to

$$\mathbb{E}[X|X > m_X] = \alpha + \frac{2(m_X)^\alpha e^{-m_X}}{\Gamma(\alpha)}.$$

Lastly, since $\mu_X = \alpha, \sigma_X = \sqrt{\alpha}$, we get

$$r = \frac{2(m_X)^\alpha e^{-m_X}}{\Gamma(\alpha)\sqrt{\alpha}}.$$

Weibull(λ,k)

Let $X \sim$ Weibull(λ,k). Then,

$$\begin{aligned}\mathbb{E}[X|X > m_X] &= \int_{m_X}^{\infty} 2x f_X(x)dx \\ &= \frac{2k}{\lambda^k}\int_{m_X}^{\infty} x^k e^{-(x/\lambda)^k}dx.\end{aligned}$$

Noting that $m_X = \lambda\log(2)^{1/k}$ and with the change of variable $y = x/\lambda$ we obtain:

$$\begin{aligned}\mathbb{E}[X|X > m_X] &= 2k\lambda\int_{\log(2)^{1/k}}^{\infty} y^k e^{-y^k}dy \\ &= 2\lambda\Gamma\left(1+\frac{1}{k},\log(2)\right),\end{aligned}$$

where the last equality was obtained through `Mathematica`. Therefore, we have

$$\begin{aligned}r &= \frac{\mathbb{E}[X|X > m_X] - \mu_X}{\sigma_X} \\ &= \frac{2\lambda\Gamma\left(1+\frac{1}{k},\log(2)\right) - \lambda\Gamma(1+\frac{1}{k})}{\sqrt{\lambda^2\left[\Gamma\left(1+\frac{2}{k}\right) - \left(\Gamma\left(1+\frac{1}{k}\right)\right)^2\right]}} \\ &= \frac{2\Gamma\left(1+\frac{1}{k},\log(2)\right) - \Gamma(1+\frac{1}{k})}{\sqrt{\Gamma\left(1+\frac{2}{k}\right) - \left(\Gamma\left(1+\frac{1}{k}\right)\right)^2}}.\end{aligned}$$

Fréchet(α) for $\alpha > 2$

Let $X \sim$ Fréchet(α). Then,

$$\begin{aligned}\mathbb{E}[X|X > m_X] &= \int_{m_X}^{\infty} 2x f_X(x) dx \\ &= 2\alpha \int_{m_X}^{\infty} x^{-\alpha} e^{-x^{-\alpha}} dx \\ &= 2\Gamma\left(1 - \frac{1}{a}\right) - 2\Gamma\left(1 - \frac{1}{a}, \log(2)\right),\end{aligned}$$

where the last equality was obtained through `Mathematica` (using the fact that $m_X = \log(2)^{-1/\alpha}$). Therefore,

$$\begin{aligned}r &= \frac{\mathbb{E}[X|X > m_X] - \mu_X}{\sigma_X} \\ &= \frac{\Gamma\left(1 - \frac{1}{\alpha}\right) - 2\Gamma\left(1 - \frac{1}{\alpha}, \log(2)\right)}{\sqrt{\Gamma\left(1 - \frac{2}{\alpha}\right) - \left(\Gamma\left(1 - \frac{1}{\alpha}\right)\right)^2}}.\end{aligned}$$

5.D Tail-heaviness: definitions and literature review

In this appendix, we review the question *what is a heavy-tailed distribution*? On a qualitative level, Resnick (2007) explains: "heavy tails are characteristic of phenomena where the probability of a huge value is relatively big". Of course, what 'huge' and 'relatively big' mean is subjective (and context dependent). While definitions and terminologies vary, the most common technical definition of what it means for a distribution F to be (right-) heavy-tailed is as follows (see, e.g. Foss et al., 2013, Definition 2.2).

Definition 5.1. *A distribution F is defined to be (right-) heavy-tailed if and only if*

$$\int_{\mathbb{R}} e^{\lambda x} F(dx) = \infty \quad \textit{for all } \lambda > 0. \tag{5.D.1}$$

By Theorem 2.6 in Foss et al. (2013), we have the following equivalent characterisation, which is perhaps easier to interpret. It utilises the survival function $\overline{F}(x) \coloneqq \mathbb{P}[X > x]$ of a random variable $X \sim F$.

Theorem 5.2. *A distribution F is (right-) heavy-tailed if and only if:*

$$\limsup_{x \to \infty} \overline{F}(x) e^{\lambda x} = \infty \quad \textit{for all } \lambda > 0.$$

Hence, clearly, tail-heaviness has to do with what happens 'far away' in the tail. If a distribution F is such that its survival function decreases more slowly than an exponential function, then it is heavy-tailed (and 'huge values' will have 'relatively big' probabilities, as the heuristic definition puts it). To sharpen this interpretation, let us also recall the closely related notion of a 'long-tailed' distribution. We give the definition as in Foss et al. (2013, Definition 2.21).

Definition 5.2. *A distribution F on $\mathbb{R}$ is called long-tailed if $\overline{F}(x) > 0$ for all x and, for any fixed $y > 0$,*

$$\frac{\overline{F}(x+y)}{\overline{F}(x)} \to 1 \quad \textit{as } x \to \infty.$$

Qualitatively, a random variable X is long-tailed if, given that X exceeds a certain high threshold x, the probability it exceeds any higher threshold is large (and tends to 1 for x very large). We note that every long-tailed distribution is heavy-tailed (but the converse is not true).

Definitions 5.1 and 5.2 provide a binary categorisation of tail-heaviness (either a distribution is heavy-tailed or it is not), but other categorisations and definitions are possible (see Rojo, 2013, for a review). In particular, a categorisation arising from Extreme Value Theory relies on the notion of *maximum domain of attraction* (see the following definition, taken from Embrechts et al., 1997, Definition 3.3.1).

Definition 5.3. *Let $X_1, \ldots, X_n$ be i.i.d non-degenerate random variables with common distribution F, and denote $M_n = \max(X_1, \ldots, X_n)$ their maximum. If there exists constants $c_n > 0, d_n \in \mathbb{R}$ and a non-degenerate distribution H such that*

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H,$$

then we say that F belongs to the maximum domain of attraction of H.

From the Fisher-Tippett Theorem (stated previously, see Theorem 4.1) we know the H in Definition 5.3 can only be of three types: Weibull, Gumbel or Fréchet. Then, from this framework, a distribution F can be said to be light-tailed, medium-tailed or long-tailed if it belongs to the maximum domain of attraction of the Weibull, Gumbel or Fréchet, respectively (Rojo, 2013). Note that, according to this classification, the Normal, Exponential and Log-normal are all classified as medium-tailed (see Embrechts et al., 1997, Section 3.3.3), which does not agree with the common idea that a Log-normal is much heavier-tailed than a Normal.

One could say that attributing distributions to a discrete number of categories (e.g., light, medium or heavy-tailed) cannot provide the full picture. For instance, using Definition 5.1 yields that the Pareto, Burr, Cauchy, Log-normal and Weibull (with shape parameter $\alpha < 1$) are all heavy-tailed. But can we declare one to be the *heavier* tailed, and based on what criteria? Furthermore, within a specific family, (e.g., Log-normal), common knowledge has it that changing some parameter(s) will alter the tail-heaviness. For example, for a Log-normal(μ, β), increasing β increases tail-heaviness. We then want to quantitatively measure tail-heaviness, and many options are available in the literature (though it seems none make consensus).

The most common measure is probably the kurtosis coefficient, which we denote κ. For any random variable X with finite fourth moment, mean μ and variance σ^2, κ is defined as

$$\kappa = \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^4\right].$$

However, the kurtosis κ has many shortcomings. As Brys et al. (2006) put it:

> The kurtosis coefficient is often regarded as a measure of the tail-heaviness of a distribution relative to that of the normal distribution. However, it also measures the peakedness of a distribution, hence there is no agreement on what kurtosis really estimates. Another disadvantage of the kurtosis is that its interpretation and consequently its use is restricted to symmetric distributions. Moreover, the kurtosis coefficient is very sensitive to outliers in the data.

Several alternative measures of tail-heaviness have been proposed. Geary (1936) introduced a measure defined for a random variable X with mean μ and standard deviation $\sigma < \infty$. Denoting $\tau = \mathbb{E}[|X - \mu|]$, this measure is given by τ/σ. Bonett and Seier (2002) proposed a modification of this measure ω through the transformation

$$\omega = 13.29\,(\log(\sigma) - \log(\tau))$$

(where the constant 13.29 is chosen so that ω is approximately 3 for a Normal distribution). Bonett and Seier (2002) note that the advantage of their ω is that it can increase without bounds as a distribution gets more heavy-tailed. They report the values of this ω for many symmetric distributions and use their ω as the basis for a normality test. Hogg (1972) proposed a measure of tail-heaviness defined as

$$Q = \frac{U_{0.05} - L_{0.05}}{U_{0.50} - L_{0.50}},$$

where U_α, and L_α are the means of the upper and lower '100α percent' tails, respectively. Moors (1988) provided a robust measure of tail-heaviness. For a continuous $X \sim F$, their measure is defined as:

$$T = \frac{(E_7 - E_5) + (E_3 - E_1)}{E_6 - E_2}, \tag{5.D.2}$$

where the E_i are the 'octiles' of distribution F, i.e., they satisfy

$$F(E_i) = i/8, \quad i = 1, \ldots, 8.$$

Like kurtosis, this T is not bounded. Furthermore, because it is entirely based on quantiles, it always exists (i.e., it does not require existence assumption on any moments).

Brys et al. (2006) proposed two other robust measures of (right-) tail-heaviness which

always exist (i.e., they do not require finite moments assumptions). Considering again a continuous distribution F with quantile function F^{-1}, the first measure is given by

$$\mathrm{RQW}_F(q) = \frac{F^{-1}((1+q)/2) + F^{-1}(1-q/2) - 2F^{-1}(3/4)}{F^{-1}((1+q)/2) - F^{-1}(1-q/2)}, \tag{5.D.3}$$

where $1/2 < q < 1$. The authors suggest the use of $q = 3/4$ or $q = 7/8$ to "retain a reasonable amount of robustness". We note that $-1 \leq \mathrm{RQW}_F(q) \leq 1$. A value close to 1 is associated with a heavy right tail, since in that case the term $F^{-1}((1+q)/2)$ (a high quantile) dominates all the others. Their second measure of (right-) tail-heaviness is a modification of the *medcouple* (Brys et al., 2004). Call m_F the median of distribution F. Let X_1 be sampled from F_1, where F_1 is a version of F truncated below the median. Likewise, let X_2 be sampled from F_2, a version of F truncated above the median. Then, the medcouple of F, denoted $\mathrm{MC}(F)$ is defined as

$$\mathrm{MC}(F) = \text{median}\left[\frac{(X_2 - m_F) - (m_F - X_1)}{X_2 - X_1}\right],$$

and is a measure of skewness. Then, for $X \sim F$, we obtain a measure of right tail-heaviness of F by computing the medcouple of the random variable $X|X > m_F$.

CHAPTER 6

CENTRAL LIMIT THEOREMS UNDER K-TUPLEWISE INDEPENDENCE

6.1 Introduction

Thus far in this thesis, we have been mostly concerned with the materiality of the difference between 'pairwise' and 'mutual' independence. In general, one can also define a notion of 'K-tuplewise independence' ($K \geq 2$). Indeed, we say a sequence of random variables $X_1, X_2, X_3, \ldots$ is 'K-tuplewise independent' if $X_{i_1}, X_{i_2}, \ldots, X_{i_K}$ are mutually independent whenever $(i_1, i_2, \ldots, i_K)$ is a K-tuple of distinct positive integers (see, e.g., Pruss, 1998). Note the case $K = 2$ corresponds to 'pairwise independence'. While mutual independence implies K-tuplewise independence (for any K), the converse is not true. That is, for any integer $K \geq 2$, we can always build a sequence of K-tuplewise independent, but still dependent, random variables.

In this chapter, we extend many of the ideas presented in the previous Chapter 5 to the case of 'triplewise independent' ($K = 3$) random variables. In particular, we present a general methodology to construct triplewise independent sequences of random variables having a common but arbitrary marginal distribution F (satisfying very mild conditions). We then investigate under which conditions such sequences satisfy a CLT, and give many specific examples.

While several examples of PIBD variables can be found in the literature (see Section 5.1 for a review), examples of K-tuplewise independent variables which are not mutually independent (for $K \geq 3$) are more scarce. This may explain why we still have an incomplete understanding of which fundamental theorems of mathematical statistics 'fail' under the weaker assumption of triplewise (or in general K-tuplewise) independence (and to what extent). This also provides motivation for this chapter.

We know from the literature that the classical CLT, arguably one of the most important results in all of statistics, need not be valid under K-tuplewise independence. Few authors have studied this question. Pruss (1998) showed that, for any integer K, one can build a sequence of K-tuplewise independent r.v.s for which no CLT holds. Bradley and Pruss (2009) further showed that even if such a sequence is *strictly stationary*, a CLT need not hold. Weakley (2013) extended this work by allowing the r.v.s in the sequence to have any symmetrical distribution (with finite variance). Takeuchi (2019) showed that K growing linearly with the sample size n is not even sufficient for a CLT to hold. In those examples, however, the asymptotic distribution of the sample mean S_n is not given explicitly, hence we cannot judge to what extent it departs from normality.

Kantorovitz (2007) does provide an example of a triplewise independent sequence for which S_n converges to a 'misbehaved' distribution —that of $Z_1 \cdot Z_2$, where Z_1 and Z_2 are independent $N(0,1)$ — but this is achieved for a very specific choice of margin, namely the Bernoulli distribution.

In Section 6.2, we present a methodology, borrowing elements from graph theory, to construct new sequences of *triplewise* independent and identically distributed (noted thereafter t.i.i.d.) r.v.s whose common marginal distribution F can be chosen arbitrarily (under very mild conditions). In Section 6.3, we provide a necessary and sufficient condition for a CLT to hold for such sequences.

In Section 6.4, we provide what we believe to be the first two examples of triplewise independent sequences with arbitrary margins for which the asymptotic distribution of the standardised sample mean is explicitly known and non-Gaussian. Those two distributions depend on the choice of the margin F and have heavier tails than a Gaussian. This allows us to assess how far away from the Gaussian distribution one can get under sole *triplewise* independence. This work thus highlights why *mutual* independence is so fundamental for the classical CLT to hold.

Lastly, in Section 6.5, we explain how our methodology can easily be extended to create new K-tuplewise independent sequences (which are not mutually independent) for any integer K. While such sequences are interesting in themselves, it appears that for $K \geq 4$ they *do* verify a CLT, and we explain heuristically why this is the case. Despite not being the focus of this thesis, we note that these sequences could prove useful to benchmark the performance of multivariate independence tests, many of which have been proposed in recent years, see, e.g., Fan et al. (2017); Jin and Matteson (2018); Yao et al. (2018); Böttcher et al. (2019); Chakraborty and Zhang (2019); Genest et al. (2019); Drton et al. (2020).

6.2 Construction of triplewise independent sequences

In this section, we present a general methodology to construct sequences $\{X_j, j \geq 1\}$ of t.i.i.d. r.v.s having a common (but arbitrary) marginal distribution F. The distribution F is not *completely* arbitrary, and must satisfy the following technical condition:

Condition 6.1. *F has finite variance and for any r.v. $W \sim F$, there exists a Borel set A with $\mathbb{P}(W \in A) = \ell^{-1}$, where $\ell \geq 2$ is an integer.*

(Although this condition may appear surprising, it is needed for our construction to hold, and we explain in Remark 6.2 why it cannot be relaxed). We begin our construction of the sequence $\{X_j, j \geq 1\}$ by letting F be a distribution satisfying Condition 6.1, with mean and variance denoted by μ and σ^2, respectively. For a r.v. $W \sim F$, let A be any Borel set such that

$$\mathbb{P}(W \in A) = \ell^{-1}, \quad \text{for some integer } \ell \geq 2. \tag{6.2.1}$$

Our construction relies on a sequence of simple graphs $\{G_m, m \geq 1\}$ with two properties (for a review of some graph theory concepts, see Section 6.B):

1. The girth of G_m is 4 (or larger), for all m;
2. The number of edges of G_m grows to infinity as $m \to \infty$.

Aside from these properties, the sequence $\{G_m, m \geq 1\}$ is left unspecified, making our construction very general. As a concrete example, consider a *complete bipartite graph* composed of two sets of m vertices, where every vertex from one set is linked by an edge to every vertex in the second set; see Figure 6.1 with $m = 4$ for an illustration. Such graphs are often denoted by $K_{m,m}$, see, e.g, Diestel (2005, p.17). Note we use them to make our construction more 'concrete' to the reader, but they are only one possible example of graphs G_m satisfying the two conditions above. Our general construction is as follows.

Let $v(m)$ be the number of vertices of G_m and let $M_1, \ldots, M_{v(m)}$ be a sequence of i.i.d. discrete uniforms on the set $\{1, 2, \ldots, \ell\}$, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Precisely, for $i = 1, 2, \ldots, \ell$, let

$$p_i := \mathbb{P}(M_1 = i) = \ell^{-1}. \tag{6.2.2}$$

Assign the uniform r.v.s $M_1, \ldots, M_{v(m)}$ to the $v(m)$ vertices of the graph (the order does

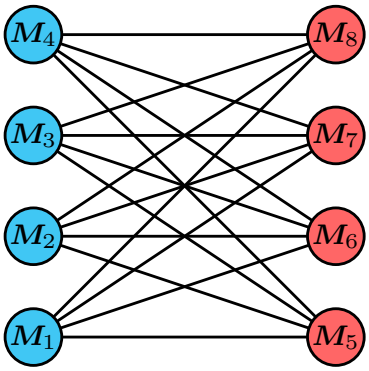

Figure 6.1: Graph $K_{4,4}$ with uniform r.v.s M_j, $1 \leq j \leq 8$, defined in (6.2.2) assigned to the vertices. The vertices on the left (colored in blue) belong to one set while the vertices on the right (colored in red) belong to another set.

not matter). Then, for every pair $(i, j), 1 \leq i < j \leq v(m)$ such that an edge connects M_i and M_j, define a r.v. $D_{i,j}$ as

$$D_{i,j} = \begin{cases} 1, & \text{if } M_i = M_j, \\ 0, & \text{otherwise.} \end{cases} \tag{6.2.3}$$

Let n be the total number of edges (in the specific example of complete bipartite graphs $K_{m,m}$ from Figure 6.1, $n = m^2$). For convenience, we relabel the n random variables in the sequence $\{D_{i,j}\}$ simply as

$$D_1, \ldots, D_n. \tag{6.2.4}$$

Importantly, the sequence $D_1, \ldots, D_n$ is triplewise independent (see Remark 6.1). However, the variables D_k have a specific distribution, i.e., they are Bernoulli, with

$$\mathbb{P}[D_k = 1] = \ell^{-1}.$$

From those D_k variables, we now construct a much more general triplewise independent sequence $X_1, \ldots, X_n$, where $X_k \sim F$, for all $k = 1, \ldots, n$. The general idea is to let every X_k be a mixture of two distributions as follows: for $W \sim F$,

- if $D_k = 1$, then X_k has the distribution of $\{W | W \in A\}$;
- if $D_k = 0$, then X_k has the distribution of $\{W | W \in A^c\}$.

Since, by construction, $\mathbb{P}(W \in A) = \ell^{-1}$, this yields that the marginal distribution of X_k is F.

More formally, define U and V, with cumulative distribution functions F_U and F_V respectively, to be the truncated versions of $W \sim F$, respectively off and on the set A:

$$U \stackrel{d}{=} W|\{W \in A^c\}, \qquad V \stackrel{d}{=} W|\{W \in A\}, \tag{6.2.5}$$

and denote

$$\mu_U := \mathbb{E}[U], \qquad \sigma_U^2 := \mathbb{V}\text{ar}[U], \qquad \mu_V := \mathbb{E}[V], \qquad \sigma_V^2 := \mathbb{V}\text{ar}[V]. \tag{6.2.6}$$

Then, consider n independent copies of U, and independently, n independent copies of V:

$$U_1, U_2, \ldots, U_n \overset{\text{i.i.d.}}{\sim} F_U, \qquad V_1, V_2, \ldots, V_n \overset{\text{i.i.d.}}{\sim} F_V, \tag{6.2.7}$$

both defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Finally, for $\omega \in \Omega$ and for all $k = 1, \ldots, n$, construct

$$X_k(\omega) = \begin{cases} U_k(\omega), & \text{if } D_k(\omega) = 0, \\ V_k(\omega), & \text{if } D_k(\omega) = 1. \end{cases} \tag{6.2.8}$$

By conditioning on D_k, it is easy to verify that

$$F_{X_k}(x) = (1 - \ell^{-1})F_{U_k}(x) + \ell^{-1}F_{V_k}(x) = F(x). \tag{6.2.9}$$

Lastly, it is not hard to see that $X_1, \ldots, X_n$ is triplewise independent. Indeed, for any given $k, k', k'' \in \{1, 2, \ldots, n\}$ with k, k', k'' all different, the r.v.s D_k, U_k, V_k, $D_{k'}$, $U_{k'}$, $V_{k'}$, $D_{k''}$, $U_{k''}$, $V_{k''}$ are mutually independent and one can write $X_k = g(D_k, U_k, V_k)$, $X_{k'} = g(D_{k'}, U_{k'}, V_{k'})$ and $X_{k''} = g(D_{k''}, U_{k''}, V_{k''})$ for g a Borel-measurable function. Since X_k, $X_{k'}$ and $X_{k''}$ are integrable, the result follows from the triplewise independence analogue of Corollary 2 in Pollard (2002, Section 4.1).

Remark 6.1. *For the sequence $D_1, \ldots, D_n$ defined in* (6.2.4) *to be t.i.i.d., the following restrictions on $p_1, p_2, \ldots, p_\ell \in (0, 1)$ must be satisfied:*

$$\begin{aligned} (1) \ &: \ p_1 + p_2 + \cdots + p_\ell = 1, \\ (2) \ &: \ p_1^2 + p_2^2 + \cdots + p_\ell^2 = w, \\ (3) \ &: \ p_1^3 + p_2^3 + \cdots + p_\ell^3 = w^2, \\ (4) \ &: \ p_1^4 + p_2^4 + \cdots + p_\ell^4 = w^3. \end{aligned} \tag{6.2.10}$$

for some $w \in (0,1)$. Now, it is straightforward that letting $p_i = \ell^{-1}$ for $i = 1, \dots, \ell$ (as we have done in our construction) satisfies (6.2.10).

To be more specific, condition (1) *is necessary for the distribution in* (6.2.2) *to be well-defined, and conditions* (2), (3) *and* (4) *are rewritings of*

$$\mathbb{P}(D_{v_1,v_2} = 1) = w, \tag{6.2.11}$$

$$\mathbb{P}(D_{v_1,v_2} = 1, D_{v_2,v_3} = 1) = \mathbb{P}(D_{v_1,v_2} = 1)\mathbb{P}(D_{v_2,v_3} = 1), \tag{6.2.12}$$

$$\mathbb{P}(D_{v_1,v_2} = 1, D_{v_2,v_3} = 1, D_{v_3,v_4} = 1) = \mathbb{P}(D_{v_1,v_2} = 1)\mathbb{P}(D_{v_2,v_3} = 1)\mathbb{P}(D_{v_3,v_4} = 1), \tag{6.2.13}$$

$$\forall\, (v_1,v_2),(v_2,v_3),(v_3,v_4) \in \mathrm{Edges}(G_m).$$

(Indeed, the edges on the path $v_1 v_2 \dots v_k$ all have the value 1 *if and only if all the corresponding values on the vertices, $M_{v_1}, M_{v_2}, \dots, M_{v_k}$, are equal. With ℓ possible choices for each vertex, this event has probability $\mathbb{P}(D_{v_{j-1},v_j} = 1\ \forall j \in \{2,3,\dots,k\}) = \sum_{i=1}^{\ell} \prod_{j=1}^{k} \mathbb{P}(M_j = i) = \sum_{i=1}^{\ell} p_i^k \quad \forall k \in \mathbb{N}$.) Lastly, conditions* (6.2.11), (6.2.12) *and* (6.2.13) *are sufficient to guarantee that the D's are identically distributed and triplewise independent.*

Remark 6.2. *In Condition 6.1, the restriction $\mathbb{P}(W \in A) = \ell^{-1}$ for some integer ℓ may seem arbitrary. Likewise, in* (6.2.2) *the choice $p_i = \ell^{-1}$ for $i = 1, \dots, \ell$ may also seem arbitrary. We establish here that none of these choices are arbitrary, simply because the solution*

$$p_i = \ell^{-1}$$

to (6.2.10) *is unique. Indeed, by squaring condition (2) in* (6.2.10) *then applying the Cauchy-Schwarz inequality, one gets*

$$w^2 = \Big(\sum_{i=1}^{\ell} p_i^{3/2} p_i^{1/2} \Big)^2 \leq \sum_{i=1}^{\ell} p_i^3 \sum_{i=1}^{\ell} p_i = \sum_{i=1}^{\ell} p_i^3 \tag{6.2.14}$$

where the last equality comes from condition (1) in (6.2.10). *Then, condition* (3) *requires that we have the equality in* (6.2.14), *and this happens if and only if $p_i^{3/2} = \lambda p_i^{1/2}$ for all $i \in \{1, \dots, \ell\}$ and for some $\lambda \in \mathbb{R}$. In turn, this implies $p_i = \lambda = \ell^{-1}$ because of* (1) *and since $p_i > 0$, which then implies $w = \ell^{-1}$ by* (2). *This unique solution also satisfies* (4). *This reasoning shows that we cannot extend our method to an arbitrary $\mathbb{P}(W \in A) \in (0,1)$ in* (6.2.1).

6.3 Main result

In the previous section, we constructed a triplewise independent sequence of random variables $\{X_k, 1 \leq k \leq n\}$ with an arbitrary common margin F. Because the construction relies on an unspecified sequence of graphs $\{G_m, m \geq 1\}$, the 'dependence structure' of this sequence is only partially specified, and the asymptotic distribution of its standardised mean (call it S_n) depends on which specific graphs $\{G_m, m \geq 1\}$ are used. In this section, we state our main result, which links the asymptotic distribution of S_n to the specific graphs G_m chosen. The result holds for any growing sequence of simple graphs $\{G_m, m \geq 1\}$ of girth at least 4 (as defined previously). Note that *specific examples* are given in the next section, two for which the standardised mean S_n is asymptotically *not* Gaussian (and two for which it *is* Gaussian).

To state our result, we first recall that

- for a fixed m, n is the number of edges in the graph G_m (n is also the number of variables in the sequence $X_1, \ldots, X_n$);
- the variables $\{X_k, 1 \leq k \leq n\}$ have finite mean μ, finite variance σ^2, and a common arbitrary margin F such that for some Borel set A and an integer $\ell \geq 2$,

$$\mathbb{P}(X_k \in A) = \ell^{-1};$$

- the sequence $X_1, \ldots, X_n$ is constructed from a sequence $D_1, \ldots, D_n$ of Bernoulli(ℓ^{-1}) random variables, see (6.2.3) and (6.2.4).

Lastly, to state our result we need to define a new quantity, Ξ_n, as being the number of 1's in the sequence $\{D_k, 1 \leq k \leq n\}$. We also let ξ_n be its standardised version, i.e.,

$$\Xi_n = \sum_{k=1}^{n} D_k, \qquad \xi_n = \frac{\Xi_n - n\ell^{-1}}{\sqrt{n\ell^{-1}(1-\ell^{-1})}}. \tag{6.3.1}$$

Note the distribution of ξ_n depends on which graphs $\{G_m, m \geq 1\}$ are chosen for the construction. We are now ready to present our main result.

Theorem 6.1. *Let $X_1, \ldots, X_n$ be random variables defined as in* (6.2.8)*, and ξ_n be as defined in* (6.3.1)*. Provided that there exists a r.v. Y such that*

$$\xi_n \xrightarrow{d} Y, \quad \text{as } m \to \infty \text{ (and thus as } n \to \infty), \tag{6.3.2}$$

then the standardised sample mean $S_n := \big(\sum_{k=1}^n X_k - n\mu\big)/\sigma\sqrt{n}$ *converges in law to the random variable*

$$S^{(\ell)} := \sqrt{1-r^2}Z + r\,Y, \tag{6.3.3}$$

where $Z \sim N(0,1)$ *and* $r := \frac{\sqrt{\ell^{-1}(1-\ell^{-1})}(\mu_V - \mu_U)}{\sigma}$.

Remark 6.3. *If* $r \neq 0$ *and* ξ_n *is asymptotically non-Gaussian (this happens for certain graphs* $\{G_m, m \geq 1\}$*, see the next section for examples), then* S_n *is asymptotically non-Gaussian. Note that the restriction* $r \neq 0$ *is not stringent, as it includes all distributions* F *(in Condition 6.1) with a non-atomic part. Indeed, if* $W \sim F$ *has a non-atomic part, then* W *has a non-atomic part on either* $(\mathbb{E}[W], \infty)$ *or* $(-\infty, \mathbb{E}[W])$*. Without loss of generality, assume that the non-atomic part is on* $(\mathbb{E}[W], \infty)$*, then we can find an integer* $\ell \geq 2$ *and a Borel set* A_0 *such that* $\mathbb{P}(W \in A) = \ell^{-1}$ *with* $A = (\mathbb{E}[W], \infty) \cap A_0$*. By construction, this yields*

$$\mathbb{E}[W|W \in A] > \mathbb{E}[W] = \mathbb{E}[W\mathbb{1}_A] + \mathbb{E}[W\mathbb{1}_{A^c}] = \mathbb{E}[W|W \in A]\,\ell^{-1} + \mathbb{E}[W|W \in A^c]\,(1-\ell^{-1}), \tag{6.3.4}$$

so that $\mathbb{E}[W|W \in A] > \mathbb{E}[W|W \in A^c]$*. The restriction* $r \neq 0$ *also includes almost all discrete distributions with at least one weight of the form* ℓ^{-1}*; see Remark 5.2 in the previous chapter for a formal argument. Also, note that, depending on* F*, many choices for* A *(with possibly different values of* ℓ*) could be available.*

Remark 6.4. *If the margin* F *satisfies Condition 6.1, and if* $r = 0$ *(i.e.,* $\mu_U = \mu_V$*) or* ξ_n *is asymptotically Gaussian, then our construction provides new triplewise independent (but not mutually independent) sequences which do satisfy a CLT (regardless of which graphs* $\{G_m, m \geq 1\}$ *are used).*

Proof of Theorem 6.1. We prove (6.3.3) by obtaining the limit of the characteristic function of S_n, and then by invoking Lévy's continuity theorem. Namely, we show that, for all $t \in \mathbb{R}$,

$$\varphi_{S_n}(t) \overset{m\to\infty}{\longrightarrow} \varphi_{\sqrt{1-r^2}Z}(t) \cdot \varphi_{rY}(t). \tag{6.3.5}$$

Recall the notation defined in (6.2.6) and let

$$\widetilde{U}_k := \frac{\sigma_U}{\sigma} \cdot \frac{U_k - \mu_U}{\sigma_U} \quad \text{and} \quad \widetilde{V}_k := \frac{\sigma_V}{\sigma} \cdot \frac{V_k - \mu_V}{\sigma_V}, \tag{6.3.6}$$

then we can write

$$\begin{aligned} S_n &= \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{\substack{k=1\\D_k=0}}^n U_k + \sum_{\substack{k=1\\D_k=1}}^n V_k - n\mu}{\sigma\sqrt{n}} \\ &= \frac{1}{\sqrt{n}}\left(\frac{(n-\Xi_n)\mu_U + \Xi_n\mu_V - n\mu}{\sigma} + \sum_{\substack{k=1\\D_k=0}}^n \frac{U_k-\mu_U}{\sigma} + \sum_{\substack{k=1\\D_k=1}}^n \frac{V_k-\mu_V}{\sigma}\right) \\ &= \frac{1}{\sqrt{n}}\left(\frac{(\mu_V-\mu_U)}{\sigma}\left[\Xi_n - n\frac{(\mu-\mu_U)}{\mu_V-\mu_U}\right] + \sum_{\substack{k=1\\D_k=0}}^n \widetilde{U}_k + \sum_{\substack{k=1\\D_k=1}}^n \widetilde{V}_k\right) \\ &= \frac{1}{\sqrt{n}}\left(r\,\frac{\left(\Xi_n - n\ell^{-1}\right)}{\sqrt{\ell^{-1}(1-\ell^{-1})}} + \sum_{\substack{k=1\\D_k=0}}^n \widetilde{U}_k + \sum_{\substack{k=1\\D_k=1}}^n \widetilde{V}_k\right), \end{aligned} \tag{6.3.7}$$

since $\Xi_n = \#\{k : D_k = 1\}$, and we know that, from (6.2.9),

$$\frac{\mu-\mu_U}{\mu_V-\mu_U} = \frac{[(1-\ell^{-1})\mu_U + \ell^{-1}\mu_V] - \mu_U}{\mu_V-\mu_U} = \ell^{-1}. \tag{6.3.8}$$

With the notation $t_n := t/\sqrt{n}$, the mutual independence between the U_k's, the V_k's and $\boldsymbol{M} := \{M_j\}_{j=1}^{v(m)}$ yields, for all $t \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}\big[e^{\mathrm{i}tS_n}|\boldsymbol{M}\big] &= e^{\mathrm{i}tr\,\frac{(\Xi_n - n\ell^{-1})}{\sqrt{n\ell^{-1}(1-\ell^{-1})}}} \prod_{\substack{k=1\\D_k=0}}^n \mathbb{E}[e^{\mathrm{i}t_n\widetilde{U}_k}|\boldsymbol{M}] \prod_{\substack{k=1\\D_k=1}}^n \mathbb{E}[e^{\mathrm{i}t_n\widetilde{V}_k}|\boldsymbol{M}] \\ &= e^{\mathrm{i}tr\,\xi_n}\ [\varphi_{\widetilde{U}}(t_n)]^{n(1-\ell^{-1})}[\varphi_{\widetilde{V}}(t_n)]^{n\ell^{-1}}\left[\frac{\varphi_{\widetilde{V}}(t_n)}{\varphi_{\widetilde{U}}(t_n)}\right]^{\Xi_n - n\ell^{-1}} \\ &= e^{\mathrm{i}tr\,\xi_n}\cdot[\varphi_{\widetilde{U}}(t_n)]^{n(1-\ell^{-1})}[\varphi_{\widetilde{V}}(t_n)]^{n\ell^{-1}} \\ &\quad\cdot\left[\frac{[\varphi_{\widetilde{V}}(t_n)]^n\cdot e^{\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2}}{[\varphi_{\widetilde{U}}(t_n)]^n\cdot e^{\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2}}\right]^{\frac{\Xi_n-n\ell^{-1}}{n}}\cdot\left[\frac{e^{-\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2}}{e^{-\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2}}\right]^{\frac{\Xi_n-n\ell^{-1}}{n}}. \end{aligned} \tag{6.3.9}$$

(The reader should note that, for n large enough, the manipulations of exponents in the second and third equality above are valid because the highest powers of the complex numbers involved have their principal argument converging to 0. This stems from the fact that $\Xi_n \le n$, and the quantities $[\varphi_{\widetilde{V}}(t_n)]^n$ and $[\varphi_{\widetilde{V}}(t_n)]^n$ both converge to real exponentials as $n \to \infty$, by the CLT.) We now evaluate the four factors on the right-hand side of (6.3.9). For the first factor in (6.3.9), the continuous mapping theorem and (6.3.2) yield

$$e^{\mathrm{i}tr\,\xi_n} \xrightarrow{d} e^{\mathrm{i}trY}, \quad \text{as } m \to \infty. \tag{6.3.10}$$

For the second factor in (6.3.9), the classical CLT yields

$$[\varphi_{\widetilde{U}}(t_n)]^{n(1-\ell^{-1})}[\varphi_{\widetilde{V}}(t_n)]^{n\ell^{-1}} \overset{m\to\infty}{\longrightarrow} \exp\Big(-\frac{1}{2}\cdot(1-\ell^{-1})\frac{\sigma_U^2}{\sigma^2}t^2\Big)\exp\Big(-\frac{1}{2}\cdot\ell^{-1}\frac{\sigma_V^2}{\sigma^2}t^2\Big) = e^{-\frac{1}{2}(1-r^2)t^2}, \tag{6.3.11}$$

where in the last equality we used the fact that, from (6.2.9),

$$\sigma^2 = \mathbb{E}[X^2] - \mu^2 = (1-\ell^{-1})\sigma_U^2 + \ell^{-1}\sigma_V^2 + \ell^{-1}(1-\ell^{-1})(\mu_U-\mu_V)^2. \tag{6.3.12}$$

For the third factor in (6.3.9), the quantity inside the bracket converges to 1 by the CLT. Hence, the elementary bound

$$|e^z-1| \le |z| + \sum_{j=2}^{\infty}\frac{|z|^j}{2} \le |z| + \frac{|z|^2}{2(1-|z|)} \le \frac{1+\ell^{-1}}{2\ell^{-1}}|z|, \quad \text{for all } |z|\le 1-\ell^{-1}, \tag{6.3.13}$$

and the fact that $\big|\frac{\Xi_n - n\ell^{-1}}{n}\big| \le 1-\ell^{-1}$ yield, as $m\to\infty$,

$$\left|\left[\frac{[\varphi_{\widetilde{V}}(t_n)]^n\cdot e^{\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2}}{[\varphi_{\widetilde{U}}(t_n)]^n\cdot e^{\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2}}\right]^{\frac{\Xi_n-n\ell^{-1}}{n}} - 1\right| \overset{\text{a.s.}}{\le} \frac{1-\ell^{-2}}{2\ell^{-1}}\left|\mathrm{Log}\left[\frac{[\varphi_{\widetilde{V}}(t_n)]^n\cdot e^{\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2}}{[\varphi_{\widetilde{U}}(t_n)]^n\cdot e^{\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2}}\right]\right| \longrightarrow 0. \tag{6.3.14}$$

For the fourth factor in (6.3.9), we note that $\frac{\Xi_n-n\ell^{-1}}{n} \overset{\mathbb{P}}{\longrightarrow} 0$ (because of the Law of Large Numbers for pairwise independent r.v.s). Then, by the continuous mapping theorem,

$$\left[\frac{e^{-\frac{1}{2}\cdot\frac{\sigma_V^2}{\sigma^2}t^2}}{e^{-\frac{1}{2}\cdot\frac{\sigma_U^2}{\sigma^2}t^2}}\right]^{\frac{\Xi_n-n\ell^{-1}}{n}} \overset{\mathbb{P}}{\longrightarrow} 1, \quad \text{as } m\to\infty. \tag{6.3.15}$$

By combining (6.3.10), (6.3.11), (6.3.14) and (6.3.15), Slutsky's lemma implies, for all $t\in\mathbb{R}$,

$$\mathbb{E}\big[e^{\mathrm{i}tS_n}|\boldsymbol{M}\big] \overset{d}{\longrightarrow} e^{\mathrm{i}trY}e^{-\frac{1}{2}(1-r^2)t^2}, \quad \text{as } m\to\infty. \tag{6.3.16}$$

Since the sequence $\{|\mathbb{E}[e^{\mathrm{i}tS_n}|\boldsymbol{M}]|\}_{m\in\mathbb{N}}$ is uniformly integrable (it is bounded by 1), Theorem 25.12 in (Billingsley, 1995) shows that we also have the mean convergence

$$\mathbb{E}\big[\mathbb{E}\big[e^{\mathrm{i}tS_n}|\boldsymbol{M}\big]\big] \longrightarrow \mathbb{E}\big[e^{\mathrm{i}trY}\big]e^{-\frac{1}{2}(1-r^2)t^2}, \quad \text{as } m\to\infty, \tag{6.3.17}$$

which proves (6.3.5). The conclusion follows. □

6.4 Examples

In Theorem 6.1, whether the standardised sample mean S_n is asymptotically Gaussian depends on the 'connectivity' of the chosen graphs $\{G_m, m \geq 1\}$. In particular, it appears that having graphs of bounded diameter is a necessary (albeit not sufficient) condition for S_n to be asymptotically non-Gaussian. To make this point explicit, we present two specific examples for which we obtain the (non-Gaussian) asymptotic distribution of ξ_n (via Theorem 6.1, this also provides the asymptotic distribution of S_n). We then present two more examples where the limiting distribution *is* Gaussian.

6.4.1 First example

Theorem 6.2. *Let $\{G_m, m \geq 1\}$ be the sequence of bipartite graphs $\{K_{m,m}\}_{m\geq 1}$ described above Figure 6.1, and consider the construction from Section 6.2 where i.i.d. discrete uniforms $M_1, \ldots, M_{2m}$ are assigned to the vertices of G_m. That is, $M_1, \ldots, M_m$ are assigned to the m vertices of set 1, and $M_{m+1}, \ldots, M_{2m}$ to the m vertices of set 2. Then,*

$$\xi_n \xrightarrow{d} \frac{\xi}{\sqrt{\ell - 1}}, \quad \text{as } m \to \infty \text{ (and thus as } n \to \infty), \tag{6.4.1}$$

where $\xi \sim \mathrm{VG}(\ell - 1, 0, 1, 0)$, and VG denotes the variance-gamma distribution (see Definition 6.1).

Remark 6.5. *Because a standardised $\mathrm{VG}(\ell-1, 0, 1, 0)$ distribution converges to a standard Gaussian as ℓ tends to infinity, we see that, in Theorem 6.1, $S^{(\ell)} \xrightarrow{d} N(0, 1)$ as $\ell \to \infty$.*

Proof. First, note that $v(m) = 2m$ and $n = m^2$. Define, for $i \in \{1, 2, \ldots, \ell\}$,

$$N_i^{(1)} = N_i^{(1)}(m), \text{ the number of } M_j\text{'s equal to } i \text{ within the sample } \{M_j\}_{j=1}^{m},$$
$$N_i^{(2)} = N_i^{(2)}(m), \text{ the number of } M_j\text{'s equal to } i \text{ within the sample } \{M_j\}_{j=m+1}^{2m}.$$

Then, $\boldsymbol{N}^{(j)} := (N_1^{(j)}, \ldots, N_\ell^{(j)}) \sim \mathrm{Multinomial}(m, (\ell^{-1}, \ldots, \ell^{-1}))$ for $j \in \{1, 2\}$, and $\boldsymbol{N}^{(1)}$ and $\boldsymbol{N}^{(2)}$ are independent. Importantly, if $\boldsymbol{N}^{(1)}$ and $\boldsymbol{N}^{(2)}$ are known, then the number of 1's in the sequence $\{D_k, 1 \leq k \leq n\}$, denoted by Ξ_n throughout, can be deduced from simple calculations as

$$\Xi_n = \sum_{i=1}^{\ell} N_i^{(1)} N_i^{(2)} = \sum_{i=1}^{\ell-1} N_i^{(1)} N_i^{(2)} + \big(m - \sum_{i=1}^{\ell-1} N_i^{(1)}\big)\big(m - \sum_{i'=1}^{\ell-1} N_{i'}^{(2)}\big)$$

$$
\begin{aligned}
&= \sum_{i=1}^{\ell-1} N_i^{(1)} N_i^{(2)} + \sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1} N_i^{(1)} N_{i'}^{(2)} - m\sum_{i=1}^{\ell-1} N_i^{(1)} - m\sum_{i'=1}^{\ell-1} N_{i'}^{(2)} + m^2 \\
&= \sum_{i=1}^{\ell-1} (N_i^{(1)} - m\ell^{-1})(N_i^{(2)} - m\ell^{-1}) + \sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1} (N_i^{(1)} - m\ell^{-1})(N_{i'}^{(2)} - m\ell^{-1}) \\
&\quad + m\ell^{-1}\sum_{i=1}^{\ell-1} N_i^{(1)} + m\ell^{-1}\sum_{i=1}^{\ell-1} N_i^{(2)} + (\ell-1)m\ell^{-1}\sum_{i=1}^{\ell-1} N_i^{(1)} + (\ell-1)m\ell^{-1}\sum_{i=1}^{\ell-1} N_i^{(2)} \\
&\quad - m\sum_{i=1}^{\ell-1} N_i^{(1)} - m\sum_{i=1}^{\ell-1} N_i^{(2)} - (\ell-1)m^2\ell^{-2} - (\ell-1)^2 m^2\ell^{-2} + m^2 \\
&= \sum_{i=1}^{\ell-1} (N_i^{(1)} - m\ell^{-1})(N_i^{(2)} - m\ell^{-1}) + \sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1} (N_i^{(1)} - m\ell^{-1})(N_{i'}^{(2)} - m\ell^{-1}) + m^2\ell^{-1} \\
&= m\ell^{-1}\sum_{i=1}^{\ell-1}\sum_{i'=1}^{\ell-1} \Big(\frac{1}{\ell^{-1}}\mathbb{1}_{\{i=i'\}} + \frac{1}{\ell^{-1}}\Big) \frac{(N_i^{(1)} - m\ell^{-1})}{\sqrt{m}} \frac{(N_i^{(2)} - m\ell^{-1})}{\sqrt{m}} + m^2\ell^{-1}, \qquad (6.4.2)
\end{aligned}
$$

where $\mathbb{1}_B$ denotes the indicator function on the set B. It is well known that the covariance matrix for the first $\ell-1$ components of a Multinomial$(m, (p_1, p_2, \ldots, p_\ell))$ vector is $m\Sigma$ where $\Sigma_{i,i'} = p_i\mathbb{1}_{\{i=i'\}} - p_i p_{i'}$, for $1 \le i, i' \le \ell-1$, and also that $(\Sigma^{-1})_{i,i'} = p_i^{-1}\mathbb{1}_{\{i=i'\}} + p_\ell^{-1},\ 1 \le i, i' \le \ell-1$, see Tanabe and Sagae (1992, eq. 21). If $\Sigma = LL^\top$ is the Cholesky decomposition of Σ when $p_i = \ell^{-1}$ for all i, and $\boldsymbol{Y}_1 := (N_i^{(1)} - m\ell^{-1})_{i=1}^{\ell-1}$ and $\boldsymbol{Y}_2 := (N_i^{(2)} - m\ell^{-1})_{i=1}^{\ell-1}$, then we have

$$
\begin{aligned}
\Xi_n - m^2\ell^{-1} &= m\ell^{-1}\boldsymbol{Y}_1^\top (m\Sigma)^{-1}\boldsymbol{Y}_2 \\
&= m\ell^{-1}(m^{-1/2}L^{-1}\boldsymbol{Y}_1)^\top (m^{-1/2}L^{-1}\boldsymbol{Y}_2).
\end{aligned} \qquad (6.4.3)
$$

By the classical multivariate CLT and Definition 6.1 in Appendix 6.A, we get the result. □

Next, we illustrate what the asymptotic distribution of S_n (the standardised sample mean) looks like in this example. By Theorem 6.1, S_n converges in law to a r.v.

$$
S^{(\ell)} \stackrel{d}{=} \sqrt{1-r^2}Z + r\,\frac{\xi}{\sqrt{\ell-1}}, \qquad (6.4.4)
$$

where the r.v.s $Z \sim N(0,1)$ and $\xi \sim \mathrm{VG}(\ell-1, 0, 1, 0)$ (see Definition 6.1 in Appendix 6.A) are independent.

For a fixed $\ell \ge 2$, the distribution of $S^{(\ell)}$ has only one parameter, r (defined in Theorem 6.1), which depends on the margin F (through the quantities A, μ_U, μ_V and σ). Note that $0 \le r^2 \le 1$, and that the critical points $r^2 = 0, 1$ are reachable for certain choices of F (for specific examples, see Example 5.1 and Appendix 5.B from Chapter 5).

Hence, when $\ell \geq 2$ is fixed, r completely determines the shape of $S^{(\ell)}$; r close to 0 means that $S^{(\ell)}$ is close to a standard Gaussian, while r close to ± 1 means that $S^{(\ell)}$ is close to a standardised VG$(\ell - 1, 0, 1, 0)$. Figure 6.2 (where $\ell = 2$ and r varies) illustrates this shift from a Gaussian distribution towards a VG$(\ell - 1, 0, 1, 0)$ distribution. On the other hand, regardless of r, if ℓ increases then $S^{(\ell)}$ gets closer to a $N(0, 1)$. This is illustrated in Figure 6.3 (where $r = 0.99$ and ℓ varies). It is clear from these figures that triplewise independence can be a very poor substitute to mutual independence as an assumption in the classical CLT.

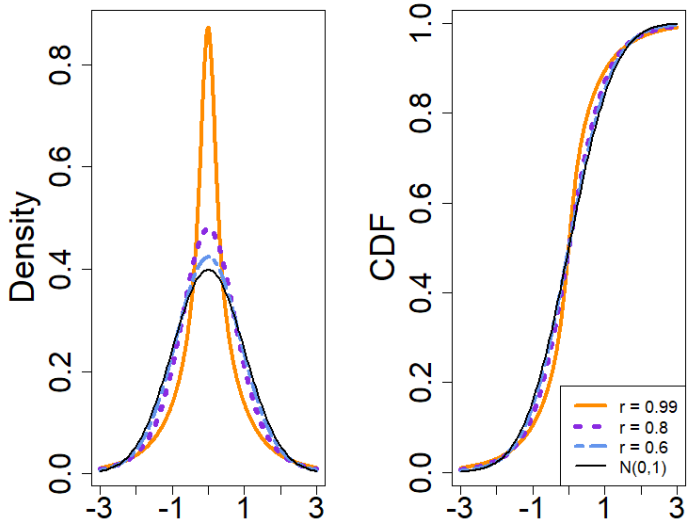

Figure 6.2: Density (left) and CDF (right) of $S^{(\ell)}$ for fixed $\ell = 2$ and varying r ($r = 0.6, 0.8, 0.99$), compared to those of a $N(0, 1)$. This illustrates that the CLT can 'fail' substantially under triplewise independence.

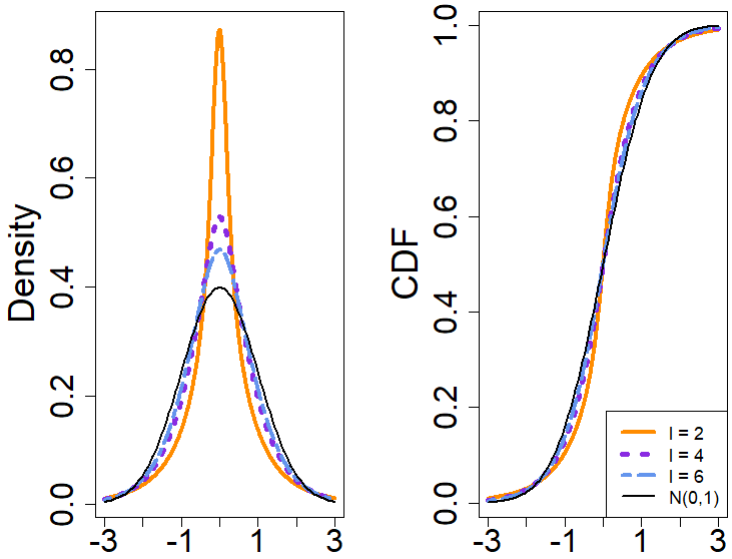

Figure 6.3: Density (left) and CDF (right) of $S^{(\ell)}$ for fixed $r = 0.99$ and varying ℓ ($\ell = 2, 4, 6$), compared to those of a $N(0, 1)$. This illustrates that $S^{(\ell)}$ converges to a $N(0, 1)$ as ℓ grows.

Lastly, the first moments of $S^{(\ell)}$ (obtained with simple calculations in `Mathematica`) are

$$\mathbb{E}[S^{(\ell)}] = 0, \quad \mathbb{E}[(S^{(\ell)})^2] = 1, \quad \mathbb{E}[(S^{(\ell)})^3] = 0 \quad \text{and} \quad \mathbb{E}[(S^{(\ell)})^4] = \frac{6r^4}{\ell - 1} + 3. \tag{6.4.5}$$

Thus, an upper bound on the kurtosis of $S^{(\ell)}$ is $6/(\ell-1)+3$, which implies that the limiting r.v. $S^{(\ell)}$ can be substantially more heavy-tailed than the standard Gaussian distribution (which is also seen in Figure 6.2).

6.4.2 Second example

Consider the sequence of graphs $\{G_m, m \geq 1\}$ as displayed in Figure 6.4 for $m = 6$, where $M_0, M_1, M_2, \ldots, M_{m+1}$ is a sequence of i.i.d. Bernoulli$(1/2)$ r.v.s assigned to the vertices. For each m, the graph G_m has $v(m) = m + 2$ vertices and $n = 2m$ edges. Every vertex in the set $\{M_1, M_2, \ldots, M_m\}$ (in the middle) is linked by an edge to the adjacent vertices M_0 (on the left) and M_{m+1} (on the right). This sequence of graphs yields Theorem 6.3.

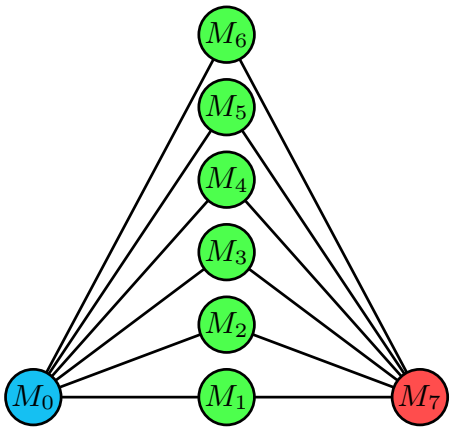

Figure 6.4: Illustration of the graph G_6 in our second example.

Theorem 6.3. *Let $\{G_m, m \geq 1\}$ be the sequence of graphs described above and consider the construction from Section 6.2 where Condition 6.1 is satisfied with $\ell = 2$. Then,*

$$\xi_n \xrightarrow{d} \sqrt{2} I \cdot Z, \quad \text{as } m \to \infty \text{ (and thus as } n \to \infty), \tag{6.4.6}$$

where the random variables $I \sim$ Bernoulli$(1/2)$ and $Z \sim N(0,1)$ are independent.

Proof. If $I \sim$ Bernoulli$(1/2)$ and $B \sim$ Binomial$(m, 1/2)$ are independent r.v.s, then Ξ_n satisfies

$$\Xi_n \stackrel{d}{=} I \cdot 2B + (1 - I) \cdot m. \tag{6.4.7}$$

Indeed, if the Bernoulli r.v.s M_0 and M_{m+1} are equal (this is represented by $I = 1$ in (6.4.7),

which has probability 1/2), then for every vertex $M_1, M_2, \ldots, M_m$ in the middle, the sum of the 1's on the two adjacent edges will be 2 with probability 1/2 and 0 with probability 1/2. By the independence of the Bernoulli r.v.s $M_1, M_2, \ldots, M_m$, we can thus represent the sum of the "m sums of 1's" that we just described by $2B$ where $B \sim \text{Binomial}(m, 1/2)$. Similarly, if the Bernoulli r.v.s M_0 and M_{m+1} are not equal (this is represented by $I = 0$ in (6.4.7), which has probability 1/2), then for every vertex $M_1, M_2, \ldots, M_m$ in the middle, the sum of the 1's on the two adjacent edges will always be 1 (either the left edge is 1 and the right edge is 0, or vice-versa, depending on whether $(M_0 = 1, M_{m+1} = 0)$ or $(M_0 = 0, M_{m+1} = 1)$). Since there are m vertices in the middle when $I = 0$, the total sum of the 1's on the edges is always m. By combining the cases $I = 1$ and $I = 0$, we get the representation (6.4.7).

Lastly, here $\mathbb{E}[\Xi_n] = m$ and $\mathbb{V}\text{ar}(\Xi_n) = \frac{m}{2}$ so that, by Lévy's continuity theorem,

$$\xi_n = \frac{\Xi_n - m}{\sqrt{\frac{m}{2}}} = \sqrt{2} I \cdot \frac{B - m/2}{\sqrt{\frac{m}{4}}} \xrightarrow{d} \sqrt{2} I \cdot Z, \quad \text{where } Z \sim N(0,1). \tag{6.4.8}$$

This ends the proof. □

Remark 6.6. *By Theorem 6.1, S_n converges in law to a random variable:*

$$S := \sqrt{1 - r^2} Z_1 + r\sqrt{2} I Z_2, \tag{6.4.9}$$

where the random variables $Z_1, Z_2 \sim N(0,1)$ and $I \sim \text{Bernoulli}(1/2)$ are all independent, and $r := \frac{\mu_V - \mu_U}{2\sigma}$. Simple calculations then yield

$$\mathbb{E}[S] = 0, \quad \mathbb{E}[S^2] = 1, \quad \mathbb{E}[S^3] = 0 \quad \textit{and} \quad \mathbb{E}[S^4] = 3(1 + r^4), \tag{6.4.10}$$

so that S in (6.4.9) is always heavier tailed than a standard Gaussian r.v. (provided $r \neq 0$, which is not a stringent requirement as seen in Remark 6.3).

6.4.3 Third example

In our construction, a CLT can hold. As a 'positive example', we consider here the sequence of m-hypercube graphs, which have $v(m) = 2^m$ vertices and $n = m2^{m-1}$ edges. Despite being 'highly connected', these graphs *do* induce a Gaussian limit for $\{S_n, n \geq 1\}$.

Theorem 6.4. *Let $\{G_m, m \geq 1\}$ be the sequence of m-hypercube graphs and consider the construction from Section 6.2 where Condition 6.1 is satisfied with $\ell = 2$. Then, ξ_n is*

asymptotically Gaussian as $m \to \infty$ (and thus as $n \to \infty$).

Proof. First, note that each vertex of G_m can be represented by a binary vector of m components. To be clear here, the hypercube graphs are all embedded in the same infinite dimensional hypercube graph, and the same goes for the Bernoulli r.v.s $M_1, M_2, \ldots, M_{2^m}$ assigned to the vertices. By definition of the m-hypercube graph, (i, j) is an edge if and only if i and j differ by only one binary component, which we write $i \sim j$ for short. In particular, we write $i \sim_d j$ if i and j differ only in the d-th binary component, where $1 \leq d \leq m$. With Ξ_n and $D_{i,j}$ defined as in (6.2.4) and (6.2.3), respectively, it will be useful here to work instead with the zero-mean r.v.s, $\widetilde{\Xi}_n$ and $\widetilde{D}_{i,j}$, defined as

$$\widetilde{\Xi}_m = 2\,\Xi_n - n = \sum_{i \sim j} \widetilde{D}_{i,j}, \quad \text{and} \quad \widetilde{D}_{i,j} = 2D_{i,j} - 1 = \begin{cases} 1, & \text{if } M_i = M_j, \\ -1, & \text{otherwise.} \end{cases} \tag{6.4.11}$$

We will prove below that $\widetilde{\Xi}_m$ is asymptotically Gaussian, which implies that ξ_n is as well. We have the following decomposition:

$$\widetilde{\Xi}_m = \sum_{d=1}^{m} \widetilde{\Xi}_m^{(d)}, \quad \text{where} \quad \widetilde{\Xi}_m^{(d)} := \sum_{i \sim_d j} \widetilde{D}_{i,j}. \tag{6.4.12}$$

Let $\mathcal{G}_d = \sigma(\widetilde{D}_{i,j} : i \sim_d j)$, and let $\mathcal{F}_m := \sigma(\cup_{d=1}^m \mathcal{G}_d)$ be the smallest σ-algebra containing the sets of all the $\mathcal{G}_d$'s, for $1 \leq d \leq m$. Then, $\mathbb{F} = \{\mathcal{F}_m, m \in \mathbb{N}_0\}$ is a filtration, where we define $\mathcal{F}_0 := \{\emptyset, \Omega\}$. We have the following preliminary result (we complete the proof of Theorem 6.4 right after).

Lemma 6.1. *If $\widetilde{\Xi}_0 := 0$, then for every $m \in \mathbb{N}_0$, the process $\{\widetilde{\Xi}_k / \sqrt{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}\}_{0 \leq k \leq m}$ is a zero-mean and bounded $\mathbb{F}$-martingale with differences $\widetilde{\Xi}_m^{(d)} / \sqrt{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}$, $1 \leq d \leq m$.*

Proof of Lemma 6.1. The process $\{\widetilde{\Xi}_m, m \in \mathbb{N}_0\}$ is trivially $\mathbb{F}$-adapted and integrable. To conclude that it is a $\mathbb{F}$-martingale, it is sufficient to show that

$$\mathbb{E}[\widetilde{\Xi}_m^{(k)} | \mathcal{F}_{k-1}] = 0, \quad \text{for all } 1 \leq k \leq m. \tag{6.4.13}$$

By symmetry of the construction, the case $k = 1$ is trivial (i.e., $\mathbb{E}[\widetilde{\Xi}_m^{(1)}] = 0$). Therefore, assume that $k \geq 2$. Consider any instance $\omega \in \Omega$ for the values of the Bernoulli r.v.s on the vertices of the m-hypercube such that $\sum_{d=1}^{k-1} \widetilde{\Xi}_m^{(d)}(\omega) = s$ and $\widetilde{\Xi}_m^{(k)}(\omega) = t$, where s, t are any specific integer values. For every such instance ω, there exists a 'conjugate'

instance $\overline{\omega}$ where $\sum_{d=1}^{k-1}\widetilde{\Xi}_m^{(d)}(\overline{\omega}) = s$ and $\widetilde{\Xi}_m^{(k)}(\overline{\omega}) = -t$. Indeed, take the configuration ω, then for every vertex that has its kth binary component equal to 1, flip the result of the Bernoulli r.v. (0 under ω becomes 1 under $\overline{\omega}$, and 1 under ω becomes 0 under $\overline{\omega}$). Since the Bernoulli r.v.s on the vertices are i.i.d., and the values 0 and 1 are equiprobable, note that $\mathbb{P}(\{\omega\}|\mathcal{F}_{k-1})(u) = \mathbb{P}(\{\overline{\omega}\}|\mathcal{F}_{k-1})(u)$ for all $u \in \Omega$ such that $\sum_{d=1}^{k-1}\widetilde{\Xi}_m^{(d)}(u) = s$. Therefore, for any summand of the form $\widetilde{\Xi}_m^{(k)}(\omega)\cdot\mathbb{P}(\{\omega\}|\mathcal{F}_{k-1})(u)$ in the calculation of $\mathbb{E}[\widetilde{\Xi}_m^{(k)}|\mathcal{F}_{k-1}](u)$, it will always be cancelled by $\widetilde{\Xi}_m^{(k)}(\overline{\omega})\cdot\mathbb{P}(\{\overline{\omega}\}|\mathcal{F}_{k-1})(u)$. Since we assumed nothing on s, we must conclude that $\mathbb{E}[\widetilde{\Xi}_m^{(k)}|\mathcal{F}_{k-1}] = 0$. □

Aside from Lemma 6.1, we also have the following three properties related to the increments of the process $\{\widetilde{\Xi}_k/\sqrt{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}\}_{0\le k\le m}$:

(a) $\max_{1\le d\le m}\frac{\widetilde{\Xi}_m^{(d)}}{\sqrt{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}} \xrightarrow{\mathbb{P}} 0$. Indeed, by a union bound and Markov's inequality with exponent 4, we have, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{1\le d\le m}\left|\frac{\widetilde{\Xi}_m^{(d)}}{\sqrt{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}}\right| > \varepsilon\right) \le m\cdot\mathbb{P}\left(\left|\frac{\widetilde{\Xi}_m^{(1)}}{\sqrt{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}}\right| > \varepsilon\right)$$
$$\le m\cdot\frac{\mathbb{E}\big[(\widetilde{\Xi}_m^{(1)})^4\big]}{\varepsilon^4 m^2\big(\mathbb{E}\big[(\widetilde{\Xi}_m^{(1)})^2\big]\big)^2} \le \frac{C}{\varepsilon^4 m} \overset{m\to\infty}{\longrightarrow} 0,$$

where $C > 0$ is a universal constant.

(b) By the weak Law of Large Numbers for weakly correlated r.v.s with finite variance, and the fact that $\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m) = m\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m^{(d)}) = m\,\mathbb{E}[(\widetilde{\Xi}_m^{(d)})^2]$ for all $1 \le d \le m$, we have

$$\sum_{d=1}^{m}\frac{(\widetilde{\Xi}_m^{(d)})^2}{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)} = \frac{1}{m}\sum_{d=1}^{m}\frac{(\widetilde{\Xi}_m^{(d)})^2}{\mathbb{E}[(\widetilde{\Xi}_m^{(d)})^2]} \xrightarrow{\mathbb{P}} 1, \quad \text{as } m\to\infty.$$

(c) $\mathbb{E}\big[\max_{1\le d\le m}\frac{(\widetilde{\Xi}_m^{(d)})^2}{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}\big]$ is bounded in m. Indeed,

$$\mathbb{E}\left[\max_{1\le d\le m}\frac{(\widetilde{\Xi}_m^{(d)})^2}{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}\right] \le \frac{\mathbb{E}\big[\sum_{d=1}^{m}(\widetilde{\Xi}_m^{(d)})^2\big]}{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)} = \frac{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)} = 1 < \infty.$$

By Lemma 6.1, (a), (b), (c), and the CLT for martingale arrays (Hall and Heyde, 1980, Theorem 3.2), we conclude that

$$\frac{\widetilde{\Xi}_m}{\sqrt{\mathbb{V}\mathrm{ar}(\widetilde{\Xi}_m)}} \xrightarrow{d} N(0,1), \quad \text{as } m\to\infty.^{1} \tag{6.4.14}$$

[1] Approximately four days after we came up with this proof, Yuval Peres provided an interesting and

This ends the proof of Theorem 6.4. □

6.4.4 Fourth example

Figure 6.5 shows a graph which can easily be made arbitrarily large (displayed here for $m = 6$). We have the following theorem, which provides another 'positive example' where a CLT *is* verified.

Theorem 6.5. *Consider the construction from Section 6.2 where Condition 6.1 is satisfied with $\ell = 2$. Let the graphs G_m be defined as described in the caption of Figure 6.5. Then, ξ_n is asymptotically Gaussian.*

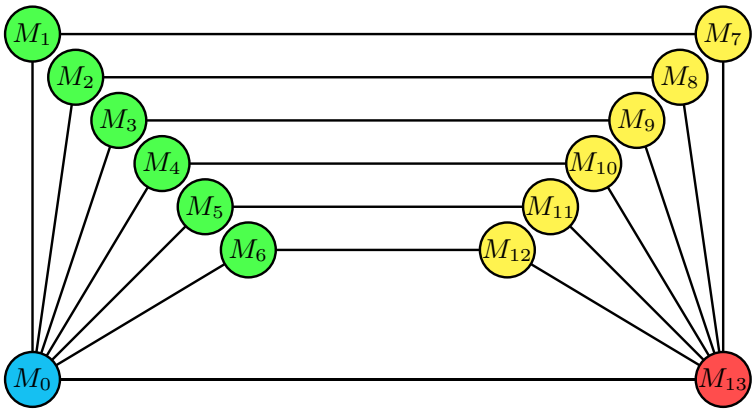

Figure 6.5: Illustration for $m = 6$ of the general construction where the vertex M_0 (on the bottom left) is linked by an edge to M_{2m+1} (on the bottom right), every vertex in the set $\{M_1, M_2, \ldots, M_m\}$ (on the top left) is linked by an edge to the vertex M_0 (on the bottom left), every vertex in the set $\{M_{m+1}, M_{m+2}, \ldots, M_{2m}\}$ (on the top right) is linked by an edge to the vertex M_{2m+1} (on the bottom right), and M_i (on the top left) is linked by an edge to M_{m+i} (on the top right) for all $1 \leq i \leq m$.

Proof. If $I \sim \text{Bernoulli}(1/2)$ and $B \sim \text{Binomial}(m, 1/4)$ are independent r.v.s, then the number of 1's on the edges satisfies

$$\Xi_n \stackrel{d}{=} I \cdot (1 + m + 2B) + (1 - I) \cdot 2(m - B). \tag{6.4.15}$$

Indeed, if the Bernoulli r.v.s M_0 and M_{2m+1} are equal in Figure 6.5 (this is represented by $I = 1$ in (6.4.15), which has probability $1/2$), then for each of the m 4-cycles in the graph, the sum of the 1's on the left, top and right edges will be 3 with probability $1/4$ and 1 with probability $3/4$. By the independence of the Bernoulli r.v.s on the top-left and top-right corners of the 4-cycles, we can thus represent the sum of the "m sums of 1's" that we just

completely different proof of (6.4.14) (not using martingales) in the following MathStackExchange post: https://math.stackexchange.com/questions/3993902/central-limit-theorem-for-dependent-bernoulli-random-variables-on-the-edges-of-a.

described by $m+2B$ where $B \sim \text{Binomial}(m, 1/4)$. We get $1+m+2B$ by including the '1' for the bottom edge (M_0, M_{2m+1}), which we only count once since this edge is common to all the 4-cycles. Similarly, if the Bernoulli r.v.s M_0 and M_{2m+1} are not equal in Figure 6.5 (this is represented by $I = 0$ in (6.4.15), which has probability $1/2$), then for each of the m 4-cycles in the graph, the sum of the 1's on the left, top and right edges will be 2 with probability $3/4$ and 0 with probability $1/4$. By the independence of the Bernoulli r.v.s on the top-left and top-right corners of the 4-cycles, we can thus represent the sum of the "m sums of 1's" that we just described by $2(m - B)$ since $m - B \sim \text{Binomial}(m, 3/4)$. By combining the cases $I = 1$ and $I = 0$, we get the representation (6.4.15).

Easy calculations then yield

$$\mathbb{E}[\Xi_n] = \frac{3m}{2} + \frac{1}{2} \quad \text{and} \quad \mathbb{V}\text{ar}(\Xi_n) = \frac{3m}{4} + \frac{1}{4}. \tag{6.4.16}$$

Hence, by Lévy's continuity theorem,

$$\xi_n = \frac{\Xi_n - \left(\frac{3m}{2} + \frac{1}{2}\right)}{\sqrt{\frac{3m}{4} + \frac{1}{4}}} = \frac{(2I-1) \cdot 2(B - \frac{m}{4}) + (I - \frac{1}{2})}{\sqrt{\frac{3m}{4} + \frac{1}{4}}} \xrightarrow{d} (2I-1) \cdot W \overset{d}{=} Z, \tag{6.4.17}$$

where $W, Z \sim N(0, 1)$. □

6.5 The general case $K \geq 4$

One can easily adapt the methodology presented in this paper to build new sequences of K-tuplewise independent random variables (with an arbitrary margin F). Indeed, all one needs to do is find a growing sequence of simple graphs of girth $K+1 \geq 5$ and then, as before, put i.i.d. discrete uniforms on the vertices and assign 1's to edges for which the r.v.s on the adjacent vertices are equal. A girth of $K+1$ guarantees K-tuplewise independence of the sequences hence created. An arbitrary margin F can be obtained as before by defining sequences $\{U_j, j \geq 1\}$ and $\{V_j, j \geq 1\}$ as in (6.2.7), and then creating the final sequence $\{X_j, j \geq 1\}$ as in (6.2.8).

Whether or not sequences created this way will satisfy a CLT is a different (and difficult) question. In Balbuena (2008), the author constructs explicitly an infinite collection of simple connected regular graphs of girth 6 and diameter 3, which we denote by G_q, where the index q runs over the possible prime powers. These graphs are obtained as the incidence graphs of projective planes of order $q = k-1$. For any given prime power q, the graph G_q is $(q+1)$-regular and has $2 \cdot (q^2+q+1)$ vertices. In particular, it is a $(k,6)$-cage because the number of vertices achieves the Moore (lower) bound, see, e.g., Biggs (1993, Chapter 23). This extremely uncommon sequence of graphs would be the perfect candidate for our construction to display a limiting non-Gaussian law for the normalised sum S_n. Indeed, in addition to having a minimal number of vertices, these graphs G_q also have a constant (and finite) diameter, which means that we do not have strong mixing of the binary random variables D_j assigned to the edges (strong mixing is the most common assumption for a CLT with dependent random variables, see, e.g., Rosenblatt (1956)). However, even in this context where the edges' dependence is, in a sense, maximised (because of the constant diameter and the minimal number of vertices), our simulations show that we cannot reject the hypothesis of a Gaussian limit for S. We applied the following normality tests with $q = 2^6$ (which corresponds to a sample of size $n = (q+1)(q^2+q+1) = 270{,}465$) and 5,000 samples:

test	Shapiro-Wilk	Anderson-Darling	Pearson chi-square
test statistic	0.9997	0.2993	67.9360
p-value	0.7148	0.5846	0.7602

For the interested reader, the code is provided in Appendix A.5.

Remark 6.7. *There seems to be a link between the fact that examples of asymptotic non-*

normality of $\{S_n, n \geq 1\}$ *exist for* $K \leq 3$ *(girth* $g \leq 4$*) but not for* $K \geq 4$ *(girth* $g \geq 5$*), and the fact that there exists growing sequences of regular graphs* G_m *of girth* $g \leq 4$ *where*

$$\liminf_{m\to\infty} \frac{\text{degree}(G_m)}{\#\ \textit{of vertices of}\ G_m} > 0, \tag{6.5.1}$$

(the $\liminf_{n\to\infty}$ *here is certainly a measure of the connectivity of the graphs* G_m*'s), whereas we always have*

$$\lim_{m\to\infty} \frac{\text{degree}(G_m)}{\#\ \textit{of vertices of}\ G_m} = 0,$$

for regular graphs of girth $g \geq 5$*, see, e.g., Biggs (1993, Proposition 23.1). This dichotomy in the statistics context (and its link to graph theory) seems to be a completely new and promising observation.*

Remark 6.8. *In contrast to the sequence of graphs in our first example (Section 6.4.1), the sequence of hypercube graphs in our third example (Section 6.4.3) do not satisfy* (6.5.1)*. The property* (6.5.1) *in a sense measures the connectivity of the graphs, and therefore the level of dependence between the r.v.s* $D_{i,j}$ *assigned to the edges in our construction. Since* (6.5.1) *cannot be satisfied for* $K \geq 4$ *when the underlying graphs are regular, the third example reinforces our intuition that, for* $K \geq 4$*, the sequence* $\{\xi_n, n \geq 1\}$ *(and thus* S_n*) will always converge to a Gaussian random variable.*

6.6 Conclusion

In this chapter, we provided a simple way to construct new sequences of dependent triplewise independent (identically distributed) r.v.s $\{X_j, j \geq 1\}$ having *any* distribution that satisfies Condition 6.1. Of course, such sequences (which are scarce in the current literature) are then also necessarily *pairwise independent*.

Our construction relies on using graphs of girth 4 (or more) and growing number of edges. We obtained an expression for the asymptotic distribution of the standardised sample mean S_n of those sequences, and saw that this distribution depends on the specific graph used (see Theorem 6.1). We then provided four different examples (different graphs) as special cases of our construction. For two of those examples, we showed that the limiting distribution of S_n is *not* Gaussian (while it *is* for the other two examples). In addition to the specific graph used, the extent to which the distribution of S_n departs from normality (when it does) also depends on the common margin of the X_j's (as was the case in the construction of the previous chapter). Those results add to the current literature on 'counterexamples to the CLT' and highlight why mutual independence is a crucial assumption we should not neglect.

We also noted that our methodology can be used to construct K-tuplewise independent sequences for arbitrary K, though it appears that for $K \geq 4$ such sequences are bound to satisfy a CLT. That said, such sequences are of independent interest. Indeed, they would prove useful to assess the performance of multivariate independence tests, many of which have been proposed in recent years. Lastly, the dichotomy that seems to exist between $K \leq 3$ and $K \geq 4$ for K-tuplewise independent sequences (constructed using our methodology) and its link to the dichotomy for the degree of regular graphs of girth $g \leq 4$ and $g \geq 5$ (see Remark 6.7) remains an interesting avenue to explore in the future.

6.A The variance-gamma distribution

Definition 6.1. *The variance-gamma distribution with parameters $\alpha > 0$, $\theta \in \mathbb{R}$, $s > 0$, $c \in \mathbb{R}$ has the density function*

$$f(x) := \frac{1}{s\sqrt{\pi}\Gamma(\alpha/2)} e^{\frac{\theta}{\alpha^2}(x-c)} \left(\frac{|x-c|}{2\sqrt{\theta^2+s^2}} \right)^{\frac{\alpha-1}{2}} K_{\frac{\alpha-1}{2}} \left(\frac{\sqrt{\theta^2+s^2}}{s^2} |x-c| \right), \quad x \in \mathbb{R}, \tag{6.A.1}$$

where K_ν is the modified Bessel function of the second kind of order ν. If a certain random variable X has this distribution, then we write $X \sim \mathrm{VG}(\alpha, \theta, s^2, c)$.

We have the following result, which is a consequence (for example) of Theorem 1 in (Gaunt, 2019).

Lemma 6.2. *Let $W_1, W_2, \ldots, W_n \overset{\text{i.i.d.}}{\sim} N(0, s^2)$ and $Z_1, Z_2, \ldots, Z_n \overset{\text{i.i.d.}}{\sim} N(0, s^2)$ be two independent sequences, then $Q_n := \sum_{i=1}^n W_i Z_i \sim \mathrm{VG}(n, 0, s^2, 0)$, following Definition 6.1, and the density function of Q_n is given by*

$$f_{Q_n}(x) = \frac{1}{s^2\sqrt{\pi}\Gamma(n/2)} \left(\frac{|x|}{2s^2} \right)^{\frac{n-1}{2}} K_{\frac{n-1}{2}} \left(\frac{|x|}{s^2} \right), \quad x \in \mathbb{R}. \tag{6.A.2}$$

It is easy to verify that the characteristic function of Q_n is given by

$$\varphi_{Q_n}(t) = (1 + s^4 t^2)^{-n/2}, \quad t \in \mathbb{R}, \tag{6.A.3}$$

and the expectation and variance are given by

$$\mathbb{E}[Q_n] = 0 \quad \textit{and} \quad \mathbb{V}\mathrm{ar}[Q_n] = n s^4. \tag{6.A.4}$$

6.B Graph theory concepts

In this section, we summarise some fundamental concepts of graphs and graph theory. The following definitions are taken from Wilson (1996).

Definition 6.2 (Simple graph)**.** *A **simple graph**, denoted generically as G, consists of a non-empty finite set $V(G)$ of elements called **vertices**, and a finite set $E(G)$ of distinct unordered pairs of distinct elements of $V(G)$, called **edges**. An edge $\{v_1, v_2\}$ is said to join the vertices v_1 and v_2, and is often abbreviated v_1v_2. In a simple graph, at most one edge joins any given pair of vertices.*

Definition 6.3 (Adjacent and incident vertices)**.** *We say that two vertices v_1 and v_2 of a graph G are **adjacent** if there is an edge v_1v_2 joining them. We say the vertices v_1 and v_2 are **incident** with such an edge.*

Definition 6.4 (Adjacent edges)**.** *Two distinct edges of a graph G are adjacent if they have a vertex in common.*

Definition 6.5 (Degree)**.** *The **degree** of a vertex v of G is the number of edges incident with v (said otherwise, it is the number of edges having that vertex v as an end-point).*

Definition 6.6 (Regular graph)**.** *A graph in which each vertex has the same degree is called a **regular graph**. If each vertex has degree r, we say the graph is **regular of degree r**.*

Definition 6.7 (Walks, trails and paths)**.** *Given a graph G, a **walk** in G is a finite sequence of edges of the form $\{v_0v_1, v_1v_2, \ldots, v_{m-1}v_m\}$, in which any two consecutive edges are adjacent or identical. A walk in which all the edges are distinct is called a **trail**. If, in addition, the vertices $v_0, v_1, \ldots, v_m$ of the trail are distinct (except, possibly, $v_0 = v_m$), then the trail is called a **path**.*

Definition 6.8 (Cycle)**.** *A path $\{v_0v_1, v_1v_2, \ldots, v_{m-1}v_m\}$ containing at least one edge and for which $v_0 = v_m$ is called a **cycle**.*

Definition 6.9 (Length)**.** *The number of edges in a walk (or trail, or path) is called its **length**.*

Definition 6.10 (Girth)**.** *The **girth** of a graph is the length of its shortest cycle.*

The next three definitions are taken from https://mathworld.wolfram.com.

Definition 6.11 (Graph distance)**.** *The* ***distance*** $d(v_1, v_2)$ *between two vertices* v_1 *and* v_2 *of a finite graph is the minimum length of the paths connecting them.*

Definition 6.12 (Graph diameter)**.** *The* ***diameter*** *of a graph is the length* $\max_{u,v} d(u, v)$ *of the "longest shortest path" between any two graph vertices* (u, v)*, where* $d(u, v)$ *is a graph distance.*

Definition 6.13 (Cage)**.** *A* (r, g)*-cage graph is a* r*-regular graph of girth* g *having the minimum possible number of vertices.*

CHAPTER 7

CONCLUSION

7.1 Summary

Numerous actuarial models rely on the assumption that some random variables within them are *mutually independent*. However, and as we covered in Chapter 1 (especially in Section 1.3), in reality, many insurance risks are dependent. This is true at the 'micro' level (e.g., the remaining lifetimes of people forming a couple can be dependent), and also at the 'macro' level (e.g., aggregate insurance losses from whole business segments can be dependent). Hence, in recent years many efforts have been made in the literature to relax the 'independence assumption', in a variety of actuarial settings. This is a topic we surveyed in Chapter 2. We also noted that, in practice, the most common way modellers assess, visualise and model dependence is via *pairs* of variables. That said, we know that *pairwise* independence of a collection of random variables does not imply their *mutual* independence. It is not clear, however, how *materially different* pairwise and mutual independence are, nor how much this difference matters in actuarial applications. It was the main goal of this thesis to learn more about this difference, and we proceeded in a few different steps.

In Chapter 3, we were concerned with *visualising* the types of dependence that are possible under pairwise independence (noting this is not something we have seen elsewhere in the literature we surveyed). In Section 3.2, we provided several theoretical examples of PIBD variables, some taken from the literature (Examples 3.1, 3.2, 3.5), and some new ones (Examples 3.3, 3.4, 3.6). Using those examples, we showcased with simple 3D scatterplots that many different types of dependence (sometimes weak, sometimes strong) are possible under pairwise independence. In Section 3.3, we developed new visualisation tools that can be used to help identify this type of dependence. We saw that by simply adding a third variable as a colour on a 2D scatterplot, dependence patterns otherwise impossible to see can suddenly 'appear' (see, e.g., Figures 3.16 and 3.17). Furthermore, we developed a colour-coding methodology which helps highlight areas of a 3D scatterplot that have a higher (or lower) concentration of points than under mutual independence (we also calibrated our method with simulations, using a MSE criteria). This 3D visualisation tool was especially useful to detect weaker forms of dependence, as that from Example 3.5 (see Figures 3.25, 3.26 and 3.27).

In Chapter 4, we showed that many results useful in actuarial science and relying on mutual independence can fail severely under sole pairwise independence. In Section 4.2.1,

we covered results about sums of random variables. We saw that the behaviour of $S = X_1 + \cdots + X_n$ under pairwise independence of the X's can be substantially different to its behaviour under mutual independence of the X's. For example, we saw that for PIBD variables, the kurtosis of S can *increase* as n increases (see Example 4.2), which is the opposite of what we expect under mutual independence. As the sum S can behave differently under pairwise independence (compared to mutual independence), so can risk measures computed on S. This is a topic we covered in Section 4.2.2 (for the risk measures VaR and TVaR). We saw that under different dependence structures (and/or different levels), the VaR is sometimes greater, sometimes smaller under pairwise independence, compared to mutual independence. In Section 4.2.3, we presented a 'counterexample' to the Fisher-Tippett theorem for PIBD variables. We saw that, under sole pairwise independence, this important theorem is not necessarily verified. In Section 4.2.4, we detailed how the bootstrap method can fail, drastically, for PIBD variables. In Section 4.3, we proved that many popular dependence models do not allow for the possibility of PIBD variables (in particular, elliptical distributions and Archimedean copulas). To our knowledge, those proofs are new. Overall, the findings of Chapter 4 shed some new light on the potential dangers of assuming mutual independence when only pairwise independence holds.

Being interested in what happens 'in the limit' (when the sample size n grows to infinity), we then covered in more detail the question of the classical CLT under pairwise independence. A fundamental result of statistics, this theorem is also crucial to actuarial science (for instance, within the Individual Risk Model, the distribution of aggregated claims is frequently approximated by a Normal, and this is done via a CLT). In Chapter 5, we constructed a new sequence of PIBD identically distributed random variables $\{X_j, j \geq 1\}$ with an arbitrary distribution (satisfying mild conditions), and for which no CLT holds. We obtained explicitly the asymptotic distribution of the standardised mean S_n of this sequence. We found that the extent of the departure from normality depends on the initial common margin of the X_j's. This observation may appear counter-intuitive, since under mutual independence, regardless of the margins, S_n always converges to a Normal. The main construction from Chapter 5 also lead us to conduct an analysis of a parameter ('r') which arises in the asymptotic distribution of the sample mean of our sequence. We explained that r is an interesting measure of tail-heaviness, and we derived its value for several distributions used in actuarial science. We saw that r depends on the shape parameter of those distributions. We also compared r to two other measures of tail-heaviness,

and saw that it roughly 'agrees' with a measure proposed by Brys et al. (2006).

Lastly, in Chapter 6 we proposed a methodology to construct *triplewise independent* sequences of random variables (again with an arbitrary marginal distribution that can be chosen under mild conditions), noting such sequences are scarce in the literature. Using this methodology, we proposed four specific triplewise independent sequences, two for which a CLT does not hold, and two for which a CLT *does* hold. Our methodology being very general, it would allow others to create additional sequences (those proposed here are only *some* possible examples). Our method can also be used more generally to create new dependent K-tuplewise independent sequences (for an arbitrary $K \geq 2$). We believe such sequences are of independent interest. In fact, we believe many results obtained in this thesis constitute useful groundwork for future research. Especially, those results could be helpful to benchmark the performance of multivariate independence tests (many of which have been proposed in recent years, see Section 2.6 and 2.A). This possible future work is briefly outlined in the next section.

7.2 Possible future work

To detect dependence in a dataset, visualisation is useful but also has its limits. For example, the three PIBD variables of Example 3.5 displayed a rather weak form of dependence, which was hard to detect on conventional 2D or 3D scatterplots (though our colour-coding method highlighting areas of higher density helped in this regard). Because of Proposition 3.1, we also know that a PIBD structure 'mixed' with mutual independence yields a new PIBD structure, one for which the dependence is potentially very subtle. Furthermore, even when *some* dependence is observed, it can be difficult from visualisation alone to judge whether this dependence is *statistically significant*. And past three (or perhaps four) random variables, it becomes hard to visualise the *multivariate* dependence that links them (if any).

To answer more formally (and systematically) the question of whether some random variables $X_1, \ldots, X_d$ are dependent (and in the general case of d variables, for d possibly large), *statistical tests* can be used (a topic we reviewed in Section 2.6). But which test should one use in practice? This question is not trivial, as one can expect that different tests have different strengths (i.e., a given test might be better at detecting certain types of dependence). For example, some tests might be better at detecting dependence for variables that are pairwise (or, more generally, K-tuplewise) independent. To help compare different tests, we think the results from this thesis can be useful, and we explain here how so.

Note that in a comparative analysis of different tests, one usually uses simulations to estimate the *power* (see Definition 7.1 below) of a test, under different dependence scenarios.

Definition 7.1 (Power)**.** *The power of a hypothesis test under H_1 is the probability that this test rejects the null hypothesis H_0 when the alternative hypothesis H_1 is true.*

The power of a given test can be estimated as follows. We can generate B (for B a large number) samples $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}$ (as in 2.6.3) under a specific dependence structure, and for each sample perform the hypothesis test (where H_0 is mutual independence, see 2.6.1). The proportion of times H_0 is rejected (out of those B trials) is then an estimate of the power of the test. Of course, the power will depend on *which* H_1 is true (i.e., on the joint distribution of $\boldsymbol{X}$). We have not encountered a systematic review and power study of *multivariate* independence tests in the literature.

Now, an important motivation for developing multivariate independence tests (perhaps the *main* motivation) is the possibility of PIBD data. Indeed, if data is *pairwise dependent*, then bivariate tests (e.g., Székely et al. (2007), Heller et al. (2012)) should suffice, at least in principle, to detect the dependence.

Their ability to detect dependence for pairwise (or, in general, K-tuplewise) independent variables being an important feature of multivariate independence tests, we think important to know which tests perform best at this specific task. To help answer this question, the results from this thesis can be useful in many ways, which we outline below. We leave the implementation for future work.

- For the case of three random variables ($d = 3$, in 2.6.1), we presented many PIBD examples, featuring a diverse range of dependence structures (see the five examples from Section 3.2). Furthermore, via Proposition 3.1, we provided a simple way to 'weaken' the dependence in those examples (also preserving pairwise independence). This can be useful when comparing tests. Because all those examples feature continuous Uniform$[0, 1]$ margins, it is easy to modify them to obtain any margins (without altering the dependence). This is relevant, since the performance of independence tests can be affected by the choice of margins (this has been observed before, see, e.g. Genest and Rémillard (2004, Section 5), Boglioni Beaulieu (2016, Chapter 3)), which is something we want to be able to assess.

- More generally, for d an arbitrary number of variables, we also provided many different PIBD examples. Examples 4.2, 4.3 and 4.5 were taken from the existing literature, while in Chapter 5 we built a new arbitrarily large sequence (again with arbitrary margins) of PIBD variables, see (5.2.8). All those examples can be used as different dependence scenarios (different H_1) when comparing different tests. Again, the fact that our sequences have arbitrary margins would allow to assess if the performance of a test is affected by the margins used.

- In Chapter 6, we provided a general methodology to build new examples of K-tuplewise independent (but dependent) random variables, for arbitary $K \geq 2$. Indeed, and as noted in Section 6.5, what matters is the girth of the graph on which the construction relies (a girth of $K + 1$ guarantees K-tuplewise independence). Therefore, our methodology can be used to generate any number of different examples. In particular, in our methodology, there are two obvious ways to alter the type and strength of the dependence:

  - by altering the girth of the graph (heuristically, the larger the girth, the weaker the dependence);
  - by choosing graphs that are more or less 'connected' (heuristically, the less connected a graph is, the weaker the dependence).

  Again, and importantly, the marginal distribution in those examples can be chosen arbitrarily (under mild conditions, see Condition 6.1).

- Lastly, we note it is perhaps the case that a test performing well for pairwise independent data would not perform well for triplewise independent (or more generally, K-tuplewise independent data, $K > 2$). It would then be interesting to assess the power of different tests, as function of both K and the sample size n.

Remark 7.1. *In closing, we remark that many tests are capable of testing the independence between two random vectors, see, e.g., Székely et al. (2007); Gretton et al. (2007); Heller et al. (2012); Zhu et al. (2017). That is, those procedure test the independence between a random vector $\boldsymbol{X} \in \mathbb{R}^p$ and another random vector $\boldsymbol{Y} \in \mathbb{R}^q$ (p, q positive integers). In principle, one can use those procedures to test for mutual independence between $d > 2$ variables (i.e., H_0 in 2.6.1). For instance, and as pointed out in Pfister et al. (2018), H_0 in* (2.6.1) *is true if and only if for every $k \in \{2, \dots, d\}$,*

$$\boldsymbol{X}_k \coloneqq X_k \quad \text{is independent of} \quad \boldsymbol{Y}_k \coloneqq (X_1, \dots, X_{k-1}). \tag{7.2.1}$$

Hence, H_0 could in effect be tested via $d-1$ bivariate tests (i.e., for $k \in \{2, \dots, d\}$, test the independence of $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$). This obviously creates an additional computational burden (performing $d-1$ tests instead of just one), but it also relies on an arbitrary choice. Indeed, in (7.2.1)*, $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$ could be defined differently (i.e., the d random variables could be split differently between the two vectors $\boldsymbol{X}_k, \boldsymbol{Y}_k$). In addition, when combining several tests into one, one must apply a Bonferroni correction in order to maintain the overall level of the test. Such a correction can be overly conservative and reduce statistical power of the test (see, e.g., Nakagawa, 2004, for a discussion of this issue).*

BIBLIOGRAPHY

Aas, K., 2016. Pair-copula constructions for financial applications: A review. Econometrics 4 (4), 43.

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Mathematics and economics 44 (2), 182–198.

Abbasi, B., Guillen, M., 2013. Bootstrap control charts in monitoring value at risk in insurance. Expert Systems with Applications 40 (15), 6125–6135.

Abdallah, A., Boucher, J.-P., Cossette, H., 2015. Modeling dependence between loss triangles with hierarchical Archimedean copulas. Astin Bulletin 45 (03), 577–599.

Abdallah, A., Boucher, J.-P., Cossette, H., Trufin, J., 2016. Sarmanov family of bivariate distributions for multivariate loss reserving analysis. North American Actuarial Journal 20 (2), 184–200.

Abdullah, M. B., 1990. On a robust correlation coefficient. The Statistician, 455–460.

Acar, E. F., Genest, C., Nešlehová, J., 2012. Beyond simplified pair-copula constructions. Journal of Multivariate Analysis 110, 74–90.

Ahn, S., Kim, J. H., Ramaswami, V., 2012. A new class of models for heavy tailed distributions in finance and insurance risk. Insurance: Mathematics and Economics 51 (1), 43–52.

Alai, D. H., Landsman, Z., Sherris, M., 2013. Lifetime dependence modelling using a truncated multivariate gamma distribution. Insurance: Mathematics and Economics 52 (3), 542–549.

Alai, D. H., Landsman, Z., Sherris, M., 2015. A multivariate Tweedie lifetime model: Censoring and truncation. Insurance: Mathematics and Economics 64, 203–213.

Alai, D. H., Landsman, Z., Sherris, M., 2016a. Modelling lifetime dependence for older ages using a multivariate Pareto distribution. Insurance: Mathematics and Economics 70, 272–285.

Alai, D. H., Landsman, Z., Sherris, M., 2016b. Multivariate Tweedie lifetimes: The impact of dependence. Scandinavian Actuarial Journal 2016 (8), 692–712.

Albrecher, H., Beirlant, J., Teugels, J. L., 2017. Reinsurance: actuarial and statistical aspects. John Wiley & Sons.

Albrecher, H., Boxma, O. J., 2004. A ruin model with dependence between claim sizes and claim intervals. Insurance: Mathematics and Economics 35 (2), 245–254.

Albrecher, H., Constantinescu, C., Loisel, S., 2011. Explicit ruin formulas for models with dependence among risks. Insurance: Mathematics and Economics 48 (2), 265–270.

Albrecher, H., Teugels, J. L., 2006. Exponential behavior in the presence of dependence in risk theory. Journal of Applied Probability 43 (1), 257–273.

Alink, S., Löwe, M., Wüthrich, M. V., 2004. Diversification of aggregate dependent risks. Insurance: Mathematics and Economics 35 (1), 77–95.

Alink, S., Löwe, M., Wüthrich, M. V., 2005. Analysis of the expected shortfall of aggregate dependent risks. Astin Bulletin 35 (1), 25–43.

Altman, D. G., Bland, J. M., 1995. Statistics notes: The normal distribution. BMJ 310 (6975), 298.

Anderson, C. J., 2010. Central Limit Theorem. In: The Corsini Encyclopedia of Psychology. American Cancer Society, pp. 1–2.

Araiza Iturria, C. A., Godin, F., Mailhot, M., 2021. Tweedie double GLM loss triangles with dependence within and across business lines. European Actuarial Journal 11 (2), 619–653.

Avanzi, B., Boglioni Beaulieu, G., Lafaye de Micheaux, P., Ouimet, F., Wong, B., 2021. A counterexample to the existence of a general central limit theorem for pairwise independent identically distributed random variables. J. Math. Anal. Appl. 499 (1), 124982.

Avanzi, B., Cassar, L. C., Wong, B., 2011. Modelling dependence in insurance claims processes with Lévy copulas. Astin Bulletin 41 (2), 575–609.

Avanzi, B., Tao, J., Wong, B., Yang, X., 2016a. Capturing non-exchangeable dependence in multivariate loss processes with nested Archimedean Lévy copulas. Annals of Actuarial Science 10 (1), 87–117.

Avanzi, B., Taylor, G., Vu, P. A., Wong, B., 2016b. Stochastic loss reserving with dependence: A flexible multivariate Tweedie approach. Insurance: Mathematics and Economics 71, 63–78.

Avanzi, B., Taylor, G., Wong, B., 2016c. Correlations between insurance lines of business: An illusion or a real phenomenon? Some methodological considerations. Astin Bulletin 46 (2), 225–263.

Avanzi, B., Taylor, G., Wong, B., 2018. Common shock models for claim arrays. Astin Bulletin 48 (3), 1109–1136.

Avanzi, B., Wong, B., Yang, X., 2016d. A micro-level claim count model with overdispersion and reporting delays. Insurance: Mathematics and Economics 71 (3), 1–14.

Avram, F., Badescu, A., Pistorius, M. R., Rabehasaina, L., 2016. On a class of dependent Sparre Andersen risk models and a bailout application. Insurance: Mathematics and Economics 71, 27–39.

Badescu, A. L., Cheung, E. C., Landriault, D., 2009. Dependent risk models with bivariate phase-type distributions. Journal of Applied Probability 46 (1), 113–131.

Badounas, I., Pitselis, G., 2020. Loss reserving estimation with correlated run-off triangles in a quantile longitudinal model. Risks 8 (1), 14.

Bagui, S. C., Bhaumik, D. K., Mehra, K. L., 2013. A few counter examples useful in teaching central limit theorems. The American Statistician 67 (1), 49–56.

Balbuena, C., 2008. Incidence matrices of projective planes and of some regular bipartite graphs of girth 6 with few vertices. SIAM J. Discrete Math. 22 (4), 1351–1363.

Barbe, P., Fougeres, A.-L., Genest, C., 2006. On the tail behavior of sums of dependent risks. Astin Bulletin 36 (02), 361–373.

Bartholomew, D. J., Knott, M., Moustaki, I., 2011. Latent variable models and factor analysis: A unified approach, 3rd Edition. John Wiley & Sons.

Bartram, S. M., Wang, Y.-H., 2015. European financial market dependence: An industry analysis. Journal of Banking & Finance 59, 146–163.

Bateup, R., Reed, I., 2001. Research and data analysis relevant to the development of standards and guidelines on liability valuation for general insurance. Tech. rep., Towers Perrin: The Institute of Actuaries of Australia and Tilinghast.

Bäuerle, N., Müller, A., 1998. Modeling and comparing dependencies in multivariate risk portfolios. Astin Bulletin 28 (01), 59–76.

Bedford, T., Cooke, R. M., 2002. Vines–a new graphical model for dependent random variables. The Annals of Statistics 30 (4), 1031–1068.

Beran, R., Bilodeau, M., Lafaye de Micheaux, P., 2007. Nonparametric tests of independence between random vectors. Journal of Multivariate Analysis 98 (9), 1805–1824.

Berlinet, A., Thomas-Agnan, C., 2004. Reproducing kernel Hilbert spaces in probability and statistics. Kluwer Academic Publishers, Boston, MA.

Bermúdez, L., Guillén, M., Karlis, D., 2018. Allowing for time and cross dependence assumptions between claim counts in ratemaking models. Insurance: Mathematics and Economics 83, 161–169.

Bernard, C., Denuit, M., Vanduffel, S., 2016. Measuring portfolio risk under partial dependence information. Journal of Risk and Insurance.

Bernard, C., Jiang, X., Wang, R., 2014. Risk aggregation with dependence uncertainty. Insurance: Mathematics and Economics 54, 93–108.

Bernard, C., Kazzi, R., Vanduffel, S., 2020. Range Value-at-Risk bounds for unimodal distributions under partial information. Insurance: Mathematics and Economics 94, 9–24.

Bernard, C., Rüschendorf, L., Vanduffel, S., 2017a. Value-at-risk bounds with variance constraints. Journal of Risk and Insurance 84 (3), 923–959.

Bernard, C., Rüschendorf, L., Vanduffel, S., Wang, R., 2017b. Risk bounds for factor models. Finance and Stochastics 21 (3), 631–659.

Bernard, C., Vanduffel, S., 2015. A new approach to assessing model risk in high dimensions. Journal of Banking & Finance 58, 166–178.

Bernšteĭn, S. N., 1927. Theory of probability (in Russian). Moscow.

Bi, J., Liang, Z., Xu, F., 2016. Optimal mean–variance investment and reinsurance problems for the risk model with common shock dependence. Insurance: Mathematics and Economics 70, 245–258.

Biggs, N., 1993. Algebraic graph theory, 2nd Edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge.

Billingsley, P., 1995. Probability and measure, 3rd Edition. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

Bilodeau, M., Nangue, A. G., 2017. Tests of mutual or serial independence of random vectors with applications. The Journal of Machine Learning Research 18 (1), 2518–2557.

Bladt, M., Nielsen, B. F., Peralta, O., 2019. Parisian types of ruin probabilities for a class of dependent risk-reserve processes. Scandinavian Actuarial Journal 2019 (1), 32–61.

Block, H. W., Basu, A., 1974. A continuous, bivariate exponential extension. Journal of the American Statistical Association 69 (348), 1031–1037.

Boglioni Beaulieu, G., 2016. A Consistent Test of Independence Between Random Vectors. Master's thesis, Université de Montréal.

Boglioni Beaulieu, G., Lafaye de Micheaux, P., Ouimet, F., 2021. Counterexamples to the classical central limit theorem for triplewise independent random variables having a common arbitrary margin. Depend. Model. 9 (1), 424–438.

Bølviken, E., Guillen, M., 2017. Risk aggregation in Solvency II through recursive log-normals. Insurance: Mathematics and Economics 73, 20–26.

Bonett, D. G., Seier, E., 2002. A test of normality with high uniform power. Computational statistics & data analysis 40 (3), 435–445.

Böttcher, B., Keller-Ressel, M., Schilling, R. L., 2019. Distance multivariance: new dependence measures for random vectors. Ann. Statist. 47 (5), 2757–2789.

Boudreault, M., Cossette, H., Landriault, D., Marceau, E., 2006. On a risk model with dependence between interclaim arrivals and claim sizes. Scandinavian Actuarial Journal 2006 (5), 265–285.

Boudreault, M., Cossette, H., Marceau, É., 2014. Risk models with dependence between claim occurrences and severities for Atlantic hurricanes. Insurance: Mathematics and Economics 54, 123–132.

Bradley, R. C., 1989. A stationary, pairwise independent, absolutely regular sequence for which the central limit theorem fails. Probab. Theory Related Fields 81 (1), 1–10.

Bradley, R. C., 2010. A strictly stationary, "causal," 5-tuplewise independent counterexample to the central limit theorem. ALEA Lat. Am. J. Probab. Math. Stat. 7, 377–450.

Bradley, R. C., Pruss, A. R., 2009. A strictly stationary, N-tuplewise independent counterexample to the central limit theorem. Stochastic Process. Appl. 119 (10), 3300–3318.

Bregman, Y., Klüppelberg, C., 2005. Ruin estimation in multivariate models with Clayton dependence structure. Scandinavian Actuarial Journal 2005 (6), 462–480.

Bretagnolle, J., Kłopotowski, A., 1995. Sur l'existence des suites de variables aléatoires s à s indépendantes échangeables ou stationnaires. Ann. Inst. H. Poincaré Probab. Statist. 31 (2), 325–350.

Brys, G., Hubert, M., Struyf, A., 2004. A robust measure of skewness. Journal of Computational and Graphical Statistics 13 (4), 996–1017.

Brys, G., Hubert, M., Struyf, A., 2006. Robust measures of tail weight. Computational statistics & data analysis 50 (3), 733–759.

Bürgi, R., Dacorogna, M. M., Iles, R., 2008. Risk aggregation, dependence structure and diversification benefit.

Busse, M., Dacorogna, M., Kratz, M., 2014. The impact of systemic risk on the diversification benefits of a risk portfolio. Risks 2 (3), 260–276.

Cacoullos, T., 1965. Estimation of a multivariate density. Annals of the Institute of Statistical Mathematics 18, 179–189.

Cai, J., Wei, W., 2012. Optimal reinsurance with positively dependent risks. Insurance: Mathematics and Economics 50 (1), 57–63.

Carriere, J. F., 2000. Bivariate survival models for coupled lives. Scandinavian Actuarial Journal 2000 (1), 17–32.

Chadjiconstantinidis, S., Vrontos, S., 2014. On a renewal risk process with dependence under a Farlie–Gumbel–Morgenstern copula. Scandinavian Actuarial Journal 2014 (2), 125–158.

Chakraborty, S., Zhang, X., 2019. Distance metrics for measuring joint dependence with application to causal inference. J. Amer. Statist. Assoc. 114 (528), 1638–1650.

Charpentier, A., Fermanian, J.-D., Scaillet, O., 2007. Copulas: from theory to application in finance. London: Risk Books, Ch. The estimation of copulas: Theory and practice, pp. 35–64.

Chen, D., Mao, T., Pan, X., Hu, T., 2012. Extreme value behavior of aggregate dependent risks. Insurance: Mathematics and Economics 50 (1), 99–108.

Chen, Y., Lin, L., Wang, R., 2022. Risk aggregation under dependence uncertainty and an order constraint. Insurance: Mathematics and Economics 102, 169–187.

Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula methods in finance. John Wiley & Sons.

Cheung, E. C., Dai, S., Ni, W., 2018. Ruin probabilities in a Sparre Andersen model with dependency structure based on a threshold window. Annals of Actuarial Science 12 (2), 269–295.

Cheung, E. C., Landriault, D., Willmot, G. E., Woo, J.-K., 2010. Structural properties of Gerber–Shiu functions in dependent Sparre Andersen models. Insurance: Mathematics and Economics 46 (1), 117–126.

Cheung, E. C., Woo, J.-K., 2016. On the discounted aggregate claim costs until ruin in dependent Sparre Andersen risk processes. Scandinavian Actuarial Journal 2016 (1), 63–91.

Cheung, K. C., Sung, K. C. J., Yam, S. C. P., 2014. Risk-Minimizing Reinsurance Protection For Multivariate Risks. The Journal of Risk and Insurance 81 (1), 219–236.

Ćmiel, B., Ledwina, T., 2020. Validation of association. Insurance: Mathematics and Economics 91, 55–67.

Coeurjolly, J.-F., Drouilhet, R., Lafaye de Micheaux, P., Robineau, J.-F., 2009. asympTest: A simple R package for classical parametric statistical tests and confidence intervals in large samples. The R Journal 1 (2), 26–30.

Collings, S., White, G., 2001. APRA risk margin analysis. In: Institute of Actuaries of Australia XIIIth General Insurance Seminar, Trowbridge Consulting.

Constantinescu, C., Dai, S., Ni, W., Palmowski, Z., 2016. Ruin probabilities with dependence on the number of claims within a fixed time window. Risks 4 (2), 17.

Constantinescu, C., Hashorva, E., Ji, L., 2011. Archimedean copulas in finite and infinite

dimensions—with application to ruin problems. Insurance: Mathematics and Economics 49 (3), 487–495.

Constantinescu, C. D., Kozubowski, T. J., Qian, H. H., 2019. Probability of ruin in discrete insurance risk model with dependent Pareto claims. Dependence Modeling 7 (1), 215–233.

Cossette, H., Côté, M.-P., Marceau, E., Moutanabbir, K., 2013. Multivariate distribution defined with Farlie–Gumbel–Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. Insurance: Mathematics and Economics 52 (3), 560–572.

Cossette, H., Gaillardetz, P., Marceau, É., Rioux, J., 2002. On two dependent individual risk models. Insurance: Mathematics and Economics 30 (2), 153–166.

Cossette, H., Larrivée-Hardy, E., Marceau, E., Trufin, J., 2015. A note on compound renewal risk models with dependence. Journal of Computational and Applied Mathematics 285, 295–311.

Cossette, H., Marceau, E., Mtalai, I., 2019. Collective risk models with dependence. Insurance: Mathematics and Economics 87, 153–168.

Cossette, H., Marceau, E., Mtalai, I., Veilleux, D., 2018. Dependent risk models with Archimedean copulas: A computational strategy based on common mixtures and applications. Insurance: Mathematics and Economics 78, 53–71.

Côté, M.-P., Genest, C., Abdallah, A., 2016. Rank-based methods for modeling dependence between loss triangles. European Actuarial Journal 6 (2), 377–408.

Cuesta, J. A., Matrán, C., 1991. On the asymptotic behavior of sums of pairwise independent random variables. Statist. Probab. Lett. 11 (3), 201–210.

Cundill, B., Alexander, N. D., 2015. Sample size calculations for skewed distributions. BMC Med. Res. Methodol. 15 (1), 28.

Czado, C., 2010. Pair-copula constructions of multivariate copulas. In: Copula theory and its applications. Springer, pp. 93–109.

Czado, C., 2019. Analyzing dependent data with vine copulas: A practical guide with R. Lecture Notes in Statistics, Springer.

Czado, C., Kastenmeier, R., Brechmann, E. C., Min, A., 2012. A mixed copula model for insurance claims and claim sizes. Scandinavian Actuarial Journal 2012 (4), 278–305.

Dacorogna, M., 2018. A change of paradigm for the insurance industry. Annals of Actuarial Science 12 (2), 211–232.

Danıelsson, J., 2002. The emperor has no clothes: Limits to risk modelling. Journal of Banking & Finance 26 (7), 1273–1296.

De Jong, P., 2012. Modeling dependence between loss triangles. North American Actuarial Journal 16 (1), 74–86.

de Lourdes Centeno, M., 2005. Dependent risks and excess of loss reinsurance. Insurance: Mathematics and Economics 37 (2), 229–238.

Denneberg, D., 1990. Premium calculation: Why standard deviation should be replaced by absolute deviation. Astin Bulletin 20 (2), 181–190.

Denuit, M., Cornet, A., 1999. Multilife premium calculation with dependent future lifetimes. Journal of Actuarial Practice 7, 147–180.

Denuit, M., Dhaene, J., Le Bailly de Tilleghem, C., Teghem, S., 2001. Measuring the impact of dependence among insured lifelengths. Belgian Actuarial Bulletin 1 (1), 18–39.

Denuit, M., Genest, C., Marceau, É., 2002. Criteria for the stochastic ordering of random sums, with actuarial applications. Scandinavian Actuarial Journal 2002 (1), 3–16.

Derriennic, Y., Kłopotowski, A., 2000. On Bernstein's example of three pairwise independent random variables. Sankhyā Ser. A 62 (3), 318–330.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., Vyncke, D., 2002a. The concept of comonotonicity in actuarial science and finance: applications. Insurance: Mathematics and Economics 31 (2), 133–161.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., Vyncke, D., 2002b. The concept of comonotonicity in actuarial science and finance: theory. Insurance: Mathematics and Economics 31 (1), 3–33.

Dhaene, J., Goovaerts, M. J., 1997. On the dependency of risks in the individual life model. Insurance: Mathematics and Economics 19 (3), 243–253.

Dhaene, J., Goovaerts, M. J., Lundin, M., Vanduffel, S., 2005. Aggregating economic capital. Belgian Actuarial Bulletin 5, 52–56.

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., Vyncke, D., 2006. Risk measures and comonotonicity: a review. Stochastic models 22 (4), 573–606.

Diestel, R., 2005. Graph theory, 3rd Edition. Vol. 173 of Graduate Texts in Mathematics. Springer-Verlag, Berlin.

Driscoll, M. F., 1978. On pairwise and mutual independence: characterizations of rectangular distributions. Journal of the American Statistical Association 73 (362), 432–433.

Drton, M., Han, F., Shi, H., 2020. High-dimensional consistent independence testing with maxima of rank correlations. Ann. Statist. 48 (6), 3206–3227.

Durante, F., Nelsen, R. B., Quesada-Molina, J. J., Úbeda-Flores, M., 2014. Pairwise and global dependence in trivariate copula models. In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer, pp. 243–251.

Dutang, C., Lefèvre, C., Loisel, S., 2013. On an asymptotic rule A+ B/u for ultimate ruin probabilities under dependence by mixing. Insurance: Mathematics and Economics 53 (3), 774–785.

D'Amato, V., Haberman, S., Piscopo, G., Russolillo, M., 2012. Modelling dependent data for longevity projections. Insurance: Mathematics and Economics 51 (3), 694–701.

E. Shemyakin, A., Youn, H., 2006. Copula models of joint last survivor analysis. Applied Stochastic Models in Business and Industry 22 (2), 211–224.

Efron, B., Tibshirani, R. J., 1994. An introduction to the bootstrap. CRC press.

EIOPA, 07 2014. The underlying assumptions in the standard formula for the Solvency Capital Requirement calculation. Tech. rep.

Eling, M., Jung, K., 2020. Risk aggregation in non-life insurance: Standard models vs. internal models. Insurance: Mathematics and Economics 95, 183–198.

Eling, M., Toplek, D., 2009. Modeling and management of nonlinear dependencies–copulas in dynamic financial analysis. Journal of Risk and Insurance 76 (3), 651–681.

Elston, R., 1991. On Fisher's method of combining p-values. Biometrical journal 33 (3), 339–345.

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

Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. Risk management: value at risk and beyond, 176–223.

Embrechts, P., Nešlehová, J., Wüthrich, M. V., 2009. Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness. Insurance: Mathematics and Economics 44 (2), 164–169.

Embrechts, P., Puccetti, G., Rüschendorf, L., 2013. Model uncertainty and VaR aggregation. Journal of Banking & Finance 37 (8), 2750–2764.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., Beleraj, A., 2014. An academic response to Basel 3.5. Risks 2 (1), 25–48.

Embrechts, P., Resnick, S. I., Samorodnitsky, G., 1999. Extreme value theory as a risk management tool. North American Actuarial Journal 3 (2), 30–41.

Emmer, S., Kratz, M., Tasche, D., 2015. What is the best risk measure in practice? A comparison of standard measures. Journal of Risk 18 (2).

England, P., Verrall, R., 1999. Analytic and bootstrap estimates of prediction errors in claims reserving. Insurance: Mathematics and Economics 25 (3), 281–293.

Englund, M., Guillén, M., Gustafsson, J., Nielsen, L. H., Nielsen, J. P., 2008. Multivariate latent risk: A credibility approach. Astin Bulletin 38 (01), 137–146.

Erdős, P., Rényi, A., 1959. On Cantor's series with convergent $\sum 1/q_n$. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 2, 93–109.

Eryilmaz, S., Gebizlioglu, O. L., 2017. Computing finite time non-ruin probability and some joint distributions in discrete time risk model with exchangeable claim occurrences. Journal of Computational and Applied Mathematics 313, 235–242.

Etemadi, N., 1981. An elementary proof of the strong law of large numbers. Z. Wahrsch. Verw. Gebiete 55 (1), 119–122.

Fan, Y., Lafaye de Micheaux, P., Penev, S., Salopek, D., 2017. Multivariate nonparametric test of independence. J. Multivariate Anal. 153, 189–210.

Fang, K.-T., Kotz, S., Ng, K. W., 1990. Symmetric Multivariate and Related Distributions. London: Chapman and Hall.

Fay, D. S., Gerow, K., 2013. A biologist's guide to statistical thinking and analysis. In: WormBook: the online review of C. elegans biology.

Feng, R., Shimizu, Y., 2016. Applications of central limit theorems for equity-linked insurance. Insurance: Mathematics and Economics 69, 138–148.

Fissler, T., Ziegel, J. F., 2021. On the elicitability of range value at risk. Statistics & Risk Modeling 38 (1-2), 25–46.

Foss, S., Korshunov, D., Zachary, S., 2013. An introduction to heavy-tailed and subexponential distributions, 2nd Edition. Vol. 6. Springer.

Frees, E. W., Carriere, J., Valdez, E., 1996. Annuity valuation with dependent mortality. Journal of risk and insurance, 229–261.

Frees, E. W., Gao, J., Rosenberg, M. A., 2011. Predicting the frequency and amount of health care expenditures. North American Actuarial Journal 15 (3), 377–392.

Frees, E. W., Wang, P., 2005. Credibility using copulas. North American Actuarial Journal 9 (2), 31–48.

Freireich, B., Kodam, M., Wassgren, C., 2010. An exact method for determining local solid fractions in discrete element method simulations. AIChE Journal 56 (12), 3036–3048.

Freund, J. E., 1961. A bivariate extension of the exponential distribution. Journal of the American Statistical Association 56 (296), 971–977.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189–1232.

Furman, E., Kye, Y., Su, J., 2021. Multiplicative background risk models: Setting a course for the idiosyncratic risk factors distributed phase-type. Insurance: Mathematics and Economics 96, 153–167.

Furman, E., Landsman, Z., 2005. Risk capital decomposition for a multivariate dependent gamma portfolio. Insurance: Mathematics and Economics 37 (3), 635–649.

Furman, E., Landsman, Z., 2008. Economic capital allocations for non-negative portfolios of dependent risks. Astin Bulletin 38 (02), 601–619.

Garrido, J., Genest, C., Schulz, J., 2016. Generalized linear models for dependent frequency and severity of insurance claims. Insurance: Mathematics and Economics 70, 205–215.

Gaunt, R. E., 2019. A note on the distribution of the product of zero-mean correlated normal random variables. Stat. Neerl. 73 (2), 176–179.

Geary, R., 1936. Moments of the ratio of the mean deviation to the standard deviation for normal samples. Biometrika 28, 295–307.

Geenens, G., Charpentier, A., Paindaveine, D., et al., 2017. Probit transformation for nonparametric kernel estimation of the copula density. Bernoulli 23 (3), 1848–1873.

Geenens, G., Lafaye de Micheaux, P., 2022. The Hellinger correlation. Journal of the American Statistical Association 117 (538), 639–653.

Geisser, S., Mantel, N., 1962. Pairwise independence of jointly dependent variables. Ann. Math. Statist. 33, 290–291.

Gel, Y. R., Gastwirth, J. L., 2008. A robust modification of the Jarque–Bera test of normality. Economics Letters 99 (1), 30–32.

Gel, Y. R., Miao, W., Gastwirth, J. L., 2007. Robust directed tests of normality against heavy-tailed alternatives. Computational Statistics & Data Analysis 51 (5), 2734–2746.

Genest, C., Boies, J.-C., 2003. Detecting dependence with Kendall plots. The American Statistician 57 (4), 275–284.

Genest, C., Nešlehová, J. G., Rémillard, B., Murphy, O. A., 2019. Testing for independence in arbitrary distributions. Biometrika 106 (1), 47–68.

Genest, C., Rémillard, B., 2004. Test of independence and randomness based on the empirical copula process. Test 13 (2), 335–369.

Gerber, H. U., 1988. Mathematical fun with the compound binomial process. Astin Bulletin 18 (2), 161–168.

Gerber, H. U., 1997. Life insurance mathematics, 3rd Edition. Springer-Verlag.

Ghasemi, A., Zahediasl, S., 2012. Normality tests for statistical analysis: A guide for non-statisticians. Int. J. Endocrinol. Metab. 10 (2), 486–489.

Ghoudi, K., Kulperger, R. J., Rémillard, B., 2001. A nonparametric test of serial independence for time series and residuals. J. Multivariate Anal. 79 (2), 191–218.

Gibbons, J. D., Chakraborti, S., 2003. Nonparametric Statistical Inference. Marcel Dekker.

Gijbels, I., Veraverbeke, N., Omelka, M., 2011. Conditional copulas, association measures and their applications. Computational Statistics & Data Analysis 55 (5), 1919–1932.

Gobbi, F., Kolev, N., Mulinacci, S., 2019. Joint life insurance pricing using extended Marshall–Olkin models. Astin Bulletin 49 (2), 409–432.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. journal of Computational and Graphical Statistics 24 (1), 44–65.

González-Estrada, E., Villaseñor, J. A., 2016. A ratio goodness-of-fit test for the Laplace distribution. Statistics & Probability Letters 119, 30–35.

GPS110, 07 2019. Prudential Standard GPS 110: "Capital Adequacy".

GPS113, 07 2019. Prudential Standard GPS 113: "Capital Adequacy: Internal Model-based Method".

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., Smola, A. J., 2007. A kernel statistical test of independence. In: Advances in neural information processing systems. pp. 585–592.

Guerra, M., de Moura, A., 2021. Reinsurance of multiple risks with generic dependence structures. Insurance: Mathematics and Economics 101, 547–571.

Gumbel, E. J., 1960. Bivariate exponential distributions. Journal of the American Statistical Association 55 (292), 698–707.

Gut, A., 2013. Probability: a graduate course. Vol. 75. Springer Science & Business Media.

Haff, I. H., Aas, K., Frigessi, A., 2010. On the simplified pair-copula construction—Simply useful or too simplistic? Journal of Multivariate Analysis 101 (5), 1296–1310.

Hahn, L., 2017. Multi-year non-life insurance risk of dependent lines of business in the multivariate additive loss reserving model. Insurance: Mathematics and Economics 75, 71–81.

Hall, P., Heyde, C. C., 1980. Martingale limit theory and its application. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London.

Han, X., Liang, Z., Zhang, C., 2019. Optimal proportional reinsurance with common shock dependence to minimise the probability of drawdown. Annals of Actuarial Science 13 (2), 268–294.

Happ, S., Wuthrich, M. V., 2013. Paid-incurred chain reserving method with dependence modeling. Astin Bulletin 43 (01), 1–20.

Hashorva, E., Ratovomirija, G., 2015. On Sarmanov mixed Erlang risks in insurance applications. Astin Bulletin 45 (1), 175–205.

Heilpern, S., 2014. Ruin measures for a compound Poisson risk model with dependence based on the Spearman copula and the exponential claim sizes. Insurance: Mathematics and Economics 59, 251–257.

Heller, R., Heller, Y., Gorfine, M., 2012. A consistent multivariate test of association based on ranks of distances. Biometrika 100 (2), 503–510.

Henshaw, K., Constantinescu, C., Menoukeu Pamen, O., 2020. Stochastic mortality modelling for dependent coupled lives. Risks 8 (1), 17.

Hernández-Bastida, A., Fernández-Sánchez, M., Gómez-Déniz, E., 2009. The net Bayes premium with dependence between the risk profiles. Insurance: Mathematics and Economics 45 (2), 247–254.

Hoeffding, Wassily; Robbins, H., 1948. The central limit theorem for dependent random variables. Duke Mathematical Journal 15, 773–780.

Hofert, M., Oldford, W., 2018. Visualizing dependence in high-dimensional data: An application to S&P 500 constituent data. Econometrics and statistics 8, 161–183.

Hogg, R. V., 1972. More light on the kurtosis and related statistics. Journal of the American Statistical Association 67 (338), 422–424.

Huberts, L. C. E., Schoonhoven, M., Goedhart, R., Diko, M. D., Does, R. J. M. M., 2018. The performance of control charts for large non-normally distributed datasets. Qual. Reliab. Eng. Int. 34 (6), 979–996.

Hušková, M., Meintanis, S. G., 2008. Testing procedures based on the empirical characteristic functions. I. Goodness-of-fit, testing for symmetry and independence. Tatra Mt. Math. Publ. 39, 225–233.

Ibragimov, R., Prokhorov, A., 2017. Heavy tails and copulas: topics in dependence modelling in economics and finance. World Scientific.

Ignatov, Z. G., Kaishev, V. K., 2012. Finite time non-ruin probability for Erlang claim

inter-arrivals and continuous inter-dependent claim amounts. Stochastics An International Journal of Probability and Stochastic Processes 84 (4), 461–485.

Janson, S., 1988. Some pairwise independent sequences for which the central limit theorem fails. Stochastics 23 (4), 439–448.

Jeong, H., Valdez, E. A., 2020. Predictive compound risk models with dependence. Insurance: Mathematics and Economics 94, 182–195.

Ji, M., Hardy, M., Li, J. S.-H., 2011. Markovian approaches to joint-life mortality. North American Actuarial Journal 15 (3), 357–376.

Jin, Z., Matteson, D. S., 2018. Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics. J. Multivariate Anal. 168, 304–322.

Joffe, A., 1974. On a set of almost deterministic k-independent random variables. Ann. Probability 2 (1), 161–162.

Josse, J., Holmes, S., 2016. Measuring multivariate association and beyond. Statistics Surveys 10, 132.

Kantorovitz, M. R., 2007. An example of a stationary, triplewise independent triangular array for which the CLT fails. Statist. Probab. Lett. 77 (5), 539–542.

Killiches, M., Kraus, D., Czado, C., 2017. Examination and visualisation of the simplifying assumption for vine copulas in three dimensions. Australian & New Zealand Journal of Statistics 59 (1), 95–117.

Klugman, S. A., Panjer, H. H., Willmot, G. E., 2012. Loss models: from data to decisions, 4th Edition. Wiley.

Kortschak, D., Albrecher, H., 2009. Asymptotic results for the sum of dependent non-identically distributed random variables. Methodology and Computing in Applied Probability 11 (3), 279–306.

Krämer, N., Brechmann, E. C., Silvestrini, D., Czado, C., 2013. Total loss estimation using copula-based regression models. Insurance: Mathematics and Economics 53 (3), 829–839.

Kresojević, B., Gajić, M., 2019. Application of the t-test in health insurance cost analysis: large data sets. Economics 7 (2), 157–167.

Kreyszig, E., 1978. Introductory functional analysis with applications. John Wiley & Sons, New York-London-Sydney.

Kruskal, W. H., 1958. Ordinal measures of association. Journal of the American Statistical Association 53 (284), 814–861.

Landriault, D., Lee, W. Y., Willmot, G. E., Woo, J.-K., 2014a. A note on deficit analysis in dependency models involving Coxian claim amounts. Scandinavian Actuarial Journal 2014 (5), 405–423.

Landriault, D., Willmot, G. E., Xu, D., 2014b. On the analysis of time dependent claims in a class of birth process claim count models. Insurance: Mathematics and Economics 58, 168–173.

Lee, G. Y., Shi, P., 2019. A dependent frequency–severity approach to modeling longitudinal insurance claims. Insurance: Mathematics and Economics 87, 115–129.

Lehmann, E. L., 1966. Some concepts of dependence. The Annals of Mathematical Statistics 37 (5), 1137–1153.

Liang, Z., Yuen, K. C., 2016. Optimal dynamic reinsurance with dependent risks: variance premium principle. Scandinavian Actuarial Journal 2016 (1), 18–36.

Liu, H., Wang, R., 2017. Collective risk models with dependence uncertainty. Astin Bulletin 47 (2), 361–389.

Loftsgaarden, D. O., Quesenberry, C. P., 1965. A nonparametric estimate of a multivariate density function. The Annals of Mathematical Statistics 36 (3), 1049–1051.

Loisel, S., 2009. A trivariate non-Gaussian copula having 2-dimensional Gaussian copulas as margins. Working Paper hal-00375715, HAL.

Lu, Y., 2017. Broken-heart, common life, heterogeneity: Analyzing the spousal mortality dependence. Astin Bulletin 47 (3), 837–874.

Luciano, E., Spreeuw, J., Vigna, E., 2008. Modelling stochastic mortality for dependent lives. Insurance: Mathematics and Economics 43 (2), 234–244.

Luciano, E., Spreeuw, J., Vigna, E., 2016. Spouses' dependence across generations and pricing impact on reversionary annuities. Risks 4 (2), 16.

Lux, T., Papapantoleon, A., 2019. Model-free bounds on Value-at-Risk using extreme

value information and statistical distances. Insurance: Mathematics and Economics 86, 73–83.

Mardia, K. V., 1962. Multivariate pareto distributions. The Annals of Mathematical Statistics, 1008–1015.

Marshall, A. W., Olkin, I., 1966. A multivariate exponential distribution. Journal of the American Statistical Association 62 (317), 30–44.

Marshall, A. W., Olkin, I., 1967. A generalized bivariate exponential distribution. Journal of Applied Probability 4, 291–302.

Matejka, J., Fitzmaurice, G., 2017. Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, pp. 1290–1294.

McNeil, A. J., Frey, R., Embrechts, P., 2015. Quantitative risk management: Concepts, techniques and tools. Princeton University Press.

McNeil, A. J., Nešlehová, J., 2009. Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. Ann. Statist. 37 (5B), 3059–3097.

Merz, M., Wüthrich, M. V., Hashorva, E., 2013. Dependence modelling in multivariate claims run-off triangles. Annals of Actuarial Science 7 (01), 3–25.

Milevsky, M. A., Promislow, S. D., Young, V. R., 2006. Killing the law of large numbers: Mortality risk premiums and the sharpe ratio. Journal of Risk and Insurance 73 (4), 673–686.

Ming, Z., Liang, Z., Zhang, C., 2016. Optimal mean–variance reinsurance with common shock dependence. The ANZIAM Journal 58 (2), 162–181.

Moors, J., 1988. A quantile alternative for kurtosis. Journal of the Royal Statistical Society: Series D (The Statistician) 37 (1), 25–32.

Mroz, T., Fuchs, S., Trutschnig, W., 2021. How simplifying and flexible is the simplifying assumption in pair-copula constructions–analytic answers in dimension three and a glimpse beyond. Electronic Journal of Statistics 15 (1), 1951–1992.

Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B., 2016. Kernel mean embedding of distributions: A review and beyond. arXiv preprint arXiv:1605.09522.

Müller, A., Pflug, G., 2001. Asymptotic ruin probabilities for risk processes with dependent increments. Insurance: Mathematics and Economics 28 (3), 381–392.

Nakagawa, S., 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behavioral ecology 15 (6), 1044–1045.

Nelsen, R. B., 1996. Nonparametric measures of multivariate association. In: Distributions with fixed marginals and related topics (Seattle, WA, 1993). Vol. 28 of IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Hayward, CA, pp. 223–232.

Nelsen, R. B., Ubeda-Flores, M., 2012. How close are pairwise and mutual independence? Statistics & Probability Letters 82 (10), 1823–1828.

Nieto-Barajas, L. E., Targino, R. S., 2021. A Gamma Moving Average Process For Modelling Dependence Across Development Years In Run-Off Triangles. Astin Bulletin 51 (1), 245–266.

Oh, R., Ahn, J. Y., Lee, W., 2021. On copula-based collective risk models: from elliptical copulas to vine copulas. Scandinavian Actuarial Journal 2021 (1), 1–33.

OSFI, 01 2019. Minimum Capital Test.

Peters, G. W., Shevchenko, P. V., Wüthrich, M. V., 2009. Dynamic operational risk: modeling dependence and combining different sources of information. arXiv preprint arXiv:0904.4074.

Peters, G. W., Wüthrich, M. V., Shevchenko, P. V., 2010. Chain ladder method: Bayesian bootstrap versus classical bootstrap. Insurance: Mathematics and Economics 47 (1), 36–51.

Pfeifer, D., Nešlehová, J., 2004. Modeling and generating dependent risk processes for IRM and DFA. Astin Bulletin 34 (02), 333–360.

Pfister, N., Bühlmann, P., Schölkopf, B., Peters, J., 2018. Kernel-based tests for joint independence. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80 (1), 5–31.

Pierce, D. A., Dykstra, R. L., 1969. Independence and the normal distribution. The American Statistician 23 (4), 39–39.

Pollard, D., 2002. A user's guide to measure theoretic probability. Vol. 8 of Cambridge

Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Pruss, A. R., 1998. A bounded N-tuplewise independent and identically distributed counterexample to the CLT. Probab. Theory Related Fields 111 (3), 323–332.

Psarakis, S., Panaretoes, J., 1990. The folded t distribution. Communications in Statistics-Theory and Methods 19 (7), 2717–2734.

Puccetti, G., Rüschendorf, L., Manko, D., 2016. VaR bounds for joint portfolios with dependence constraints. Dependence Modeling 4 (1).

Puccetti, G., Rüschendorf, L., Small, D., Vanduffel, S., 2017. Reduction of Value-at-Risk bounds via independence and variance information. Scandinavian Actuarial Journal 2017 (3), 245–266.

Pukthuanthong, K., Roll, R., 2009. Global market integration: An alternative measure and its application. Journal of Financial Economics 94 (2), 214–232.

Rényi, A., 1959. On measures of dependence. Acta Math. Acad. Sci. Hungar. 10, 441–451.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., Sabeti, P. C., 2011. Detecting novel associations in large data sets. Science 334 (6062), 1518–1524.

Resnick, S., 2004. The extremal dependence measure and asymptotic independence. Stochastic Models 20 (2), 205–227.

Resnick, S. I., 1999. A probability path. Birkhäuser Boston, Inc., Boston, MA.

Resnick, S. I., 2007. Heavy-tail phenomena: probabilistic and statistical modeling. Springer Science & Business Media.

Révész, P., Wschebor, M., 1965. On the statistical properties of the Walsh functions. Magyar Tud. Akad. Mat. Kutató Int. Közl. 9, 543–554.

Ribas, C., Marín-Solano, J., Alegre, A., 2003. On the computation of the aggregate claims distribution in the individual life model with bivariate dependencies. Insurance: Mathematics and Economics 32 (2), 201–215.

Richter, A., Wilson, T. C., 2020. Covid-19: implications for insurer risk management and the insurability of pandemic risk. The Geneva Risk and Insurance Review 45 (2), 171–199.

Rodríguez-Lallena, J. A., Úbeda-Flores, M., 2009. Some new characterizations and properties of quasi-copulas. Fuzzy Sets and Systems 160 (6), 717–725.

Rojo, J., 2013. Heavy-tailed densities. Wiley Interdisciplinary Reviews: Computational Statistics 5 (1), 30–40.

Romano, J. P., Siegel, A., 1986. Counterexamples in Probability And Statistics. Wadsworth, London.

Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. U.S.A. 42, 43–47.

Sanders, L., Melenberg, B., 2016. Estimating the joint survival probabilities of married individuals. Insurance: Mathematics and Economics 67, 88–106.

Sarabia, J. M., Gómez-Déniz, E., Prieto, F., Jordá, V., 2016. Risk aggregation in multivariate dependent Pareto distributions. Insurance: Mathematics and Economics 71, 154–163.

Sarabia, J. M., Gómez-Déniz, E., Prieto, F., Jordá, V., 2018. Aggregation of dependent risks in mixtures of exponential distributions and extensions. Astin Bulletin 48 (3), 1079–1107.

Scherer, M., Stahl, G., 2021. The standard formula of Solvency II: a critical discussion. European Actuarial Journal 11 (1), 3–20.

Schmidli, H., 2017. Risk Theory. Springer.

Shi, H., Drton, M., Han, F., 2022. Distribution-free consistent independence tests via center-outward ranks and signs. Journal of the American Statistical Association 117 (537), 395–410.

Shi, P., Frees, E. W., 2011. Dependent loss reserving using copulas. Astin Bulletin 41 (02), 449–486.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231.

Smith, M. L., Kane, S. A., 1994. The law of large numbers and the strength of insurance. In: Insurance, risk management, and public policy. Springer, pp. 1–27.

Smith, V. K., 1975. A simulation analysis of the power of several tests for detecting heavy-tailed distributions. Journal of the American Statistical Association 70 (351a), 662–665.

Spanhel, F., Kurz, M. S., 2019. Simplified vine copula models: Approximations based on the simplifying assumption. Electronic Journal of Statistics 13 (1), 1254–1291.

Spreeuw, J., 2006. Types of dependence and time-dependent association between two lifetimes in single parameter copula models. Scandinavian Actuarial Journal 2006 (5), 286–309.

Spreeuw, J., Wang, X., 2008. Modelling the short-term dependence between two remaining lifetimes. Cass Business School Discussion Paper 2 (3).

Stöber, J., Joe, H., Czado, C., 2013. Simplified pair copula constructions—limitations and extensions. Journal of Multivariate Analysis 119, 101–118.

Stoyanov, J. M., 2013. Counterexamples in probability, 3rd Edition. Dover Publications, Inc., Mineola, NY.

Székely, G. J., Rizzo, M. L., Bakirov, N. K., 2007. Measuring and testing dependence by correlation of distances. The Annals of Statistics 35 (6), 2769–2794.

Takahasi, K., 1965. Note on the multivariate Burr's distribution. Annals of the Institute of Statistical Mathematics 17 (1), 257–260.

Takeuchi, K., 2019. A family of counterexamples to the central limit theorem based on binary linear codes. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E102.A (5), 738–740.

Tanabe, K., Sagae, M., 1992. An exact Cholesky decomposition and the generalized inverse of the variance-covariance matrix of the multinomial distribution, with applications. J. Roy. Statist. Soc. Ser. B 54 (1), 211–219.

Tang, A., Valdez, E. A., 2009. Economic capital and the aggregation of risks using copulas.

Taylor, G. C., 2018. Observations on industry practice in the construction of large correlation structures for risk and capital margins. European Actuarial Journal 8 (2), 517–543.

Ting Lee, M.-L., 1996. Properties and applications of the Sarmanov family of bivariate distributions. Communications in Statistics-Theory and Methods 25 (6), 1207–1222.

Tjøstheim, D., Hufthammer, K. O., 2013. Local Gaussian correlation: a new measure of dependence. Journal of Econometrics 172 (1), 33–48.

Tjøstheim, D., Otneim, H., Støve, B., 2022. Statistical Dependence: Beyond Pearson's ρ. Statistical Science 37 (1), 90–109.

Valdez, E. A., 2014. Empirical investigation of insurance claim dependencies using mixture models. European Actuarial Journal 4 (1), 155–179.

Vernic, R., Bolancé, C., Alemany, R., 2022. Sarmanov distribution for modeling dependence between the frequency and the average severity of insurance claims. Insurance: Mathematics and Economics 102, 111–125.

Wang, X., Jiang, B., Liu, J., 2017. Generalized R-squared for detecting dependence. Biometrika 104 (1), 129–139.

Weakley, L. M., 2013. Some strictly stationary, N-tuplewise independent counterexamples to the central limit theorem. ProQuest LLC, Ann Arbor, MI, thesis (Ph.D.)–Indiana University.

Webster, R., Oliver, M. A., 2007. Geostatistics for environmental scientists. John Wiley & Sons.

Wen, L., Wu, X., Zhou, X., 2009. The credibility premiums for models with dependence induced by common effects. Insurance: Mathematics and Economics 44 (1), 19–25.

Wilcox, R. R., 2011. Introduction to Robust Estimation and Hypothesis Testing. Academic Press.

Willmot, G. E., Woo, J.-K., 2012. On the analysis of a general class of dependent risk processes. Insurance: Mathematics and Economics 51 (1), 134–141.

Wilson, R. J., 1996. Introduction to graph theory, 4th Edition. Longman, Harlow.

Woo, J.-K., Cheung, E. C., 2013. A note on discounted compound renewal sums under dependency. Insurance: Mathematics and Economics 52 (2), 170–179.

Wüthrich, M. V., 2003. Asymptotic value-at-risk estimates for sums of dependent random variables. Astin Bulletin 33 (01), 75–92.

Xiao, Y., Valdez, E. A., 2015. A Black–Litterman asset allocation model under Elliptical distributions. Quantitative Finance 15 (3), 509–519.

Yan, J., 2007. Enjoy the joy of copulas: with a package copula. Journal of Statistical Software 21, 1–21.

Yao, S., Zhang, X., Shao, X., 2018. Testing mutual independence in high dimension via distance covariance. J. R. Stat. Soc. Ser. B. Stat. Methodol. 80 (3), 455–480.

Yuen, K. C., Liang, Z., Zhou, M., 2015. Optimal proportional reinsurance with common shock dependence. Insurance: Mathematics and Economics 64, 1–13.

Zhang, L., Hu, X., Duan, B., 2015. Optimal reinsurance under adjustment coefficient measure in a discrete risk model based on Poisson MA (1) process. Scandinavian Actuarial Journal 2015 (5), 455–467.

Zhang, Z., Yang, H., Yang, H., 2012. On a Sparre Andersen risk model with time-dependent claim sizes and jump-diffusion perturbation. Methodology and computing in applied probability 14 (4), 973–995.

Zhu, L., Xu, K., Li, R., Zhong, W., 2017. Projection correlation between two random vectors. Biometrika 104 (4), 829–843.

Ziegel, J. F., 2016. Coherence and elicitability. Mathematical Finance 26 (4), 901–918.

Appendices

APPENDIX A

COMPUTING CODES

Computing codes (in the `R` language) which produced the results from this thesis are available on GitHub, at https://github.com/gboglioni/PhD_thesis.

For reference, we also include them in this Appendix, presenting them chapter by chapter.

A.1 Codes for Chapter 1

Below is the R code producing the results from Example 1.1:

```
# 0-correlation, but VaR(X + Y) different to VaR(X.ind + Y.ind)
set.seed(1)
n       <- 5000
X       <- rexp(n)
Y       <- rnorm(n, 1, X)
X.indep <- rexp(n)

# Save scatterplot of X versus Y
pdf(file = "0_correl_scatter.pdf", width = 7, height = 7)
par(mar=c(5,5,1,1))
plot(rank(X)/(n+1), rank(Y)/(n+1), cex = 0.5, cex.axis = 1.25, cex.lab = 1.5)
dev.off()

# Theoretical values,
# (numbers obtained from calculations in Mathematica)
VaR.X          <- qexp(0.995)
VaR.Y          <- 6.607
VaR.sum.indep <- 8.351
VaR.sum.dep    <- 10.405
sum.VaR        <- VaR.X + VaR.Y

# Diversification under independence
1 - VaR.sum.indep/sum.VaR
# Diversification under dependence
1 - VaR.sum.dep/sum.VaR
```

Below is the R code producing the scatterplots from Figure 1.2:

```
# Generating the plots of the "datasauRus" dataset

# Load packages
library('datasauRus')
library('dplyr')
library('ggplot2')

# Generate scatterplots (and save them)
pdf(file = "datasauRusPlots.pdf", width = 7, height = 9)
ggplot(datasaurus_dozen, aes(x=x, y=y, colour=dataset))+
  geom_point()+
  coord_fixed(ratio=0.6)+
  theme(legend.position = "none")+
  facet_wrap(~dataset, ncol=3)
dev.off()
```

```
# Check the summary statistics (they are very similar)
  datasaurus_dozen %>%
    group_by(dataset) %>%
    summarize(
      mean_x    = mean(x),
      mean_y    = mean(y),
      std_dev_x = sd(x),
      std_dev_y = sd(y),
      corr_x_y  = cor(x, y)
    )
```

Below is the R code producing Figure 1.3:

```
# Individual Risk Model with three PIBD variables

# Load packages
library(ggplot2)
library(gridExtra)
# Set "rate" of the Exponentials
l <- 1

# Independent case
df  <- function(x){
  return(3*dgamma(x,shape=1, rate=l)/8 + 3*dgamma(x,shape=2, rate=l)/8
         + dgamma(x,shape=3, rate=l)/8)}
cdf <- function(x){
  return(1/8*pbinom(x+1,1,1) + 3*pgamma(x,shape=1, rate=l)/8
         +3*pgamma(x,shape=2, rate=l)/8 + pgamma(x,shape=3, rate=l)/8)}

# PIBD case
df.p  <- function(x){3*dgamma(x,shape=1, rate=l)/4 + dgamma(x,shape=3, rate=l)/4}
cdf.p <- function(x){3*pgamma(x,shape=1, rate=l)/4 + pgamma(x,shape=3, rate=l)/4}
x      <- seq(0.01, 7.5, by = 0.01)
pdf.x  <- df(x)
cdf.x  <- cdf(x)
pdf.p.x  <- df.p(x)
cdf.p.x  <- cdf.p(x)

#CDF comparison
df <- data.frame(x, pdf.x, cdf.x, pdf.p.x, cdf.p.x)

#PDF comparison
mp=data.frame(x=c(0), y=c(0), vx=c(0), vy=c(1/8))

df.plot <- ggplot()+
  geom_line(data=df,aes(x, y=pdf.x, colour="darkblue"),size=1.25)+
  geom_line(data=df,aes(x, y=pdf.p.x, colour="red"),size=1.25)+
  geom_segment(data=mp, mapping=aes(x=x, y=y, xend=vx, yend=vy),
  size=1.5, color="#F8766D")+  geom_point()+
  geom_point(aes(x=0, y=1/8), size = 3, colour="#F8766D")+
```

```
  labs(y= "f(x)", x = "x") + theme_grey(base_size = 16)+
  theme(legend.position = "none")+ theme(legend.title = element_blank())


#CDF
cdf.plot <- ggplot()+
  geom_line(data=df,aes(x, y=cdf.x, colour="darkblue"),size=1.25)+
  geom_line(data=df,aes(x, y=cdf.p.x, colour="red"),size=1.25)+
  geom_point()+
  geom_point(aes(x=0, y=1/8), size = 3, colour="#F8766D")+
  geom_point(aes(x=0, y=0), size = 3, colour="#00BFC4")+
  scale_color_discrete(labels = c("mutual indep.", "PIBD"))+
  labs(y= "F(x)", x = "x") + theme_grey(base_size = 16)+
  theme(legend.position = c(0.75, 0.1)) + theme(legend.title = element_blank())

# Put two plots side by side
grid.arrange(df.plot, cdf.plot, ncol=2)
```

Below is the R code producing Figure 1.4:

```
# Example of three N(0,1) PIBD with X+Y+Z NOT a N(0,3)

# Set seed
set.seed(1)

#Sample size
n <- 1000000

#Generate a sample of (X,Y,Z)
X <- rnorm(n)
Y <- rnorm(n)
W <- rnorm(n)
Z <- abs(W)*sign(X*Y)

#Sum of the PIBD Normals
S <- X + Y + Z

# Plot (and save) density and CDF (with, as comparison, those of a N(0, sqrt(3)))
pdf(file = "density_CDF_S_romano.pdf", width = 14, height = 7)
par(mfrow=c(1,2))
par(mar=c(3,5,1,1))

# Density
my.d <- density(S)
plot(my.d, lwd= 4, col = 'brown3', cex.axis = 1.5, cex.lab = 1.75, main = "",
     xlab = "",  bty="n")
curve(dnorm(x, 0, sqrt(3)), col="black", lty = 1, lwd=2, add=TRUE)

# CDF (we take a subset of "only" 100000 points)
plot(ecdf(S[1:100000]), col='brown3', lwd=4, xlab ="",
```

```
      ylab = "CDF", cex.axis = 1.5, cex.lab = 1.75, main = "", col.01line = NULL)
curve(pnorm(x, 0, sqrt(3) ), lwd=2,add=TRUE)
legend("bottomright", NULL, ncol=1, cex = 1.75, legend=c("S", "N(0, 3)"),
       col=c("brown3", "black"), lty = c(1,1), lwd = c(4,2))
dev.off()
```

A.2 Codes for Chapter 3

Below is the R code defining a function, `generate.PI`, which allows to generate PIBD data according to all examples from Section 3.2.

```
# Function to simulate PIBD variables (U1,U2,U3) under various examples

# Arguments of function 'generate.PI '
#   n:          Sample size
#   type:       Key word for the name of the example
#   alpha:      Parameter (used in some examples only)

generate.PI <- function(n, type = 'indep', alpha=1){

  # Two independent samples of U(0,1)
  U1 <- runif(n, 0, 1)
  U2 <- runif(n, 0, 1)

  # when "mixing" a PIBD structure with mutual independence,
  # "n.dep" is the number of observations stemming from the dependent structure
  if (type == 'bernstein' || type == 'triangles'){
    n.dep <- rbinom(1, n, alpha)
  }

  # mutual independence
  if (type == 'indep'){
    U3 <- runif(n, 0, 1)
  }

  # Examples 3.1 and 3.6
  if (type == 'triangles'){
    U3 <- c((U1[1:n.dep]+U2[1:n.dep])%%1, runif(n-n.dep))
  }

  # Example 3.2
  if (type == 'bernstein'){
    X <- rbinom(n.dep, 1, 1/2)
    Y <- rbinom(n.dep, 1, 1/2)
    Z <- -abs(X-Y)+1
    U1 <- c((X+runif(n.dep))/2, runif(n-n.dep))
    U2 <- c((Y+runif(n.dep))/2, runif(n-n.dep))
    U3 <- c((Z+runif(n.dep))/2, runif(n-n.dep))
  }

  # Example 3.3
  if (type == 'tetra'){
    n1 <- rbinom(1, n, 1/2)
    U3 <- rep(NA, n)
```

```
    for (i in 1:n1){
      ifelse(U1[i]+U2[i] >= 1, U3[i]<- U1[i]+U2[i]-1, U3[i] <- 1-(U1[i]+U2[i]))
    }
    for (i in (n1+1):n){
      ifelse(U1[i]-U2[i] >= 0, U3[i]<- 1 - (U1[i]-U2[i]), U3[i] <- 1+(U1[i]-U2[i]))
    }
  }

  # Example 3.4
  if (type == 'cosine'){
    U3 <- pbeta((cos(2*pi*(U1+U2))+1)/2, 1/2, 1/2)
  }

  # Example 3.5
  if (type == 'copula'){
    Tval  <- runif(n)
    a  <- alpha*(1-2*U1)*(1-2*U2)
    U3 <- (1+a - sqrt((1+a)^2-4*a*Tval))/(2*a)
  }

  # Example 3.7
  if (type == 'bernstein4D'){
    M1 <- rmultinom(n, 1, rep(1/alpha, alpha))
    M2 <- rmultinom(n, 1, rep(1/alpha, alpha))
    M3 <- rmultinom(n, 1, rep(1/alpha, alpha))
    M4 <- rmultinom(n, 1, rep(1/alpha, alpha))
    X1 <- apply(M1 == M3, 2, prod)
    X2 <- apply(M1 == M4, 2, prod)
    X3 <- apply(M2 == M3, 2, prod)
    X4 <- apply(M2 == M4, 2, prod)
    U1 <- X1 * runif(n, (alpha-1)/alpha,1) + ifelse(X1==0,1,0)*runif(n,0,(alpha-1)/alpha)
    U2 <- X2 * runif(n, (alpha-1)/alpha,1) + ifelse(X2==0,1,0)*runif(n,0,(alpha-1)/alpha)
    U3 <- X3 * runif(n, (alpha-1)/alpha,1) + ifelse(X3==0,1,0)*runif(n,0,(alpha-1)/alpha)
    U4 <- X4 * runif(n, (alpha-1)/alpha,1) + ifelse(X4==0,1,0)*runif(n,0,(alpha-1)/alpha)
    data <- as.data.frame(cbind(U1,U2,U3,U4))
  }

  # Return data.frame
  if(type != 'bernstein4D'){
    data <- as.data.frame(cbind(U1,U2,U3))
  }
  return(data)
}
```

Below are all the `R` functions, needed to generate the Figures from Chapter 3. The main function is called `p.i.figures`, but it uses several other functions we define first.

```
# Load needed package
library('pdist')
```

```
# Functions to find fraction of volume of a ball sitting OUTSIDE the [0,1]^3 "cube"
# (when the ball extends outside the cube in either 1, 2 or 3 directions)
# The arguments d1,d2,d3 are as defined in the article by Freireich et al. (2010)

f.face <- function(d,r){
  dr <- d/r
  return((3*dr^2 - dr^3)/4)
}

f.edge <- function(d1,d2,r){
  a  <- 1-d1/r
  b  <- 1-d2/r
  x2 <- 1 - a^2 - b^2
  if(x2 <= 0){
    return(0)}
  else{
    x <- sqrt(x2)
    return((2*a*b*x-(3*a-a^3)*atan(x/b)-(3*b-b^3)*atan(x/a)+2*atan(x*a/b)
            +2*atan(x*b/a))/(4*pi))
  }
}

f.corner <- function(d1,d2,d3,r){
  a  <- 1-d1/r
  b  <- 1-d2/r
  c  <- 1-d3/r

  if((a^2+b^2+c^2) >= 1){
    return(0)}
  else{
    A  <- sqrt(1-a^2-c^2)
    B  <- sqrt(1-b^2-c^2)
    return(f.edge(d1,d2,r)/2 -
          (6*a*b*c-2*a*A*c-2*b*B*c-(3*b-b^3)*atan(c/B)-(3*a-a^3)*atan(c/A)+
          (3*c-c^3)*(atan(A/a)-atan(b/B))+2*(atan(c*a/A)+atan(c*b/B)))/(8*pi))
  }
}

# Functions for the fraction of volume of the ball comprised inside the "cube"
# (different functions apply, depending on in how many directions (1,2 or 3)
# the ball extends outside the cube)
v.1inter <- function(d,r){
  1-(f.face(d=d,r=r))
}

v.2inter <- function(d1,d2,r){
  1-(f.face(d=d1,r=r) + f.face(d=d2,r=r) - f.edge(d1=d1,d2=d2,r=r))
}
```

```
v.3inter <- function(d1,d2,d3,r){
  1-(f.face(d1,r)+f.face(d2,r)+f.face(d3,r)
     -f.edge(d1,d2,r)-f.edge(d1,d3,r)-f.edge(d2,d3,r)
     +f.corner(d1,d2,d3,r))
}

# Function 'v.star' returns the proportion (volume wise) of a Ball[p,r]
# which is inside the unit cube [0,1]^3
# Arguments of function 'v.star':
#   p:  Vector of coordinates (x,y,z) of the point at the center of the ball
#   r:  Radius of the Ball

v.star <- function(p,r){
  p1 <- p[1]
  p2 <- p[2]
  p3 <- p[3]
  p_pm_r <- c(p1-r,p1+r,p2-r,p2+r,p3-r,p3+r)

  # Count in how many directions does the Ball[p,r] falls outside [0,1]^3
  count  <- sum(p_pm_r < 0) + sum(p_pm_r > 1)
  # Distances from p to edges of [0,1]^3
  # (for the directions for which the Ball[p,r] extends outside [0,1]^3)
  bounds <- c(p[p < r], 1 - p[p + r > 1])
  # (d is (d1, d2, d3) with d1, d2, d3 corresponding to Freireich et al. (2010))
  d <- r - bounds

  if (count == 0){
    delta <- 1
  }
  # Case the Ball[p,r] extends outside [0,1]^3 in one direction
  if (count == 1){
    delta <- v.1inter(d=d[1], r)
  }
  # Case the Ball[p,r] extends outside [0,1]^3 in two directions
  if (count == 2){
    delta <- v.2inter(d1=d[1], d2=d[2], r)
  }
  # Case the Ball[p,r] extends outside [0,1]^3 in three directions
  if (count == 3){
    delta <- v.3inter(d1=d[1], d2=d[2], d3=d[3], r)
  }
  if (count > 3){
    print("Error: 'count' cannot be >3")
  }

  #Return result
  return(delta)
}
```

```
# Function to create a 3D grid in [0,1]^3
create.grid <- function(g=10){
U1 <- NULL
U2 <- NULL
for (i in 1:g){
  U1 <- c(U1, rep(i,g^2))
  U2 <- c(U2, rep(i,g))
}
U3 <- rep(1:g,g^2)
return(cbind(U1,U2,U3)/(g+1))
}

# Function that finds the 'concentration index' h() for all points of a dataset
# (data assumed in [0,1]^3)
# Arguments:
#     data:          (nx3) matrix containing the data
#     rad:           radius of the small balls around each point
#     scaled:        TRUE to have 0-to-1 scaling on a RELATIVE basis
#                    (i.e. full colour palette is used)
#     return.h:      TRUE to return the value of 'h', otherwise returns 'f(h)'
#     grid.method:   TRUE to use "fixed grid" method
#     g:             Size of grid, if "fixed grid" method is used

pts.concentration <- function(data, rad, scaled=F, return.h = F,
                              grid.method = F, g=10){
  # Sample size
  n          <- length(data[,1])
  # Initial (not smoothed) concentration
  distances <- as.matrix(dist(data, upper = TRUE))
  # Count number of points within radius "rad" of any point
  hp       <- (rowSums(distances <= rad)-1)/(n-1) #"-1" excludes the point itself
  hp       <- hp / apply(data, 1, v.star, r = rad)

  # Grid (g*g*g)
  grid <- create.grid(g=g)

  # Values that depend on whether we use the 'fixed grid' method or not
  if (grid.method){
    # Number of balls is the number of points of the grid, g^3
    n.balls <- g^3
    # Distances between each point of the grid and each point of the sample
    grid.dist <- as.matrix(pdist(X = grid, Y = data))
  } else{
    n.balls <- n
    grid.dist <- distances
  }

  # Find the final estimates of "h"
```

```
  # (each point's value is a weighted average of its nearest neighbours)
  final.hp <- rep(NA, n.balls)
  for(i in 1:n.balls){
    closest.pt  <- grid.dist[i,] <= rad
    dist        <- grid.dist[i,][closest.pt]
    # If there are ZERO points in the ball, hp should just be 0
    if (length(dist) == 0){final.hp[i] <- 0}
    # If there is only ONE point in the ball, hp should be that of this point
    if (length(dist) == 1){final.hp[i] <- hp[closest.pt]}
    # If there are at least TWO points
    if (length(dist)  > 1){
      # If one point is EXACTLY the center of the ball,
      # we give it a "distance" equal to the distance with THE closest neighbour
      if(min(dist)==0){dist[which.min(dist)] <- sort(dist)[-1][1]}
      # Weights are simply the reverse of the distances
      weights     <- 1/dist
      # Compute final value: a weighted average of all points inside the ball
      final.hp[i] <- sum(hp[closest.pt] * weights / sum(weights))
    }
  }

  # Parameters for scaling to a color scale (0-to-1)
  v = 4*pi*rad^3/3
  # Case where we want the entire color palette to be used (i.e. 'relative' scale)
  if (scaled){
    a <- min(final.hp)
    c <- max(final.hp)
    if (a > v){
      print("Warning: 'a' is bigger than 'v'. Replacing 'v' by (a+c)/2")
      v <- (a+c)/2
    }
    if (c < v){
      print("Error: 'c' must be larger than 'v'.")
    }
    # Case where the color palette is absolute (not relative to the given sample)
  } else{
    a <- 0
    c <- 2 * rad/sqrt(3)
  }
  beta       <- -log(2)/log((v-a)/(c-a))

  # Put the points on a 0-to-1 scale
  pt.scaled <- ((final.hp-a)/(c-a))^beta
  # Cap values at '1'
  ind <- pt.scaled > 1
  pt.scaled[ind] <- 1
  # Return results
  ifelse(return.h, return(final.hp), return(pt.scaled))
}
```

```
# Function that assigns a range of colors to a range of values
# Values must be between 0 and 1
myColorRamp <- function(colors, values) {
  x <- colorRamp(colors)(values)
  rgb(x[,1], x[,2], x[,3], maxColorValue = 255)
}

# Function 'scatter2D.plus' creates scatterplots of two variables,
# where a third variable is represented by a colour
scatter2D.plus <- function(x,y,z, legend = TRUE,
                           my.colors = c("magenta", "white", "cornflowerblue"),
                           X.lab = "X", Y.lab = "Y", nb.levels = 50,
                           pt.size=3, lab.size=1){
  Rx <- diff(range(x))
  Ry <- diff(range(y))
  levelplot(z ~ x + y, colorkey = legend, panel = panel.levelplot.points,
            col.regions = colorRampPalette(my.colors)(nb.levels),
            xlab=list(label = X.lab, cex=lab.size),
            ylab=list(label = Y.lab, cex=lab.size), cex = pt.size,
            col =  "black", pch = 21, xlim = c(min(x)-0.05*Rx, max(x)+0.05*Rx),
            ylim = c(min(y)-0.05*Ry, max(y)+0.05*Ry))
}
```

We now include the main function `p.i.figures()`.

```
##############################################################################
# Author: Guillaume Boglioni Beaulieu
# Description: Function to generate the data and 3D Figures of PIBD examples
# Last update: 06/09/2022
##############################################################################

# Libraries
library('rgl')
library('latticeExtra')
library('gridExtra')
library('grid')
library('ggplot2')

# Load external functions
source("3D_visualisation_functions.R")
source("generate_PI.R")

# Arguments of function 'p.i.figures'
#   seed:             Seed for simulations
#   type:             Identifier of the type of dependence
#   n:                Sample size
#   my.col:           Colour scale of the points on the scatterplot
#   alpha:            Parameter used in certain examples
#                     ('triangles', 'bernstein', 'bernstein4D' and 'copula')
```

```
#   pt.size:           Size of the points on the plots
#   plot.type:         Type visualisation: '2D.black','2D.colour', '2D.matrix','3D'
#   var.ind:           For plot.type = '2D, black', '2D.colour' or '3D.matrix',
#                      order of appearance of the variables on the scatterplots
#   emp.colors:        TRUE to have empirical estimation of point concentration
#                      on 3D scatterplots
#   scale.col:         TRUE for a RELATIVE color scale
#                      (i.e. full color palette of 'my.col' is used)
#   rad:               Radius of balls in empirical estimation of points concentration
#   grid.method:       TRUE if we want the 3D color coding using a fixed grid
#   g:                 Integer for the size of the grid
#   adjust.pts.size:   TRUE to have the size of points proportional
#                      to the concentration around them
#   type.3d:           Type of points: 'p' is for points, 's' is for 3D spheres
#   legend:            For 2D coloured scatterplots, should a legend be displayed
#   margins:           If "TRUE", then simulated data will have marginals given by
#                      arguments mar1, mar2, mar3 (specifying quantile functions)

p.i.figures <- function(seed=1, type = 'bernstein', n = 1000,
                        my.col = c("blue", "white", "red"), alpha = 1,
                        pt.size = 5, plot.type = '2D', var.ind = c(1,2,3),
                        emp.colors = F, scale.col = F, rad=1/10, grid.method=F,
                        g=8, adjust.pts.size=F, type.3d = 'p', legend=T,
                        margins=F, mar1 = qnorm, mar2 = qexp, mar3 = qlnorm){

  # Set seed for random data generation
  set.seed(seed)

  # Generate data (according to 'type')
  data <- generate.PI(n, type, alpha)
  # At first, we set the "colors" of points to be a single color
  color <- my.col[1]

  # Color for the '4D example' (Example 3.7)
  if (type == 'bernstein4D'){
    color <- myColorRamp(my.col, data$U4)
  }

  # Update margins
  if(margins){
    data <- as.data.frame(
      cbind(X1=mar1(data[,1]), X2=mar2(data[,2]), X3=mar3(data[,3]))
    )
  }

  # Empirical estimation of 'h' for all points
  # ('h' calculated with function 'pts.concentration')
  if (emp.colors){
    # Option to have 3D plots where points vary in size (based on 'h' values)
```

```
  if(adjust.pts.size){
    # Values of the concentration index
    h.values <- pts.concentration(data, rad, scaled=scale.col,
                                  grid.method=grid.method, g=g, return.h = T)
    # Values on the [0,1] scale (for color coding)
    f.values <- pts.concentration(data, rad, scaled=scale.col,
                                  grid.method=grid.method, g=g)
    # For the grid method, the 'data' needs to be the data points of the grid
    if(grid.method){data <- as.data.frame(create.grid(g=g))}
    n.balls <- length(data[,1])
    data <- cbind(data, h.values, f.values)
    # Create "factors" for the size of the points
    # (with max value the max of "h.values" in the sample)
    size <- as.numeric(cut(c(h.values,0, max(h.values)), 100))
    dataList <- split(data, size[-c(n.balls+1,n.balls+2)])
    # Find colors for the different 'size' factors
    # (we take the average colour of points having the same 'size')
    colors <- rep(NA, length(dataList))
    for(i in seq_along(dataList)){
      colors[i] <- myColorRamp(colors=my.col, mean(dataList[[i]]$f.values))
    }

  }else{
    color  <- myColorRamp(colors=my.col,
                          values=pts.concentration(data, rad, scaled=scale.col,
                                                   grid.method=grid.method, g=g))
  }
}

# Plot 2D scatterplot
if(plot.type=='2D.black'){
    par(mfrow=c(1,2))
    par(mar = c(5, 4.5, 1.5, 1.5))
    plot(data[,var.ind[1]], data[,var.ind[3]], cex=pt.size, pch = 20,
         xlab = colnames(data)[var.ind[1]],
         ylab =colnames(data)[var.ind[3]], cex.axis = 1.25, cex.lab = 1.75)
    plot(data[,var.ind[2]], data[,var.ind[3]], cex=pt.size, pch = 20,
         xlab = colnames(data)[var.ind[2]],
         ylab =colnames(data)[var.ind[3]], cex.axis = 1.25, cex.lab = 1.75)
    }
if(plot.type=='2D.colour'){
    return(scatter2D.plus(data[,var.ind[1]], data[,var.ind[2]],
                          data[,var.ind[3]], my.colors = my.col,
                          lab.size=1.75, pt.size=pt.size,
                          X.lab = colnames(data)[var.ind[1]],
                          Y.lab =colnames(data)[var.ind[2]]))
}

# Plot matrix of 2D scatterplot
```

```
if(plot.type=='2D.matrix'){

  # Transform data to the ECDF
  ecdf.data <- apply(data, 2, rank)/n

  p12 <- scatter2D.plus(ecdf.data[,var.ind[1]], ecdf.data[,var.ind[2]],
         ecdf.data[,var.ind[3]], pt.size = pt.size, my.color=my.col,
         legend=legend, X.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[1]])*")"),
         Y.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[2]])*")"))
  p13 <- scatter2D.plus(ecdf.data[,var.ind[1]], ecdf.data[,var.ind[3]],
         ecdf.data[,var.ind[2]], pt.size = pt.size, my.color=my.col,
         legend=legend, X.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[1]])*")"),
         Y.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[3]])*")"))
  p23 <- scatter2D.plus(ecdf.data[,var.ind[2]], ecdf.data[,var.ind[3]],
         ecdf.data[,var.ind[1]], pt.size = pt.size, my.color=my.col,
         legend=legend, X.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[2]])*")"),
         Y.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[3]])*")"))

  p11 <- ggplot(data, aes(x=data[,var.ind[1]])) + geom_histogram(aes(y=..density..),
         colour="black", fill="lightblue", binwidth=2*IQR(data[,var.ind[1]])/n^(1/3))+
         labs(x = colnames(data)[var.ind[1]], y = "") + theme_classic()
  p22 <- ggplot(data, aes(x=data[,var.ind[2]])) + geom_histogram(aes(y=..density..),
         colour="black", fill="lightblue", binwidth=2*IQR(data[,var.ind[2]])/n^(1/3))+
         labs(x = colnames(data)[var.ind[2]], y = "") + theme_classic()
  p33 <- ggplot(data, aes(x=data[,var.ind[3]])) + geom_histogram(aes(y=..density..),
         colour="black", fill="lightblue", binwidth=2*IQR(data[,var.ind[3]])/n^(1/3))+
         labs(x = colnames(data)[var.ind[3]], y = "") + theme_classic()

  p21 <- scatter2D.plus(ecdf.data[,var.ind[1]], ecdf.data[,var.ind[2]],
         ecdf.data[,var.ind[3]], pt.size = pt.size/2,
         X.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[1]])*")"),
         Y.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[2]])*")"),
         my.colors = "black", legend = F)
  p31 <- scatter2D.plus(ecdf.data[,var.ind[1]], ecdf.data[,var.ind[3]],
         ecdf.data[,var.ind[2]], pt.size = pt.size/2,
         X.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[1]])*")"),
         Y.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[3]])*")"),
         my.colors = "black", legend = F)
  p32 <- scatter2D.plus(ecdf.data[,var.ind[2]], ecdf.data[,var.ind[3]],
         ecdf.data[,var.ind[1]], pt.size = pt.size/2,
         X.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[2]])*")"),
         Y.lab = bquote(hat(F)*"("*.(colnames(data)[var.ind[3]])*")"),
         my.colors = "black", legend = F)

  print(grid.arrange(p11, p12,p13, p21, p22, p23, p31, p32, p33, ncol=3, nrow = 3))

}

# Plot 3D scatterplot
```

```
  if(plot.type=='3D'){

    # Option where we alter the size of each point
    if(adjust.pts.size){
      # Create plot (points are added in a second step)
      with(data, plot3d(NULL, NULL, NULL,
                        colnames(data)[1], ylab = colnames(data)[2],
                        zlab = colnames(data)[3], col='white', size=0))
      par3d(windowRect = 50 + c(0, 0, 500, 500)) # adjust size of window

      # Use separate calls of points3d() to plot points of each size
      for(i in seq_along(dataList)){
        if(type.3d == 'p'){with(dataList[[i]], points3d(U1, U2, U3,  col=colors[i],
                           size=pt.size*n*mean(dataList[[i]]$h.values)))}
        if(type.3d == 's'){with(dataList[[i]], spheres3d(U1, U2, U3, col=colors[i],
                           radius=pt.size*n*mean(dataList[[i]]$h.values)/500))}
      }

    } else{

      if (grid.method){
        data <- create.grid(g=g)
      }
      open3d()
      plot3d(data[,1], data[,2], data[,3], col=color, size=pt.size, type=type.3d,
             xlab=colnames(data)[1], ylab=colnames(data)[2], zlab=colnames(data)[3])
      par3d(windowRect = 50 + c(0, 0, 500, 500)) # adjust size of window
    } # End 'else' for adjust.pts.size
  } # End 'else' for 3D scatterplot
} # End function
```

Below is the R code producing the results from Section 3.C. Note this code uses function `pts.concentration` defined previously.

```
# Load external functions
source("3D_visualisation_functions.R")

# Set path to where results (MSE values and graphs) should be saved
graphs.direct <- "C:/Users/zzzzzzzz/Dropbox/Apps/Overleaf/PhD_Thesis/code3"
mse.direct    <- "C:/Users/zzzzzzzz/Dropbox/Apps/Overleaf/PhD_Thesis/code3"

# Function that runs simulations
# Arguments:
#     n.sim:     number of samples generated
#     my.n:      sample size
#     r1,r2:     parameters for the max and min values of the radius 'r'
#     spacing:   interval length between different values of 'r'
tuning.sim <- function(n.sim=100, my.n=200, r1=0.1, r2=0.2, spacing=0.002){
set.seed(1)
```

```
r <- seq(from = r1, to = r2, by = spacing)
l <- length(r)
# Matrix to contain MSE:
# rows represent different samples, columns represent different r's
MSE <- matrix(NA, nrow=n.sim, ncol=l)

for (k in 1:n.sim){
  X <- runif(my.n, 0, 1)
  Y <- runif(my.n, 0, 1)
  Z <- runif(my.n, 0, 1)
  my.data <- as.data.frame(cbind(X,Y,Z))
  # For different value of 'r', estimate f(h) at every point in the sample
  # Then compute the MSE
  for (j in 1:l){
    fh        <- pts.concentration(my.data, rad=r[j], return.h = F)
    MSE[k,j]  <- sqrt(mean((fh - 1/2)^2))
  }
}
# Mean MSE (for every r, averaged across all simulated samples)
mse     <- apply(MSE, 2, mean)
results <- cbind(r, mse, c(NA, 100*(mse[-1] - mse[-l])/mse[-l]))

# Save results
write.csv(results,
          file = paste(mse.direct, "/", "nsim", n.sim, "n", my.n, ".csv", sep=""))

# Plots of MSE versus 'r'
pdf(paste(graphs.direct, "/", "nsim", n.sim, "n", my.n, ".pdf", sep = ""),
    width = 12, height = 10)
par(mfrow=c(1,1))
par(mar = c(5, 5, 2, 2))
plot(r, results[,2], type = 'l', lty = 1, lwd = 3, col='brown3', xlab="r",
     ylab="MSE", cex.lab = 2, cex.axis = 2)
dev.off()

return(list(results, r[which.min(results[,2])], which.min(results[,2])))
}

# Launch simulations (takes many hours)
tuning.sim(n.sim=2000, my.n = 200,   r1=0.15, r2=.35)
tuning.sim(n.sim=2000, my.n = 300,   r1=0.10, r2=.30)
tuning.sim(n.sim=2000, my.n = 400,   r1=0.10, r2=.30)
tuning.sim(n.sim=2000, my.n = 500,   r1=0.10, r2=.40)
tuning.sim(n.sim=2000, my.n = 1000,  r1=0.07, r2=.25)
tuning.sim(n.sim=1000, my.n = 2000,  r1=0.05, r2=.23)
tuning.sim(n.sim=1000, my.n = 3000,  r1=0.05, r2=.23)
```

A.3 Codes for Chapter 4

Below is the R code producing Figures 4.1 and 4.2.

```
# Theoretical PMF of the sum S = X_1 + ... X_n
pois.S <- function(s,m){
  p.s <- 0
  for (k in 0:m){
    p.k <- (2*k^2+m^2-2*m*k-m)/2
    j   <- 0:p.k
    partial.sum <- choose(m,k)*ifelse(p.k<=s,1,0)*sum((-1)^j*(p.k-j)^s*choose(p.k,j))
    p.s = p.s + partial.sum
  }
  return(log(2)^s/factorial(s)*(1/2)^m*p.s)
}

# Function to  plot the PMF of S, compared to that of a Pois(n*log(2))
plot.pois <-function(m=3, max=m^2, legend=F, col.2='brown3', col.1='darkgoldenrod1'){
  domain <- 0:max
  probs  <- sapply(domain, pois.S, m = m)
  probs.pois <- sapply(domain, dpois, lambda = (m^2-m)/2*log(2))
  plot(domain,probs,type="h", lwd = 4, col= col.1,xlab="s",ylab="p(s)",
       cex.lab=1.5, cex.axis=1.5)
  points(domain,probs,col=col.1, cex = 1.5, pch = 19)
  points(domain,probs.pois, col=col.2,cex = 2.5, pch=18)
  abline(h=0,col='black')
  if(legend){legend("topright", NULL, ncol = 1, cex = 2, legend=c("S", "Poisson"),
         col=c(col.1, col.2), lwd = c(4,0), lty = c(1,0), pch = c(19,18),
         pt.cex = c(3,3.5))}
}

# Create (and save) three plots (for different sample sizes)
pdf(file = "poisson_pmf.pdf", width = 8, height = 11)
par(mfrow=c(3,1))
par(mar=c(3,6,1,1))
plot.pois(m=3, 7)
plot.pois(m=5, 18)
plot.pois(m=7, 34, legend=T)
dev.off()

# Density and CDF of the standardised sample mean (if sample size goes to infinity)
f.s <- function(s){
  s <- s + r*sqrt((ell-1)/2)
  k = (ell-1)/2
  theta = r*sqrt(2/(ell-1))
  b <- (2*(1-r^2))
  l.S <- function(x){x^(k-1)*exp(x*(2*s/b - 1/theta) - x^2/b)}
  return(exp(-s^2/b)/(sqrt(pi*b)*gamma(k)*theta^k)
```

```
        *integrate(l.S, lower = 0, upper = Inf, abs.tol = 10^(-20))$value)
}

F.S <- function(s){
  s <-  s + r*sqrt((ell-1)/2)
  k = (ell-1)/2
  theta = r*sqrt(2/(ell-1))
  l.S <- function(x){x^(k-1)*exp(-x/theta)*pnorm((s-x)/sqrt(1-r^2))}
  1/(gamma(k)*theta^k)*integrate(l.S, lower=0, upper = Inf, abs.tol=10^(-20))$value
}

sd <- 5
x <- matrix(seq(-2*sd/3, sd, by = 0.005), ncol = 1)
r    <- sqrt(log(2))
ell <- 2
hx.r08    <- apply(x, 1, f.s)
hxCDF.r08 <- apply(x, 1, F.S)

# Plot density and CDF, also compared to a N(0,1)
pdf(file = "poisson_CLT.pdf", width = 16, height = 8)
par(mfrow=c(1,2))
par(mar=c(3,6,1,1))
#Density
plot(x, hx.r08, type="l", lty=1, col = "brown3", lwd = 4, xlab="",
     ylab="Density", cex.lab = 1.5, cex.axis = 1.25)
curve(dnorm(x, 0, 1), col="black", lty = 1, lwd=2, add=TRUE)
#CDF
plot(x, hxCDF.r08, type="l", lty=1, col = "brown3", lwd = 4, xlab="",
     ylab="CDF", cex.lab = 1.5, cex.axis = 1.25)
curve(pnorm(x, 0, 1), col="black", lty = 1, lwd=2, add=TRUE)
legend("bottomright", NULL, ncol = 1, cex = 1.5, legend=c("Z", "N(0,1)"),
       col=c("brown3", "black"), lty = c(1, 1), lwd = c(7,3))
dev.off()
```

Below is the R code producing the results from Example 4.2.

```
# Load packages
library('moments')

# Function to generate the PIBD sample
piid.seq <- function(m=10){
  e   <- runif(1)
  eta <- runif(1)
  Z <- rep(NA, m)
  for (j in 1:m){Z[j] <- (eta + j*e)%%1}
  return(Z)
}

# Function to generate samples of S, for various choices of sample sizes 'n'
samples.of.S <- function(n = c(10, 100, 1000, 10000), B = 10000){
```

```
  set.seed(1)
  # Matrix to contain all samples of S
  # (rows represent samples: from one row to the next, 'n' increases)
  S    <- matrix(NA, nrow = length(n), ncol = B)
  kurt <- rep(NA, length(n))
  for (i in 1:length(n)){ # for all rows
    for (j in 1:B){ # for all columns within that row
      U <- piid.seq(n[i])
      S[i,j] <- sum(U)
    }
    kurt[i] <- kurtosis(S[i,])
  }
  plot(log(n), log(kurt), type="b", lty=1, col = "brown3", lwd = 6,
       xlab="log(n)", ylab="log(kurtosis)", cex.lab = 1.5, cex.axis = 1.25)
  return(summary(lm(kurt ~ n)))
}

# Launch function and save plot
# (warning: running time is many hours)
pdf(file = "increasing_kurtosis.pdf", width = 14, height = 8)
par(mar=c(5,5,1,1))
samples.of.S(n = c(10, 100, 1000, 10000, 100000), B = 3000000)
dev.off()
```

Below is the R code producing all simulation results from Section 4.2.2. Note this code uses function `generate.PI`, as defined already in Section A.2.

```
# Source needed function
source('generate_PI.R')

# Function to find the empirical TVaR of a sample ("data")
emp.TVaR <- function(data, alpha = c(0.75,0.95)){
  l    <- length(alpha)
  TVAR <- rep(NA, l)
  VAR  <- quantile(data, alpha)
  for (i in 1:l){TVAR[i] <- mean(data[data > VAR[i]])}
  return(TVAR)
}

# Function 'VaR.TVaR.S' estimates VaR, TVaR of X1+X2+X3 (for many PIBD examples)
# Arguments of this function are:
#   B          Number of simulated samples
#   alpha:     Vector of levels (for which VaR and TVaR are computed)
#   dep.type:  Name of the dependence example (as defined in function 'generate.PI')
#   param:     Value of the parameter within the example (if one is needed)
#   relative:  TRUE to divide all values by their corresponding values
#              under mutual independence
#   seed:      seed of the random generation
```

```
VaR.TVaR.S <- function(B=200000, alpha = c(0.70, 0.90, 0.95, 0.99, 0.995),
                       dep.type = 'indep', param=1, relative=T, seed=1){

  # Set seed
  set.seed(seed)

  # Data with uniform margins
  Udata <- as.matrix(generate.PI(n = B, type = dep.type, alpha = param))

  # Transform the data to different margins (always such that E[X]=1, Var[X]=1)
  U.data  <- qunif(Udata, min = 1-sqrt(3), max = 1+sqrt(3))
  N.data  <- qnorm(Udata, mean = 1, sd = 1)
  G.data  <- qgamma(Udata, shape = 1, rate = 1)
  LN.data <- qlnorm(Udata, -log(2)/2, sqrt(log(2)))

  # Get samples of "S", for all margins
  S.piid.U  <- apply(U.data, MARGIN = 1, sum)
  S.piid.N  <- apply(N.data, MARGIN = 1, sum)
  S.piid.G  <- apply(G.data, MARGIN = 1, sum)
  S.piid.LN <- apply(LN.data, MARGIN = 1, sum)

  # VaR
  VaR.1  <- quantile(S.piid.U, alpha)-3
  VaR.2  <- quantile(S.piid.N, alpha)-3
  VaR.3  <- quantile(S.piid.G, alpha)-3
  VaR.4  <- quantile(S.piid.LN, alpha)-3

  #TVaR
  TVaR.1 <- emp.TVaR(S.piid.U, alpha)-3
  TVaR.2 <- emp.TVaR(S.piid.N, alpha)-3
  TVaR.3 <- emp.TVaR(S.piid.G, alpha)-3
  TVaR.4 <- emp.TVaR(S.piid.LN, alpha)-3

  result <- c(VaR.1, VaR.2, VaR.3, VaR.4, TVaR.1, TVaR.2, TVaR.3, TVaR.4)
  ifelse(relative, return(round(result/VaR.TVaR.S(B, relative=F, seed=2),2)),
         return(result))
}

# Run the function for all examples
# (warning: running time is many hours)
VaR.TVaR.S(B=10^7, dep.type = 'triangles')
VaR.TVaR.S(B=10^7, dep.type = 'bernstein')
VaR.TVaR.S(B=10^7, dep.type = 'tetra')
VaR.TVaR.S(B=10^7, dep.type = 'cosine')
VaR.TVaR.S(B=10^7, dep.type = 'copula', param = 1)
VaR.TVaR.S(B=10^7, dep.type = 'copula', param = -1)
```

Below is the R code producing the results from Example 4.3.

```
# Generate a random sample as in the sequence in Janson1988 (from Remark 2)
```

```
janson.seq <- function(n=10){
e <- runif(1)
eta <- runif(1)
Z <- rep(NA,n)
for (j in 1:n){
    Z[j] <- (cos(2*pi*(eta+j*e))+1)/2
}
return(Z)
}

# Find distribution of the standardised maxima in the Janson1988 sequence
# We do this in a function, which returns the histogram and ECDF of -log(-M),
# (compared to the 'prediction' from the Fisher-Tippett THM)
# We use the log scale, because the distribution itself is very heavy-tailed

# Arguments of function 'EVT.comparison'
#   n            Vector of sample sizes
#   B:           Number of simulations (i.e. number of samples generated)
#   low/up:      Lower and upper limits for the histogram
#   seed:        Seed to reproduce simulation results
#   br:          Number of breaks for the histogram
#   compare:     TRUE for a comparison of the CDF for two sample sizes
#                (the last two values of vector n are used)
#   plot.means: TRUE to plot -log(-mean(H.n)) as function of log(n)

EVT.comparison <- function(n=c(10,1000,10000), B=50000, low = -15, up=15,
                           seed=1, br=200, compare=F, plot.means=T){
set.seed(seed)
# Number of different sample sizes
n.n <- length(n)
# Matrix to contain the samples
M   <- matrix(NA, nrow = B, ncol = n.n)

# Generation of the samples (of maxima)
# (different rows represent different samples)
# (different columns represent different sample sizes)
for (i in 1:B){
  U     <- janson.seq(tail(n,1))
  for (j in 1:n.n){
    M[i,j]  <- (max(U[1:n[j]])-1)*(2*n[j]/pi)^2
  }
}
# Log transformation
logH  <- -log(-M)

# Mean of M.n (for all sample sizes)
mean    <- apply(M, 2, mean)

# Plotting (and saving plot) log-transformed mean of M.n versus sample size
```

```
if(plot.means){
  pdf(file = paste("evt_mean_B", B, "n", n[n.n], ".pdf", sep=""),
      width = 14, height = 8)
  par(mfrow=c(1,1))
  par(mar=c(5,5,1,1))
  plot(log(n), -log(-mean), type="b", lty=1, col = "brown3", lwd = 6,
       xlab="log(n)", ylab=bquote(-log(-E(H[n]))), cex.lab = 1.5, cex.axis=1.25)
  dev.off()
}

# Plots of empirical distribution (histogram & ECDF)
# (vs. what is 'predicted' by Fisher-Tippet)
pdf(file = paste("evt_df_B", B, "n", n[n.n], ".pdf", sep=""), width=16, height=8)
par(mfrow=c(1,2))
par(mar=c(3,5,1,1))
hist(logH[,n.n], xlim = c(low, up), breaks = br, prob=T,
     cex.lab = 1.5, cex.axis =1.25, main = "", xlab="")
curve(0.5*exp(-exp(-x/2))*exp(-x/2), lwd = 2, n = 500,
      col = "brown3", add  = TRUE) # Density of -Log(-Weibull(1/2))

plot(ecdf(logH[,n.n]),  cex.lab = 1.5, cex.axis =1.25, do.points = F,
     col.01line = NULL, xlim = c(low, up), main = "", ylab="CDF", xlab="")
if(compare){
  plot(ecdf(logH[,n.n-1]), do.points = F, add = T, col = 'darkblue', lty = 2,
       xlim = c(low, up), main = "", ylab="CDF", xlab="")
}
else{
  curve(exp(-exp(-x/2)), lwd = 2, n = 500,
        col = "brown3", add  = TRUE) # Density of -Log(-Weibull(1/2))
  legend("bottomright", NULL, ncol = 1, cex = 1.5, legend=c("F-T_THM"),
         col=c("brown3"), lty = 1, lwd = 3)
}
dev.off()
}

# (warning: running time is many hours)
EVT.comparison(n=10^(1:5), B=3000000, seed=1)
```

Below is the `R` code producing all results from Section 4.2.4. Note this code uses function `piid.generator`, which is defined first (and is necessary to generate random samples from Example 4.1).

```
# Generator of pairwise independent observations
piid.generator <- function(m = 3, randF = rnorm,
                           indA = function(x) ifelse(x <= 0, FALSE, TRUE),
                           ell = 2) {

  # Check that the value of 'ell' provided is coherent
  #  with the function 'indA' provided
```

```
  # This check is valid only for moderate values of 'ell'
  if (round(1 / mean(indA(randF(10 ^ 5)))) != ell)
    warning("Is 'ell' consistent with your A?")

  # Sample size
  n <- choose(m, 2)

  # Generate the 'initial' multinomial sample of size m
  M <- rmultinom(m, 1, rep(1 / ell, ell))

  # Find all possible pairs out of the m multinomials
  # Use those pairs to create the 'n' D's in Eq. (2.4)
  combin <- combn(1:m, 2)
  D <- apply(M[, combin[1,]] == M[, combin[2,]], 2, all)
  D <- as.integer(D)
  # Compute the number of 1's among the D's
  pN <- sum(D)

  # Generate pN r.v.s with distribution F restricted to A
  # and (n - pN) r.v.s with distribution F restricted to A^c
  nU <- 0
  nV <- 0
  U <- rep(NA, n - pN)
  V <- rep(NA, pN)
  while((nU < n - pN) | (nV < pN)) {
    W <- randF(1)
    indAW <- indA(W)
    if (indAW & (nV < pN)) {
      nV <- nV + 1
      V[nV] <- W
    } else if ((indAW == 0) & (nU < n - pN)) {
      nU <- nU + 1
      U[nU] <- W
    }
  }
  # Return the resulting random generated variables
  X <- rep(NA, n)
  X[which(D == 0)] <- U
  X[which(D == 1)] <- V
  return(X)
}

# Source function 'piid.generator' (which allows to generate PIBD variables)
source('piid-generator.R')

# Functions needed for the generation of PIBD Poisson(log(2))
randF <- function(m) rpois(m, log(2))
indA  <- function(x) ifelse(x >= 1, TRUE, FALSE)
```

```
# Histogram for the empirical distribution of q.hat for the PIBD sequence
m     <- 10 # Controls the sample size: sample size = m(m-1)/2
B     <- 10000 # Number of samples
q     <- rep(NA, B)
set.seed(1)
for (i in 1:B){
  sample <- piid.generator(randF = randF, indA = indA, ell = 2L, m = m)
  q[i]   <- sum(sample == 0)/length(sample)
}

# Plot histogram
pdf(file = "piid_poisson_q_dist.pdf", width = 14, height = 8)
par(mar=c(5,5,1,1))
hist(q, probability = T, breaks = 20, main = '', xlab =  bquote(hat(q)),
     cex.lab = 1.75, cex.axis = 1.5, col = 'lightgreen')
abline(v = 1/2, col = 'black', lwd = 4, lty = 2)
legend("topleft", NULL, ncol = 1, cex = 2, legend=c("true q"),
       col=c("black"), lwd = 4, lty = c(2))
dev.off()

# Function 'boot.piid' produces bootstrapped re-samples of a dataset ('data')
# It computes a statistic (% of 0's) from the bootstrap samples and creates a CI
# It also checks whether the value (q=0.5) is inside the bootstrap CI

# Arguments of function 'boot.piid':
#   m:          For simulated data, controls the sample size n, with n = m(m-1)/2
#   B:          Number of bootstrapped samples
#   data:       Original sample of data to use
#               (if unspecified, a random mutually independent sample is generated)
#   histo:      TRUE to plot a histogram of the bootstrapped statistics
#   br:         Number of 'bins' for the histograms
#   alpha:      Level for the confidence interval

boot.piid <- function(m = 3, B = 500, data = '',
                      histo = F, br=30, alpha = 0.10){

  # For unspecified 'data', generate a random sample of independent Pois(log(2))
  if(length(data)==1){data <- rpois(m*(m-1)/2, log(2))}

  # Sample size of bootstrapped samples
  boot.size=length(data)

  # Matrix of bootstrapped samples (col = observations, row = bootstrap replicates)
  S <- matrix(NA, ncol = boot.size, nrow = B)
  for (i in 1:B){
    S[i,] <- sample(data, replace = T, size = boot.size)
  }

  # Bootstrapped replicates of the statistic
```

```
  q <- rep(NA, B) # initiate vector of bootstrapped statistics
  for (i in 1:B){
      q[i] = sum(S[i,] == 0)/boot.size
  }

  # Indicator equal to 1 if the real value is outside the bootstrap CI (L, 1)
  ind <- ifelse(1/2 < quantile(q, alpha), 1,0)

  # Histogram of the bootstrapped statistic 'q'
  if (histo == T){
    par(mar=c(5,5,1,1))
    hist(q, probability = T, breaks = br, main = '', xlab = bquote(hat(q)),
         cex.lab = 1.75, cex.axis = 1.5, col = 'lightgreen')
    abline(v = quantile(q, alpha), col = 'brown3', lwd = 4)
    abline(v = 1/2, col = 'black', lwd = 4, lty = 2)
    legend("topright", NULL, ncol = 1, cex = 2, legend=c("L", "true q"),
           col=c("brown3", "black"), lwd = 4, lty = c(1,2))
  }
  return(ind)

} # End function

# Call function 'boot.piid' to obtain the histogram of one bootstrapped sample
# (and save the plot)
# IID sample
pdf(file = "bootstrapCI_iid.pdf", width = 14, height = 8)
set.seed(1)
boot.piid(m=10, B = 5000, histo=T)
dev.off()
# PIBD sample
pdf(file = "bootstrapCI_piid.pdf", width = 14, height = 8)
set.seed(1)
boot.piid(data = piid.generator(randF = randF, indA = indA, ell = 2L, m = 10),
          B = 5000, histo = T)
dev.off()

# Function that calls 'boot.piid' a large number of times
# (to estimate the empirical level of the BCI)
boot.level <- function(nb.trials=10000, B=5000, m=3, indep=TRUE, seed=1){
  set.seed(seed)
  count<-0
  for (i in 1:nb.trials){
    if(indep){sample <- rpois(m*(m-1)/2, lambda = log(2))} #iid
    else{
      sample  <- piid.generator(randF = randF, indA = indA, ell = 2L, m = m) #PIBD
    }
    count <- count + boot.piid(m = m, B = B, data = sample, alpha = 0.1)
  }
  count/nb.trials
```

```
}

# Launch simulations
# (warning: takes a few hours)

# IID case
boot.level(nb=10000, B=5000, m=10)
boot.level(nb=10000, B=5000, m=45)

# PIBD case
boot.level(nb=10000, B=5000, m=10, indep=F)
boot.level(nb=10000, B=5000, m=45, indep=F)
```

A.4 Codes for Chapter 5

Below is the R code producing Figures 5.1 and 5.2.

```
# Density and CDF of S (for diffent values of r and \ell)
# They are found via the convolution:
# S = Normal(0, 1-r^2) + Gamma(shape = (ell-1)/2, scale = r*sqrt(2/(ell-1)) )
#     - r*sqrt((ell-1)/2)
f.s <- function(s){
  s <- s + r*sqrt((ell-1)/2)
  k = (ell-1)/2
  theta = r*sqrt(2/(ell-1))
  b <- (2*(1-r^2))
  I.S <- function(x){x^(k-1)*exp(x*(2*s/b - 1/theta) - x^2/b)}
  return(exp(-s^2/b)/(sqrt(pi*b)*gamma(k)*theta^k)
         * integrate(I.S, lower=0, upper=Inf, abs.tol=10^(-20))$value)
}

F.S <- function(s){
  s <-  s + r*sqrt((ell-1)/2)
  k = (ell-1)/2
  theta = r*sqrt(2/(ell-1))
  I.S <- function(x){x^(k-1)*exp(-x/theta)*pnorm((s-x)/sqrt(1-r^2))}
  1/(gamma(k)*theta^k) * integrate(I.S, lower=0, upper=Inf, abs.tol=10^(-20))$value
}

# Plot df and CDF for many values of 'r' or many values of 'ell'
par(mfrow=c(1,2))
par(mar=c(3,6,1,1))
sd <- 5
x <- matrix(seq(-2*sd/3, sd, by = 0.005), ncol = 1)

# Creating figures: either fix 'r' and change 'ell', or the other way around
#r   <- .95
r    <- .9
ell <- 3
hx.r095    <- apply(x, 1, f.s)
hxCDF.r095 <- apply(x, 1, F.S)
#r   <- .8
ell <- 6
hx.r08    <- apply(x, 1, f.s)
hxCDF.r08 <- apply(x, 1, F.S)
#r   <- .6
ell <- 15
hx.r06    <- apply(x, 1, f.s)
hxCDF.r06 <- apply(x, 1, F.S)

#Density
```

```
plot(x, hx.r095, type="l", lty=1, col = "darkorange", lwd = 4, xlab="",
     ylab="Density", cex.lab = 2.25, cex.axis = 2)
lines(x, hx.r08, type="l", lty=3, col = "blueviolet", lwd = 5)
lines(x, hx.r06, type="l", lty=6, col = "cornflowerblue", lwd = 4)
curve(dnorm(x, 0, 1), col="black", lty = 1, lwd=2, add=TRUE)

#CDF
plot(x, hxCDF.r095, type="l", lty=1, col = "darkorange", lwd = 4, xlab="",
     ylab="CDF", cex.lab = 2.25, cex.axis = 2)
lines(x, hxCDF.r08, type="l", lty=3, col = "blueviolet", lwd = 5)
lines(x, hxCDF.r06, type="l", lty=6, col = "cornflowerblue", lwd = 4)
curve(pnorm(x, 0, 1), col="black", lty = 1, lwd=2, add=TRUE)
legend("bottomright", NULL, ncol = 1, cex = 2,
       legend=c("l = 3", "l = 6", "l = 15", "N(0,1)"),
       col=c("darkorange", "blueviolet", "cornflowerblue", "black"),
       lty = c(1,3,6, 1), lwd = c(7,7,7,3))
```

We present below `R` codes that allow one to generate pseudorandom samples from all our pairwise independent examples (i.e., Example 5.1 and the additional examples of Section 5.B). The main function for the random generation is called `piid.generator`, and was defined already in Section A.3. The first argument of this function `m` determines the sample size n through $n = m(m-1)/2$. Its other arguments `randF`, `indA` and `ell` all depend on the specific example considered. We provide below those arguments for all our examples.

```
# Specific values of randF, indA and ell for all examples

# Example 5.1
randF <- function(m, beta) rlnorm(m, 0, beta)
indA  <- function(x) ifelse(x >= 1, TRUE, FALSE)
piid.generator(randF = function(m) randF(m, beta = 2), indA = indA, ell = 2L)

# Example 5.3
r.ex6 <- function(rand, ell) {
  if (rand < 1 / (2 * ell)) {
    res <- -1L
  } else if (rand < 1 / ell) {
    res <- 1L
  } else if (rand < 1 / (2 * ell) + 1 / 2) {
    res <- -2L
  } else {
    res <- 2L
  }
  return(res)
}
randF <- function(m, ell) sapply(runif(m), FUN = r.ex6, ell = ell)
indA  <- function(x) ifelse((x == 1L) | (x == -1L), TRUE, FALSE)
```

```
piid.generator(randF = function(m) randF(m, ell = 2L), indA = indA, ell = 2L)

# Example 5.4
randF <- function(m) runif(m, -1, 1)
indA  <- function(x, ell) ifelse((x >= -1 / ell) & (x <= 1 / ell), TRUE, FALSE)
piid.generator(randF = randF, indA = function(x) indA(x, ell = 2L), ell = 2L)

# Example 5.5
randF <- function(m, ell) as.integer(2 * rbinom(m, 1, 1 / ell) - 1)
indA  <- function(x) ifelse(x == 1L, TRUE, FALSE)
piid.generator(randF = function(m) randF(m, ell = 10L), indA = indA, ell = 10L)

# Example 5.6
F <-  function(x, ell, sigma) sum(c(1 - 1 / ell, 1 / ell) *
                        pnorm(x, mean = c(-1 / ell, 1 - 1 / ell),
                              sd = c(sigma, sigma)))
Finv  <- function(p, ell, sigma = 1/(4*ell)){
  G = function(x) F(x, ell, sigma) - p
  return(uniroot(G, c(-100,100))$root)
}
r.ex9 <- function(rand, ell, sigma) {
  if (rand < 1 / ell) {
    res <- rnorm(1, 1 - 1 / ell, sigma)
  } else {
    res <- rnorm(1, -1 / ell, sigma)
  }
  return(res)
}
randF <- function(m, ell, sigma = 1/(4*ell))
          sapply(runif(m), FUN = r.ex9, ell = ell, sigma = sigma)
indA  <- function(x, ell, sigma = 1/(4*ell))
          ifelse(x >= Finv(1 - 1 / ell, ell = ell, sigma = sigma), TRUE, FALSE)
piid.generator(randF = function(m) randF(m, ell = 3L),
               indA = function(x) indA(x, ell = 3L), ell = 3L)

# Example 5.7
randF <- function(m, mu, sigma) rnorm(m, mu, sigma)
indA  <- function(x, mu) ifelse(x >= mu, TRUE, FALSE)
piid.generator(randF = function(m) randF(m, mu = 2, sigma = 1),
              indA = function(x) indA(x, mu = 2), ell = 2L)
```

Below is the R code producing Figure 5.3.

```
# Plot a shifted Log-norm with different values of sigma
# (standardised to have mean 0, sd 1)
# This highlights what it means for 'r' to be close to 0 or close to 1
Erf    <- function(x){2 * pnorm(x * sqrt(2)) - 1}
# Find r parameter (as function of sigma)
r.lnorm      <- function(sig){Erf(sig/sqrt(2))/sqrt(exp(sig^2)-1)}
med.lnorm    <- function(sig){(exp(sig^2)-1)^(-1/2) * (exp(-sig^2/2)-1)}
```

```
# Function to graph density, as function of 'sig'
graph.logN <- function(sig = sqrt(0.1480239), legend = T){

# 'mu' parameter of the Log-normal (as function of 'sig') (such that Var[X]=1)
mu      <- (- sig^2 -log(exp(sig^2)-1))/2
# EX (also function of 'sig'), such that (X - EX) has mean 0 and unit variance
EX      <- sqrt(1/(exp(sig^2)-1))
# median
med           <- exp(mu) - EX
r             <- r.lnorm(sig)
# because we deal with a standard distribution, mu_V = r
mu_V  <- r
x       <- seq(-2, 2.5, by = 0.002)
df      <- dlnorm(x+EX, meanlog = mu, sdlog = sig)
cdf     <- plnorm(x+EX, meanlog = mu, sdlog = sig)

par(mar = c(2, 5, 2, 1))
plot(x, df, type = "l", col="brown3", lwd=3,
     main = paste("r =", round(r,3)), ylim = c(-max(df)/15, max(df)),
     xlab="", ylab="Density", cex.lab = 2.25, cex.axis = 2)
# lines at the median and mean
abline(v = med, lwd = 2, col = 'darkgoldenrod1')
abline(v = 0, lty = 3, lwd = 2)
abline(h=0)

ps <- matrix(c(0, 0, mu_V, 0), ncol = 2, byrow = T)
points(ps, col=c("steelblue4"), pch=c("|"), cex=1.7)
text(ps, col=c("steelblue4"), labels= c("", "r"), adj = c(.25,-.6), cex = 2.5)
arrows(x0 = 0, y0 = 0, x1 = mu_V, y1 = 0, length = 0.2, code = 2, lwd = 3,
       col = 'steelblue4')
ifelse(legend, legend("topright", NULL, ncol = 1,
               cex = 1.75, legend=c("median", "mean"),
               col = c("darkgoldenrod1", "black"), lty = c(1,3), lwd = 5), "")
}

par(mfrow=c(3,1))

graph.logN(sig = sqrt(0.1480239))
graph.logN(sig = sqrt(0.6733226), legend = F)
graph.logN(sig = sqrt(1.596642),  legend = F)
```

Below is the R code producing Figures 5.4, 5.5 and 5.6.

```
# Analysing the "r" parameter as measure of tail-heaviness for common distributions
# The analysis is done for r* (i.e. 1-r) so that larger r* imply heavier-tail

# Log Normal(mu, s) (s is like sigma^2)
# min r: 1-sqrt(2/pi) = 0.2022
# max r: 1
```

```
erf         <- function(x){2 * pnorm(x * sqrt(2)) - 1}
r.logN      <- function(s){1-erf(sqrt(s/2))/sqrt(exp(s)-1)}
r.logN.inv <- function(r){uniroot(function(x){r.logN(x) - r}, lower = 0.00001,
                                  upper=100)$root}
k.logN      <- function(s){exp(4*s) + 2*exp(3*s) + 3*exp(2*s) - 6  +3}
k.logN.inv <- function(k){uniroot(function(x){k.logN(x) - k}, lower = 0.0001,
                                  upper=100)$root}
RQW.logN    <- function(s,q=7/8){(qlnorm(0.5+q/2, sdlog=sqrt(s))
                                 +qlnorm(1-q/2, sdlog=sqrt(s))
                                 -2*qlnorm(3/4, sdlog=sqrt(s)))/(qlnorm(0.5+q/2,
                                  sdlog=sqrt(s))-qlnorm(1-q/2, sdlog=sqrt(s)))}
RQW.logN.inv <- function(k){uniroot(function(x){RQW.logN(x) - k},
                                    lower = 0.0001, upper=1000)$root}

# Gamma with shape parameter a
# min r: 1-sqrt(2/pi) = 0.2030 (seems to have numerical problems to evaluate at lower r)
# max r: 1
r.Gam <- function(a){
  mx <- qgamma(1/2, shape = a, rate = 1)
  1-2*mx^a*exp(-mx)/(gamma(a)*sqrt(a))
}
r.Gam.inv <- function(r){uniroot(function(x){r.Gam(x) - r},
                                 lower = .0001, upper=130)$root}
k.Gam       <- function(a){6/a + 3}
k.Gam.inv <- function(k){uniroot(function(x){k.Gam(x) - k},
                                 lower = .0001, upper=130)$root}
RQW.Gam     <- function(a,q=7/8){(qgamma(0.5+q/2, shape=a)+qgamma(1-q/2, shape=a)
                                 -2*qgamma(3/4, shape=a))/(qgamma(0.5+q/2, shape=a)
                                                           -qgamma(1-q/2, shape=a))}
RQW.Gam.inv <- function(r){uniroot(function(x){RQW.Gam(x) - r},
                                   lower = .001, upper=500)$root}

# Weibull with shape parameter k
# min r: 0.191 (roughly, numerically)
# max r: 1
r.Wei       <- function(k){1-(2*igamma(1+1/k,log(2)) - gamma(1+1/k))/sqrt(gamma(1+2/k)
                                                                         - gamma(1+1/k)^2)}
r.Wei.inv <- function(r){uniroot(function(x){r.Wei(x) - r}, lower = 0.05, upper=2.9)$root}
G1          <- function(k){gamma(1+1/k)}
G2          <- function(k){gamma(1+2/k)}
G3          <- function(k){gamma(1+3/k)}
G4          <- function(k){gamma(1+4/k)}
k.Wei       <- function(k){(-6*G1(k)^4+12*G1(k)^2*G2(k)-3*G2(k)^2-4*G1(k)*G3(k)
                           +G4(k))/(G2(k)-G1(k)^2)^2 + 3}
k.Wei.inv <- function(k){uniroot(function(x){k.Wei(x) - k}, lower = 0.05, upper=2.9)$root}
RQW.Wei     <- function(k,q=7/8){(qweibull(0.5+q/2, shape=k)+qweibull(1-q/2, shape=k)
                                 -2*qweibull(3/4, shape=k))/(qweibull(0.5+q/2, shape=k)
                                                             -qweibull(1-q/2, shape=k))}
RQW.Wei.inv <- function(r){uniroot(function(x){RQW.Wei(x) - r}, lower = 0.05,
```

```
                                        upper=2.9)$root}

# Frechet
# min r: 0.2453 (roughly, numerically)
# max r: 1
# (kurtosis does not exist for r > 0.4013)
igamma <- function(a, x){gamma(a) * (1 - pgamma(x,a,1))}
r.Fre  <- function(a){1-(gamma(1-1/a) -
                         2*igamma(1-1/a, log(2)))/sqrt(gamma(1-2/a)-gamma(1-1/a)^2)}
r.Fre.inv <- function(r){uniroot(function(x){r.Fre(x) - r}, lower = 2.0001,
                                 upper=10000)$root}
k.Fre  <- function(a){(gamma(1-4/a)-4*gamma(1-3/a)*gamma(1-1/a)
                       +3*gamma(1-2/a)^2)/(gamma(1-2/a)-gamma(1-1/a)^2)^2-6  +3}
k.Fre.inv <- function(k){uniroot(function(x){k.Fre(x) - k}, lower = 4.0001,
                                 upper=1000)$root}
qfrechet <- function(q,a){log(1/q)^{-1/a}}
RQW.Fre  <- function(a,q=7/8){(qfrechet(0.5+q/2, a)+qfrechet(1-q/2, a)
                               -2*qfrechet(3/4, a))/(qfrechet(0.5+q/2, a)
                                                     -qfrechet(1-q/2, a))}
RQW.Fre.inv <- function(rqw){uniroot(function(x){RQW.Fre(x) - rqw}, lower = 0.2,
                                     upper=1000)$root}

# Find a in Pareto(a, lambda) (a>2)
# min r: 1-log(2) = 0.3069
# max r: 1
# (Kurtosis does not exist past r > 0.4648)
r.Par     <- function(a){1-sqrt(a*(a-2))*(2^(1/a)-1) }
r.Par.inv <- function(r){uniroot(function(x){r.Par(x) - r}, lower = 2.00001,
                                 upper=10000)$root}
k.Par     <- function(a){6*(a^3+a^2-6*a-2)/(a*(a-3)*(a-4))  +3}
k.Par.inv <- function(k){uniroot(function(x){k.Par(x) - k}, lower = 4.00001,
                                 upper=10000)$root}
qpareto   <- function(q,a){(1-q)^(-1/a)-1}
RQW.Par   <- function(a,q=7/8){(qpareto (0.5+q/2, a)+qpareto (1-q/2, a)
                                -2*qpareto (3/4, a))/(qpareto (0.5+q/2, a)
                                                      -qpareto(1-q/2, a))}
RQW.Par.inv <- function(r){uniroot(function(x){RQW.Par(x) - r}, lower = .1,
                                   upper=1000)$root}

# Student (v) (v > 2)
# min r: 1-sqrt(2/pi) = 0.2028 (seems to have numerical problems to evaluate at lower r)
# max r: 1
# (Kurtosis does not exist past r > 0.2928)
r.t     <- function(v){1-2*sqrt((v-2)/pi)*gamma((v+1)/2)/((v-1)*gamma(v/2))}
r.t.inv <- function(r){uniroot(function(x){r.t(x) - r}, lower = 2.00001,
                               upper=340)$root}
# Excess Kurtosis
k.t     <- function(v){6/(v-4)  +3}
k.t.inv <- function(k){uniroot(function(x){k.t(x) - k}, lower = 4.00001,
```

```
                                    upper=340)$root}
RQW.t    <- function(v,q=7/8){(qt(0.5+q/2, df=v)+qt(1-q/2, df=v)
                                    -2*qt(3/4, df=v))/(qt(0.5+q/2, df=v)-qt(1-q/2, df=v))}
RQW.t.inv <- function(r){uniroot(function(x){RQW.t(x) - r}, lower = 0.05,
                                    upper=340)$root}

# Sequence of values of Kurtosis (NOT excess kurtosis)
n          <- 1000
range.kurt <- seq(from = 10, to = 300, length.out=n)

# Sequence of values of r
r.end      <- 0.95
# maximal ranges for all distributions
range.Fre  <- seq(from=0.2453,  to = r.end, length.out=n)
range.LogN <- seq(from=0.2022,  to = r.end, length.out=n)
range.Gam  <- seq(from=0.2030,  to = r.end, length.out=n)
range.Wei  <- seq(from=0.191,   to = r.end, length.out=n)
range.t    <- seq(from=0.2028,  to = r.end, length.out=n)
range.Par    <- seq(from=0.3069,  to = r.end, length.out=n)

# additional ranges such that the kurtosis exists
range.2.Fre  <- seq(from=0.2453,  to = 0.4013, length.out=n)
range.2.t    <- seq(from=0.2028,  to = 0.2928, length.out=n)
range.2.Par    <- seq(from=0.3069,  to = 0.4648, length.out=n)

# range of RQW
range.RQW <- seq(from=0.3,  to = .995, length.out=n)

# Colors to be used for plots
my.col = c("brown3","cornflowerblue", "blueviolet", "darkgoldenrod1", "darkgreen",
           "pink")

# Initialise sequences of parameters and kurtosis
logN.para  <- rep(NA, n)
k.logN.vec <- rep(NA, n)
RQW.logN.vec <- rep(NA, n)

Wei.para  <- rep(NA, n)
k.Wei.vec <- rep(NA, n)
RQW.Wei.vec <- rep(NA, n)

Gam.para  <- rep(NA, n)
k.Gam.vec <- rep(NA, n)
RQW.Gam.vec <- rep(NA, n)

t.para  <- rep(NA, n)
k.t.vec <- rep(NA, n)
RQW.t.vec <- rep(NA, n)
```

```
Par.para  <- rep(NA, n)
k.Par.vec <- rep(NA, n)
RQW.Par.vec <- rep(NA, n)

Fre.para  <- rep(NA, n)
k.Fre.vec <- rep(NA, n)
RQW.Fre.vec <- rep(NA, n)

# Finding the "k:kurtosis" corresponding to specific values of "RQW"
for (i in 1:n){
  logN.para[i]  <- RQW.logN.inv(range.RQW[i])
  k.logN.vec[i] <- k.logN(logN.para[i])

  Wei.para[i]   <- RQW.Wei.inv(range.RQW[i])
  k.Wei.vec[i]  <- k.Wei(Wei.para[i])

  Gam.para[i]   <- RQW.Gam.inv(range.RQW[i])
  k.Gam.vec[i]  <- k.Gam(Gam.para[i])

}

# Plots: kurtosis versus RQW (log scale)
pdf(file = "k.vs.RQW0875.pdf", width = 8, height = 6)
y.range <- c(min(log(k.logN.vec),log(k.Wei.vec), log(k.Gam.vec)),
             max(log(k.logN.vec),log(k.Wei.vec), log(k.Gam.vec)))
par(mar = c(5, 5, 2, 1))
plot(range.RQW,   log(k.logN.vec), type = 'l', lwd=3, ylim = y.range,
     lty = 1, col=my.col[1], xlab="RQW", ylab="log(kurtosis)", cex.lab = 1.5,
     cex.axis = 1.25)
lines(range.RQW, log(k.Wei.vec), lwd = 3, lty=2, col=my.col[2])
lines(range.RQW, log(k.Gam.vec), lwd = 3, lty=3, col=my.col[3])
legend('topleft', ncol=1, legend=c("LogN","Weibull","Gamma"), col=my.col[1:3],
       lty = 1:3, lwd = 5, cex = 1.5)
# Saving plot
dev.off()

# Finding the "k:kurtosis" corresponding to specific values of "r*"
for (i in 1:n){
  logN.para[i]  <- r.logN.inv(range.LogN[i])
  k.logN.vec[i] <- k.logN(logN.para[i])

  Wei.para[i]   <- r.Wei.inv(range.Wei[i])
  k.Wei.vec[i]  <- k.Wei(Wei.para[i])

  Gam.para[i]   <- r.Gam.inv(range.Gam[i])
  k.Gam.vec[i]  <- k.Gam(Gam.para[i])

  Par.para[i]   <- r.Par.inv(range.2.Par[i])
  k.Par.vec[i]  <- k.Par(Par.para[i])
```

```
  Fre.para[i]   <- r.Fre.inv(range.2.Fre[i])
  k.Fre.vec[i]  <- k.Fre(Fre.para[i])

  t.para[i]     <- r.t.inv(range.2.t[i])
  k.t.vec[i]    <- k.t(t.para[i])
}


# Plots: kurtosis versus r (log scale)
pdf(file = "k.vs.r.pdf", width = 8, height = 6)

y.range <- c(min(log(k.logN.vec),log(k.Wei.vec), log(k.Gam.vec), log(k.Par.vec),
                 log(k.Fre.vec)),
             max(log(k.logN.vec),log(k.Wei.vec), log(k.Gam.vec), log(k.Par.vec),
                 log(k.Fre.vec)))
par(mar = c(5, 5, 2, 1))
plot(range.LogN,   log(k.logN.vec), type = 'l', lwd=3, ylim = y.range,
     lty = 1, col=my.col[1], xlab="r*", ylab="log(kurtosis)", cex.lab = 1.5,
     cex.axis = 1.25)
lines(range.Wei, log(k.Wei.vec), lwd = 3, lty=2, col=my.col[2])
lines(range.Gam, log(k.Gam.vec), lwd = 3, lty=3, col=my.col[3])
legend('topleft', ncol=1, legend=c("LogN","Weibull","Gamma"), col=my.col[1:5],
       lty = 1:5, lwd = 5, cex = 1.5)
# Saving plot
dev.off()

# Finding the "RQW" corresponding to specific values of "r*"
for (i in 1:n){
  logN.para[i]    <- r.logN.inv(range.LogN[i])
  RQW.logN.vec[i] <- RQW.logN(logN.para[i])

  Wei.para[i]     <- r.Wei.inv(range.Wei[i])
  RQW.Wei.vec[i]  <- RQW.Wei(Wei.para[i])

  Gam.para[i]     <- r.Gam.inv(range.Gam[i])
  RQW.Gam.vec[i]  <- RQW.Gam(Gam.para[i])

  Par.para[i]     <- r.Par.inv(range.Par[i])
  RQW.Par.vec[i]  <- RQW.Par(Par.para[i])

  Fre.para[i]     <- r.Fre.inv(range.Fre[i])
  RQW.Fre.vec[i]  <- RQW.Fre(Fre.para[i])

  t.para[i]       <- r.t.inv(range.t[i])
  RQW.t.vec[i]    <- RQW.t(t.para[i])
}

# Plots: RQW versus r
```

```
pdf(file = "RQW.vs.r.pdf", width = 8, height = 6)
y.range <- c(0.0,1)
par(mar = c(5, 5, 2, 1))
plot(range.LogN,    (RQW.logN.vec), type = 'l', lwd=3, ylim = y.range,
     lty = 1, col=my.col[1], xlab="r*", ylab="RQW", cex.lab = 1.5, cex.axis = 1.25)
lines(range.Wei, (RQW.Wei.vec), lwd = 3, lty=2, col=my.col[2])
lines(range.Gam, (RQW.Gam.vec), lwd = 3, lty=3, col=my.col[3])
lines(range.Par, (RQW.Par.vec), lwd = 3, lty=4, col=my.col[4])
lines(range.Fre, (RQW.Fre.vec), lwd = 3, lty=5, col=my.col[5])
lines(range.t, RQW.t.vec, lwd = 3, lty=6, col=my.col[6])
legend('topleft', ncol=1,
       legend=c("LogN","Weibull","Gamma","Pareto","Frechet", "Student"),
       col=my.col[1:6], lty = 1:6, lwd = 5, cex = 1.5)
# Saving plot
dev.off()
```

A.5 Codes for Chapter 6

Below is the R code producing Figures 6.2 and 6.3.

```
# density function for the VG(n, 0, 1, 0) distribution
density_VG <- function (x, n) {
  return (1 / (sqrt(pi) *  gamma(n / 2)) * (abs(x) / 2) ^ ((n - 1) / 2) *
            besselK(abs(x), (n - 1) / 2))
}

# density of the convolution in the main theorem
f.s <- function (s) {
  convolution_density <- function(y) {dnorm(y / sqrt(1 - r ^ 2)) / sqrt(1 - r ^ 2)*
      density_VG((s - y) / (r / sqrt(ell - 1)), ell - 1) / (r / sqrt(ell - 1))}
  return (integrate(convolution_density, lower = -Inf, upper = s,
                    abs.tol = 10 ^ (-20))$value +
            integrate(convolution_density, lower = s, upper = Inf,
                      abs.tol = 10 ^ (-20))$value)
}

# cdf of the convolution in the main theorem
F.S <- function (s) {
  convolution_cdf <- function(y) {pnorm(y / sqrt(1 - r ^ 2)) *
      density_VG((s - y) / (r / sqrt(ell - 1)), ell - 1) / (r / sqrt(ell - 1))}
  return (integrate(convolution_cdf, lower = -Inf, upper = s,
                    abs.tol = 10 ^ (-20))$value +
            integrate(convolution_cdf, lower = s, upper = Inf,
                      abs.tol = 10 ^ (-20))$value)
}


# Plot df and CDF for many values of 'r' or many values of 'ell'
par(mfrow = c(1, 2))
par(mar = c(3, 6, 1, 1))
x <- matrix(seq(-3, 3, by = 0.005), ncol = 1)

# To create Figures: either fix r and change 'ell', or the other way around
r <- 0.99
ell <- 2
hx.r099    <- apply(x, 1, f.s)
hxCDF.r099 <- apply(x, 1, F.S)

#r <- 0.8
ell <- 4
hx.r08    <- apply(x, 1, f.s)
hxCDF.r08 <- apply(x, 1, F.S)

#r <- 0.6
```

```
ell <- 6
hx.r06    <- apply(x, 1, f.s)
hxCDF.r06 <- apply(x, 1, F.S)

#Density
plot(x, hx.r099, type = "l", lty = 1, col = "darkorange", lwd = 4,
     xlab="", ylab="Density", cex.lab = 2.25, cex.axis = 2)
lines(x, hx.r08, type = "l", lty = 3, col = "blueviolet", lwd = 5)
lines(x, hx.r06, type = "l", lty = 6, col = "cornflowerblue", lwd = 4)
curve(dnorm(x, 0, 1), col="black", lty = 1, lwd=2, add=TRUE)

#CDF
plot(x, hxCDF.r099, type = "l", lty = 1, col = "darkorange", lwd = 4,
     xlab="", ylab="CDF", cex.lab = 2.25, cex.axis = 2)
lines(x, hxCDF.r08, type = "l", lty = 3, col = "blueviolet", lwd = 5)
lines(x, hxCDF.r06, type = "l", lty = 6, col = "cornflowerblue", lwd = 4)
curve(pnorm(x, 0, 1), col="black", lty = 1, lwd=2, add=TRUE)
legend("bottomright", NULL, ncol = 1, cex = 1.1,
       legend=c("r = 0.99", "r = 0.8", "r = 0.6", "N(0,1)"),
       col=c("darkorange", "blueviolet", "cornflowerblue", "black"),
       lty = c(1,3,6,1), lwd = c(4,5,4,2))
```

Below is the R codes which produced the simulations results of Section 6.5.

```
##############################################
## Incidence matrix of projective planes ##
##############################################

require(nortest) # for the Anderson-Darling and Pearson tests at the end

## the graph is bipartite, (q+1)-regular, has girth 6, diameter 3
## and 2*(q^2 + q + 1) vertices
## (q must be a prime power)

## Sigma matrix

sigma <- function (u, q) {
  res <- (u * (matrix(rep(0:(q - 1), q), nrow = q)
           + t(matrix(rep(0:(q - 1), q), nrow = q)))) %% q
  return(replace(res, res == 0, q))
}

## Position matrix of A

pm <- function (A, q) {
  res <- matrix(as.integer(A == 1), nrow = q)
  for (i in 2:q) {
    res <- cbind(res, matrix(as.integer(A == i), nrow = q))
  }
  return(res)
```

```
}

## Position matrix of a family F

pmf <- function (q) {
  res <- pm(sigma(1, q), q)
  for (i in 2:(q - 1)) {
    res <- rbind(res, pm(sigma(i, q), q))
  }
  return(res)
}

## Incidence matrix of the bipartite graph

inc_mat <- function (q) {
  res1 <- rbind(pmf(q), do.call(cbind, replicate(q, diag(q), simplify=FALSE)),
                pm(matrix(rep(1:q, q), nrow = q), q), rep(0, q ^ 2))
  res2 <- rbind(t(pm(matrix(rep(1:(q + 1), q), nrow = q), q + 1)), rep(1, q + 1))
  return(cbind(res1, res2))
}

## Adjacency matrix

adj_mat <- function (q) {
  n <- q ^ 2 + q + 1
  zz <- matrix(0, nrow = n, ncol = n)
  im <- inc_mat(q)
  return(rbind(cbind(zz, im), cbind(t(im), zz)))
}

## Computation of the standardized rv Z

stand_rand <- function (q) {
  n_vert <- 2 * (q ^ 2 + q + 1) ## number of vertices in the graph
  n_edges <- n_vert * (q + 1) / 2 ## number of edges in the graph
  vec <- rbinom(n_vert, 1, 1 / 2)
  mat <- matrix(rep(vec, n_vert), n_vert)
  res <- 1 - ((mat + t(mat)) %% 2)
  x <- res * adj_mat(q) ## generates rvs on the edges
  return((sum(x) / 2 - n_edges / 2) / sqrt(n_edges / 4))
}

## Monte-Carlo histogram

q <- 2 ^ 6
sim <- 5000
z <- rep(0, sim)

for (i in 1:sim) {
```

```
  z[i] <- stand_rand(q)
}
hist(z)
qqnorm(z, pch = 1, frame = FALSE)
qqline(z, col = "steelblue", lwd = 2)

shapiro.test(z)
ad.test(z)
pearson.test(z)
```