KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0216

THE CONCEPT OF COMONOTONICITY IN ACTUARIAL SCIENCE AND FINANCE: THEORY

by
J. DHAENE
M. DENUIT
M. GOOVAERTS
R. KAAS
D. VYNCKE

D/2002/2376/16

The Concept of Comonotonicity in Actuarial Science and Finance: Theory

J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, D. Vyncke

December 5, 2001

Abstract

In an insurance context, one is often interested in the distribution function of a sum of random variables. Such a sum appears when considering the aggregate claims of an insurance portfolio over a certain reference period. It also appears when considering discounted payments related to a single policy or a portfolio at different future points in time. The assumption of mutual independence between the components of the sum is very convenient from a computational point of view, but sometimes not realistic. We will determine approximations for sums of random variables, when the distributions of the terms are known, but the stochastic dependence structure between them is unknown or too cumbersome to work with. In this paper, the theoretical aspects are considered. Applications of this theory are considered in a subsequent paper. Both papers are to a large extent an overview of recent research results obtained by the authors, but also new theoretical and practical results are presented.

1 Introduction

In traditional risk theory, the individual risks of a portfolio are usually assumed to be mutually independent. Standard techniques for determining the distribution function of aggregate claims, such as Panjer's recursion, De Pril's recursion, convolution or moment based approximations, are based on the independence assumption. Insurance 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.

A trend in actuarial science is to combine the (actuarial) technical risk with the (financial) investment risk. Assume that the nominal random payments X_i are due at fixed and known times t_i, $i = 1, 2, \ldots, n$. Let Y_t denote the nominal discount factor over the interval $[0, t]$, $t \geq 0$. This means that the amount one needs to invest at time 0 to get an amount 1 at time t is the random variable Y_t. By nominal we mean that there is no correction for inflation. In this case, a random variable of interest will be the scalar product of two random vectors:

$$S = \sum_{i=1}^{n} X_i Y_{t_i}.$$

If the payments X_i at time t_i are independent of inflation, then the vectors $\underline{X} = (X_1, X_2, \ldots, X_n)$ and $\underline{Y} = (Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ can be assumed to be mutually independent. On the other hand, if the payments are adjusted for inflation, the vectors $\underline{X}$ and $\underline{Y}$ are not mutually independent anymore. Denoting the inflation factor over the period $[0, t]$ by Z_t, the random variable S can in this case be rewritten as $S = \sum_{i=1}^{n} X_i^{'} Y_{t_i}^{'}$ where the real payments $X_i^{'}$ and the real discount factors $Y_{t_i}^{'}$ are given by $X_i^{'} = X_i / Z_{t_i}$ and $Y_{t_i}^{'} = Y_{t_i} \; Z_{t_i}$ respectively. Hence, in this case S is the scalar product of the two mutually independent random vectors $(X_1^{'}, X_2^{'}, \ldots, X_n^{'})$ and $(Y_{t_1}^{'}, Y_{t_2}^{'}, \ldots, Y_{t_n}^{'})$.

In general however, each vector on its own will have dependent components. Especially the factors of the discount vector will possess a strong positive dependence.

Introduction of the stochastic financial aspects in actuarial models immediately reveals the necessity of determining distribution functions of sums of dependent random variables. Hereafter we describe some situations where random variables, which are scalar products of two vectors, arise.

First, consider the random variable $S = \sum_{i=1}^{n} X_i Y_i$, where the X_i represent the claim amounts of one policy (or one portfolio) at different times i, $i = 1, 2, \ldots, n$. Even if the discount factors Y_i are deterministic, S will often be a sum of dependent random variables in this case. An example is a life annuity on a single life (x) which pays an amount equal to 1 at times $1, 2, \ldots, n$ provided the insured (x) is alive at that time. It is clear that the stochastic payments X_i possess a strong positive dependence in this case. Another example is the case of an individual automobile insurance policy where X_i represents the loss in year i of the policy under consideration. High values of X_1 and X_2 might indicate that the insured is a bad risk with high claim frequencies and/or severities also in the coming years.

In case of stochastic discount factors Y_i, the sum $S = \sum_{i=1}^{n} X_i Y_i$ will be a sum of strongly positive dependent random variables, where the dependence is also caused by the dependence between the Y_i. Consider for instance Y_i and Y_{i+j}, with j small. Discounting over the period $[0, i+j]$ is equal to discounting over the period $[0, i] \cup (i, i+j]$. Hence, in any realistic model these discount factors Y_i and Y_{i+j} will possess a strong positive dependence.

Intuitively, in the presence of positive dependencies, large values of one term in a sum of random variables tend to go hand in hand with large values of the other terms. The Law of Large Numbers will not hold and the aggregate risk S will exhibit greater variation than in the case of a sum of mutually independent random variables. So in this case, the independence assumption tends to underestimate the tails of the distribution function of S.

Second, consider the case where the X_i represent the claims or gains/losses of the different policies in an insurance portfolio and that all t_i are identical and equal to t. The random variable $S = \sum_{i=1}^{n} X_i Y_t$ can then be interpreted as the aggregate claims of the portfolio over a certain reference period, for instance one year.

If the discount factor Y_t is stochastic, then S is a sum of strongly positive dependent random variables as each individual random variable $X_i Y_t$ contains the same discount factor Y_t.

If the discount factor Y_t is assumed to be deterministic, then the independence assumption will often be not too far from reality, and can be used for determining the distribution of S. Moreover, one can force a portfolio of risks to satisfy the independence assumption as much as possible by diversifying, not including too many related risks like the fire risks of different floors of the same building or the risks concerning several layers of the same large

reinsured risk.
In certain situations, however, the individual risks X_i will not be mutually independent because they are subject to the same claim generating mechanism or are influenced by the same economic or physical environment. The independence assumption is then violated and just isn't an adequate way to describe the relations between the different random variables involved.The individual risks of an earthquake or flooding risk portfolio which are located in the same geographic area are correlated, since individual claims are contingent on the occurrence and severity of the same earthquake or flood. On a foggy day all cars of a region have higher probability to be involved in an accident. During dry hot summers, all wooden cottages are more exposed to fire. More generally, one can say that if the density of insured risks in a certain area or organization is high enough, then catastrophes such as storms, explosions, earthquakes, epidemics and so on can cause an accumulation of claims for the insurer. As a financial example, consider a bond portfolio. Individual bond default experience may be conditionally independent for given market conditions. However, the underlying economic environment (for instance interest rates) affects all individual bonds in the market in a similar way. In life insurance, there is ample evidence that the lifetimes of husbands and their wives are positively associated. There may be certain selection mechanisms in the matching of couples ("birds of a feather flock together"): both partners often belong to the same social class and have the same life style. Further, it is known that the mortality rate increases after the passing away of one's spouse (the "broken heart syndrome"). These phenomena have implications on the valuation of aggregate claims in life insurance portfolios. Another example in a life insurance context is a pension fund that covers the pensions of persons working for the same company. These persons work at the same location, they take the same flights. It is evident that the mortality of these persons will be dependent, at least to a certain extent.

As a theoretical example, consider an insurance portfolio consisting of n risks. The payments to be made by the insurer are described by a random vector $(X_1, X_2, \ldots, X_n)$, where X_i is the claim amount of policy i during the insurance period. We assume that all payments have to be done at the end of the insurance period $[0, 1]$. In a deterministic financial setting, the present value at time 0 of the aggregate claims $X_1 + X_2 + \ldots + X_n$ to be paid by the insurer at time 1 is determined by

$$S = (X_1 + X_2 + \ldots + X_n)\ v,$$

where $v = (1+r)^{-1}$ is the deterministic discount factor and r is the technical interest rate. This will be chosen in a conservative way (i.e. sufficiently low), if the insurer doesn't want to underestimate his future obligations. To demonstrate the effect of introducing random interest on insurance business, we look at the following special case. Assume all risks X_i to be non-negative, independent and identically distributed, and let $X \stackrel{d}{=} X_i$, where the symbol $\stackrel{d}{=}$ is used to indicate equality in distribution. The average payment $\frac{S}{n}$ has mean and variance

$$E\left[\frac{S}{n}\right] = v\ E[X]; \quad Var\left[\frac{S}{n}\right] = \frac{v^2}{n} Var[X].$$

The stability necessary for both insureds and insurer is maintained by the Law of Large Numbers, provided that n is indeed 'large' and that the risks are mutually independent and rather well-behaved, not describing for instance risks of catastrophic nature for which the variance might be very large or even infinite.

Now let us examine the consequences of introducing stochastic discounting. Replacing the fixed discount factor v by a random variable Y, representing the stochastic amount to be invested at time 0 with value 1 at the end of the period $[0, 1]$, the present value of the aggregate claims becomes

$$S = (X_1 + X_2 + \ldots + X_n)\ Y.$$

If we assume that the discount factor is independent of the payments, we find that the average payment per policy $\frac{S}{n}$ has mean and variance

$$E\left[\frac{S}{n}\right] = E[X]\ E[Y]; \quad Var\left[\frac{S}{n}\right] = \frac{Var[X]}{n} E\left[Y^2\right] + (E[X])^2\ Var[Y]$$

Assuming that $E[X]$ and $Var[Y]$ are positive, the Law of Large Numbers no longer eliminates the risk involved. This is because for $n \to \infty$, $Var[\frac{S}{n}]$ converges to its second term. So to evaluate the total risk, both the distributions of insurance risk and financial risk are needed. Risk pooling and large portfolios are no longer sufficient tools to eliminate or reduce the average risk associated with a portfolio. This observation implies that the introduction of stochastic financial aspects in actuarial models immediately leads to the necessity of determining distribution functions of sums of dependent random variables.

Under the assumption that the vectors $\underline{X} = (X_1, X_2, \ldots, X_n)$ and $\underline{Y} = (Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ are mutually independent and that the marginal distributions of the X_i and the Y_{t_i} are given, the problem of determining bounds for the distribution function of $S = \sum_{i=1}^n X_i Y_{t_i}$ can be reduced to determining bounds for the distribution function of a sum

$$S = Z_1 + Z_2 + \ldots + Z_n$$

of random variables $Z_1, Z_2, \ldots, Z_n$ with given marginal distributions, but of which the joint distribution is either unspecified or too cumbersome to work with. The unknown or complex nature of the dependence between the random variables Z_i is the reason why it is impossible to derive the distribution function of S exactly.

Recently, several authors in the actuarial literature have derived stochastic lower and upper bounds for sums S of this type. These bounds are bounds in the sense of convex order. The concept of convex order is closely related to the notion of stop-loss order which is more familiar in actuarial circles. Both stochastic orders express which of two random variables is the "less dangerous" one. Replacing S by a less attractive random variable S' will be a safe strategy from the viewpoint of the insurer. Considering also "more attractive" random variables will help to give an idea of the degree of overestimation of the real risk.

In this paper, we will describe how to make safe decisions in case we have a sum of random variables with given marginal distribution functions but of which the stochastic dependent structure is unknown. We will give an overview of the recent actuarial literature on this topic. This paper is partly based on the results described in Dhaene & Goovaerts (1996, 1997), Wang & Dhaene (1998), Goovaerts & Redant (1999), Goovaerts & Dhaene (1999), Goovaerts & Kaas (2001), Dhaene, Wang, Young & Goovaerts (2000), Goovaerts, Dhaene & De Schepper (2000), Simon, Goovaerts & Dhaene (2000), Vyncke, Goovaerts & Dhaene (2001), Kaas, Dhaene & Goovaerts (2000), Denuit, Dhaene, Le Bailly De Tilleghem & Teghem (2001), De Vylder & Dhaene (2001), Kaas, Dhaene, Vyncke, Goovaerts, Denuit (2001). It is the first text integrating these results in a consistent way. The paper also contains several new results and simplified proofs of existing results. Actuarial-financial applications, demonstrating the practical usability of this theory, are considered in Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002). Dependence in portfolios and related stochastic orders are also considered in

Denuit & Lefèvre (1997), Müller (1997), Bäuerle & Müller (1998), Wang & Young (1998), Denuit, De Vijlder & Lefèvre (1999), Denuit and Cornet (1999), Denuit, Genest & Marceau (1999, 2001), Dhaene & Denuit (1999), Embrechts, Mc.Neil and Straumann (1999), Cossette, Denuit & Marceau (2000), Dhaene, Vanneste & Wolthuis (2000), Cossette, Denuit, Dhaene & Marceau (2001), Denuit, Dhaene & Ribas (2001), amongst others.

2 Ordering random variables

In the sequel, we will always consider random variables with finite mean. This implies that for any random variable X we have that $\lim_{x\to\infty} x\,(1 - F_X(x)) = \lim_{x\to-\infty} xF_X(x) = 0$, where $F_X(x) = \Pr[X \leq x]$ is used to denote the cumulative distribution function (cdf) of X. Using the technique of integration by parts on both terms of the right-hand side in $E[X] = \int_{-\infty}^{0} x\ dF_X(x) - \int_0^\infty x\ d\,(1 - F_X(x))$, we find the following expression for $E[X]$:

$$E[X] = -\int_{-\infty}^{0} F_X(x)dx + \int_0^\infty (1 - F_X(x))\,dx. \tag{1}$$

In the actuarial literature it is common practice to replace a random variable by a "less attractive" random variable which has a simpler structure, making it easier to determine its distribution function, see e.g. Goovaerts, Kaas, Van Heerwaarden & Bauwelinckx (1990), Kaas, Van Heerwaarden & Goovaerts (1994) or Denuit, de Vylder & Lefèvre (1999). Performing the computations (of premiums, reserves and so on) with the less attractive random variable is a prudent strategy. Of course, we have to clarify what we mean by a less attractive random variable. For this purpose we first introduce the notion of "stop-loss premium". The stop-loss premium with retention d of random variable X is defined by $E[(X-d)_+]$, with the notation $(x-d)_+ = \max(x-d, 0)$. Using an integration by parts, we immediately find that

$$E[(X-d)_+] = \int_d^\infty (1 - F_X(x))\,dx, \qquad -\infty < d < +\infty, \tag{2}$$

from which we see that the stop-loss premium with retention d can be considered as the weight of an upper tail of (the distribution function of) X: it is the surface between the cdf F_X of X and the constant function 1, from d on. Also useful is the observation that $E[(X-d)_+]$ is a decreasing continuous

function of d, with derivative $F_X(d) - 1$ at d, which vanishes at $+\infty$.
Now, we are able to define the stop-loss order between random variables.

Definition 1 *Consider two random variables X and Y. Then X is said to precede Y in the stop-loss order sense, notation $X \leq_{sl} Y$, if and only if X has lower stop-loss premiums than Y:*

$$E[(X-d)_+] \leq E[(Y-d)_+], \qquad -\infty < d < +\infty. \tag{3}$$

Hence, $X \leq_{sl} Y$ means that X has uniformly smaller upper tails than Y, which in turn means that a payment Y is indeed less attractive than a payment X. Stop-loss order has a natural economic interpretation in terms of expected utility. Indeed, it can be shown that $X \leq_{sl} Y$ if and only if $E\left[u\left(-X\right)\right] \geq E\left[u\left(-Y\right)\right]$ holds for all non-decreasing concave real functions u for which the expectations exist. This means that any risk-averse decision maker would prefer to pay X instead of Y, which implies that acting as if the obligations X are replaced by Y indeed leads to conservative/prudent decisions. The characterization of stop-loss order in terms of utility functions is equivalent to $E\left[v\left(X\right)\right] \leq E\left[v\left(Y\right)\right]$ holding for all non-decreasing convex functions v for which the expectations exist. Therefore, stop-loss order is often called "increasing convex order" and denoted by $\leq_{icx}$. For more details and properties of stop-loss order in a general context, see Shaked & Shanthikumar (1994) or Kaas, Van Heerwaarden & Goovaerts (1994), where stochastic orders are considered in an actuarial context.

Stop-loss order between random variables X and Y implies a corresponding ordering of their means. To prove this, assume that $d < 0$. From the expression (2) of stop-loss premiums as upper tails, we immediately find the following equality:

$$d + E[(X-d)_+] = -\int_d^0 F_X(x)dx + \int_0^\infty \left(1 - F_X(x)\right)dx,$$

and also, letting $d \to -\infty$,

$$\lim_{d \to -\infty} \left(d + E[(X-d)_+]\right) = E\left[X\right].$$

Hence, adding d to both members of the inequality (3) in Definition 1, and taking the limit for $d \to -\infty$, we get $E[X] \leq E[Y]$.

A sufficient condition for $X \leq_{sl} Y$ to hold is that $E[X] \leq E[Y]$, together with the condition that their cumulative distribution functions only cross once. This means that there exists a real number c such that $F_X(x) \leq F_Y(x)$ for $x < c$, but $F_X(x) \geq F_Y(x)$ for $x \geq c$. Indeed, considering the function $f(d) = E[(Y-d)_+] - E[(X-d)_+]$, we have that $\lim_{d \to -\infty} f(d) = E[Y] - E[X] \geq 0$, and $\lim_{d \to +\infty} f(d) = 0$. Further, $f(d)$ first increases, and then decreases (from c on) but remains non-negative.

Recall that our original problem was to replace a random payment X by a less favorable random payment Y, for which the distribution function is easier to obtain. If $X \leq_{sl} Y$, then also $E[X] \leq E[Y]$, and it is intuitively clear that the best approximations arise in the borderline case where $E[X] = E[Y]$. This leads to the so-called convex order.

Definition 2 *Consider two random variables X and Y. Then X is said to precede Y in the convex order sense, notation $X \leq_{cx} Y$, if and only if*

$$\begin{aligned} E[X] &= E[Y], \\ E[(X-d)_+] &\leq E[(Y-d)_+], \qquad -\infty < d < +\infty. \end{aligned} \tag{4}$$

From $E[(X-d)_+] - E[(d-X)_+] = E[X] - d$, we find

$$X \leq_{cx} Y \Leftrightarrow \begin{cases} E[X] = E[Y], \\ E[(d-X)_+] \leq E[(d-Y)_+], \qquad -\infty < d < +\infty. \end{cases} \tag{5}$$

Note that partial integration leads to

$$E[(d-X)_+] = \int_{-\infty}^{d} F_X(x)\, dx, \tag{6}$$

which means that $E[(d-X)_+]$ can be interpreted as the weight of a lower tail of X: it is the surface between the constant function and the cdf of X, from $-\infty$ to d. We have seen that stop-loss order entails uniformly heavier upper tails. The additional condition of equal means implies that convex order also leads to uniformly heavier lower tails.

Let $d > 0$. From (6) we find

$$d - E[(d-X)_+] = -\int_{-\infty}^{0} F_X(x)dx + \int_{0}^{d} (1 - F_X(x))\, dx,$$

and also

$$\lim_{d \to +\infty} (d - E[(d-X)_+]) = E[X].$$

This implies that convex order can also be characterized as follows:

$$X \leq_{cx} Y \Leftrightarrow \begin{cases} E[(X-d)_+] \leq E[(Y-d)_+], & -\infty < d < +\infty, \\ E\left[(d-X)_+\right] \leq E\left[(d-Y)_+\right], & -\infty < d < +\infty. \end{cases} \tag{7}$$

Indeed, the $\Leftarrow$ implication follows from observing that the upper tail inequalities imply $E[X] \leq E[Y]$, while the lower tail inequalities imply $E[X] \geq E[Y]$, hence $E[X] = E[Y]$ must hold.

Note that with stop-loss order, we are concerned with large values of a random loss, and call the random variable Y less attractive than X if the expected values of all top parts $(Y-d)_+$ are larger than those of X. Negative values for these random variables are actually gains. With stability in mind, excessive gains might also be unattractive for the decision maker, for instance for tax reasons. In this situation, X could be considered to be more attractive than Y if both the top parts$(X-d)_+$ and the bottom parts $(d-X)_+$ have a lower expected value than for Y. Both conditions just define the convex order introduced above.

A sufficient condition for $X \leq_{cx} Y$ to hold is that $E[X] = E[Y]$, together with the condition that their cdf's only cross once. This once-crossing condition can be observed to hold in most natural examples, but it is of course easy to construct examples with $X \leq_{cx} Y$ and cdf's that cross more than once.

It can be proven that $X \leq_{cx} Y$ if and only $E\left[v\left(X\right)\right] \leq E\left[v\left(Y\right)\right]$ for all convex functions v, provided the expectations exist. This explains the name "convex order". Note that when characterizing stop-loss order, the convex functions v are additionally required to be non-decreasing. Hence, stop-loss order is weaker: more pairs of random variables are ordered.

We also find that $X \leq_{cx} Y$ if and only $E\left[X\right] = E\left[Y\right]$ and $E\left[u\left(-X\right)\right] \geq E\left[u\left(-Y\right)\right]$ for all non-decreasing concave functions u, provided the expectations exist. Hence, in a utility context, convex order represents the common preferences of all risk-averse decision makers between random variables with equal mean.

In case $X \leq_{cx} Y$, the upper tails as well as the lower tails of Y eclipse the corresponding tails of X, which means that extreme values are more likely to occur for Y than for X. This observation also implies that $X \leq_{cx} Y$ is equivalent to $-X \leq_{cx} -Y$. Hence, the interpretation of the random variables as payments or as incomes is irrelevant for the convex order.

As the function v defined by $v(x) = x^2$ is a convex function, it follows immediately that $X \leq_{cx} Y$ implies $Var\left[X\right] \leq Var\left[Y\right]$. The reverse implication does not hold in general.

Note that comparing variances is meaningful when comparing stop-loss premiums of convex ordered random variables, see, e.g. Kaas, Van Heerwaarden & Goovaerts (1994, p. 68). The following relation links variances and stop-loss premiums:

$$\frac{1}{2}Var[X] = \int_{-\infty}^{\infty} \left(E[(X-t)_+] - (E[X]-t)_+\right) dt. \tag{8}$$

To prove this relation, write

$$\int_{-\infty}^{\infty} \left(E[(X-t)_+] - (E[X]-t)_+\right) dt = \int_{-\infty}^{E[X]} E[(t-X)_+]\, dt + \int_{E[X]}^{\infty} E[(X-t)_+]\, dt.$$

Interchanging the order of the integrations and using partial integration, one finds

$$\int_{-\infty}^{E[X]} E[(t-X)_+]\, dt = \int_{-\infty}^{E[X]} \int_{-\infty}^{t} F_X(x)\, dx\, dt = \frac{1}{2}\int_{-\infty}^{E[X]} (x-E[X])^2\, dF_X(x).$$

Similarly,

$$\int_{E[X]}^{\infty} E[(X-t)_+]\, dt = \frac{1}{2}\int_{E[X]}^{\infty} (x-E[X])^2\, dF_X(x).$$

This proves (8). From (8) we deduce that if $X \leq_{cx} Y$,

$$\int_{-\infty}^{\infty} |E[(Y-t)_+] - (E[(X-t)_+]|\, dt = \frac{1}{2}\{Var[Y] - Var[X]\} \tag{9}$$

A graphical interpretation of relations (8) and (9) is given in Figure 1.

Thus, if $X \leq_{cx} Y$, their stop-loss distance, i.e. the integrated absolute difference of their respective stop-loss premiums, equals half the variance difference between these two random variables. The integrand above is non-negative, so if in addition $Var[X] = Var[Y]$, then X and Y must necessary have equal stop-loss premiums, which implies that they are equal in distribution. We also find that if $X \leq_{cx} Y$, and X and Y are not equal in distribution, then $Var[X] < Var[Y]$ must hold. Note that (8) and (9) have been derived under the additional conditions that both $\lim_{x\to\infty} x^2(1-F_X(x))$ and $\lim_{x\to-\infty} x^2 F_X(x)$ are equal to 0 (and similar for Y). A sufficient condition for these requirements is that X and Y have finite second moments.

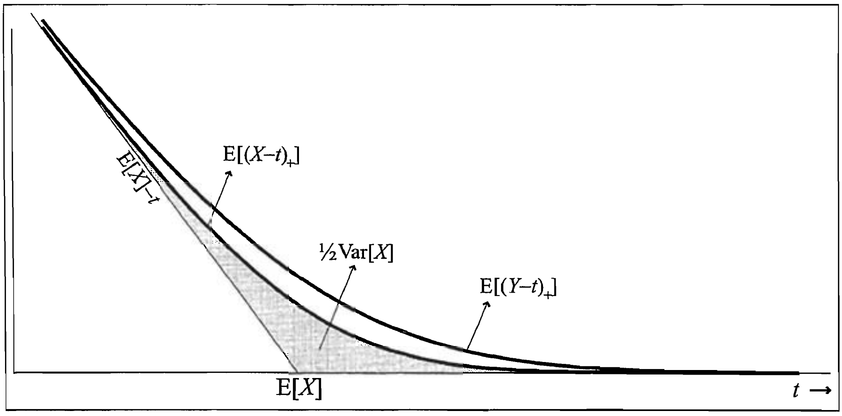

Figure 1: Two stop-loss transforms $\pi_X(t) = E[(X-t)_+]$ and $\pi_Y(t) = E[(Y-t)_+]$ where $X \leq_{cx} Y$.

3 Inverse distribution functions

The cdf $F_X(x) = P[X \leq x]$ of a random variable X is a right-continuous (further abbreviated as r.c.) non-decreasing function with

$$F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0, \qquad F_X(+\infty) = \lim_{x\to+\infty} F_X(x) = 1.$$

The usual definition of the inverse of a distribution function is the non-decreasing and left-continuous (l.c.) function $F_X^{-1}(p)$ defined by

$$F_X^{-1}(p) = \inf\{x \in \mathbb{R} \mid F_X(x) \geq p\}, \qquad p \in [0,1] \tag{10}$$

with $\inf \emptyset = +\infty$ by convention. For all $x \in \mathbb{R}$ and $p \in [0,1]$, we have

$$F_X^{-1}(p) \leq x \Leftrightarrow p \leq F_X(x). \tag{11}$$

In this paper, we will use a more sophisticated definition for inverses of distribution functions. For any real $p \in [0,1]$, a possible choice for the inverse of F_X in p is any point in the closed interval

$$\left[\inf\{x \in \mathbb{R} \mid F_X(x) \geq p\},\ \sup\{x \in \mathbb{R} \mid F_X(x) \leq p\}\right],$$

where, as before, $\inf \emptyset = +\infty$, and also $\sup \emptyset = -\infty$. Taking the left hand border of this interval to be the value of the inverse cdf at p, we get $F_X^{-1}(p)$. Similarly, we define $F_X^{-1+}(p)$ as the right hand border of the interval:

$$F_X^{-1+}(p) = \sup \{x \in \mathbb{R} \mid F_X(x) \leq p\}, \qquad p \in [0,1] \tag{12}$$

which is a non-decreasing and r.c. function. Note that $F_X^{-1}(0) = -\infty$, $F_X^{-1+}(1) = +\infty$ and that all the probability mass of X is contained in the interval $\left[F_X^{-1+}(0),\ F_X^{-1}(1)\right]$. Also note that $F_X^{-1}(p)$ and $F_X^{-1+}(p)$ are finite for all $p \in (0,1)$. In the sequel, we will always use p as a variable ranging over the open interval $(0,1)$, unless stated otherwise.

For any $\alpha \in [0,1]$, we define the α-mixed inverse function of F_X as follows:

$$F_X^{-1(\alpha)}(p) = \alpha\ F_X^{-1}(p) + (1-\alpha)\ F_X^{-1+}(p), \qquad p \in (0,1), \tag{13}$$

which is a non-decreasing function. In particular, we find $F_X^{-1(0)}(p) = F_X^{-1+}(p)$ and $F_X^{-1(1)}(p) = F_X^{-1}(p)$. One immediately finds that for all $\alpha \in [0,1]$,

$$F_X^{-1}(p) \leq\ F_X^{-1(\alpha)}(p) \leq F_X^{-1+}(p), \qquad p \in (0,1). \tag{14}$$

Note that only values of p corresponding to a horizontal segment of F_X lead to different values of $F_X^{-1}(p)$, $F_X^{-1+}(p)$ and $F_X^{-1(\alpha)}(p)$. This phenomenon is illustrated in Figure 2.

Now let d be such that $0 < F_X(d) < 1$. Then $F_X^{-1}\left(F_X(d)\right)$ and $F_X^{-1+}\left(F_X(d)\right)$ are finite, and $F_X^{-1}\left(F_X(d)\right) \leq d \leq F_X^{-1+}\left(F_X(d)\right)$. So for some value $\alpha_d \in [0,1]$, d can be expressed as $d = \alpha_d\ F_X^{-1}\left(F_X(d)\right) + (1-\alpha_d)\ F_X^{-1+}\left(F_X(d)\right) = F_X^{-1(\alpha_d)}\left(F_X(d)\right)$. This implies that for any random variable X and any d with $0 < F_X(d) < 1$, there exists an $\alpha_d \in [0,1]$ such that $F_X^{-1(\alpha_d)}\left(F_X(d)\right) = d$.

In the following theorem, we state the relation between the inverse distribution functions of the random variables X and $g(X)$ for a monotone function g.

Theorem 1 *Let X and $g(X)$ be real-valued random variables, and let $0 < p < 1$.*
(a) If g is non-decreasing and l.c., then

$$F_{g(X)}^{-1}(p) = g\left(F_X^{-1}(p)\right). \tag{15}$$

(b) If g is non-decreasing and r.c., then

$$F_{g(X)}^{-1+}(p) = g\left(F_X^{-1+}(p)\right). \tag{16}$$

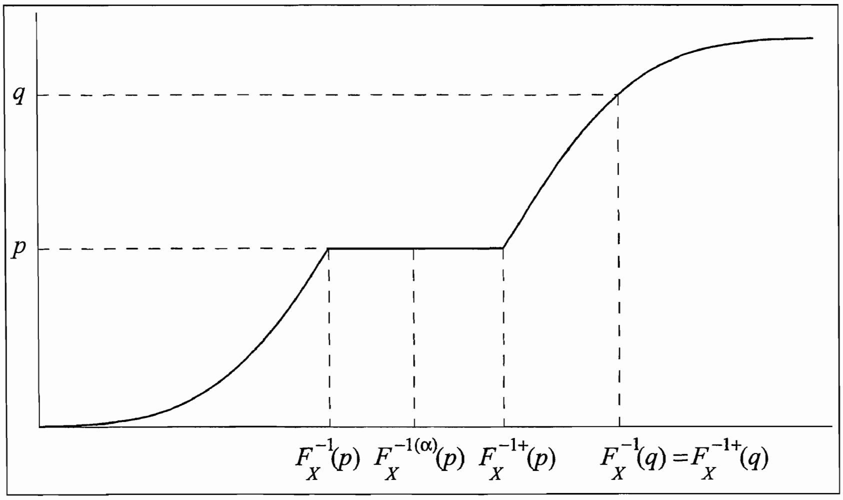

Figure 2: Graphical definition of F_X^{-1}, F_X^{-1+} and $F_X^{-1(\alpha)}$.

(c) If g is non-increasing and l.c., then

$$F_{g(X)}^{-1+}(p) = g\left(F_X^{-1}(1-p)\right). \tag{17}$$

(d) If g is non-increasing and r.c., then

$$F_{g(X)}^{-1}(p) = g\left(F_X^{-1+}(1-p)\right). \tag{18}$$

Proof. We will prove (a). The other results can be proven similarly. Let $0 < p < 1$ and consider a non-decreasing and left-continuous function g. For any real x we find from (11) that

$$F_{g(X)}^{-1}(p) \leq x \Leftrightarrow p \leq F_{g(X)}(x).$$

As g is l.c., we have that

$$g(z) \leq x \Leftrightarrow z \leq \sup\{y \mid g(y) \leq x\}$$

holds for all real z and x. Hence,

$$p \leq F_{g(X)}(x) \Leftrightarrow p \leq F_X \left[\sup \{y \mid g(y) \leq x\}\right]$$

If $\sup \{y \mid g(y) \leq x\}$ is finite then we find from (11) and the equivalence above

$$p \leq F_X \left[\sup \{y \mid g(y) \leq x\}\right] \Leftrightarrow F_X^{-1}(p) \leq \sup \{y \mid g(y) \leq x\}.$$

In case $\sup \{y \mid g(y) \leq x\}$ is $+\infty$ or $-\infty$, we cannot use (11), but one can verify that the equivalence above also holds in this case. Indeed, if the supremum equals $-\infty$, then the equivalence becomes $p \leq 0 \Leftrightarrow F_X^{-1}(p) \leq -\infty$. If the supremum equals $+\infty$, then the equivalence becomes $p \leq 1 \Leftrightarrow F_X^{-1}(p) \leq +\infty$.
Because g is non-decreasing and l.c., we get that

$$F_X^{-1}(p) \leq \sup \{y \mid g(y) \leq x\} \Leftrightarrow g\left(F_X^{-1}(p)\right) \leq x$$

Combining the equivalences, we finally find that

$$F_{g(X)}^{-1}(p) \leq x \Leftrightarrow g\left(F_X^{-1}(p)\right) \leq x$$

holds for all values of x, which means that (a) must hold. ■

For the special cases that g and F_X are continuous and strictly increasing on $\left[F_X^{-1+}(0),\ F_X^{-1}(1)\right]$, a simpler proof is possible. Indeed, in this case we have that $F_{g(X)}(x) = \left(F_X \circ g^{-1}\right)(x)$, which is a continuous and strictly increasing function of x. The results (a) and (b) then follow by inversion of this relation. A similar proof holds for (c) and (d) if g and F_X are both continuous, while g is strictly decreasing and F_X is strictly increasing.

Hereafter, we will reserve the notation U for a uniform$(0,1)$ random variable, i.e. $F_U(p) = p$ and $F_U^{-1}(p) = p$ for all $0 < p < 1$. One can prove that for all $\alpha \in [0,1]$,

$$X \stackrel{d}{=} F_X^{-1}(U) \stackrel{d}{=} F_X^{-1+}(U) \stackrel{d}{=} F_X^{-1(\alpha)}(U). \tag{19}$$

The first distributional equality is known as the quantile transform theorem and follows immediately from (11). It states that a sample of random numbers from a general distribution function F_X can be generated from a sample of uniform random numbers. Note that F_X has at most a countable number of horizontal segments, implying that the last three random variables in (19) only differ in a null-set of values of U. This implies that these random variables are equal with probability one.

4 Comonotonicity

4.1 Comonotonic sets and random vectors

As mentioned in the introduction, quite often in financial actuarial situations one encounters random variables of the type $S = \sum_{i=1}^{n} X_i$ where the terms X_i are not mutually independent, but the multivariate distribution function of the random vector $\underline{X} = (X_1, X_2, \ldots, X_n)$ is not completely specified because one only knows the marginal distribution functions of the random variables X_i. In such cases, to be able to make decisions, it may be helpful to find the dependence structure for the random vector $(X_1, \ldots, X_n)$ producing the least favorable aggregate claims S with the given marginals. Therefore, given the marginal distributions of the terms in a random variable $S = \sum_{i=1}^{n} X_i$, we will look for the joint distribution with the largest sum, in the convex order sense. As we will prove in Section 5.1, the convex-largest sum of the components of a random vector with given marginals will be obtained in the case that the random vector $(X_1, \ldots, X_n)$ has the *comonotonic* distribution, which means that each two possible outcomes $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ of $(X_1, \ldots, X_n)$ are ordered componentwise.

We start by defining comonotonicity of a set of n-vectors in $\mathbb{R}^n$. A n-vector $(x_1, x_2, \ldots, x_n)$ will be denoted by $\underline{x}$. For two n-vectors $\underline{x}$ and $\underline{y}$, the notation $\underline{x} \leq \underline{y}$ will be used for the componentwise order which is defined by $x_i \leq y_i$ for all $i = 1, 2, \ldots, n$.

Definition 3 *The set $A \subseteq \mathbb{R}^n$ is said to be comonotonic if for any $\underline{x}$ and $\underline{y}$ in A, either $\underline{x} \leq \underline{y}$ or $\underline{y} \leq \underline{x}$ holds.*

So, a set $A \subseteq \mathbb{R}^n$ is comonotonic if for any $\underline{x}$ and $\underline{y}$ in A, if $x_i < y_i$ for some i, then $\underline{x} \leq \underline{y}$ must hold. Hence, a comonotonic set is simultaneously non-decreasing in each component. Notice that a comonotonic set is a 'thin' set: it cannot contain any subset of dimension larger than 1. Any subset of a comonotonic set is also comonotonic.

We will denote the (i, j)-projection of a set A in $\mathbb{R}^n$ by $A_{i,j}$. It is defined by

$$A_{i,j} = \{(x_i, x_j) \mid \underline{x} \in A\} \tag{20}$$

Lemma 2 *$A \subseteq \mathbb{R}^n$ is comonotonic if and only if $A_{i,j}$ is comonotonic for all $i \neq j$ in $\{1, 2, \ldots, n\}$.*

The proof of Lemma 2 is straightforward.

For a general set A, comonotonicity of the $(i, i+1)$-projections $A_{i,i+1}$, $(i = 1, 2, \ldots, n-1)$, will not necessarily imply that A is comonotonic. As an example, consider the set $A = \{(x_1, 1, x_3) \mid 0 < x_1, x_3 < 1\}$. This set is not comonotonic, although $A_{1,2}$ and $A_{2,3}$ are comonotonic.

Next, we will define the notion of support of an n-dimensional random vector $\underline{X} = (X_1, \ldots, X_n)$. Any subset $A \subseteq \mathbb{R}^n$ will be called a support of $\underline{X}$ if $\Pr[\underline{X} \in A] = 1$ holds true. In general we will be interested in supports which are "as small as possible". Informally, the smallest support of a random vector $\underline{X}$ is the subset of $\mathbb{R}^n$ that is obtained by subtracting of $\mathbb{R}^n$all points which have a zero-probability neighborhood (with respect to $\underline{X}$). This support can be interpreted as the set of all possible outcomes of $\underline{X}$.

Next, we will define comonotonicity of random vectors.

Definition 4 *A random vector $\underline{X} = (X_1, \ldots, X_n)$ is said to be comonotonic if it has a comonotonic support.*

From the definition, we can conclude that comonotonicity is a very strong positive dependency structure. Indeed, if $\underline{x}$ and $\underline{y}$ are elements of the (comonotonic) support of $\underline{X}$, i.e. $\underline{x}$ and $\underline{y}$ are possible outcomes of $\underline{X}$, then they must be ordered componentwise. This explains why the term comonotonic (common monotonic) is used.

Comonotonicity of a random vector $\underline{X}$ implies that the higher the value of one component X_j, the higher the value of any other component X_k. This means that comonotonicity entails that no X_j is in any way a "hedge", perfect or imperfect, for another component X_k.

In the following theorem, some equivalent characterizations are given for comonotonicity of a random vector.

Theorem 3 *A random vector $\underline{X} = (X_1, X_2, \ldots, X_n)$ is comonotonic if and only if one of the following equivalent conditions holds:*
(1) $\underline{X}$ has a comonotonic support;
(2) For all $\underline{x} = (x_1, x_2, \ldots, x_n)$, we have

$$F_{\underline{X}}(\underline{x}) = \min\{F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\}; \tag{21}$$

(3) For $U \sim$ Uniform(0,1), we have

$$\underline{X} \stackrel{d}{=} (F_{X_1}^{-1}(U), F_{X_2}^{-1}(U), \ldots, F_{X_n}^{-1}(U)); \tag{22}$$

(4) There exist a random variable Z and non-decreasing functions f_i, $(i = 1, 2, \ldots, n)$, such that

$$\underline{X} \stackrel{d}{=} (f_1(Z), f_2(Z), \ldots, f_n(Z)). \tag{23}$$

Proof. **(1)⇒(2):**Assume that $\underline{X}$ has comonotonic support B. Let $\underline{x} \in \mathbb{R}^n$ and let A_j be defined by

$$A_j = \left\{ \underline{y} \in B \mid y_j \leq x_j \right\}, \qquad j = 1, 2, \ldots, n.$$

Because of the comonotonicity of B, there exists an i such that $A_i = \cap_{j=1}^n A_j$. Hence, we find

$$\begin{aligned} F_{\underline{X}}(\underline{x}) &= \Pr\left(\underline{X} \in \cap_{j=1}^n A_j\right) = \Pr\left(\underline{X} \in A_i\right) = F_{X_i}(x_i) \\ &= \min\left\{F_{X_1}(x_1), F_{X_2}(x_2), \ \ldots \ , F_{X_n}(x_n)\right\}. \end{aligned}$$

The last equality follows from $A_i \subset A_j$ so that $F_{X_i}(x_i) \leq F_{X_j}(x_j)$ holds for all values of j.

(2)⇒(3): Now assume that $F_{\underline{X}}(\underline{x}) = \min\left\{F_{X_1}(x_1), F_{X_2}(x_2), \ \ldots \ , F_{X_n}(x_n)\right\}$ for all $\underline{x} = (x_1, x_2, \ldots, x_n)$. Then we find by (11)

$$\begin{aligned} &\Pr[F_{X_1}^{-1}(U) \leq x_1, \ldots, F_{X_n}^{-1}(U) \leq x_n] \\ &= \Pr[U \leq F_{X_1}(x_1), \ldots, U \leq F_{X_n}(x_n)] \\ &= \Pr[U \leq \min_{j=1,\ldots,n} \left\{F_{X_j}(x_j)\right\}] \\ &= \min_{j=1,\ldots,n} \left\{F_{X_j}(x_j)\right\}. \end{aligned}$$

(3)⇒(4): straightforward.

(4)⇒(1): Assume that there exists a random variable Z with support B, and non-decreasing functions f_i, $(i = 1, 2, \ldots, n)$, such that

$$\underline{X} \stackrel{d}{=} (f_1(Z), f_2(Z), \ldots, f_n(Z)).$$

The set of possible outcomes of $\underline{X}$ is $\{(f_1(z), f_2(z), \ldots, f_n(z)) \mid z \in B\}$ which is obviously comonotonic, which implies that $\underline{X}$ is indeed comonotonic. ■

From (21) we see that, in order to find the probability of all the outcomes of n comonotonic risks X_i being less than x_i, $(i = 1, \ldots, n)$, one simply takes the probability of the least likely of these n events. It is obvious that for any random vector $(X_1, \ldots, X_n)$, not necessarily comonotonic, the following inequality holds:

$$\Pr\left[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\right] \leq \min\left\{F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\right\}, \tag{24}$$

and since Hoeffding (1940) and Fréchet (1951) it is known that the function $\min\{F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\}$ is indeed the multivariate cdf of a random vector, i.e. $(F_{X_1}^{-1}(U), F_{X_2}^{-1}(U), \ldots, F_{X_n}^{-1}(U))$, which has the same marginals as $(X_1, \ldots, X_n)$. The inequality (24) states that in the class of all random vectors $(X_1, \ldots, X_n)$ with the same marginals, the probability that all X_i simultaneously realize 'small' values is maximized if the vector is comonotonic, suggesting that comonotonicity is indeed a very strong positive dependency structure.

From (22) we find that in the special case that all marginal distribution functions F_{X_i} are identical, comonotonicity of $\underline{X}$ is equivalent to saying that $X_1 = X_2 = \ldots = X_n$ holds almost surely.

A standard way of modelling situations where individual random variables $X_1, \ldots, X_n$ are subject to the same external mechanism is to use a secondary mixing distribution. The uncertainty about the external mechanism is then described by a structure variable z, which is a realization of a random variable Z, and acts as a (random) parameter of the distribution of $\underline{X}$. The aggregate claims can then be seen as a two-stage process: first, the external parameter $Z = z$ is drawn from the distribution function F_Z of z. The claim amount of each individual risk X_i is then obtained as a realization from the conditional distribution function of X_i given $Z = z$. A special type of such a mixing model is the case where given $Z = z$, the claim amounts X_i are degenerate on x_i, where the $x_i = x_i(z)$ are non-decreasing in z. This means that $(X_1, \ldots, X_n) \stackrel{d}{=} (f_1(Z), \ldots, f_n(Z))$ where all functions f_i are non-decreasing. Hence, $(X_1, \ldots, X_n)$ is comonotonic. Such a model is in a sense an extreme form of a mixing model, as in this case the external parameter $Z = z$ completely determines the aggregate claims.

As the random vectors $(F_{X_1}^{-1}(U),\ F_{X_2}^{-1}(U),\ \ldots,\ F_{X_n}^{-1}(U))$ and $(F_{X_1}^{-1(\alpha_1)}(U), F_{X_2}^{-1(\alpha_2)}(U), \ldots, F_{X_n}^{-1(\alpha_n)}(U))$ are equal with probability one, we find that comonotonicity of $\underline{X}$ can be characterized by

$$\underline{X} \stackrel{d}{=} (F_{X_1}^{-1(\alpha_1)}(U), F_{X_2}^{-1(\alpha_2)}(U), \ldots, F_{X_n}^{-1(\alpha_n)}(U)) \tag{25}$$

for $U \sim$ Uniform(0,1) and given real numbers $\alpha_i \in [0,1]$.

If $U \sim$ Uniform(0,1), then also $1 - U \sim$ Uniform(0,1). This implies that comonotonicity of $\underline{X}$ can also be characterized by

$$\underline{X} \stackrel{d}{=} (F_{X_1}^{-1}(1-U), F_{X_2}^{-1}(1-U), \ldots, F_{X_n}^{-1}(1-U)). \tag{26}$$

One can prove that $\underline{X}$ is comonotonic if and only if there exist a random variable Z and non-increasing functions f_i, $(i = 1, 2, \ldots, n)$, such that

$$\underline{X} \stackrel{d}{=} (f_1(Z), f_2(Z), \ldots, f_n(Z)). \tag{27}$$

The proof is similar to the proof of the characterization (4) in Theorem 3.

In the sequel, for any random vector $(X_1, \ldots, X_n)$, the notation $(X_1^c, \ldots, X_n^c)$ will be used to indicate a comonotonic random vector with the same marginals as $(X_1, \ldots, X_n)$. From (22), we find that for any random vector $\underline{X}$ the outcome of its comonotonic counterpart $\underline{X}^c = (X_1^c, \ldots, X_n^c)$ is with probability 1 in the following set

$$\left\{ (F_{X_1}^{-1}(p), F_{X_2}^{-1}(p), \ldots, F_{X_n}^{-1}(p)) \mid 0 < p < 1 \right\}. \tag{28}$$

This support of $\underline{X}^c$ is not necessarily a connected curve. Indeed, all horizontal segments of the cdf of X_i lead to "missing pieces" in this curve. This support can be seen to be a series of ordered connected curves. Now by connecting the endpoints of consecutive curves by straight lines, we obtain a comonotonic connected curve in $\mathbb{R}^n$. Hence, it may be traversed in a direction which is upwards for all components simultaneously. We will call this set the connected support of $\underline{X}^c$. It might be parameterized as follows:

$$\left\{ \left(F_{X_1}^{-1(\alpha)}(p), F_{X_2}^{-1(\alpha)}(p), \ldots, F_{X_n}^{-1(\alpha)}(p) \right) \mid 0 < p < 1,\, 0 \leq \alpha \leq 1 \right\}. \tag{29}$$

Observe that this parameterization is not necessarily unique: there may be elements in the connected support which can be characterized by different values of α.

Theorem 4 *A random vector $\underline{X}$ is comonotonic if and only if (X_i, X_j) is comonotonic for all $i \neq j$ in $\{1, 2, \ldots, n\}$.*

Proof. The proof of the "$\Rightarrow$"-implication is straightforward.
For the proof of the "$\Leftarrow$"-implication, consider the set A in $\mathbb{R}^n$ defined by

$$A = \left\{ (F_{X_1}^{-1}(p), F_{X_2}^{-1}(p), \ldots, F_{X_n}^{-1}(p)) \mid 0 < p < 1 \right\}.$$

Its (i, j)-projections are given by

$$A_{i,j} = \left\{ \left(F_{X_i}^{-1}(p), F_{X_j}^{-1}(p) \right) \mid 0 < p < 1 \right\}.$$

The event "$\underline{X} \in A$" is equivalent with the event "$(X_i, X_j) \in A_{i,j}$ for all (i,j)". Because of the comonotonicity of the pairs (X_i, X_j), the latter event is the certain event. Hence we find that $\Pr[\underline{X} \in A] = 1$, so that the comonotonic set A is a support of $\underline{X}$. This implies that $\underline{X}$ is a comonotonic random vector.
■

The theorem states that comonotonicity of a random vector is equivalent with pairwise comonotonicity.

Consider the random vector $(U, 1, V)$ where U and V are mutually independent random variables that are both uniformly distributed on the unit-interval $(0,1)$. It is clear that $(U,1)$ and $(1,V)$ are both comonotonic pairs, but $(U,1,V)$ isn't comonotonic. Hence, for a general random vector $\underline{X}$, comonotonicity of the pairs (X_i, X_{i+1}), $(i = 1, 2, \ldots, n-1)$, will not necessary imply comonotonicity of $\underline{X}$.

4.2 Some examples

First, we give an example with continuous distributions. Let $X \sim$ Uniform on the set $(0, \frac{1}{2}) \cup (1, \frac{3}{2})$, $Y \sim$ Beta(2,2), hence $F_Y(y) = 3y^2 - 2y^3$ on $(0,1)$, and $Z \sim$ Normal$(0,1)$.
If X, Y and Z are mutually independent, then the support of (X,Y,Z) is the set

$$\{(x,y,z)) \mid x \in (0, \frac{1}{2}) \cup (1, \frac{3}{2}),\ y \in (0,1),\ z \in \mathbb{R}\}.$$

The support of the comonotonic random vector (X^c, Y^c, Z^c) is given by

$$\{(F_X^{-1}(p), F_Y^{-1}(p), F_Z^{-1}(p)) \mid 0 < p < 1\},$$

see Figure 3. Actually, not all of this support is depicted. The part left out corresponds to $p \notin (\Phi(-2), \Phi(2))$ and extends along the vertical asymptotes $(0,0,z)$ and $(\frac{3}{2}, 1, z)$. The thick continuous line is the support of $\underline{X}^c$, while the dotted line is the straight line needed to transform this support into the connected support. Note that F_X has a horizontal segment between $\frac{1}{2}$ and 1. The projection of the connected curve along the z-axis can also be seen to constitute an increasing curve, as do projections along the other axes.

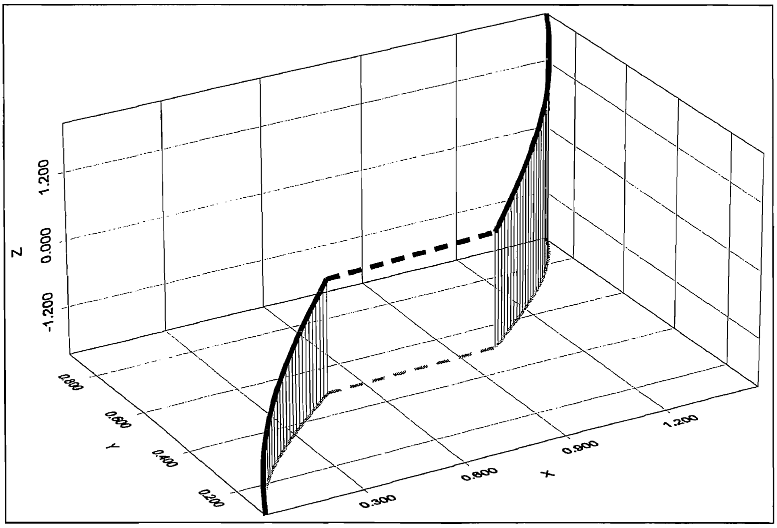

Figure 3: A continuous example with $n = 3$.

For an example with discrete distributions, take $X \sim \text{Uniform}\{0, 1, 2, 3\}$ and $Y \sim \text{Binomial}(3, \frac{1}{2})$. It is easy to verify that

$$
\begin{aligned}
(F_X^{-1}(p), F_Y^{-1}(p)) &= (0,0) \text{ for } 0 < p \leq \frac{1}{8}, \\
&= (0,1) \text{ for } \frac{1}{8} < p \leq \frac{2}{8}, \\
&= (1,1) \text{ for } \frac{2}{8} < p \leq \frac{4}{8}, \\
&= (2,2) \text{ for } \frac{4}{8} < p \leq \frac{6}{8}, \\
&= (3,2) \text{ for } \frac{6}{8} < p \leq \frac{7}{8}, \\
&= (3,3) \text{ for } \frac{7}{8} < p < 1.
\end{aligned}
$$

The support of (X^c, Y^c) is just these six points, and the connected support arises by simply connecting them consecutively with straight lines, the dotted lines in Figure 4. The straight line connecting $(1, 1)$ and $(2, 2)$ is not along

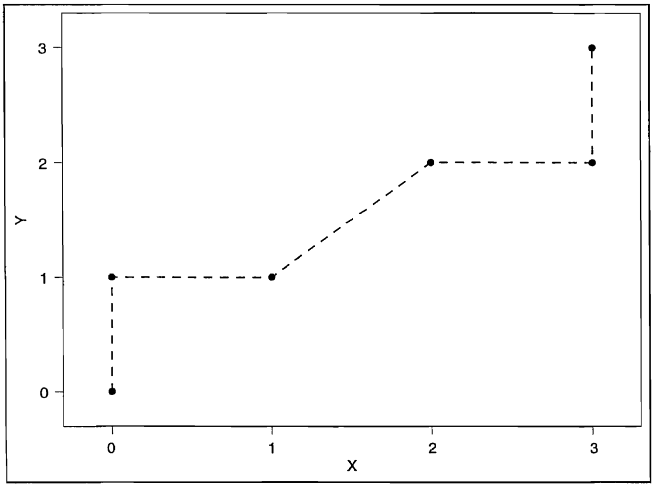

Figure 4: A discrete example.

one of the axes. This happens because at level $p = \frac{1}{2}$, both $F_X(y)$ and $F_Y(y)$ have horizontal segments. Note that any non-decreasing curve connecting $(1,1)$ and $(2,2)$ would have led to a feasible connected curve. These two points have probability $\frac{2}{8}$, the other points $\frac{1}{8}$.

4.3 Location-scale families of distribution functions

For a random couple (X, Y), Pearson's correlation coefficient is defined by

$$r(X,Y) = \frac{Cov\,[X,Y]}{\sqrt{Var\,[X]\;Var\,[Y]}}$$

where

$$Cov\,[X,Y] = E\,[(X - E\,[X])\,(Y - E\,[Y])]$$

is the covariance of X and Y. Recall that $r(X,Y) = 1$ if and only if real numbers $a > 0$ and b exist such that $Y = aX + b$ holds with probability one. Hence, $r(X,Y) = 1$ implies comonotonicity of the couple (X,Y). In this case the connected support is a straight line. In this sense, comonotonicity is an extension of the concept of positive perfect correlation.

As is shown in Theorem 3, in the class of all n-dimensional random variables with given marginal distribution functions F_i, $i = 1, 2, \ldots, n$, the comonotonic upper bound is reached by $\left(F_1^{-1}(U), F_2^{-1}(U), \ldots, F_n^{-1}(U)\right)$. On the other hand, it is only rarely possible to find a pair (X, Y) with $r(X, Y) = 1$ in the class of all bivariate random variables with given marginals F_1 and F_2, since for this to hold, $a > 0$ and b must exist such that $F_2(y) = F_1(\frac{y-b}{a})$ for all y, which means that F_1 and F_2 belong to the same location-scale family of distributions.

Definition 5 *The random vector $\underline{X}$ has marginal cdf's F_{X_i} that belong to the same location-scale family of distributions, if there exist a random variable Y, positive real constants a_i and real constants b_i such that the relation*

$$X_i \stackrel{d}{=} a_i Y + b_i \tag{30}$$

holds for $i = 1, 2, \ldots, n$.

Note that the condition in the definition above is equivalent with saying that there exists a cdf F_Y, positive real constants a_i and real constants b_i such that $F_{X_i}(x) = F_Y\left(\frac{x-b_i}{a_i}\right)$ holds for $i = 1, 2, \ldots, n$.

For a random vector $\underline{X}$ with marginal cdf's F_{X_i} belonging to the same location-scale family, one finds from Theorem 1 that

$$F_{X_i}^{-1}(p) = a_i F_Y^{-1}(p) + b_i, \qquad p \in (0, 1). \tag{31}$$

In this case, we also find that the comonotonic sum

$$X_1^c + \ldots + X_n^c \stackrel{d}{=} \sum_{i=1}^{n} a_i \; F_Y^{-1}(U) + \sum_{i=1}^{n} b_i \tag{32}$$

has a distribution function that also belongs to the same location-scale family.

Theorem 5 *A random vector $\underline{X}$ with marginal cdf's F_{X_i} belonging to the same location-scale family is comonotonic if and only if $r(X_i, X_j) = 1$ for all $i, j \in \{1, 2, \ldots, n\}$.*

Proof. From (31) and Theorem 3, we find that $\underline{X}$ is comonotonic if and only if

$$\underline{X} \stackrel{d}{=} \left(a_1 F_Y^{-1}(U) + b_1, \ldots, a_n F_Y^{-1}(U) + b_n\right).$$

Hence, comonotonicity of $\underline{X}$ implies that $r(X_i, X_j) = 1$ for all pairs (i,j). Conversely, if all correlations are equal to 1, then all couples (X_i, X_j) are comonotonic, which means that $\underline{X}$ is a comonotonic random vector by Theorem 4. ■

Example 1 ***(Uniform marginals)***

Consider a random vector $\underline{X}$ with uniform marginals F_{X_i}: for each X_i we assume that $X_i \sim \text{Uniform}(\alpha_i, \beta_i)$, with $\alpha_i < \beta_i$. In this case, the marginals belong to the same location-scale family of distributions since for each X_i, we have that

$$X_i \stackrel{d}{=} \alpha_i + (\beta_i - \alpha_i)\ U. \tag{33}$$

We also have that

$$F_{X_i}^{-1}(p) = \alpha_i + (\beta_i - \alpha_i)\ \ p, \qquad 0 < p < 1, \tag{34}$$

from which we find that

$$S^c = X_1^c + \ldots + X_n^c \stackrel{d}{=} \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} (\beta_i - \alpha_i)\ U. \tag{35}$$

Hence, S^c is uniform on the interval $(\sum_{i=1}^n \alpha_i,\ \sum_{i=1}^n \beta_i)$.▼

Example 2 ***(Normal marginals)***

Consider a random vector $\underline{X}$ with normal marginals F_{X_i}: for each X_i we have that $X_i \sim N(\mu_i, \sigma_i^2)$. In this case, the marginals belong to the same location-scale family of distributions since

$$X_i \stackrel{d}{=} \sigma_i\ Z + \mu_i \tag{36}$$

where $Z \sim N(0,1)$. We find that

$$F_{X_i}^{-1}(p) = \sigma_i \Phi^{-1}(p) + \mu_i, \qquad p \in (0,1), \tag{37}$$

where Φ is the standard normal cdf. From Theorem 5, we find that $\underline{X}$ is comonotonic if and only if $r(X_i, X_j) = 1$ for all $i, j \in \{1, 2, \ldots, n\}$. We also have that $X_1^c + \ldots + X_n^c$ is normally distributed with mean $\sum_{i=1}^n \mu_i$ and variance $(\sum_{i=1}^n \sigma_i)^2$. Note that if the X_i were independent, we would get the normal distribution with mean $\sum_{i=1}^n \mu_i$ and variance $\sum_{i=1}^n \sigma_i^2 \leq (\sum_{i=1}^n \sigma_i)^2$.▼

4.4 Sums of comonotonic random variables

In the sequel, the notation S^c will be used for the sum of the components of the comonotonic counterpart $(X_1^c, X_2^c, \ldots, X_n^c)$ of a random vector $(X_1, X_2, \ldots, X_n)$:

$$S^c = X_1^c + X_2^c + \ldots + X_n^c. \tag{38}$$

Further on in this paper, we will prove that approximating the distribution function of $S = X_1 + X_2 + \ldots + X_n$ by the distribution function of the comonotonic sum S^c is a prudent strategy in the sense that $S \leq_{cx} S^c$. Performing this approximation will only be meaningful if we can easily determine the distribution function and the stop-loss premiums of S^c. In the two next theorems, we will prove that these quantities can indeed easily be determined from the marginal distribution functions of the terms in the sum.

In the next theorem we prove that the inverse distribution function of a sum of comonotonic random variables is simply the sum of the inverse distribution functions of the marginal distributions.

Theorem 6 *The α-inverse distribution function $F_{S^c}^{-1(\alpha)}$ of a sum S^c of comonotonic random variables $(X_1^c, X_2^c, \ldots, X_n^c)$ is given by*

$$F_{S^c}^{-1(\alpha)}(p) = \sum_{i=1}^{n} F_{X_i}^{-1(\alpha)}(p), \qquad 0 < p < 1, \quad 0 \leq \alpha \leq 1. \tag{39}$$

Proof. Consider the random vector $(X_1, X_2, \ldots, X_n)$ and its comonotonic counterpart $(X_1^c, X_2^c, \ldots, X_n^c)$. Then $S^c = X_1^c + X_2^c + \ldots + X_n^c \stackrel{d}{=} g\,(U)$, with U uniformly distributed on $(0,1)$ and with the function g defined by

$$g\,(u) = \sum_{i=1}^{n} F_{X_i}^{-1}(u), \qquad 0 < u < 1.$$

It is clear that g is non-decreasing and left-continuous. Application of Theorem 1(a) leads to

$$F_{S^c}^{-1}(p) = F_{g(U)}^{-1}(p) = g\left(F_U^{-1}(p)\right) = g(p), \qquad 0 < p < 1,$$

so the inverse distribution function of S^c can be computed from

$$F_{S^c}^{-1}(p) = \sum_{i=1}^{n} F_{X_i}^{-1}(p), \qquad 0 < p < 1.$$

Similarly, from Theorem 1(b), we find that

$$F_{S^c}^{-1+}(p) = \sum_{i=1}^{n} F_{X_i}^{-1+}(p), \qquad 0 < p < 1.$$

Multiplying the last two equalities by α and $1 - \alpha$ respectively, and adding up, we find the desired result. ■

Note that

$$S^c \stackrel{d}{=} \sum_{i=1}^{n} F_{X_i}^{-1(\alpha)}(U). \tag{40}$$

By the theorem above, we find that the connected support of S^c is given by

$$\begin{aligned} & \left\{ F_{S^c}^{-1(\alpha)}(p) \mid 0 < p < 1,\, 0 \leq \alpha \leq 1 \right\} \\ = & \left\{ \sum_{i=1}^{n} F_{X_i}^{-1(\alpha)}(p) \mid 0 < p < 1,\, 0 \leq \alpha \leq 1 \right\}. \end{aligned}$$

This implies

$$F_{S^c}^{-1+}(0) = \sum_{i=1}^{n} F_{X_i}^{-1+}(0), \tag{41}$$

$$F_{S^c}^{-1}(1) = \sum_{i=1}^{n} F_{X_i}^{-1}(1). \tag{42}$$

Hence, the minimal value of the comonotonic sum equals the sum of the minimal values of each term. Similarly, the maximal value of the comonotonic sum equals the sum of the maximal values of each term. The number $\sum_{i=1}^{n} F_{X_i}^{-1+}(0)$, which is either finite or $-\infty$ (if any of the terms in the sum is $-\infty$), is the minimum possible value of S^c, and $\sum_{i=1}^{n} F_{X_i}^{-1}(1)$ is the maximum. Also note that

$$\begin{aligned} F_{S^c}^{-1+}(1) &= \sum_{i=1}^{n} F_{X_i}^{-1+}(1) = +\infty, \\ F_{S^c}^{-1}(0) &= \sum_{i=1}^{n} F_{X_i}^{-1+}(0) = -\infty. \end{aligned}$$

For any $(X_1, X_2, \ldots, X_n)$, we have that $S = X_1+X_2+\ldots+X_n \geq \sum_{i=1}^n F_{X_i}^{-1+}(0)$ must hold with probability 1. This implies

$$\sum_{i=1}^n F_{X_i}^{-1+}(0) \leq F_S^{-1+}(0). \tag{43}$$

Similarly, we find

$$F_S^{-1}(1) \leq \sum_{i=1}^n F_{X_i}^{-1}(1). \tag{44}$$

This means that the sum S of the components of any random vector $(X_1, X_2, \ldots, X_n)$ has a support that is contained in the interval $\left[\sum_{i=1}^n F_{X_i}^{-1+}(0),\ \sum_{i=1}^n F_{X_i}^{-1}(1)\right]$. The minimal value of S is larger than or equal to the one of S^c, since by comonotonicity all terms of the latter are small simultaneously.

Given the inverse functions $F_{X_i}^{-1}$, the cdf of $S^c = X_1^c + X_2^c + \ldots + X_n^c$ can be determined as follows:

$$\begin{aligned} F_{S^c}(x) &= \sup\{p \in (0,1) \mid F_{S^c}(x) \geq p\} \\ &= \sup\{p \in (0,1) \mid F_{S^c}^{-1}(p) \leq x\} \\ &= \sup\left\{p \in (0,1) \mid \sum_{i=1}^n F_{X_i}^{-1}(p) \leq x\right\}. \end{aligned} \tag{45}$$

In the sequel, for any random variable X, the expression "F_X is strictly increasing" should always be interpreted as "F_X is strictly increasing on $\left(F_X^{-1+}(0),\ F_X^{-1}(1)\right)$".
Observe that for any random variable X, the following equivalences hold:

$$F_X \text{ is strictly increasing} \iff F_X^{-1} \text{ is continuous on } (0,\ 1), \tag{46}$$

and also

$$F_X \text{ is continuous} \iff F_X^{-1} \text{ is strictly increasing on } (0,\ 1). \tag{47}$$

Now assume that the marginal distribution functions $F_{X_i}, i = 1, \ldots, n$ of the comonotonic random vector $(X_1^c, X_2^c, \ldots, X_n^c)$ are strictly increasing and continuous. Then each inverse distribution function $F_{X_i}^{-1}$ is continuous on $(0,\ 1)$, which implies that $F_{S^c}^{-1}$ is continuous on $(0,\ 1)$ because $F_{S^c}^{-1}(p) = \sum_{i=1}^n F_{X_i}^{-1}(p)$ holds for $0 < p < 1$. This in turn implies that F_{S^c} is

strictly increasing on $\left(F_{S^c}^{-1+}(0),\ F_{S^c}^{-1}(1)\right)$. Further, by a similar reasoning we find that F_{S^c} is continuous.
Hence, in case of strictly increasing and continuous marginals, for any $F_{S^c}^{-1+}(0) < x < F_{S^c}^{-1}(1)$, the probability $F_{S^c}(x)$ is uniquely determined by $F_{S^c}^{-1}(F_{S^c}(x)) = x$, or equivalently,

$$\sum_{i=1}^{n} F_{X_i}^{-1}(F_{S^c}(x)) = x, \qquad F_{S^c}^{-1+}(0) < x < F_{S^c}^{-1}(1). \tag{48}$$

It suffices thus to solve the latter equation to get $F_{S^c}(x)$.

In the following theorem, we prove that also the stop-loss premiums of a sum of comonotonic random variables can be obtained from the stop-loss premiums of the terms.

Theorem 7 *The stop-loss premiums of the sum S^c of the components of the comonotonic random vector $(X_1^c, X_2^c, \ldots, X_n^c)$ are given by*

$$E\left[(S^c - d)_+\right] = \sum_{i=1}^{n} E\left[(X_i - d_i)_+\right], \qquad (F_{S^c}^{-1+}(0) < d < F_{S^c}^{-1}(1)), \tag{49}$$

with the d_i given by

$$d_i = F_{X_i}^{-1(\alpha_d)}(F_{S^c}(d)), \qquad (i = 1, 2, \ldots, n) \tag{50}$$

and $\alpha_d \in [0,1]$ determined by

$$F_{S^c}^{-1(\alpha_d)}(F_{S^c}(d)) = d. \tag{51}$$

Proof. Let $d \in (F_{S^c}^{-1+}(0),\ F_{S^c}^{-1}(1))$, hence $0 < F_{S^c}(d) < 1$.
As the connected support of $\underline{X}^c$ as defined in (29) is comonotonic, it can have at most one point of intersection with the hyperplane $\{\underline{x} \mid x_1 + \ldots + x_n = d\}$. This is because the hyperplane contains no different points $\underline{x}$ and $\underline{y}$ such that $\underline{x} \leq \underline{y}$ or $\underline{x} \geq \underline{y}$ holds.
Now we will prove that the vector $\underline{d} = (d_1, d_2, \ldots, d_n)$ as defined above is the unique point of this intersection. As $0 < F_{S^c}(d) < 1$ must hold, we know from Section 3 that there exists an $\alpha_d \in [0,1]$ that fulfils condition (51). Also note that by Theorem 6, we have that $\sum_{i=1}^{n} d_i = d$. Hence, the vector $\underline{d}$ with the d_i defined in (50) and (51) is an element of both the connected support of $\underline{X}^c$ and the hyperplane $\{\underline{x} \mid x_1 + \ldots + x_n = d\}$.

We can conclude that $\underline{d}$ is the unique element of the intersection of the connected support and the hyperplane.

Let $\underline{x}$ be an element of the connected support of $\underline{X}^c$. Then the following equality holds:

$$(x_1 + x_2 + \ldots + x_n - d)_+ \equiv (x_1 - d_1)_+ + (x_2 - d_2)_+ + \ldots + (x_n - d_n)_+.$$

This is because $\underline{x}$ and $\underline{d}$ are both elements of the connected support of $\underline{X}^c$, and hence, if there exists any j such that $x_j > d_j$ holds, then we also have $x_k \geq d_k$ for all k, and the left hand side equals the right hand side because $\sum_{i=1}^n d_i = d$. On the other hand, when all $x_j \leq d_j$, obviously the left hand side is 0 as well.
Now replacing constants by the corresponding random variables in the equality above and taking expectations, we find (49). ■

Note that we also find that

$$E\left[(S^c - d)_+\right] = \sum_{i=1}^n E\left[X_i\right] - d, \qquad \text{if } d \leq F_{S^c}^{-1+}(0) \tag{52}$$

and

$$E\left[(S^c - d)_+\right] = 0, \qquad \text{if } d \geq F_{S^c}^{-1}(1). \tag{53}$$

So from (41), (42), (52), (53) and Theorem 7 we can conclude that for any real d, there exist d_i with $\sum_{i=1}^n d_i = d$, such that $E\left[(S^c - d)_+\right] = \sum_{i=1}^n E\left[(X_i - d_i)_+\right]$ holds.

The expression for the stop-loss premiums of a comonotonic sum S^c can also be written in terms of the usual inverse distribution functions. Indeed, for any retention $d \in \left(F_{S^c}^{-1+}(0), F_{S^c}^{-1}(1)\right)$, we have

$$\begin{aligned} & E\left[\left(X_i - F_{X_i}^{-1(\alpha_d)}\left(F_{S^c}(d)\right)\right)_+\right] \\ = \;& E\left[\left(X_i - F_{X_i}^{-1}\left(F_{S^c}(d)\right)\right)_+\right] - \left(F_{X_i}^{-1(\alpha_d)}\left(F_{S^c}(d)\right) - F_{X_i}^{-1}\left(F_{S^c}(d)\right)\right)\left(1 - F_{S^c}(d)\right) \end{aligned}$$

Summing over i, and taking into account the definition of α_d, we find the expression derived in Dhaene, Wang, Young & Goovaerts (2000), where the random variables were assumed to be non-negative. This expression holds for any retention $d \in \left(F_{S^c}^{-1+}(0), F_{S^c}^{-1}(1)\right)$:

$$\begin{aligned} E\left[(S^c - d)_+\right] \;&=\; \sum_{i=1}^n E\left[\left(X_i - F_{X_i}^{-1}\left(F_{S^c}(d)\right)\right)_+\right] \\ &\quad - \left(d - F_{S^c}^{-1}\left(F_{S^c}(d)\right)\right)\left(1 - F_{S^c}(d)\right). \end{aligned} \tag{54}$$

In case the marginal cdf's F_{X_i} are strictly increasing, (54) reduces to

$$E[(S^c - d)_+] = \sum_{i=1}^{n} E\left[\left(X_i - F_{X_i}^{-1}\left(F_{S^c}(d)\right)_+\right], \qquad d \in \left(F_{S^c}^{-1+}(0), F_{S^c}^{-1}(1)\right). \tag{55}$$

From Theorem 7, we can conclude that any stop-loss premium of a sum of comonotonic random variables can be written as the sum of stop-loss premiums for the individual random variables involved. The theorem provides an algorithm for directly computing stop-loss premiums of sums of comonotonic random variables, without having to compute the entire distribution function of the sum itself. Indeed, in order to compute the stop-loss premium with retention d, we only need to know $F_{S^c}(d)$, which can be computed directly from (45).

Application of the relation $E[(X-d)_+] = E\left[(d-X)_+\right] + E\left[X\right] - d$ for S^c and the X_i in relation (49) leads to the following expression for the lower tails of a sum of comonotonic random variables:

$$E\left[(d - S^c)_+\right] = \sum_{i=1}^{n} E\left[(d_i - X_i)_+\right], \qquad F_{S^c}^{-1+}(0) < d < F_{S^c}^{-1}(1), \tag{56}$$

with the d_i as defined in (50) and (51).

Example 3 ***(Exponential marginals)***

Consider a random vector $\underline{X}$ with exponential marginals: $X_i \sim \text{Exp}(1/\beta_i)$. Then

$$F_{X_i}(x) = 1 - e^{-\frac{x}{\beta_i}}, \qquad \beta_i > 0,\ x \geq 0. \tag{57}$$

We find the following expression for the inverse distribution function:

$$F_{X_i}^{-1}(p) = -\beta_i \ \ln\left(1-p\right), \qquad 0 < p < 1. \tag{58}$$

One can easily verify that the stop-loss premium with retention d is given by

$$E[(X_i - d)_+] = \beta_i \, e^{-\frac{d}{\beta_i}}, \qquad 0 < d < \infty \tag{59}$$

The inverse distribution function of the comonotonic sum S^c is given by

$$F_{S^c}^{-1}(p) = -\left(\sum_{i=1}^{n} \beta_i\right) \ln\left(1-p\right), \qquad 0 < p < 1. \tag{60}$$

This means that the comonotonic sum of exponentially distributed random variables is again exponentially distributed with parameter $\beta = \sum_{i=1}^{n} \beta_i$. The stop-loss premiums of S^c are given by

$$E\left[S^c - d\right]_+ = \beta\, e^{-\frac{d}{\beta}}, \qquad 0 < d < \infty. \tag{61}$$

The case $n = 2$ is considered in Heilmann (1986).▼

Example 4 ***(Pareto marginals)***

Consider a random vector $\underline{X}$ with Pareto distributed marginals: $X_i \sim$Pareto$(\alpha,\ x_i)$. Then

$$F_{X_i}(x) = 1 - \left(\frac{x_i}{x}\right)^{\alpha}, \qquad \alpha > 0,\ x > x_i > 0. \tag{62}$$

The inverse cdf is given by

$$F_{X_i}^{-1}(p) = \frac{x_i}{(1-p)^{\frac{1}{\alpha}}}, \qquad 0 < p < 1. \tag{63}$$

One can easily verify that

$$E[(X_i - d)_+] = \left(\frac{x_i}{d}\right)^{\alpha - 1} \frac{x_i}{\alpha - 1}, \qquad x_i < d < \infty,\ \alpha > 1. \tag{64}$$

The inverse distribution function of the comonotonic sum S^c is given by

$$F_{S^c}^{-1}(p) = \frac{\sum_{i=1}^{n} x_i}{(1-p)^{\frac{1}{\alpha}}}, \qquad 0 < p < 1. \tag{65}$$

This means that the comonotonic sum of Pareto distributed random variables (with identical first parameter) is again Pareto distributed.▼

Similarly, one can prove that the comonotonic sum of Inverse Gaussian distributed random variables has an Inverse Gaussian distribution, see Dhaene, Wang, Young & Goovaerts (2000). Also the comonotonic sum of Gamma distributed random variables with fixed first parameter is Gamma distributed.

5 Convex bounds for sums of random variables

5.1 The comonotonic upper bound for $\sum_{i=1}^{n} X_i$

In this section we will derive bounds for sums $S = X_1 + X_2 + \ldots + X_n$ of random variables $X_1, X_2, \ldots, X_n$ of which the marginal distributions are given. The bounds are random variables that are larger (or smaller) than S in the sense of convex order. Therefore, we will call these bounds convex bounds. The reason we will resort to convex bounds is that the joint distribution of the random vector $(X_1, X_2, \ldots, X_n)$ is either unspecified or too cumbersome to work with.
The upper bound that we will derive in this subsection is actually attainable in the class of all random vectors with given marginals, it is reached by the comonotonic random vectors in this class. So, the upper bound is a supremum in the sense of convex order.

Theorem 8 *For any random vector $(X_1, X_2, \ldots, X_n)$ we have*

$$X_1 + X_2 + \ldots + X_n \leq_{cx} X_1^c + X_2^c + \ldots + X_n^c. \tag{66}$$

Proof. It suffices to prove stop-loss order, since it is obvious that the means of these two sums are equal. Hence, we have to prove that

$$E[(X_1 + X_2 + \ldots + X_n - d)_+] \leq E[(X_1^c + X_2^c + \ldots + X_n^c - d)_+]$$

holds for all retentions d with $d \in \left(F_{S^c}^{-1+}(0), F_{S^c}^{-1}(1)\right)$, since the stop-loss premiums can be seen to be equal for other values of d.
The following holds for all $(x_1, x_2, \ldots, x_n)$ when $d_1 + d_2 + \ldots + d_n = d$:

$$\begin{aligned}
& (x_1 + x_2 + \ldots + x_n - d)_+ \\
= \; & ((x_1 - d_1) + (x_2 - d_2) + \ldots + (x_n - d_n))_+ \\
\leq \; & ((x_1 - d_1)_+ + (x_2 - d_2)_+ + \ldots + (x_n - d_n)_+)_+ \\
= \; & (x_1 - d_1)_+ + (x_2 - d_2)_+ + \ldots + (x_n - d_n)_+.
\end{aligned}$$

Now replacing constants by the corresponding random variables in the inequality above and taking expectations, we get that

$$E[(X_1 + X_2 + \ldots + X_n - d)_+] \leq E[(X_1 - d_1)_+] + E[(X_2 - d_2)_+] + \ldots + E[(X_n - d_n)_+] \tag{67}$$

holds for all d and d_i such that $\sum_{i=1}^n d_i = d$.
By choosing $d \in \left(F_{S^c}^{-1+}(0), F_{S^c}^{-1}(1)\right)$ and the d_i as in Theorem 7, the above inequality becomes the one that was to be proven. ∎

The theorem above states that the least attractive random vector $(X_1, \ldots, X_n)$ with given marginals F_i, in the sense that the sum of their components is largest in the convex order, has the *comonotonic* joint distribution, which means that it has the joint distribution of $\left(F_1^{-1}(U), F_2^{-1}(U), \ldots, F_n^{-1}(U)\right)$. The components of this random vector are maximally dependent, all components being non-decreasing functions of the same random variable. Several proofs gave been given for this result, see e.g. Denneberg (1994), Dhaene & Goovaerts (1996), Müller (1997) or Dhaene, Wang, Young & Goovaerts (2000).

Note that the inequality (67) holds in particular if $(X_1, \ldots, X_n)$ is comonotonic. From the Theorems 7 and 8, we find that for any random vector $\underline{X}$ the inequalities

$$
\begin{aligned}
E[(X_1 + X_2 + \ldots + X_n - d)_+] &\leq \sum_{i=1}^n E[(X_i - F_{X_i}^{-1(\alpha_d)}(F_{S^c}(d)))_+] \\
&\leq \sum_{i=1}^n E[(X_i - d_i)_+] \qquad (68)
\end{aligned}
$$

holds for all $d \in \left(F_{S^c}^{-1+}(0), F_{S^c}^{-1}(1)\right)$ such that $\sum_{i=1}^n d_i = d$. Hence, the smallest upper bound of the form $\sum_{i=1}^n E[(X_i - d_i)_+]$ with $\sum_{i=1}^n d_i = d$ for the stop-loss premium $E[(X_1 + X_2 + \ldots + X_n - d)_+]$ is the comonotonic upper bound.

We can generalize Theorem 8 above as follows.

Corollary 9 *Consider the random vectors $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_n)$. If $X_i \leq_{sl} Y_i$ holds for all $i = 1, \ldots, n$, then*

$$X_1 + X_2 + \ldots + X_n \leq_{sl} Y_1^c + Y_2^c + \ldots + Y_n^c. \qquad (69)$$

Proof. Since $Y_1^c + \ldots + Y_n^c$ is comonotonic, for any real d, one can find $d_1, \ldots, d_n$ with $d = d_1 + \ldots + d_n$ and $E\left[(Y_1^c + \ldots + Y_n^c - d)_+\right] = E\left[(Y_1 - d_1)\right]_+ + \ldots + E\left[(Y_n - d_n)_+\right]$. Hence

$$
\begin{aligned}
E[(X_1 + X_2 + \ldots + X_n - d)_+] &\leq E[(X_1 - d_1)_+] + \ldots + E[(X_n - d_n)_+] \\
&\leq E[(Y_1 - d_1)_+] + \ldots + E[(Y_n - d_n)_+] \\
&= E\left[(Y_1^c + \ldots + Y_n^c - d)_+\right].
\end{aligned}
$$

■

In Theorem 5, we proved that a random vector with marginals that belong to the same location-scale family of distributions is comonotonic if and only if the correlation of each pair of marginal components equals 1. Using the fact that in the class of all random vectors with given marginals the comonotonic sum is the largest in the sense of convex order, we can prove that comonotonicity can be characterized by maximal correlations of all pairs of random variables involved. In order to prove this result, we need an expression for the stop-loss premiums of a sum of two random variables in terms of the bivariate distribution function.

Lemma 10 *For any bivariate random variable (X,Y) and any real number d, the stop-loss premium of $X+Y$ at retention d is given by*

$$E\left[(X+Y-d)_+\right] = E\left[X\right] + E\left[Y\right] - d + \int_{-\infty}^{+\infty} F_{X,Y}(x,d-x)\ dx \tag{70}$$

Proof. By reversing the order of the integration, we find

$$\begin{aligned} E\left[(d-X-Y)_+\right] &= \int_{x=-\infty}^{+\infty}\int_{y=-\infty}^{d-x}\int_{t=x}^{d-y} dt\ dF_{X,Y}(x,y) \\ &= \int_{t=-\infty}^{+\infty}\int_{x=-\infty}^{t}\int_{y=-\infty}^{d-t} dF_{X,Y}(x,y)\ dx \\ &= \int_{t=-\infty}^{+\infty} F_{X,Y}(t,d-t)\ dt, \end{aligned}$$

from which we find the desired result. ■

Theorem 11 *A random vector $\underline{X}$ is comonotonic if and only if $r(X_i,X_j) = r\left(X_i^c, X_j^c\right)$ for all $i,j \in \{1,2,\ldots,n\}$.*

Proof. Because comonotonicity is equivalent with pairwise comonotonicity, it suffices to give the proof for a two-dimensional random vector (X_i,X_j). The proof of the "$\Rightarrow$"-implication is straightforward.
In order to prove the "$\Leftarrow$"- implication, note that $r(X_i,X_j) = r\left(X_i^c,X_j^c\right)$ is equivalent to $Var\left[X_i+X_j\right] = Var\left[X_i^c + X_j^c\right]$. As we have that $X_i + X_j \leq_{cx} X_i^c + X_j^c$, this implies $X_i + X_j \overset{d}{=} X_i^c + X_j^c$. Hence, for all real d, we must have that

$$E\left[(X_i+X_j-d)_+\right] = E\left[\left(X_i^c+X_j^c-d\right)_+\right].$$

Using Lemma 10, this equality can be written as

$$\int_{-\infty}^{+\infty} \left[F_{X_i^c, X_j^c}(x, d-x) - F_{X_i, X_j}(x, d-x) \right] \, dx = 0,$$

From (24), we have that the integrand is non-negative,which implies that

$$F_{X_i, X_j}(x, d-x) = F_{X_i^c, X_j^c}(x, d-x)$$

must hold for all values of x. As this must hold for all values of d, we have proven the theorem. ■

From the proof of Theorem 11 we also find that random vector $\underline{X}$ is comonotonic if and only if $Var(X_i + X_j) = Var\left(X_i^c + X_j^c\right)$ for all $i, j \in \{1, 2, \ldots, n\}$.

From the convex ordering relation in Theorem 8, we find that for any random vector (X_1, X_2) the following inequality holds:

$$Var\left[X_1 + X_2\right] \leq Var\left[X_1^c + X_2^c\right], \tag{71}$$

which is equivalent with

$$r(X_1, X_2) \leq r\left(X_1^c, X_2^c\right), \tag{72}$$

with strict inequalities when (X_1, X_2) is not comonotonic. As a special case of (72), we find that $r\left(X_1^c, X_2^c\right) \geq 0$ always hold. Also note that a random vector (X_1, X_2) is comonotonic and has mutual independent components if and only if X_1 or X_2 is degenerate, see Luan (2001).

Example 5 ***(Lognormal marginals)***

Consider a random vector $(\alpha_1 X_1, \alpha_2 X_2, \ldots, \alpha_n X_n)$ of which the α_i are non-zero real numbers and the X_i are lognormal distributed: $\ln(X_i) \sim N\left(\mu_i, \sigma_i^2\right)$. We have that

$$\begin{aligned} E\left[X_i\right] &= e^{\mu_i + \frac{1}{2}\sigma_i^2}, &(73)\\ Var\left[X_i\right] &= e^{2\mu_i + \sigma_i^2}\left(e^{\sigma_i^2} - 1\right). &(74) \end{aligned}$$

Consider e.g. the situation where the α_i are deterministic payments at times i, and the X_i are the corresponding lognormal distributed discount factors. Then $(\alpha_1 X_1, \alpha_2 X_2, \ldots, \alpha_n X_n)$ is the vector of the stochastically discounted deterministic payments. As $\Phi^{-1}(1-p) = -\Phi^{-1}(p)$, we find from Theorem 1 that

$$F^{-1}_{\alpha_i X_i}(p) = \alpha_i \; e^{\mu_i + sign(\alpha_i)\;\sigma_i \Phi^{-1}(p)}, \quad 0 < p < 1, \tag{75}$$

where $sign\,(\alpha_i)$ equals 1 if $\alpha_i > 0$ and -1 if $\alpha_i < 0$. In particular, we find that the product of n comonotonic lognormal random variables is again lognormal:

$$\Pi_{i=1}^{n} F^{-1}_{X_i}(U) \stackrel{d}{=} e^{\sum_{i=1}^{n}\mu_i + \sum_{i=1}^{n}\sigma_i \Phi^{-1}(U)}. \tag{76}$$

The stop-loss premiums of a lognormal distributed random variable are given by

$$E[(X_i - d_i)_+] = e^{\mu_i + \frac{\sigma_i^2}{2}}\;\Phi(d_{i,1}) - d_i\;\Phi(d_{i,2}), \qquad d_i > 0. \tag{77}$$

where $d_{i,1}$ and $d_{i,2}$ are determined by

$$d_{i,1} = \frac{\mu_i + \sigma_i^2 - \ln{(d_i)}}{\sigma_i}, \quad d_{i,2} = d_{i,1} - \sigma_i. \tag{78}$$

This result can easily be proven. Indeed, by differentiating both sides of (77) with respect to d_i, one sees that both derivatives are equal to $F_X(d_i) - 1$. Also, for $d_i \to \infty$, both sides tend to zero.
For the lower tails we find

$$E[(d_i - X_i)_+] = -e^{\mu_i + \frac{\sigma_i^2}{2}}\;\Phi(-d_{i,1}) + d_i\;\Phi(-d_{i,2}), \qquad d_i > 0. \tag{79}$$

As $E[(\alpha_i(X_i - d_i))_+] = -\alpha_i\;E[(d_i - X_i)_+]$ if α_i is negative, we find from (78) and (79)

$$E[(\alpha_i(X_i - d_i))_+] = \alpha_i\; e^{\mu_i + \frac{\sigma_i^2}{2}}\;\Phi(sign(\alpha_i)\; d_{i,1}) - \alpha_i\; d_i\;\Phi(sign(\alpha_i)\; d_{i,2}), \qquad d_i > 0, \tag{80}$$

with $d_{i,1}$ and $d_{i,2}$ as defined above.
Let $S = \alpha_1 X_1 + \ldots + \alpha_n X_n$, and S^c its comonotonic counterpart: $S^c = F^{-1}_{\alpha_1 X_1}(U) + \ldots + F^{-1}_{\alpha_n X_n}(U)$. Then $S \leq_{cx} S^c$. As the marginal distribution functions are strictly increasing and continuous, by (48) we find that the distribution function $F_{S^c}(x)$ is implicitly defined by $F^{-1}_{S^c}(F_{S^c}(x)) = x$, or equivalently,

$$\sum_{i=1}^{n} \alpha_i\; e^{\mu_i + sign(\alpha_i)\;\sigma_i \Phi^{-1}(F_{S^c}(x))} = x, \qquad F^{-1+}_{S^c}(0) < x < F^{-1}_{S^c}(1). \tag{81}$$

For $F_{S^c}^{-1+}(0) < d < F_{S^c}^{-1}(1)$, the stop-loss premium of S^c at retention d follows from (55):

$$\begin{aligned} E[(S^c - d)_+] &= \sum_{i=1}^{n} E[(\alpha_i X_i - F_{\alpha_i X_i}^{-1}(F_{S^c}(d)))_+] \\ &= \sum_{i=1}^{n} E[\left(\alpha_i \left(X_i - e^{\mu_i + sign(\alpha_i)\ \sigma_i \Phi^{-1}(F_{S^c}(d))}\right)\right)_+]. \end{aligned}$$

Using (80) and (81), we finally find the following expression for the stop-loss premium at retention d with $F_{S^c}^{-1+}(0) < d < F_{S^c}^{-1}(1)$:

$$E[(S^c - d)_+] = \sum_{i=1}^{n} \alpha_i\ e^{\mu_i + \frac{\sigma_i^2}{2}} \Phi\left(sign\left(\alpha_i\right)\ \sigma_i - \Phi^{-1}\left(F_{S^c}(d)\right)\right) - d\ \left(1 - F_{S^c}(d)\right). \tag{82}$$

The lower tails are given by

$$E[(d - S^c)_+] = -\sum_{i=1}^{n} \alpha_i\ e^{\mu_i + \frac{\sigma_i^2}{2}} \Phi\left(-sign\left(\alpha_i\right)\ \sigma_i + \Phi^{-1}\left(F_{S^c}(d)\right)\right) + d\ F_{S^c}(d). \tag{83}$$

We also find the following expression for the correlation coefficient of two comonotonic lognormally distributed random variables with variances given by σ_i^2 and σ_j^2 respectively:

$$r\left(F_{X_i}^{-1}(U), F_{X_j}^{-1}(U)\right) = \frac{e^{\sigma_i \sigma_j} - 1}{\sqrt{e^{\sigma_i^2} - 1}\sqrt{e^{\sigma_j^2} - 1}}. \tag{84}$$

As in Embrechts, McNeil & Straumann (2001), consider the special case that $\ln X_1 \sim N(0,1)$ and $\ln X_2 \sim N(0, \sigma^2)$, then the correlation coefficient becomes

$$r\left(F_{X_1}^{-1}(U), F_{X_2}^{-1}(U)\right) = \frac{e^{\sigma} - 1}{\sqrt{e^{\sigma^2} - 1}\sqrt{e - 1}}$$

which approaches 0 if $\sigma \to \infty$, see Figure 5. As a consequence, there exist comonotonic random couples of which the correlation is almost 0. As comonotonicity leads to the highest correlation possible for a given pair of marginals, this observation clearly demonstrates that a correlation coefficient equal to 1 is not always attainable in the class of random vectors with given marginals. ▼

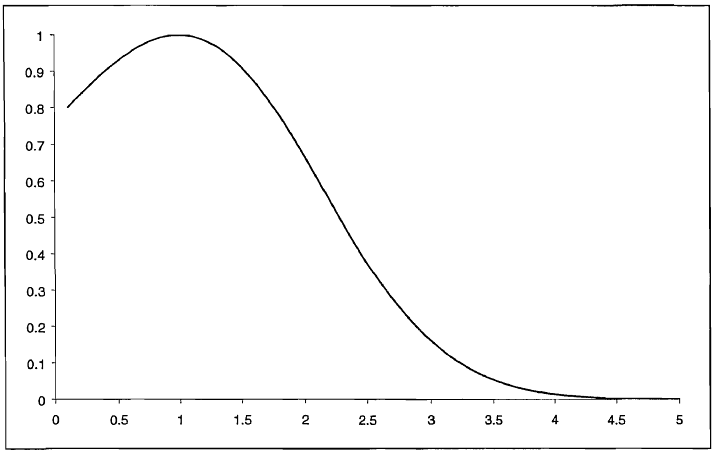

Figure 5: The correlation coefficient of the comonotonic random couple (X_1, X_2) as a function of σ.

We end this section by summarizing the main advantages of using $S^c = X_1^c + \ldots + X_n^c$ instead of $S = X_1 + \ldots + X_n$:

- Replacing the cdf of S by the cdf of S^c is a prudent strategy in the framework of utility theory: the real cdf is replaced by a less attractive one.

- The random variables S and S^c have the same expected value. As these random variables are ordered in the convex order sense, we have that the moment of order $2k$ $(k = 1, 2, \ldots)$ of S is smaller than the corresponding moment of S^c. Many actuarially relevant quantities reflect convex order, for instance both the ruin probability and the Lundberg upper bound for it increase when the claim size distribution is replaced by a convex larger one. Other examples are zero-utility premiums such as the exponential premium, and of course stop-loss premiums for any retention d.

- The cdf of S^c is easily obtained; essentially, S^c has a one-dimensional distribution, depending only on the random variable U. The cdf of S can only be obtained if the dependency structure is known. Even if this dependency structure is known, it can be hard to determine the cdf of S from it.

- The stop-loss premiums of S^c follow from stop-loss premiums of the marginal random variables involved. Computing the stop-loss premiums of S can only be carried out when the dependency structure is known, and in general requires n integrations to be performed.

5.2 Improved upper bounds for $\sum_{i=1}^{n} X_i$

If the only information available concerning the multivariate distribution function of the random vector $(X_1, \ldots, X_n)$ consists of the marginal distribution functions of the X_i, then the distribution function of $S^c = F_{X_1}^{-1}(U) + F_2^{-1}(U) + \ldots + F_n^{-1}(U)$ is a prudent choice for approximating the unknown distribution function of $S = X_1 + \ldots + X_n$. It is a supremum in terms of convex order, hence it is the best upper bound that can be derived under the given conditions.

Let us now assume that we have some additional information available concerning the stochastic nature of $(X_1, \ldots, X_n)$. More precisely, we assume that there exists some random variable Λ with a given distribution function, such that we know the conditional cdf's, given $\Lambda = \lambda$, of the random variables X_i, for all possible values of λ. We will show that in this case we can derive improved upper bounds in terms of convex order for S, which are smaller in convex order than the upper bound S^c. Essentially, the idea of this subsection is to determine comonotonic upper bounds for the sum S, conditionally given $\Lambda = \lambda$. Next, we mix the resulting distributions with weights $dF_\Lambda(\lambda)$. By this procedure, convex order is maintained. The upper bound obtained in this way turns out to be sharper than the comonotonic upper bound S^c because it still has the right marginal cdf's for its terms.

In the following theorem, we introduce the notation $F_{X_i|\Lambda}^{-1}(U)$ for the random variable $f_i(U, \Lambda)$, where the function f_i is defined by $f_i(u, \lambda) = F_{X_i|\Lambda=\lambda}^{-1}(u)$.

Theorem 12 *Let U be uniform(0,1), and independent of the random variable Λ. Then we have*

$$X_1 + X_2 + \ldots + X_n \leq_{cx} F_{X_1|\Lambda}^{-1}(U) + F_{X_2|\Lambda}^{-1}(U) + \ldots + F_{X_n|\Lambda}^{-1}(U). \qquad (85)$$

Proof. From Theorem 8, we get for any convex function v,

$$\begin{aligned} E\left[v\left(X_1+\ldots+X_n\right)\right] &= \int_{-\infty}^{+\infty} E\left[v\left(X_1+\ldots+X_n\right) \mid \Lambda=\lambda\right]\, dF_\Lambda(\lambda) \\ &\leq \int_{-\infty}^{+\infty} E\left[v\left(f_1(U,\lambda)+\ldots+f_n(U,\lambda)\right)\right] dF_\Lambda(\lambda) \\ &= E\left[v\left(f_1(U,\Lambda)+\ldots+f_n(U,\Lambda)\right)\right] \end{aligned}$$

from which the stated result follows directly. ■

Note that the random vector $\left(F^{-1}_{X_1|\Lambda}(U), F^{-1}_{X_2|\Lambda}(U), \ldots, F^{-1}_{X_n|\Lambda}(U)\right)$ has marginals $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$, because

$$\begin{aligned} F_{X_i}(x) &= \int_{-\infty}^{\infty} \Pr\left[X_i \leq x \mid \Lambda=\lambda\right] dF_\Lambda(\lambda) \\ &= \int_{-\infty}^{\infty} \Pr\left[F^{-1}_{X_i|\Lambda=\lambda}(U) \leq x\right] dF_\Lambda(\lambda) \\ &= \int_{-\infty}^{\infty} \Pr\left[f_i(U,\lambda) \leq x\right] dF_\Lambda(\lambda) \\ &= \Pr\left[f_i(U,\Lambda) \leq x\right]. \end{aligned}$$

In view of Theorem 8 this implies

$$F^{-1}_{X_1|\Lambda}(U)+\ldots+F^{-1}_{X_n|\Lambda}(U) \leq_{cx} F^{-1}_{X_1}(U)+\ldots+F^{-1}_{X_n}(U), \tag{86}$$

which means that the upper bound derived in this subsection is indeed an improved upper bound.

If Λ is independent of all $X_1, X_2, \ldots, X_n$, then we actually do not have any extra information at all and the improved upper bound reduces to the comonotonic upper bound derived in Theorem 8.

Let S and S^u be defined by

$$S = X_1 + X_2 + \ldots + X_n \tag{87}$$

and

$$S^u = F^{-1}_{X_1|\Lambda}(U) + F^{-1}_{X_2|\Lambda}(U) + \ldots + F^{-1}_{X_n|\Lambda}(U). \tag{88}$$

If the random vector $(X_1, \ldots, X_n)$ is comonotonic, any choice of Λ is optimal as it leads to the exact distribution function for the sum. We also find that if for any possible outcome λ, conditionally on $\Lambda = \lambda$, the random vector $(X_1, X_2, \ldots, X_n,)$ is comonotonic, then $S \overset{d}{=} S^u$.

In general, to judge the quality of the stochastic upper bound S^u, we might compare its variance with the variance of S. As we have that $Var\left[E\left(S^u \mid \Lambda\right)\right] = Var\left[E\left(S \mid \Lambda\right)\right]$, we find that $Var\left[S^u\right] = Var\left[S\right]$ if and only if $E\left[Var\left(S^u \mid \Lambda\right)\right] = E\left[Var\left(S \mid \Lambda\right)\right]$. This condition will hold if for any outcome λ of Λ, we have that $Var\left(S^u \mid \Lambda = \lambda\right) = Var\left(S \mid \Lambda = \lambda\right)$. Hence, if we find a conditioning random variable Λ such that for any outcome λ of Λ, we have that conditionally given $\Lambda = \lambda$, the vector $(X_1, \ldots, X_n)$ is comonotonic, then the distribution function of the improved upper bound coincides with the exact distribution function.

As a special case, assume for the moment that $S = X_1 + X_2$. In this case the optimal choice is to take $\Lambda \equiv X_1$ (or $\Lambda \equiv X_2$), since then the cdf's of S and S^u coincide. This example illustrates the fact that the optimal conditioning random variable Λ will in general not be S.
It is clear that in general, the optimal choice for the conditioning random variable Λ will strongly depend on the dependency structure of the random vector $(X_1, \ldots, X_n)$.

In order to obtain the distribution function of S^u, observe that given the event $\Lambda = \lambda$, the random variable S^u is a sum of comonotonic random variables. Hence,

$$F^{-1}_{S^u|\Lambda=\lambda}(p) = \sum_{i=1}^{n} F^{-1}_{X_i|\Lambda=\lambda}(p), \quad p \in (0,1)\,. \tag{89}$$

Given $\Lambda = \lambda$, the cdf of S^u follows from (45):

$$F_{S^u|\Lambda=\lambda}(x) = \sup\left\{ p \in (0,1) \mid \sum_{i=1}^{n} F^{-1}_{X_i|\Lambda=\lambda}(p) \leq x \right\}. \tag{90}$$

The cdf of S^u then follows from

$$F_{S^u}(x) = \int_{-\infty}^{+\infty} F_{S^u|\Lambda=\lambda}(x)\; dF_\Lambda(\lambda). \tag{91}$$

If the marginal cdf's $F_{X_i|\Lambda=\lambda}$ are strictly increasing and continuous, then $F_{S^u|\Lambda=\lambda}(x)$ follows by solving

$$\sum_{i=1}^{n} F^{-1}_{X_i \mid \Lambda=\lambda}\left(F_{S^u \mid \Lambda=\lambda}(x)\right) = x, \qquad F^{-1+}_{S^u \mid \Lambda=\lambda}(0) < x < F^{-1}_{S^u \mid \Lambda=\lambda}(1), \tag{92}$$

see (48). In this case, we also find from (55) that for any $d \in \left(F^{-1+}_{S^u|\Lambda=\lambda}(0), F^{-1}_{S^u|\Lambda=\lambda}(1)\right)$:

$$E\left[(S^u - d)_+ \mid \Lambda = \lambda\right] = \sum_{i=1}^{n} E\left[\left(X_i - F^{-1}_{X_i|\Lambda=\lambda}\left(F_{S^u|\Lambda=\lambda}(d)\right)\right)_+ \mid \Lambda = \lambda\right], \tag{93}$$

from which the stop-loss premium at retention d of S^u can be determined. An application of the results presented in this subsection to lognormal marginals X_i is considered in Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002).

5.3 Lower bounds for $\sum_{i=1}^{n} X_i$

Let $\underline{X} = (X_1, \ldots, X_n)$ be a random vector with given marginal cdf's $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$. As in the previous subsection, we assume that there exists some random variable Λ with a given distribution function, such that we know the conditional cdf's, given $\Lambda = \lambda$, of the random variables X_i, for all possible values of λ. We will show how to obtain a lower bound, in the sense of convex order, for $S = X_1 + X_2 + \ldots + X_n$ by conditioning on this random variable. Considering a more attractive random variable than S will help to give an idea of the degree of overestimation of the risk involved by replacing S by the less attractive random variables S^u or S^c.

The idea of this subsection is to observe that the expectation of a random variable is always smaller than or equal in convex order than the random variable itself, and also that convex order is maintained under mixing.

Theorem 13 *For any random vector $\underline{X}$ and any random variable Λ, we have*

$$E[X_1 \mid \Lambda] + E[X_2 \mid \Lambda] + \ldots + E[X_n \mid \Lambda] \leq_{cx} X_1 + X_2 + \ldots + X_n. \tag{94}$$

Proof. By Jensen's inequality, we find that for any convex function v, the following inequality holds:

$$\begin{aligned} E[v(X_1 + X_2 + \ldots + X_n)] &= E_\Lambda\, E[v(X_1 + X_2 + \ldots + X_n) \mid \Lambda] \\ &\geq E_\Lambda\, [v(E[X_1 + X_2 + \ldots + X_n \mid \Lambda])] \\ &= E_\Lambda\, [v(E[X_1 \mid \Lambda] + \ldots + E[X_n \mid \Lambda])]. \end{aligned}$$

This proves the stated result. ■

Let S be defined as above, and let S^l be defined by

$$S^l = E[S \mid \Lambda] \tag{95}$$

Note that if Λ and S are mutually independent, we find the trivial result

$$E[S] \leq_{cx} S. \tag{96}$$

On the other hand, if Λ and S have a one-to-one relation (i.e. Λ completely determines S), the lower bound coincides with S. Note further that $E[E[X_i \mid \Lambda]] = E[X_i]$ always holds, but $Var[E[X_i \mid \Lambda]] < Var[X_i]$ unless $E[Var[X_i \mid \Lambda]] = 0$ which means that X_i, given $\Lambda = \lambda$, is degenerate for each λ. This implies that the random vector $(E[X_1 \mid \Lambda], E[X_2 \mid \Lambda], \ldots, E[X_n \mid \Lambda])$ will in general not have the same marginal distribution functions as $\underline{X}$. But if we can find a conditioning random variable Λ with the property that all random variables $E[X_i \mid \Lambda]$ are non-increasing functions of Λ (or all are non-decreasing functions of Λ), the lower bound S^l is a sum of n comonotonic random variables. The cdf of this sum can then be obtained by previous results. Applications of Theorem 13 in the case of lognormal marginals X_i is considered Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002).

To judge the quality of the stochastic lower bound $E[S \mid \Lambda]$, we might look at its variance. To maximize it, i.e. to make it as close as possible to $Var[S]$, the average value of $Var[S \mid \Lambda = \lambda]$ should be minimized. In other words, to get the best lower bound, Λ and S should be as alike as possible.

Let's further assume that the random variable Λ is such that all $g_i(\lambda) \equiv E[X_i \mid \Lambda = \lambda]$ are non-increasing and continuous functions of λ. The quantiles of the lower bound S^l then follow from

$$\begin{aligned} F_{S^l}^{-1}(p) &= \sum_{i=1}^{n} F_{E[X_i|\Lambda]}^{-1}(p) = \sum_{i=1}^{n} F_{g_i(\Lambda)}^{-1}(p) \\ &= \sum_{i=1}^{n} E\left[X_i \mid \Lambda = F_{\Lambda}^{-1+}(1-p)\right], \quad p \in (0,1). \end{aligned} \tag{97}$$

Further, the cdf of S^l follows from (45):

$$F_{S^l}(x) = \sup\left\{p \in (0,1) \mid \sum_{i=1}^{n} E\left[X_i \mid \Lambda = F_{\Lambda}^{-1+}(1-p)\right] \leq x\right\} \tag{98}$$

If we now additionally assume that the cdf's of the random variables $E[X_i \mid \Lambda]$ are strictly increasing and continuous, then the cdf of S^l is also strictly increasing and continuous, and from (48) we get for all $x \in \left(F^{-1+}_{E[S|\Lambda]}(0)\,,\ F^{-1}_{E[S|\Lambda]}(1)\right)$,

$$\sum_{i=1}^{n} F^{-1}_{E[X_i|\Lambda]}\left(F_{S^l}(x)\right) = x,$$

or equivalently,

$$\sum_{i=1}^{n} E\left[X_i \mid \Lambda = F_\Lambda^{-1+}\left(1 - F_{S^l}(x)\right)\right] = x, \tag{99}$$

which unambiguously determines the cdf of the convex order lower bound $S^l = E[S \mid \Lambda]$ for S.

Under the same assumptions, the stop-loss premiums of S^l can be determined from (55):

$$E\left[\left(S^l - d\right)_+\right] = \sum_{i=1}^{n} E\left[\left(E[X_i \mid \Lambda] - E\left[X_i \mid \Lambda = F_\Lambda^{-1+}\left(1 - F_{S^l}(d)\right)\right]\right)_+\right], \tag{100}$$

which holds for all retentions $d \in \left(F^{-1+}_{S^l}(0)\,,\ F^{-1}_{S^l}(1)\right)$.

So far, we considered the case that all $E[X_i \mid \Lambda]$ are non-increasing functions of Λ. The case where all $E[X_i \mid \Lambda]$ are non-decreasing functions of Λ also leads to a comonotonic vector $(E[X_1 \mid \Lambda], E[X_2 \mid \Lambda], \ldots, E[X_n \mid \Lambda])$, and can be treated in a similar way.

Let us now consider the general case where not all $E[X_i \mid \Lambda]$ are non-increasing (or not all are non-decreasing). In this case the lower bound is not a sum of comonotonic random variables, making the determination of the distribution function of the lower bound more complicated. The cdf and the stop-loss premiums of S^l can be determined as follows in this case:

$$\begin{aligned} F_{S^l}(x) &= \int_{-\infty}^{+\infty} \Pr\left[\sum_{i=1}^{n} E[X_i \mid \Lambda] \leq x \mid \Lambda = \lambda\right] dF_\Lambda(\lambda) \\ &= \int_{-\infty}^{+\infty} I\left(\sum_{i=1}^{n} E[X_i \mid \Lambda = \lambda] \leq x\right) dF_\Lambda(\lambda)\,; \end{aligned} \tag{101}$$

$$E\left[\left(S^l - d\right)_+\right] = \int_{-\infty}^{+\infty} \left(\sum_{i=1}^{n} E[X_i \mid \Lambda = \lambda] - d\right)_+ dF_\Lambda(\lambda)\,. \tag{102}$$

A somewhat different procedure can be used when F_Λ is continuous and strictly increasing. In this case, define the random variable $U \equiv F_\Lambda(\Lambda)$ which is uniformly distributed on the unit interval. We have that $U = u \Leftrightarrow \Lambda = F_\Lambda^{-1}(u)$ holds for all $0 < u < 1$. Hence, the cdf and the stop-loss premiums of S^l then follow from

$$\begin{aligned} F_{S^l}(x) &= \int_0^1 \Pr\left[\sum_{i=1}^n E\left[X_i|\Lambda\right] \leq x \mid U = u\right] du \\ &= \int_0^1 I\left(\sum_{i=1}^n E\left[X_i|\Lambda = F_\Lambda^{-1}(u)\right] \leq x\right) du \end{aligned} \qquad (103)$$

$$E\left[\left(S^l - d\right)_+\right] = \int_0^1 \left(\sum_{i=1}^n E\left[X_i|\Lambda = F_\Lambda^{-1}(u)\right] - d\right)_+ du. \qquad (104)$$

The technique for deriving lower bounds as explained in this section is also considered (for some special cases) in Vyncke, Goovaerts & Dhaene (2000). The idea of this technique stems from mathematical physics, and was applied by Rogers & Shi (1995) to derive approximate values for the price of Asian options.

6 Conclusions

In this paper, we presented some simple yet powerful techniques to deal with sums of dependent random variables whose marginal distributions are known but with an unknown or complicated joint distribution. The central idea consists in replacing the original sum by another one, with a simpler dependence structure, and which is considered to be less favorable by all risk-averse decision makers. This extremal sum involves the components of the comonotonic version of the original random vector.

The main advantage of this approach is that it leads to easily computable distribution functions and stop-loss premiums, while the evaluations are always conservative. Moreover, considering comonotonic random vectors essentially reduces the multidimensional problem to a univariate one.

In some cases, improved approximations can be obtained when additional information is available. Specifically, if the marginal distributions of the summands, given some other random variable, are known, more accurate bounds on actuarial quantities can be derived.

The present paper aimed to describe the theoretical aspects of the problem. In a subsequent paper, Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002), we propose several applications of the techniques considered in this paper to various financial-actuarial problems.

7 Acknowledgements

Michel Denuit, Jan Dhaene and Marc Goovaerts would like to acknowledge the financial support of the Committee on Knowledge Extension Research of the Society of Actuaries for the project "Actuarial Aspects of Dependencies in Insurance Portfolios". The current paper and also Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002) result from this project.
Marc Goovaerts and Jan Dhaene also acknowledge the financial support of the Onderzoeksfonds K.U. Leuven (GOA/02: Actuariële, financiële en statistische aspecten van afhankelijkheden in verzekerings- en financiële portefeuilles).

References

[1] Bäuerle, N.; Müller, A. (1998). "Modeling and comparing dependencies in multivariate risk portfolios", *ASTIN Bulletin* 28, 59-76.

[2] Cossette, H.; Denuit, M.; Dhaene, J.; Marceau, E. (2001). "Stochastic approximations of present value functions", *Bulletin of the Swiss Association of Actuaries*, to appear.

[3] Cossette, H.; Denuit, M.; Marceau, E. (2000). "Impact of dependence among multiple claims in a single loss", *Insurance: Mathematics & Economics*, 26, 213-222.

[4] Denneberg, D. (1994). "Non-additive measure and integral", Kluwer Academic Publishers, Boston, pp. 184.

[5] Denuit, M; Cornet, A. (1999). "Premium calculation with dependent time-until-death random variables: the widow's pension", *Journal of Actuarial Practice*, 7, 147-180.

[6] Denuit, M.; De Vylder, F.; Lefèvre, C. (1999). "Extrema with respect to s-convex orderings in moment spaces: a general solution", *Insurance: Mathematics & Economics* 24, 201-217.

[7] Denuit, M.; Dhaene, J.; Lebailly De Tilleghem, C.; Teghem, S. (2001). "Measuring the impact of a dependence among insured lifelengths", *Belgian Actuarial Bulletin*, to be published.

[8] Denuit, M; Dhaene, J.; Ribas, C. (2001). "Does positive dependence between individual risks increase stop-loss premiums?", *Insurance: Mathematics & Economics*, to appear.

[9] Denuit, M.; Genest, Ch.; Marceau, E. (1999). "Stochastic bounds on sums of dependent risks", *Insurance: Mathematics & Economics, 25, 85-104.*

[10] Denuit, M.; Genest, Ch.; Marceau, E. (2001). "A criterion of the stochastic ordering of random sums", *Scandinavian Actuarial Journal*, to appear.

[11] Denuit, M.; Lefèvre, C. (1997). "Stochastic product orderings, with applications in actuarial sciences", *Bulletin Français d'Actuariat* 1, 61-82.

[12] De Vijlder, F.; Dhaene, J. (2001). "An introduction to the theory of comonotonic risks", to appear.

[13] Dhaene, J.; Denuit, M. (1999). "The safest dependency structure among risks", *Insurance: Mathematics & Economics* 25, 11-21.

[14] Dhaene, J.; Denuit, M.; Goovaerts, M.J.; Kaas, R.; Vyncke, D. (2002). "The concept of comonotonicity in actuarial science and finance: Applications", in preparation.

[15] Dhaene, J.; Goovaerts, M.J. (1996). "Dependency of risks and stop-loss order", *ASTIN Bulletin* 26, 201-212.

[16] Dhaene, J.; Goovaerts, M.J. (1997). "On the dependency of risks in the individual life model", *Insurance: Mathematics & Economics* 19, 243-253.

[17] Dhaene, J.; Vanneste, M.; Wolthuis H. (2000). "A note on dependencies in multiple life statuses", *Mitteilungen der Schweiz. Aktuarvereinigung*, 2000(1), 19-34.

[18] Dhaene, J.; Wang, S.; Young, V.; Goovaerts, M.J. (2000). "Comonotonicity and maximal stop-loss premiums", *Mitteilungen der Schweiz. Aktuarvereinigung*, 2000(2), 99-113.

[19] Embrechts, P.; Mc.Neil, A.; Straumann, D. (2001). "Correlation and dependency in risk management: properties and pitfalls", in "Risk Management: Value at Risk and Beyond", edited by Dempster, M. and Moffatt, H.K., Cambridge University Press.

[20] Fréchet, M. (1951). "Sur les tableaux de corrélation dont les marges sont donnés", *Ann. Univ. Lyon Sect.A,* Series 3, 14, 53-77.

[21] Goovaerts, M.J.; Dhaene, J.(1999). "Supermodular ordering and stochastic annuities", *Insurance: Mathematics & Economics*, 24, 281-290.

[22] Goovaerts, M.J.; Dhaene, J;. De Schepper, A. (2000). "Stochastic Upper Bounds for Present Value Functions", *Journal of Risk and Insurance Theory*, 67.1, 1-14.

[23] Goovaerts, M.J.; Kaas, R. (2001). "Some problems in actuarial finance involving sums of dependent risks", *Statistica Neerlandica*, to appear.

[24] Goovaerts, M.J.; Kaas, R.; Van Heerwaarden, A.; Bauwelinckx, T. (1990). Effective Actuarial Methods. Insurance Series, vol. 3, North-Holland, Amsterdam, New York, Oxford, Tokyo.

[25] Goovaerts, M.J.; Redant, R. (1999). "On the distribution of IBNR reserves", *Insurance: Mathematics & Economics* 25, 1-9.

[26] Heilmann, W.R. (1986). "On the impact of the independence of risks on stop-loss premiums", *Insurance: Mathematics & Economics*, 5, 197-199.

[27] Hoeffding, W. (1940). "Masstabinvariante Korrelationstheorie", *Schriften des mathematischen Instituts und des Instituts für angewandte Mathematik der Universität Berlin* 5, 179-233.

[28] Kaas, R.; Dhaene, J.; Vyncke, D.; Goovaerts, M.J.; Denuit, M. "A simple geometric proof that comonotonic risks have the convex-largest sum", Research Report Department of Applied Economics, K.U. Leuven.

[29] Kaas, R., Van Heerwaarden, A.E., Goovaerts, M.J. (1994). "Ordering of actuarial risks", Institute for Actuarial Science and Econometrics, University of Amsterdam, Amsterdam.

[30] Kaas, R.; Dhaene, J.; Goovaerts, M.J. (2000). "Upper and lower bounds for sums of random variables", *Insurance: Mathematics & Economics*, 27, 151-168.

[31] Luan, C. (2001). "Insurance premium calculations with anticipated utility theory", *ASTIN Bulletin*, 31(1), 23-35.

[32] Müller, A. (1997). "Stop-loss order for portfolios of dependent risks", *Insurance: Mathematics & Economics*, 21, 219-223.

[33] Rogers, L.C.G.; Shi, Z. (1995). "The value of an Asian option", *Journal of Applied Probability* 32, 1077-1088.

[34] Shaked, M.; Shanthikumar, J.G. (1994). "Stochastic orders and their applications", Academic Press, pp. 545.

[35] Simon, S.; Goovaerts, M.J.; Dhaene, J. (2000). "An easy computable upper bound for the price of an arithmetic Asian option", *Insurance: Mathematics & Economics*, 26.2-3, 175-184.

[36] Vyncke, D.; Goovaerts, M.J.; Dhaene, J. (2001): "Convex upper and lower bounds for present value functions", *Applied Stochastic Models in Business and Industry*, 17, 149-164.

[37] Wang, S.; Dhaene, J. (1998). "Comonotonicity, correlation order and stop-loss premiums", *Insurance: Mathematics & Economics*, 22, 235-243.

[38] Wang, S.; Young, V. (1998). "Ordering risks: expected utility versus Yaari's dual theory of choice under risk", *Insurance: Mathematics & Economics* 22, 145-162.