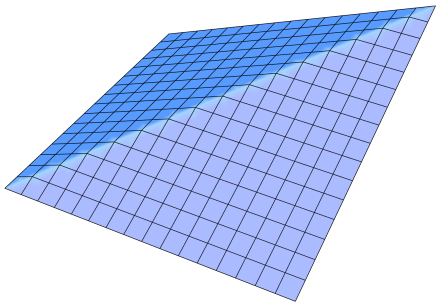

STA 4528H: Dependence Modelling

Prof. Silvana Pesenti • Winter 2021 • University of Toronto
T 3:00–6:00 PM (Online)

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is https://rob-zimmerman.github.io/coursenotes.

Contents

1 Distributional Foundations and Copulas **3**
Univariate Distributions . . . 3
Quantile Functions . . . 6
Multivariate Distributions . . . 8
Conditional Distributions . . . 15
Conditional Copulas . . . 15
Rosenblatt Transform . . . 19
Sampling from a Copula . . . 21

2 Dependence Structures and Copula Families **22**
Comonotonicity and Countermonotonicity . . . 22
Bounds on Copulas . . . 25
Examples of Copulas . . . 26
Gaussian and t-Copulas . . . 30
Archimedean Copulas . . . 31
Constructing Multivariate Copulas . . . 35

3 Dependence Measures and Tail Dependence **36**
Pearson Linear Correlation . . . 36
Rank Correlation . . . 38
Kendall's Tau and Spearman's Rho . . . 39
Tail Dependence . . . 41
Survival Copulas . . . 42

4 Risk Measures and Expected Shortfall **44**
Risk Measures . . . 44
Desirable Properties . . . 44
Convex Risk Measures . . . 46
Expected Shortfall . . . 51
Representation Theorems . . . 53
Distortion Risk Measures . . . 56

5 Dependence Uncertainty and Risk Bounds **62**
Risk Versus Uncertainty . . . 62
Risk Measurement Under Uncertainty . . . 62
Stochastic Ordering . . . 63
Complete Dependence Uncertainty . . . 67
Bounds on VaR . . . 70
Rearrangement Algorithm . . . 78
Bounds with Moment Information . . . 82

Appendix: Lecture and Tutorial Index **86**

Lecture 1 — Tuesday, March 09, 2021

1. Distributional Foundations and Copulas

Univariate Distributions

Throughout, all random variables are assumed to be defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 1.1 The **distribution function** of a random variable X is the function $F_X : \mathbb{R} \to [0,1]$ defined by

$$F_X(x) := \mathbb{P}(X \leq x) \quad \text{for all } x \in \mathbb{R}.$$

Proposition 1.2 A function $F : \mathbb{R} \to [0,1]$ is a distribution function if and only if it has the following properties:

1. The function F is nondecreasing: $F(x) \leq F(y)$ whenever $x \leq y$.
2. The function F is right-continuous:
$$\lim_{h \downarrow 0} F(x+h) = F(x) \quad \text{for all } x \in \mathbb{R}.$$
3. The left tail limit satisfies $\lim_{x \to -\infty} F(x) = 0$.
4. The right tail limit satisfies $\lim_{x \to +\infty} F(x) = 1$.

The jump in Figure 1 illustrates right-continuity: the value of the distribution function at the jump is the upper endpoint.

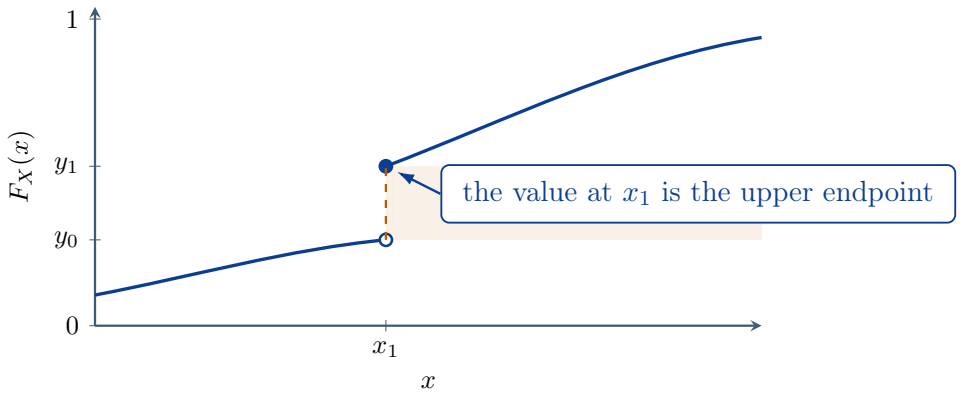

FIGURE 1

The flat portion in Figure 2 shows that a distribution function need not attain every value in $[0, 1]$.

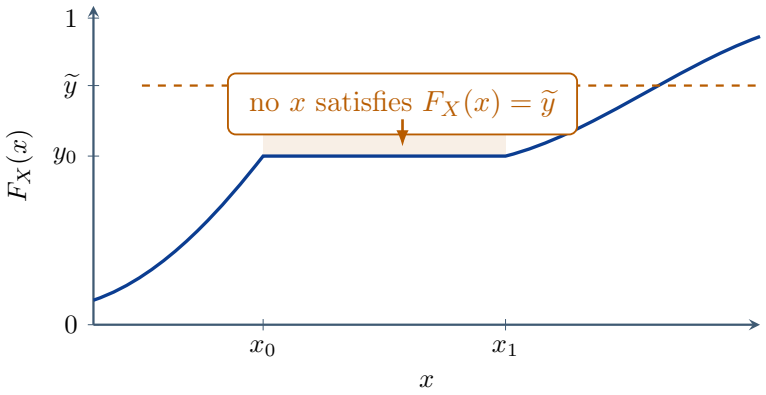

FIGURE 2

At the highlighted level in Figure 2, there is no $x \in \mathbb{R}$ such that $F_X(x) = \widetilde{y}$.

Definition 1.3 Let $F_X : \mathbb{R} \to [0, 1]$ be the distribution function of a random variable X. The **left-continuous generalized inverse** of F_X is defined by

$$\begin{aligned} F_X^{-1}(u) &:= \inf\{y \in \mathbb{R} \mid F_X(y) \geq u\} \\ &= \inf\{y \in \mathbb{R} \mid \mathbb{P}(X \leq y) \geq u\} \quad \text{for all } u \in [0, 1], \end{aligned}$$

with the convention that $\inf \emptyset = +\infty$. The function $F_X^{-1} : [0, 1] \to \overline{\mathbb{R}}$ is called the **quantile function** of X, or of F_X.

The horizontal guide in Figure 3 shows how the infimum in Definition 1.3 selects the leftmost point at which the distribution function reaches a prescribed level.

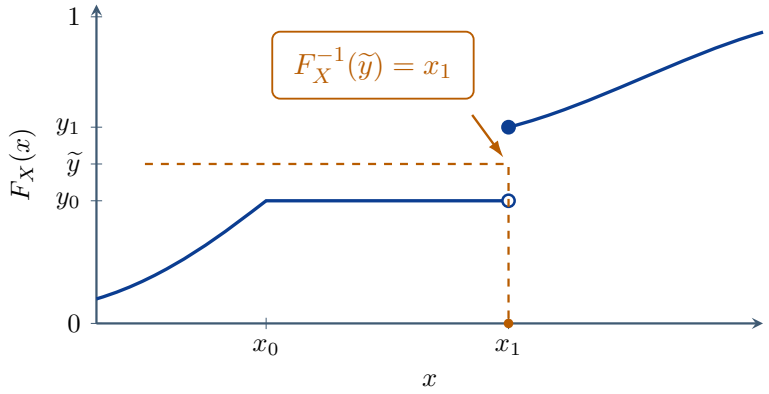

FIGURE 3

Interchanging the roles of the horizontal and vertical axes produces the generalized inverse shown in Figure 4.

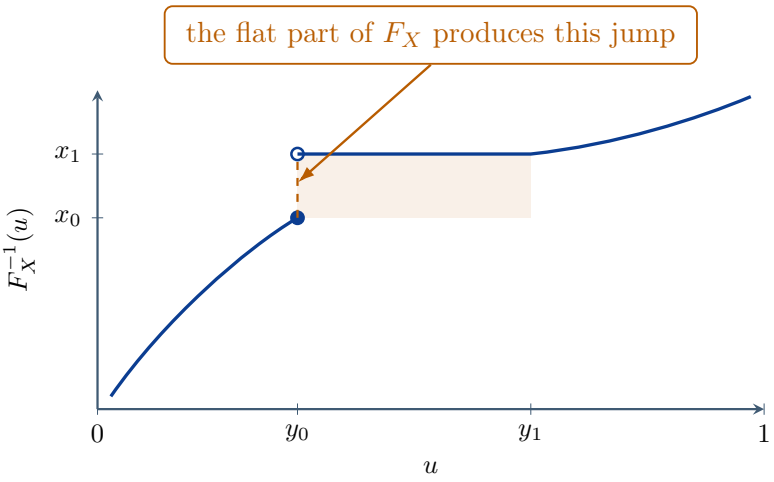

FIGURE 4

Remark 1.4 Flat portions of a distribution function F correspond to jumps of F^{-1}, whereas jumps of F correspond to flat portions of F^{-1}.

Quantile Functions

Proposition 1.5 Let F be a distribution function with generalized inverse F^{-1}. Then the following properties hold:

1. The function F^{-1} is nondecreasing.
2. The function F^{-1} is left-continuous.
3. $F^{-1}(F(x)) \leq x$ for all $x \in \mathbb{R}$; if F^{-1} is strictly increasing at the point $F(x)$, then equality holds, that is $F^{-1}(F(x)) = x$.
4. For $u \in (0,1)$ and $x \in \mathbb{R}$, the inequality $F(x) \geq u$ holds if and only if $x \geq F^{-1}(u)$.
5. For $u \in (0,1)$, one has $F(F^{-1}(u)) \geq u$. If, in addition, $u \in \mathrm{im}(F)$, then $F(F^{-1}(u)) = u$. Endpoint identities are understood through limits because $F^{-1}(0)$ or $F^{-1}(1)$ may be infinite.
6. The function F is continuous if and only if F^{-1} is strictly increasing on
$$[\inf \mathrm{im}(F), \sup \mathrm{im}(F)) .$$
7. The function F is strictly increasing if and only if F^{-1} is continuous on $\mathrm{im}(F)$.

Here
$$\mathrm{im}(F) := \{u \in [0,1] \mid \text{there exists } x \in \mathbb{R} \text{ such that } F(x) = u\}.$$

By (5), a random variable X with distribution function F_X satisfies
$$F_X(F_X^{-1}(u)) = u \quad \text{for all } u \in \mathrm{im}(F_X).$$
If F_X is continuous and strictly increasing, then
$$F_X^{-1}(F_X(x)) = x \quad \text{for all } x \in \mathbb{R}.$$

Theorem 1.6 (Inverse transform theorem) Let F be a distribution function, and let X have distribution function F.

1. If $U \sim \mathrm{Unif}[0,1]$, then $F^{-1}(U)$ has distribution function F.
2. If F is continuous, then $F(X) \sim \mathrm{Unif}[0,1]$.

Proof. For (1), let $x \in \mathbb{R}$. Since $U \in (0,1)$ almost surely, (4) gives

$$\mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F(x),$$

because $U \sim \mathrm{Unif}[0,1]$. More explicitly, the relevant events are equal:

$$\{\omega \in \Omega \mid F^{-1}(U(\omega)) \le x\} = \{\omega \in \Omega \mid U(\omega) \le F(x)\}.$$

For (2), fix $u \in (0,1)$ and set $q = F^{-1}(u)$. Continuity gives $F(q) = u$. Moreover, $\{F(X) \le u\}$ and $\{X \le q\}$ can differ only when X lies in a flat portion of F at level u, an event of probability zero. Hence

$$\mathbb{P}(F(X) \le u) = \mathbb{P}(X \le q) = F(q) = u.$$

The endpoint values follow automatically, so $F(X) \sim \mathrm{Unif}[0,1]$. □

Corollary 1.7 For any random variable Y and any $U \sim \mathrm{Unif}[0,1]$,

$$Y \overset{d}{=} F_Y^{-1}(U).$$

Proposition 1.8 Let X have continuous distribution function F. Then there exists $U \sim \mathrm{Unif}[0,1]$ such that

$$X = F^{-1}(U) \quad \mathbb{P}\text{-almost surely.}$$

In particular, we may take $U \coloneqq F(X)$.

Proof. By (2), $U = F(X)$ is uniformly distributed on $[0,1]$. Moreover,

$$F^{-1}(F(X)) \le X \quad \mathbb{P}\text{-almost surely.}$$

The inverse-transform theorem also gives $F^{-1}(F(X)) \overset{d}{=} X$. Two identically distributed real random variables that are almost surely ordered must be equal almost surely, which proves the required identity. Notice that pointwise equality can fail inside a flat portion of F; such a portion carries no probability under F. □

The inverse transform may be extended to distribution functions that are not continuous by randomizing within their jumps.

Proposition 1.9 Let Y have distribution function F_Y, and let $V \sim \text{Unif}[0,1]$ be independent of Y. Define the **distributional transform**

$$\overline{F}_Y(y, \lambda) \coloneqq \mathbb{P}(Y < y) + \lambda \mathbb{P}(Y = y).$$

Then

1. $\overline{F}_Y(Y, V) \sim \text{Unif}[0,1]$.
2. $F_Y^{-1}(\overline{F}_Y(Y,V)) = Y$ almost surely, and hence $F_Y^{-1}(\overline{F}_Y(Y,V))$ has distribution function F_Y.

Proposition 1.10 For every random variable Y, there exists $U \sim \text{Unif}[0,1]$ such that

$$F_Y^{-1}(U) = Y \quad \mathbb{P}\text{-almost surely.}$$

In particular, we may take $U \coloneqq \overline{F}_Y(Y,V)$ as in Proposition 1.9.

Multivariate Distributions

Definition 1.11 The **multivariate distribution function** of a random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is the function $F_{\boldsymbol{X}} \colon \mathbb{R}^n \to [0,1]$ defined by

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \coloneqq \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n.$$

Notation 1.12

1. Capital letters denote random variables.
2. Lowercase letters denote their realizations.
3. Bold letters denote random vectors or their realizations, with capitalization distinguishing the two cases.

For vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, the inequality $\boldsymbol{x} \leq \boldsymbol{y}$ is understood componentwise.

Proposition 1.13 A function $F\colon \mathbb{R}^n \to [0,1]$ is a multivariate distribution function if and only if it has the following properties:

1. The function F is nondecreasing on $\mathbb{R}^n$:
$$F(\boldsymbol{x}) \le F(\boldsymbol{y}) \quad \text{whenever } \boldsymbol{x} \le \boldsymbol{y}.$$
2. The function F is right-continuous on $\mathbb{R}^n$:
$$\lim_{\Delta\boldsymbol{x} \searrow \boldsymbol{0}} F(\boldsymbol{x} + \Delta\boldsymbol{x}) = F(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n,$$
where $\Delta\boldsymbol{x} \searrow \boldsymbol{0}$ means that $\Delta x_i \searrow 0$ for every $i \in \{1,\dots,n\}$.
3. For each $i \in \{1,\dots,n\}$,
$$\lim_{x_i \to -\infty} F(\boldsymbol{x}) = 0.$$
4. The joint right tail limit satisfies
$$\lim_{\boldsymbol{x} \to +\boldsymbol{\infty}} F(\boldsymbol{x}) = 1.$$
5. The function F is n-increasing. More precisely, for $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^n$ with $\boldsymbol{x} \le \boldsymbol{x}'$, define
$$\Delta_{x_i, x_i'} F(\boldsymbol{y}) \coloneqq F(y_1,\dots,y_{i-1}, x_i', y_{i+1},\dots,y_n) - F(y_1,\dots,y_{i-1}, x_i, y_{i+1},\dots,y_n).$$
Then
$$\Delta_{x_1,x_1'} \Delta_{x_2,x_2'} \cdots \Delta_{x_n,x_n'} F(\boldsymbol{y}) \ge 0.$$

The final quantity does not depend on $\boldsymbol{y}$.

Remark 1.14 Condition (5) is equivalent to
$$\mathbb{P}\big(X_1 \in (x_1, x_1'], \dots, X_n \in (x_n, x_n']\big) \ge 0 \quad \text{for all } \boldsymbol{x} \le \boldsymbol{x}'.$$
If, for example, $x_n = x_n'$, then the displayed n-dimensional increment is simply zero. The corresponding $(n-1)$-dimensional nonnegativity statement is obtained instead by applying the increment operators to the $(n-1)$-dimensional marginal, equivalently by sending the omitted coordinate to $+\infty$.

Example 1.15 For $n = 2$ and $(x_1, x_2) \leq (x_1', x_2')$, the difference in the second coordinate is

$$\Delta_{x_2, x_2'} F(\boldsymbol{y}) \coloneqq F(y_1, x_2') - F(y_1, x_2).$$

Solution. Taking the difference in the first coordinate gives

$$\Delta_{x_1, x_1'} \Delta_{x_2, x_2'} F(\boldsymbol{y}) = F(x_1', x_2') - F(x_1', x_2) - F(x_1, x_2') + F(x_1, x_2) \geq 0.$$

In probabilistic terms, this identity becomes

$$\begin{aligned}
&\mathbb{P}(X_1 \leq x_1', X_2 \leq x_2') - \mathbb{P}(X_1 \leq x_1', X_2 \leq x_2) \\
&\quad - \big(\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2') - \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)\big) \\
&= \mathbb{P}(X_1 \leq x_1', X_2 \in (x_2, x_2']) - \mathbb{P}(X_1 \leq x_1, X_2 \in (x_2, x_2']) \\
&= \mathbb{P}(X_1 \in (x_1, x_1'], X_2 \in (x_2, x_2']) \geq 0.
\end{aligned}$$

□

Proposition 1.16 Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ have joint distribution function $F_{\boldsymbol{X}}$. For every $i \in \{1, \ldots, n\}$,

$$F_{X_i}(x) = F_{\boldsymbol{X}}(\infty, \ldots, \infty, x, \infty, \ldots, \infty),$$

where x appears in the ith coordinate. Similarly, for distinct i and j,

$$F_{X_i, X_j}(x, y) = F_{\boldsymbol{X}}(\infty, \ldots, x, \ldots, y, \ldots, \infty),$$

where x and y appear in the ith and jth coordinates, respectively.

Remark 1.17 The operation in Proposition 1.16 is the analogue, for distribution functions, of integrating out coordinates from a density. If (X_1, X_2) has joint density $f_{\boldsymbol{X}}$, then

$$f_{X_1}(x) = \int_{\mathbb{R}} f_{\boldsymbol{X}}(x, y)\, \mathrm{d}y = \frac{\partial}{\partial x} F_{X_1}(x) = \frac{\partial}{\partial x} F_{\boldsymbol{X}}(x, \infty).$$

Definition 1.18 An n-dimensional **copula** $C\colon [0,1]^n \to [0,1]$ is a multivariate distribution function with uniform marginals.

Equivalently, if $\boldsymbol{U} = (U_1, \ldots, U_n)$ has distribution function C, then each U_i is uniformly distributed

on $[0,1]$ and

$$\mathbb{P}(U_1 \leq u_1, \ldots, U_n \leq u_n) = C(\boldsymbol{u}) \quad \text{for all } \boldsymbol{u} \in [0,1]^n.$$

No assumption of independence is imposed on the coordinates.

Proposition 1.19 A function $C\colon [0,1]^n \to [0,1]$ is a copula if and only if it satisfies the following conditions:

1. The function C is right-continuous on $[0,1]^n$.
2. The function C is nondecreasing on $[0,1]^n$.
3. For every $i \in \{1, \ldots, n\}$,
$$C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_n) = 0.$$
4. For every $i \in \{1, \ldots, n\}$,
$$C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i.$$
5. The function C is n-increasing: for $\boldsymbol{u} \leq \boldsymbol{u}'$,
$$\Delta_{u_1, u_1'} \cdots \Delta_{u_n, u_n'} C(\boldsymbol{v}) \geq 0.$$

In particular, $C(1, \ldots, 1) = 1$.

Remark 1.20 Condition (4) gives the uniform marginals directly. If $\boldsymbol{U} \sim C$, then

$$\mathbb{P}(U_i \leq u_i) = C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i.$$

Proposition 1.21 Let C be an n-dimensional copula and $i \in \{1, \ldots, n\}$. Then

$$\widetilde{C}(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) \coloneqq C(u_1, \ldots, u_{i-1}, 1, u_{i+1}, \ldots, u_n)$$

is an $(n-1)$-dimensional copula. More generally, every k-dimensional marginal obtained by setting $n-k$ arguments equal to 1 is a k-dimensional copula.

Theorem 1.22 (Sklar's theorem)

1. If F is an n-dimensional distribution function with univariate marginals $F_1, \ldots, F_n$, then there exists a copula C such that

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n.$$

2. Conversely, if C is an n-dimensional copula and $F_1, \ldots, F_n$ are univariate distribution functions, then

$$F(x_1, \ldots, x_n) \coloneqq C(F_1(x_1), \ldots, F_n(x_n))$$

defines a multivariate distribution function with marginals $F_1, \ldots, F_n$.

Moreover, if $F_1, \ldots, F_n$ are continuous, then C is unique. Otherwise, C is uniquely determined on $\operatorname{im}(F_1) \times \cdots \times \operatorname{im}(F_n)$.

The decomposition and reconstruction in Theorem 1.22 are summarized in Figure 5. The marginal models determine the behaviour of the individual coordinates, whereas the copula records how their probability levels move together.

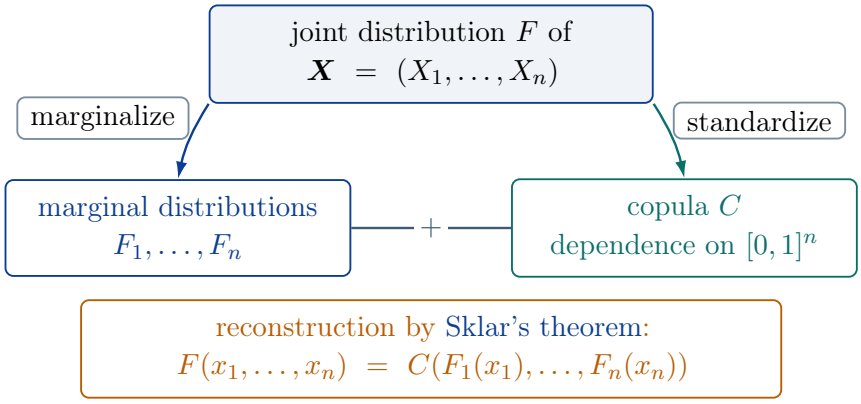

FIGURE 5

Remark 1.23 Sklar's theorem separates a multivariate distribution into its univariate marginals and its dependence structure. The marginals are often easier to estimate because each coordinate supplies its own observations. Estimating a joint tail probability such

as

$$\mathbb{P}(X_1 > x_1, \dots, X_n > x_n)$$

is much more difficult, particularly in high dimensions, because very few observations may lie in the joint tail.

Proof. 1. Let $\boldsymbol{X} \sim F$, enlarge the probability space by independent standard uniforms $V_1, \dots, V_n$, and define the coordinatewise distributional transforms

$$U_i \coloneqq F_i(X_i-) + V_i\big(F_i(X_i) - F_i(X_i-)\big), \quad i \in \{1, \dots, n\}.$$

Here $F_i(x-) = \lim_{t \uparrow x} F_i(t) = \mathbb{P}(X_i < x)$. By Proposition 1.9, every U_i is standard uniform and $X_i = F_i^{-1}(U_i)$ almost surely. Let C be the joint distribution function of $\boldsymbol{U} = (U_1, \dots, U_n)$. Since its marginals are uniform, C is a copula. Therefore,

$$\begin{aligned} F(\boldsymbol{x}) &= \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) \\ &= \mathbb{P}(F_1^{-1}(U_1) \le x_1, \dots, F_n^{-1}(U_n) \le x_n) \\ &= \mathbb{P}(U_1 \le F_1(x_1), \dots, U_n \le F_n(x_n)) \\ &= C(F_1(x_1), \dots, F_n(x_n)). \end{aligned}$$

2. Let $\boldsymbol{U} \sim C$ and define $X_i = F_i^{-1}(U_i)$. The inverse-transform theorem gives $X_i \sim F_i$, while the generalized-inverse event identity gives

$$\begin{aligned} \mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) &= \mathbb{P}(U_1 \le F_1(x_1), \dots, U_n \le F_n(x_n)) \\ &= C(F_1(x_1), \dots, F_n(x_n)). \end{aligned}$$

Thus this composition is a multivariate distribution function with the prescribed marginals.

If every F_i is continuous, its image is all of $[0,1]$ and substitution of the generalized inverses determines C everywhere. Without continuity, the identity determines C only on $\operatorname{im}(F_1) \times \cdots \times \operatorname{im}(F_n)$; the distributional-transform construction supplies one extension to the full cube. □

Proposition 1.24 Let $\boldsymbol{X} = (X_1, \dots, X_n)$ have copula C, and let $g_1, \dots, g_n \colon \mathbb{R} \to \mathbb{R}$ be strictly increasing. Then $(g_1(X_1), \dots, g_n(X_n))$ also has copula C. Moreover, for y in the range of g_i,

$$\mathbb{P}(g_i(X_i) \le y) = F_i(g_i^{-1}(y)) \quad \text{for } y \in g_i(\mathbb{R}).$$

Proof. For $y_i \in g_i(\mathbb{R})$, strict monotonicity gives

$$\begin{aligned}\mathbb{P}(g_1(X_1) \le y_1, \dots, g_n(X_n) \le y_n) &= \mathbb{P}(X_1 \le g_1^{-1}(y_1), \dots, X_n \le g_n^{-1}(y_n)) \\ &= C(F_1(g_1^{-1}(y_1)), \dots, F_n(g_n^{-1}(y_n))).\end{aligned}$$

The ith marginal is

$$\mathbb{P}(g_i(X_i) \le y_i) = C(1, \dots, 1, F_i(g_i^{-1}(y_i)), 1, \dots, 1) = F_i(g_i^{-1}(y_i)).$$

For values outside the range, the transformed marginal is 0 or 1 and the same copula identity follows by a one-sided limit. □

Remark 1.25 By Proposition 1.24, every random vector $\boldsymbol{X}$ with strictly increasing marginals has the same copula as $(F_1(X_1), \dots, F_n(X_n))$. Thus, after estimating the marginals from data, we may transform the observations to uniform marginals and study their copula separately.

Proposition 1.26 Let F be a multivariate distribution function with continuous marginals $F_1, \dots, F_n$, and let C be its copula. Then

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \quad \text{for all } \boldsymbol{u} \in [0,1]^n.$$

Proof. By (1),

$$F(\boldsymbol{x}) = C(F_1(x_1), \dots, F_n(x_n)).$$

Substituting $x_i = F_i^{-1}(u_i)$ and using (5) gives

$$\begin{aligned}F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) &= C(F_1(F_1^{-1}(u_1)), \dots, F_n(F_n^{-1}(u_n))) \\ &= C(u_1, \dots, u_n).\end{aligned}$$

□

for $\boldsymbol{u} \in (0,1)^n$. The formula at boundary points follows by continuity of copulas and the corresponding one-sided limits.

Conditional Distributions

Definition 1.27 Let $F_{\boldsymbol{X}}$ be the joint distribution function of $\boldsymbol{X} = (X_1, X_2)$, and fix a regular conditional distribution of X_1 given X_2. The **conditional distribution function** of X_1 given $X_2 = x_2$ is a measurable version of

$$F_{X_1|X_2}(x_1 \mid x_2) \coloneqq \mathbb{P}(X_1 \leq x_1 \mid X_2 = x_2).$$

Remark 1.28 For F_{X_2}-almost every x_2, the function $F_{X_1|X_2}(\cdot \mid x_2)$ is a distribution function in its first argument. Values on the remaining F_{X_2}-null set may be chosen as part of the version. In contrast, $F_{X_1|X_2}(x_1 \mid \cdot)$ need not be a distribution function in x_2.

Conditional Copulas

Let $C_{1,2}\colon [0,1]^2 \to [0,1]$ be the copula of (U_1, U_2). The conditional distribution function associated with the copula is

$$C_{1|2}(u_1 \mid u_2) \coloneqq \mathbb{P}(U_1 \leq u_1 \mid U_2 = u_2).$$

For every fixed $u_2 \in [0,1]$, the function $C_{1|2}(\cdot \mid u_2)$ is a distribution function.

As in Definition 1.3, we may take generalized inverses in the first argument. For (X_1, X_2) with joint distribution function F_{X_1,X_2}, define

$$\begin{aligned} F^{-1}_{X_1|X_2}(u \mid x_2) &\coloneqq \inf\{y \in \mathbb{R} \mid F_{X_1|X_2}(y \mid x_2) \geq u\} \\ &= \inf\{y \in \mathbb{R} \mid \mathbb{P}(X_1 \leq y \mid X_2 = x_2) \geq u\}. \end{aligned}$$

For fixed x_2, the map $F^{-1}_{X_1|X_2}(\cdot \mid x_2)\colon [0,1] \to \overline{\mathbb{R}}$ has the usual properties of a generalized inverse. Likewise, if $(U_1, U_2) \sim C_{1,2}$, define

$$C^{-1}_{1|2}(u \mid v) \coloneqq \inf\{y \in [0,1] \mid C_{1|2}(y \mid v) \geq u\}.$$

Remark 1.29 In each of these definitions, the inverse is taken with respect to the first argument. The same construction extends to additional conditioning variables. For example,

$$F^{-1}_{X_1|X_2,X_3}(u \mid x_2, x_3)$$

is the generalized inverse of the conditional distribution function of X_1 given $X_2 = x_2$ and $X_3 = x_3$.

Proposition 1.30 For every fixed y for which the conditional distribution function is defined, continuity of $F_{X_1|X_2}(\cdot \mid y)$ gives

$$F_{X_1|X_2}\big(F^{-1}_{X_1|X_2}(u \mid y) \mid y\big) = u \quad \text{for all } u \in (0,1).$$

If $F_{X_1|X_2}(\cdot \mid y)$ is strictly increasing as well, then

$$F^{-1}_{X_1|X_2}\big(F_{X_1|X_2}(x \mid y) \mid y\big) = x.$$

Analogously, continuity of $C_{1|2}(\cdot \mid v)$ gives the first identity below, whereas strict increase gives the second:

$$C_{1|2}\big(C^{-1}_{1|2}(u \mid v) \mid v\big) = u,$$
$$C^{-1}_{1|2}\big(C_{1|2}(u \mid v) \mid v\big) = u,$$

for all $u \in (0,1)$ and every conditioning value v at which the stated version has these properties.

Theorem 1.31 Let (X, Y) be a bivariate random vector with uniform marginals and copula $C_{X,Y}$. Fix a jointly measurable regular conditional distribution $C_{Y|X}$, and assume that $C_{Y|X}(\cdot \mid x)$ is continuous and strictly increasing for F_X-almost every $x \in [0,1]$. Then there exists $V \sim \mathrm{Unif}[0,1]$, independent of X, such that

$$Y = C^{-1}_{Y|X}(V \mid X) \quad \mathbb{P}\text{-almost surely.}$$

In particular, we may take

$$V := C_{Y|X}(Y \mid X).$$

Remark 1.32 The expression $C_{Y|X}(Y \mid X)$ means that the function of two variables $C_{Y|X}(y \mid x)$ is evaluated at the random pair (Y, X).

Proof. It suffices to verify the following claims:

1. The random variable V is uniformly distributed on $[0,1]$.
2. V is independent of X.
3. The random variable $C^{-1}_{Y|X}(V \mid X)$ is uniformly distributed on $[0,1]$, and hence has the same marginal distribution as Y.
4. The pair $(X, C^{-1}_{Y|X}(V \mid X))$ has copula $C_{X,Y}$, the same copula as (X,Y).
5. The almost-sure representation holds.

For (1), the law of total probability gives

$$\begin{aligned}\mathbb{P}(V \le v) &= \mathbb{P}(C_{Y|X}(Y \mid X) \le v) \\ &= \mathbb{E}\left[\mathbb{P}(C_{Y|X}(Y \mid X) \le v \mid \sigma(X))\right].\end{aligned}$$

For F_X-almost every $x \in [0,1]$,

$$\begin{aligned}\mathbb{P}(C_{Y|X}(Y \mid X) \le v \mid X = x) &= \mathbb{P}(Y \le C^{-1}_{Y|X}(v \mid x) \mid X = x) \\ &= C_{Y|X}(C^{-1}_{Y|X}(v \mid x) \mid x) \\ &= v.\end{aligned}$$

Therefore,

$$\mathbb{P}(V \le v) = \mathbb{E}[v] = v \quad \text{for all } v \in [0,1],$$

which proves that $V \sim \text{Unif}[0,1]$.

For (2),

$$\begin{aligned}\mathbb{P}(X \le x, V \le v) &= \mathbb{P}(X \le x, C_{Y|X}(Y \mid X) \le v) \\ &= \mathbb{E}\left[\mathbb{1}_{\{X \le x\}} \mathbb{P}(C_{Y|X}(Y \mid X) \le v \mid \sigma(X))\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{X \le x\}} v\right] \\ &= \mathbb{P}(X \le x)\mathbb{P}(V \le v).\end{aligned}$$

Thus X and V are independent.

For (3), independence gives

$$\begin{aligned}\mathbb{P}(C^{-1}_{Y|X}(V \mid X) \le y) &= \int_0^1 \mathbb{P}(C^{-1}_{Y|X}(V \mid x) \le y)\, \mathrm{d}F_X(x) \\ &= \int_0^1 \mathbb{P}(V \le C_{Y|X}(y \mid x))\, \mathrm{d}F_X(x)\end{aligned}$$

$$
\begin{aligned}
&= \int_0^1 C_{Y|X}(y \mid x)\, \mathrm{d}F_X(x) \\
&= \int_0^1 \mathbb{P}(Y \le y \mid X = x)\, \mathrm{d}F_X(x) \\
&= \mathbb{P}(Y \le y) = y,
\end{aligned}
$$

because $Y \sim \text{Unif}[0,1]$. Thus Y and $C^{-1}_{Y|X}(V \mid X)$ have the same distribution.

For (4), note first that the order of the coordinates matters: $(C^{-1}_{Y|X}(V \mid X), X)$ need not have copula $C_{X,Y}$. In the stated order,

$$
\begin{aligned}
&\mathbb{P}(X \le x, C^{-1}_{Y|X}(V \mid X) \le v) \\
&= \int_0^x \mathbb{P}(C^{-1}_{Y|X}(V \mid t) \le v)\, \mathrm{d}F_X(t) \\
&= \int_0^x \mathbb{P}(V \le C_{Y|X}(v \mid t))\, \mathrm{d}F_X(t) \\
&= \int_0^x C_{Y|X}(v \mid t)\, \mathrm{d}F_X(t) \\
&= \int_0^x \mathbb{P}(Y \le v \mid X = t)\, \mathrm{d}F_X(t) \\
&= \mathbb{P}(X \le x, Y \le v) = C_{X,Y}(x, v).
\end{aligned}
$$

Finally, (5) follows from the inverse identity

$$
C^{-1}_{Y|X}(C_{Y|X}(y \mid x) \mid x) = y.
$$

□

The conditional inverse transform in Theorem 1.31 has a direct distributional analogue.

Corollary 1.33 Let (X, Y) have joint distribution function $F_{X,Y}$, and assume that the relevant conditional distribution functions are continuous and strictly increasing in their first arguments. Then $F_{Y|X}(Y \mid X)$ is uniformly distributed on $[0,1]$ and independent of X. Similarly, $F_{X|Y}(X \mid Y)$ is uniformly distributed on $[0,1]$ and independent of Y.

To interpret $F_{Y|X}(Y \mid X)$, begin with the deterministic function

$$
F_{Y|X}(y \mid x) = \mathbb{P}(Y \le y \mid X = x).
$$

Evaluating this function at the random pair (Y, X) produces a random variable. It should not be read as the ill-formed expression $\mathbb{P}(Y \le Y \mid X = X)$. To make the distinction precise, let $K(x, \cdot)$ be the chosen regular conditional law of Y given $X = x$. Then

$$F_{Y|X}(Y \mid X) = K(X, (-\infty, Y]),$$

which evaluates a deterministic probability kernel at the random pair (X, Y); it does not condition on the generally null event that an independent copy satisfies $X' = X$.

Theorem 1.34 Let (X, Y) have copula $C_{X,Y}$ and continuous, strictly increasing marginals F_X and F_Y. Assume also that a jointly measurable version of the conditional distribution of $F_Y(Y)$ given $F_X(X) = u$ is continuous and strictly increasing in its first argument for almost every u. Then there exists $V \sim \mathrm{Unif}[0, 1]$, independent of X, such that

$$Y = F_Y^{-1}\left(C_{Y|X}^{-1}(V \mid F_X(X))\right) \quad \mathbb{P}\text{-almost surely.}$$

Moreover,

$$V = C_{Y|X}(F_Y(Y) \mid F_X(X)).$$

Proof. The pair $(F_X(X), F_Y(Y))$ has uniform marginals and copula $C_{X,Y}$ by Proposition 1.24. Apply Theorem 1.31 to this pair, and then apply F_Y^{-1} to recover Y. □

Rosenblatt Transform

Theorem 1.35 Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random vector. On an auxiliary atomless probability space, there exist measurable functions $\psi^{(i)} \colon [0, 1]^n \to \mathbb{R}$ and a vector $\boldsymbol{U} = (U_1, \ldots, U_n)$ of independent standard uniform random variables such that

$$\boldsymbol{X} \overset{d}{=} \left(\psi^{(1)}(\boldsymbol{U}), \ldots, \psi^{(n)}(\boldsymbol{U})\right).$$

If the original probability space already supports the required independent randomization, the variables may be constructed there; otherwise the assertion is a distributional representation.

The standard **Rosenblatt transform** uses the marginal and conditional distribution functions of

$\boldsymbol{X}$ to define

$$\begin{aligned} U_1 &\coloneqq F_1(X_1), \\ U_2 &\coloneqq F_{2|1}(X_2 \mid X_1), \\ U_3 &\coloneqq F_{3|1,2}(X_3 \mid X_1, X_2), \\ &\vdots \\ U_n &\coloneqq F_{n|1,\ldots,n-1}(X_n \mid X_1, \ldots, X_{n-1}). \end{aligned}$$

If every marginal and conditional distribution function appearing in the construction is continuous for almost every value of its conditioning variables, then $U_1, \ldots, U_n$ are independent standard uniform random variables.

Conversely, let $\boldsymbol{U} = (U_1, \ldots, U_n)$ have independent standard uniform coordinates and define the **inverse Rosenblatt transform** recursively by

$$\begin{aligned} Y_1 &\coloneqq F_1^{-1}(U_1), \\ Y_2 &\coloneqq F_{2|1}^{-1}(U_2 \mid Y_1), \\ Y_3 &\coloneqq F_{3|1,2}^{-1}(U_3 \mid Y_1, Y_2), \\ &\vdots \\ Y_n &\coloneqq F_{n|1,\ldots,n-1}^{-1}(U_n \mid Y_1, \ldots, Y_{n-1}). \end{aligned}$$

Then $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ has the same distribution as $\boldsymbol{X}$. If $\boldsymbol{U}$ is the Rosenblatt transform of $\boldsymbol{X}$ under the preceding continuity assumptions, then $\boldsymbol{Y} = \boldsymbol{X}$ almost surely.

Remark 1.36 The inverse construction has the form $\boldsymbol{Y} = (\psi^{(1)}(\boldsymbol{U}), \ldots, \psi^{(n)}(\boldsymbol{U}))$ almost surely on the auxiliary space, and $\boldsymbol{Y} \overset{d}{=} \boldsymbol{X}$, as asserted in Theorem 1.35.

Notation 1.37 The Rosenblatt transform is denoted by $\boldsymbol{X} \rightsquigarrow \boldsymbol{U}$, where $\boldsymbol{U}$ has independent standard uniform coordinates. The inverse Rosenblatt transform is denoted by $\boldsymbol{U} \rightsquigarrow \boldsymbol{X}$.

Remark 1.38 The Rosenblatt transform constructs independent uniforms from a random vector. Its inverse writes

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{\psi}(\boldsymbol{U}),$$

where $\boldsymbol{\psi} = (\psi^{(1)}, \ldots, \psi^{(n)})$ need not be unique. The probabilistic dependence among the

coordinates of $\boldsymbol{X}$ is thereby encoded as functional dependence on $\boldsymbol{U}$.

Sampling from a Copula

If the inverse Rosenblatt transform of a random vector $\boldsymbol{X}$ is known, samples from the distribution of $\boldsymbol{X}$ can be generated from independent standard uniform samples.

Example 1.39 Let $\boldsymbol{X} = (X_1, \ldots, X_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\mathbf{A}$ is a Cholesky factor of $\boldsymbol{\Sigma}$, so that

$$\mathbf{A}\mathbf{A}^\mathsf{T} = \boldsymbol{\Sigma},$$

then

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{\mu} + \mathbf{A}\boldsymbol{Z},$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ has independent standard normal coordinates. Define $U_i = \Phi(Z_i)$ for $i \in \{1, \ldots, n\}$, where Φ is the standard normal distribution function. The inverse Rosenblatt representation becomes

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{\mu} + \mathbf{A}\big(\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_n)\big)^\mathsf{T}.$$

Example 1.40 Suppose that (X_1, X_2) is bivariate normal with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Then $\mathrm{Corr}(X_1, X_2) = \rho$, and a Cholesky factor is

$$\mathbf{A} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}.$$

Indeed,

$$\begin{aligned} \mathbf{A}\mathbf{A}^\mathsf{T} &= \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} \sigma_1 & \rho\sigma_2 \\ 0 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \rho^2\sigma_2^2 + \sigma_2^2(1-\rho^2) \end{pmatrix} \\ &= \boldsymbol{\Sigma}. \end{aligned}$$

Consequently, with independent $Z_1, Z_2 \sim \mathcal{N}(0,1)$,

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 Z_1 = \mu_1 + \sigma_1 \Phi^{-1}(U_1), \\ X_2 &= \mu_2 + \rho\sigma_2 Z_1 + \sigma_2\sqrt{1-\rho^2} Z_2 \\ &= \mu_2 + \sigma_2 \left(\rho\Phi^{-1}(U_1) + \sqrt{1-\rho^2}\Phi^{-1}(U_2)\right), \end{aligned}$$

where $U_i = \Phi(Z_i)$ for $i \in \{1,2\}$. These expressions have the required marginal moments, since

$$\begin{aligned} \mathbb{E}[X_2] &= \mu_2, \\ \mathrm{Var}(X_2) &= \rho^2\sigma_2^2 + \sigma_2^2(1-\rho^2) = \sigma_2^2, \end{aligned}$$

and their correlation is

$$\begin{aligned} \mathrm{Corr}(X_1, X_2) &= \frac{\mathbb{E}[(X_1-\mu_1)(X_2-\mu_2)]}{\sigma_1\sigma_2} \\ &= \mathbb{E}\left[Z_1\left(\rho Z_1 + \sqrt{1-\rho^2} Z_2\right)\right] \\ &= \rho \underbrace{\mathbb{E}[Z_1^2]}_{=1} + \sqrt{1-\rho^2}\underbrace{\mathbb{E}[Z_1 Z_2]}_{=0} = \rho. \end{aligned}$$

Lecture 2 — Tuesday, March 16, 2021

2. Dependence Structures and Copula Families

Comonotonicity and Countermonotonicity

Definition 2.1 Random variables $X_1, \ldots, X_n$ with marginal distribution functions $F_1, \ldots, F_n$ are **comonotonic** if any, and hence all, of the following equivalent conditions hold:

1. Their joint distribution function satisfies

$$F(\boldsymbol{x}) = \min_{1 \le i \le n} F_i(x_i) \quad \text{for all } \boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

2. There exist a random variable Z and nondecreasing functions $h_1, \dots, h_n \colon \mathbb{R} \to \mathbb{R}$ such that

$$(X_1, \dots, X_n) \overset{d}{=} (h_1(Z), \dots, h_n(Z)).$$

3. For every $U \sim \text{Unif}[0,1]$,

$$(X_1, \dots, X_n) \overset{d}{=} (F_1^{-1}(U), \dots, F_n^{-1}(U)).$$

Remark 2.2 If (X_1, X_2) is comonotonic and F_1, F_2 are continuous, then

$$(X_1, X_2) \overset{d}{=} (X_1, F_2^{-1}(F_1(X_1))) \overset{d}{=} (F_1^{-1}(F_2(X_2)), X_2).$$

Indeed, if $U = F_1(X_1)$, then $U \sim \text{Unif}[0,1]$ and (3) gives

$$(X_1, X_2) \overset{d}{=} (F_1^{-1}(U), F_2^{-1}(U)) \overset{d}{=} (X_1, F_2^{-1}(F_1(X_1))).$$

The other representation follows symmetrically.

Remark 2.3 Every random variable X is comonotonic with every degenerate random variable Y. If $\mathbb{P}(Y = a) = 1$, then $Y = f(X)$ almost surely for the constant, and hence nondecreasing, function $f(x) = a$.

Definition 2.4 A pair (X_1, X_2) with marginal distribution functions F_1, F_2 is **countermonotonic** if any, and hence all, of the following equivalent conditions hold:

1. Its joint distribution function satisfies

$$F(x_1, x_2) = \max\{F_1(x_1) + F_2(x_2) - 1, 0\} \quad \text{for all } (x_1, x_2) \in \mathbb{R}^2.$$

2. There exist a random variable Z, a nondecreasing function h_1, and a nonincreasing function h_2 such that

$$(X_1, X_2) \overset{d}{=} (h_1(Z), h_2(Z)).$$

3. For every $U \sim \text{Unif}[0,1]$,

$$(X_1, X_2) \stackrel{d}{=} (F_1^{-1}(U), F_2^{-1}(1-U)).$$

Remark 2.5 Every random variable X is countermonotonic with every degenerate random variable Y. Indeed, if $\mathbb{P}(Y = a) = 1$, then $Y = f(X)$ almost surely for the constant, and hence nonincreasing, function $f(x) = a$.

Example 2.6 Let $U \sim \text{Unif}[0,1]$. The pair (U, U) is comonotonic, whereas $(U, 1-U)$ is countermonotonic. For $u, v \in [0,1]$, their joint distribution functions are

$$\mathbb{P}(U \leq u, U \leq v) = \mathbb{P}(U \leq \min\{u,v\}) = \min\{u,v\}$$

and

$$\mathbb{P}(U \leq u, 1-U \leq v) = \mathbb{P}(1-v \leq U \leq u) = (u-(1-v))\mathbb{1}_{\{u>1-v\}} = \max\{u+v-1, 0\}.$$

More generally, if X has distribution function F, then $(X, \ldots, X)$ is comonotonic. Under the usual continuity assumption, $U_X = F(X)$ is uniformly distributed, and

$$(X, F^{-1}(1-F(X))) \stackrel{d}{=} (F^{-1}(U_X), F^{-1}(1-U_X))$$

is countermonotonic.

Remark 2.7 If $U_X = F_X(X)$, then (X, U_X) is comonotonic because F_X is nondecreasing.

Comonotonicity means that all components move in the same direction because they can be represented as nondecreasing functions of one underlying random variable. In dimension two, countermonotonicity means that one component moves in the opposite direction from the other. There is no direct analogue of countermonotonicity in dimensions greater than two.

Lemma 2.8 When the marginals are continuous, so that the copula is unique, comonotonicity and countermonotonicity depend only on the copula.

Proof. Suppose first that $\boldsymbol{X} = (X_1, \ldots, X_n)$ is comonotonic with continuous marginals

$F_1, \dots, F_n$. There is a standard uniform random variable U such that

$$\boldsymbol{X} \stackrel{d}{=} (F_1^{-1}(U), \dots, F_n^{-1}(U)).$$

For arbitrary marginal distribution functions $G_1, \dots, G_n$, the vector

$$(G_1^{-1}(F_1(X_1)), \dots, G_n^{-1}(F_n(X_n))) \stackrel{d}{=} (G_1^{-1}(U), \dots, G_n^{-1}(U))$$

is again comonotonic. Thus changing the marginals by coordinatewise increasing transformations does not alter comonotonicity.

For a countermonotonic pair, the same argument begins with

$$(X_1, X_2) \stackrel{d}{=} (F_1^{-1}(U), F_2^{-1}(1-U))$$

and yields $(G_1^{-1}(U), G_2^{-1}(1-U))$. Hence countermonotonicity is likewise determined by the copula. □

Bounds on Copulas

The Fréchet–Hoeffding bounds describe the pointwise extremes of the class of copulas. In dimension two, every copula $C\colon [0,1]^2 \to [0,1]$ satisfies

$$C^{\mathrm{L}}(u_1, u_2) := \max\{u_1 + u_2 - 1, 0\} \le C(u_1, u_2) \le \min\{u_1, u_2\} =: C^{\mathrm{U}}(u_1, u_2)$$

for all $(u_1, u_2) \in [0,1]^2$. The lower bound C^{L} is the countermonotonic copula, and the upper bound C^{U} is the comonotonic copula. They therefore represent the two extreme forms of bivariate dependence introduced in Definitions 2.1 and 2.4.

Theorem 2.9 Let (X_1, X_2) be a bivariate random vector with continuous marginal distribution functions.

1. The pair is countermonotonic if and only if its copula is the Fréchet–Hoeffding lower bound C^{L}.
2. The pair is comonotonic if and only if its copula is the Fréchet–Hoeffding upper bound C^{U}.

In dimension $n > 2$, every copula $C\colon [0,1]^n \to [0,1]$ satisfies

$$\max\left\{\sum_{i=1}^{n} u_i - (n-1), 0\right\} \leq C(\boldsymbol{u}) \leq \min_{1\leq i\leq n} u_i \quad \text{for all } \boldsymbol{u} \in [0,1]^n.$$

The upper bound $C^{\mathrm{U}}(\boldsymbol{u}) = \min_{1\leq i\leq n} u_i$ is a copula in every dimension. By contrast, when $n > 2$, the function

$$\max\left\{\sum_{i=1}^{n} u_i - (n-1), 0\right\}$$

is not a copula. This failure is another manifestation of the fact that countermonotonicity has no direct higher-dimensional generalization.

Examples of Copulas

For the comonotonic copula, or Fréchet–Hoeffding upper bound, a pair $(U_1, U_2) \sim C^{\mathrm{U}}$ has standard uniform marginals and

$$C^{\mathrm{U}}(u_1, u_2) = \min\{u_1, u_2\} = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2).$$

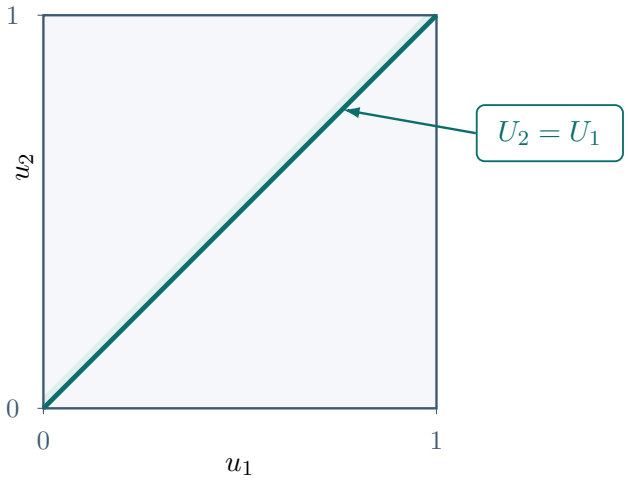

FIGURE 6

As illustrated in Figure 6, the support is concentrated on the diagonal. Thus C^{U} is singular with respect to two-dimensional Lebesgue measure.

In the following derivative calculation, differentiation is understood in the distributional sense.

The Dirac measure in the final line is concentrated at $u = v$; it is not an ordinary Lebesgue density:

$$c^{\mathrm{U}}(u,v) = \frac{\partial^2 C^{\mathrm{U}}(u,v)}{\partial u \partial v} = \frac{\partial}{\partial u}\left[\frac{\partial}{\partial v}\left(u\mathbb{1}_{\{u<v\}} + v\mathbb{1}_{\{u\geq v\}}\right)\right] = \frac{\partial}{\partial u}\mathbb{1}_{\{u\geq v\}} = \delta_v(u).$$

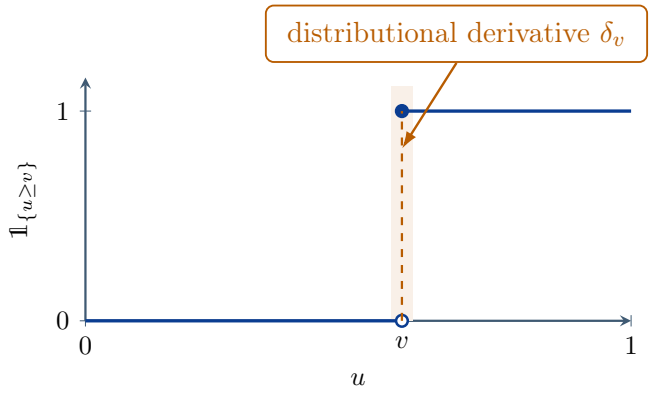

FIGURE 7

The step in Figure 7 locates the singular mass at $u = v$.

For the countermonotonic copula, or Fréchet–Hoeffding lower bound, $(U_1, U_2) \sim C^{\mathrm{L}}$ has standard uniform marginals and

$$C^{\mathrm{L}}(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}.$$

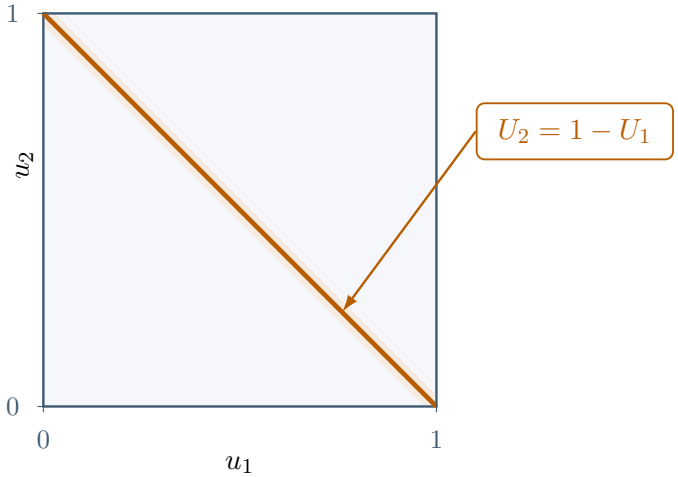

FIGURE 8

$$c^{\mathrm{L}}(u_1, u_2) = \frac{\partial^2}{\partial u_1 \, \partial u_2} \max\{u_1 + u_2 - 1, 0\} = \frac{\partial}{\partial u_1} \mathbb{1}_{\{u_1 > 1-u_2\}} = \delta_{1-u_2}(u_1),$$

where the last expression is again interpreted distributionally, now as singular mass concentrated on the antidiagonal displayed in Figure 8.

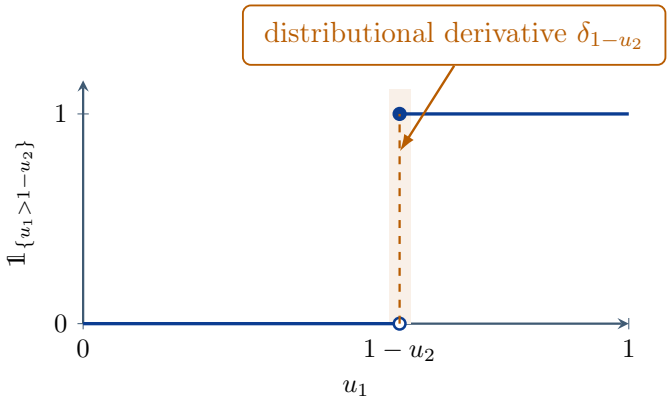

FIGURE 9

The analogous step in Figure 9 locates the singular mass at $u = 1 - u_2$.

The **independence copula** is

$$C^{\perp}(u_1, u_2) = \mathbb{P}(U_1 \le u_1, U_2 \le u_2) = \mathbb{P}(U_1 \le u_1)\mathbb{P}(U_2 \le u_2) = u_1 u_2,$$

for $(u_1, u_2) \in [0, 1]^2$. Its density is $c^{\perp}(u_1, u_2) = 1$, so a scatterplot fills the unit square uniformly, as in Figure 10.

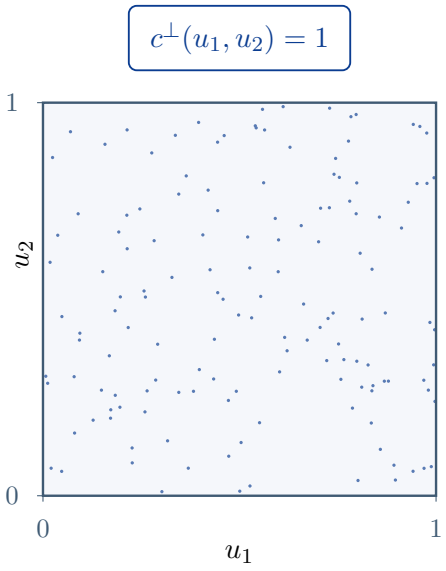

FIGURE 10

The preceding support plots describe the probability measures generated by the three copulas. The graphs of the copula functions themselves are compared in Figure 11. The countermonotonic copula is zero on the lower triangle and rises linearly on the upper triangle, the independence copula forms the bilinear saddle $u_1 u_2$, and the comonotonic copula consists of the two planar faces determined by $\min\{u_1, u_2\}$.

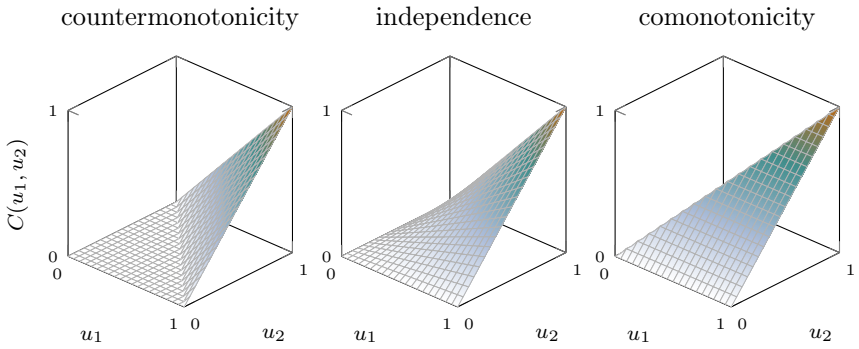

FIGURE 11

Gaussian and t-Copulas

Let $\boldsymbol{X} = (X_1, \ldots, X_n) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a correlation matrix. Then $X_i \sim \mathcal{N}(0,1)$ for every i. The **Gaussian copula** (or **Gauss copula**) with correlation matrix $\boldsymbol{\Sigma}$ is the distribution function of $(\Phi(X_1), \ldots, \Phi(X_n))$ and is given by

$$\begin{aligned} C^{\mathrm{G}}_{\boldsymbol{\Sigma}}(\boldsymbol{u}) &= \mathbb{P}(\Phi(X_1) \le u_1, \ldots, \Phi(X_n) \le u_n) \\ &= \mathbb{P}(X_1 \le \Phi^{-1}(u_1), \ldots, X_n \le \Phi^{-1}(u_n)) \\ &= \Phi_{\boldsymbol{\Sigma}}\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\big), \end{aligned}$$

where $\Phi_{\boldsymbol{\Sigma}}$ denotes the n-dimensional standard normal distribution function with correlation matrix $\boldsymbol{\Sigma}$. Coordinatewise quantile transformations may then replace the normal marginals by arbitrary desired marginals without changing this copula.

For the bivariate Gaussian copula, Figure 12 shows how the sign of the correlation parameter rotates the region in which the density is largest. Both panels use the same colour scale, truncated at density value five; warmer colours represent larger values. The boundary is omitted because the density may be unbounded near two corners of the unit square.

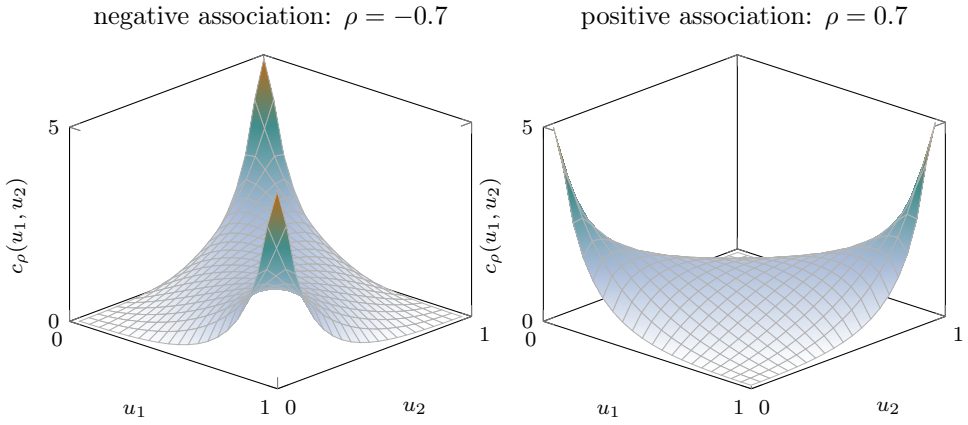

FIGURE 12

Reflecting either coordinate about $1/2$ transforms one panel into the other, in agreement with the change from ρ to $-\rho$.

Similarly, let $\boldsymbol{X} \sim t_{\nu,\boldsymbol{\Sigma}}$, where ν is the number of degrees of freedom and $\boldsymbol{\Sigma}$ has unit diagonal. If T_ν is the univariate Student t distribution function and $T_{\nu,\boldsymbol{\Sigma}}$ is its multivariate counterpart, the

Student t copula is

$$C_{\nu,\mathbf{\Sigma}}(\boldsymbol{u}) = T_{\nu,\mathbf{\Sigma}}\big(T_\nu^{-1}(u_1),\ldots,T_\nu^{-1}(u_n)\big), \quad \boldsymbol{u}\in[0,1]^n.$$

At correlation $\rho = 0.7$, Figure 13 compares the bivariate Student t copula density for two choices of ν. The common colour scale is the same one used in Figure 12 and is truncated at density value five.

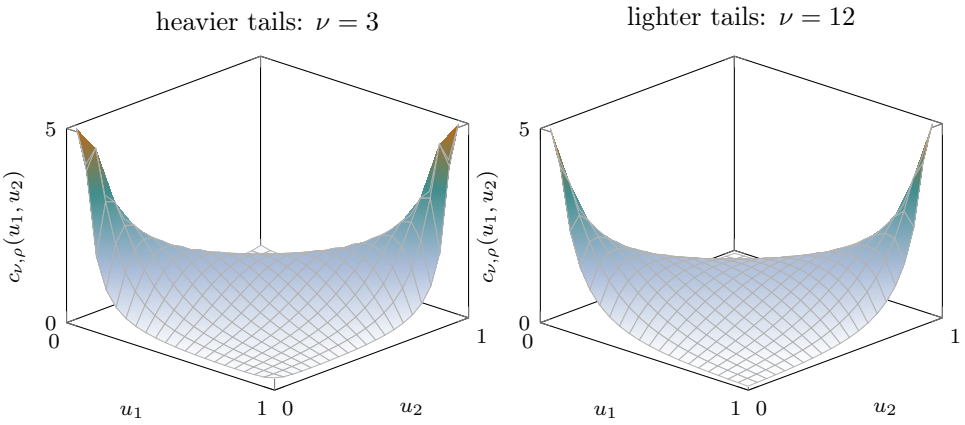

FIGURE 13

As ν increases, the density approaches the positively associated Gaussian copula density shown in Figure 12. For finite ν, the lower-left and upper-right corners retain more pronounced concentration.

Archimedean Copulas

Definition 2.10 A function $\psi\colon [0,\infty) \to [0,1]$ is an **Archimedean generator** if it is continuous, satisfies

$$\psi(0) = 1 \qquad \text{and} \qquad \lim_{t\to\infty}\psi(t) = 0,$$

and is strictly decreasing on $[0, \inf\{t \geq 0 : \psi(t) = 0\}]$.

An n-dimensional copula C is **Archimedean** if there is a generator ψ such that

$$C(\boldsymbol{u}) = \psi\left(\sum_{i=1}^{n}\psi^{-1}(u_i)\right), \quad \boldsymbol{u}\in[0,1]^n,$$

where ψ^{-1} is the generalized inverse of ψ.

Because ψ is decreasing, this generalized inverse is

$$\psi^{-1}(u) = \inf\{t \geq 0 : \psi(t) \leq u\}, \quad u \in (0, 1],$$

with the endpoint value at $u = 0$ defined by the corresponding limit. This is different from the increasing generalized inverse used for distribution functions.

Remark 2.11 Some references define $\varphi \coloneqq \psi^{-1}$ and write

$$C(\boldsymbol{u}) = \varphi^{-1}\left(\sum_{i=1}^{n} \varphi(u_i)\right).$$

The convention in Definition 2.10 agrees with the notation used by the `copula` package in R.

Proposition 2.12 A generator ψ defines a d-dimensional Archimedean copula if and only if it is d-monotone, meaning that

1. ψ is $(d-2)$ times differentiable and
$$(-1)^{\ell}\psi^{(\ell)}(t) \geq 0 \quad \text{for all } \ell \in \{0, \ldots, d-2\} \text{ and } t \in (0, \infty);$$
2. the function $(-1)^{d-2}\psi^{(d-2)}$ is decreasing and convex on $(0, \infty)$.

The **Clayton copula** has generator

$$\psi(t) = (1+t)^{-1/\theta}, \quad t \geq 0 \text{ and } \theta \in (0, \infty).$$

The **Gumbel copula** has generator

$$\psi(t) = \exp\left(-t^{1/\theta}\right), \quad t \geq 0 \text{ and } \theta \in [1, \infty).$$

The density surfaces in Figure 14 compare the two families at $\theta = 2$. The common colour scale is truncated at density value five. Clayton dependence concentrates toward the lower left corner, whereas Gumbel dependence concentrates toward the upper right corner.

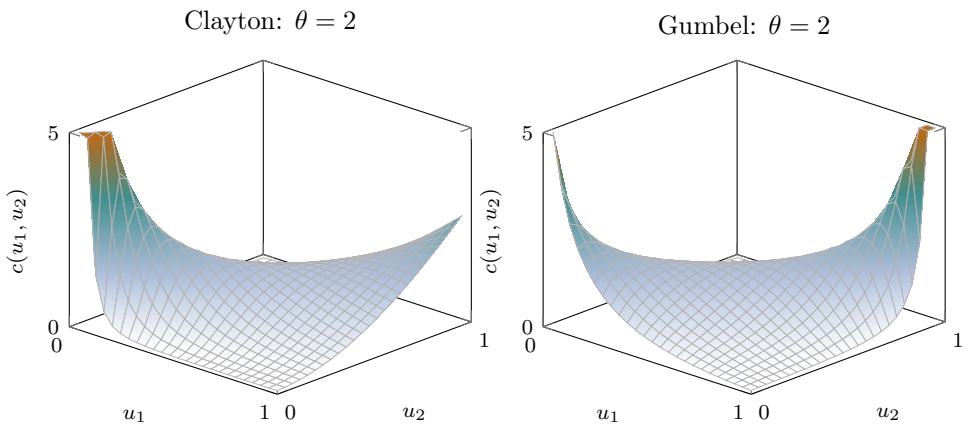

FIGURE 14

The **Marshall–Olkin copula** is

$$\begin{aligned} C_{\alpha,\beta}(u,v) &= \min\{u^{1-\alpha}v, uv^{1-\beta}\} \\ &= \begin{cases} u^{1-\alpha}v, & u^\alpha \geq v^\beta, \\ uv^{1-\beta}, & u^\alpha < v^\beta, \end{cases} \\ &= uv\min\{u^{-\alpha}, v^{-\beta}\}, \end{aligned}$$

where $0 \leq \alpha, \beta \leq 1$. Observe that

$$C_{\alpha,0}(u,v) = C_{0,\beta}(u,v) = uv,$$

so either zero parameter yields the independence copula.

For $0 < \alpha < 1$ and $0 < u < 1$, one has $u^{1-\alpha} > u$. The exponent $u^{1-\alpha}$ is compared with u in Figure 15.

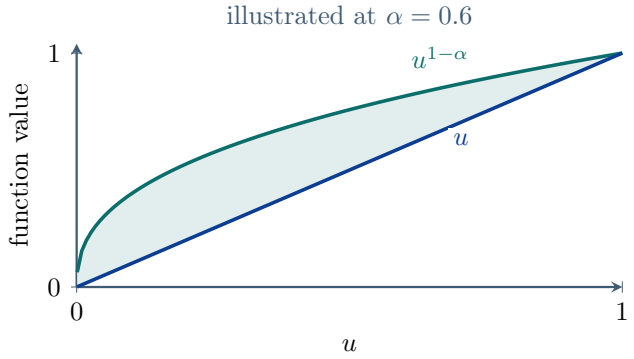

FIGURE 15

At the parameter endpoints, $u^{1-\alpha} = u$ when $\alpha = 0$, whereas $u^{1-\alpha} = 1$ on $(0, 1]$ when $\alpha = 1$. Consequently, $C_{\alpha,0} = C_{0,\beta} = C^{\perp}$, while $C_{1,1}(u, v) = \min\{u, v\}$ is the Fréchet–Hoeffding upper bound. The two endpoint profiles are shown in Figure 16.

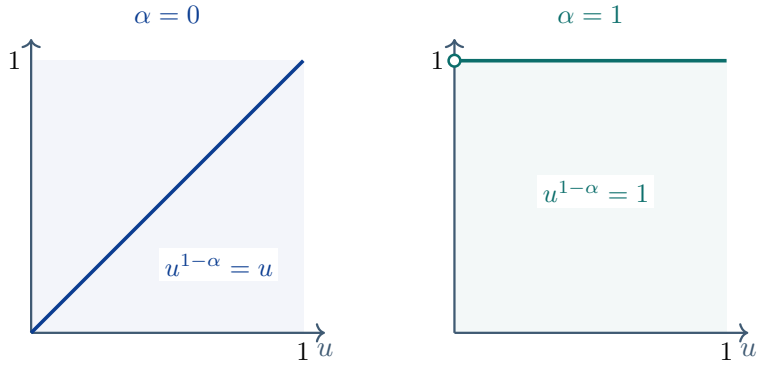

FIGURE 16

For $0 < \alpha, \beta < 1$, the Marshall–Olkin copula has support that is neither absolutely continuous nor singular. It has a part that is absolutely continuous and a part that is singular.

Remark 2.13 Gaussian and Student t copulas with positive-definite parameter matrices are absolutely continuous. At singular parameter matrices they can instead be singular. Comonotonic and countermonotonic copulas are singular.

For the Marshall–Olkin copula, the singular part is concentrated on the curve $u^{\alpha} = v^{\beta}$. Thus, if $(U_1, U_2) \sim C_{\alpha,\beta}$, then

$$\mathbb{P}(U_1^{\alpha} = U_2^{\beta}) = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta}.$$

The mixed geometry is displayed in Figure 17. The absolutely continuous component occupies the unit square away from the curve, while the singular component places positive probability on the curve itself.

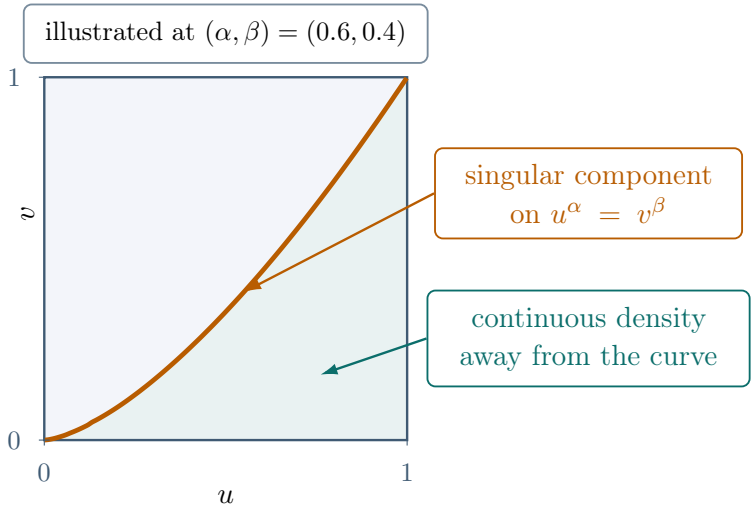

Figure 17

Constructing Multivariate Copulas

Hierarchical or nested constructions combine bivariate copulas to form multivariate models. For example, let X_1, X_2, X_3, X_4 denote automobile insurance losses in Alberta, Ontario, Quebec, and British Columbia, respectively. The tree in Figure 18 groups geographically related risks while allowing an additional positive dependence relation between Ontario and Quebec.

Vine copulas provide another systematic method for building multivariate copulas from bivariate components.

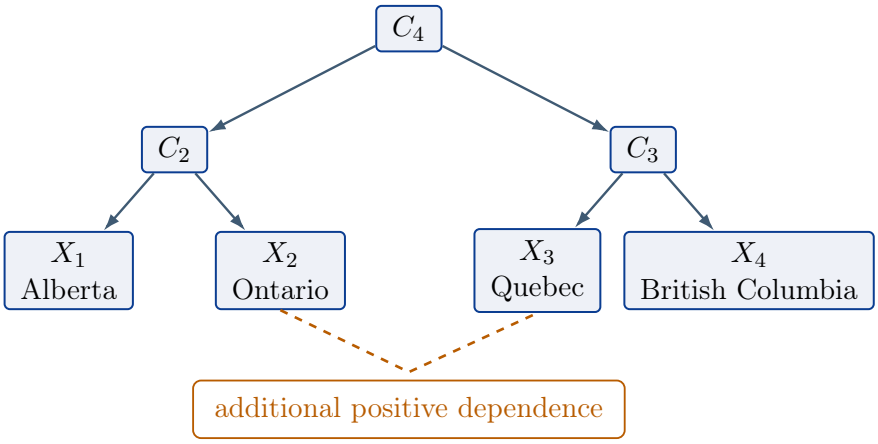

Figure 18

Lecture 3 — Tuesday, March 23, 2021

3. Dependence Measures and Tail Dependence

A **dependence measure** summarizes the dependence between the components of a bivariate random vector (X_1, X_2) by a real number.

Pearson Linear Correlation

Provided both variances are positive, the Pearson linear correlation is

$$\rho(X_1, X_2) = \mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}}.$$

Remark 3.1 The Pearson correlation has the following features:

1. It depends on both the marginals and the copula.
2. It measures linear dependence.

Theorem 3.2 Suppose that X_1 and X_2 have finite, positive variances.

1. If X_1 and X_2 are independent, then $\rho(X_1, X_2) = 0$. The reverse implication fails in general, although it holds for jointly normal random variables.
2. The equality $\rho(X_1, X_2) = 1$ holds if and only if $X_1 = a + bX_2$ almost surely for some $a \in \mathbb{R}$ and $b > 0$.
3. The equality $\rho(X_1, X_2) = -1$ holds if and only if $X_1 = a + bX_2$ almost surely for some $a \in \mathbb{R}$ and $b < 0$.
4. Comonotonicity need not imply correlation 1, and countermonotonicity need not imply correlation -1.

Proof. For (1), independence gives

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2],$$

so $\mathrm{Cov}(X_1, X_2) = 0$ and hence $\rho(X_1, X_2) = 0$.

For (2), the two implications are proved separately.

Forward implication. Suppose that $\rho(X_1, X_2) = 1$. The Cauchy–Schwarz inequality gives

$$\begin{aligned}
\rho(X_1, X_2) &= \frac{\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])]}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}} \\
&\leq \frac{\sqrt{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2]\mathbb{E}[(X_2 - \mathbb{E}[X_2])^2]}}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}} \\
&= 1.
\end{aligned}$$

Equality holds precisely when the centred variables are linearly dependent almost surely. The positive sign of the correlation forces the slope to be positive, so $X_1 = a + bX_2$ almost surely with $b > 0$.

Reverse implication. Suppose that $X_1 = a + bX_2$ almost surely with $b > 0$. Then

$$\rho(X_1, X_2) = \mathrm{Corr}(a + bX_2, X_2) = 1.$$

Statement (3) follows from (2) and the identity $\rho(X_1, X_2) = -\rho(X_1, -X_2)$.

For (4), let $U \sim \text{Unif}[0,1]$. The pair (U, U^2) is comonotonic, but its coordinates are not related by an affine function, so (2) shows that its correlation is not 1. Similarly, $(U, (1-U)^2)$ is countermonotonic but does not have correlation -1 by (3). □

Rank Correlation

Rank correlation depends only on the ranks of observed values. If $(x_1, \ldots, x_n)$ is a sample from X, sorting the observations replaces their magnitudes by relative positions, as illustrated in Figure 19.

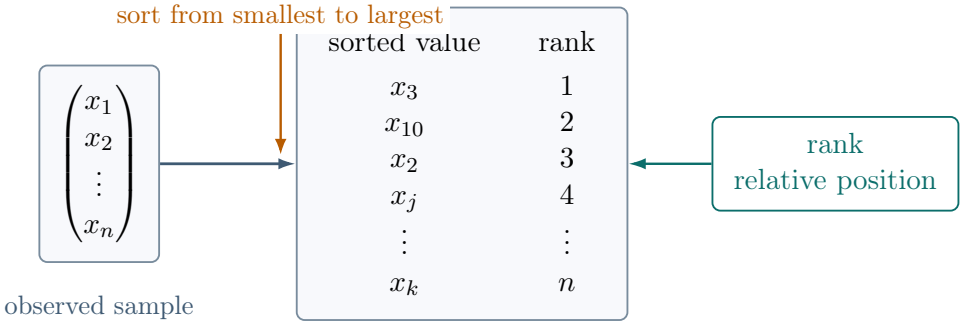

FIGURE 19

Because ranks are invariant under strictly increasing transformations, rank correlations depend only on the copula. The coordinatewise probability integral transform in Figure 20 therefore leaves them unchanged.

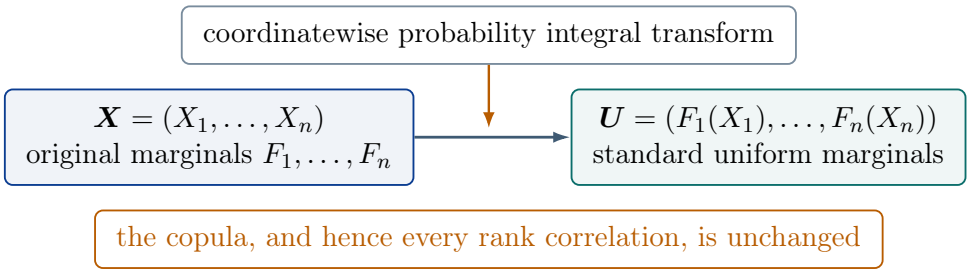

FIGURE 20

Kendall's Tau and Spearman's Rho

Definition 3.3 Let (X_1, X_2) be a bivariate random vector, and let $(\widetilde{X}_1, \widetilde{X}_2)$ be an independent copy. **Kendall's tau** is

$$\begin{aligned}\rho_\tau(X_1, X_2) &= \mathbb{P}((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0) \\ &\quad - \mathbb{P}((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) < 0) \\ &= \mathbb{E}\left[\operatorname{sign}((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2))\right].\end{aligned}$$

If F_1 and F_2 are continuous, **Spearman's rho** is

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)).$$

Proposition 3.4 Suppose that X_1 and X_2 have continuous distribution functions and copula C. Then

$$\rho_\tau(X_1, X_2) = 4 \int_{[0,1]^2} C(u_1, u_2)\, \mathrm{d}C(u_1, u_2) - 1, \tag{3.1}$$

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 \big(C(u_1, u_2) - u_1 u_2\big)\, \mathrm{d}u_1\, \mathrm{d}u_2. \tag{3.2}$$

Proof. Continuity implies that ties occur with probability zero. Concordance and discordance are therefore complementary, and symmetry of the two concordant events gives

$$\begin{aligned}\rho_\tau(X_1, X_2) &= 2\mathbb{P}((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0) - 1 \\ &= 4\mathbb{E}\left[\mathbb{P}(X_1 < \widetilde{X}_1, X_2 < \widetilde{X}_2 \mid \widetilde{X}_1, \widetilde{X}_2)\right] - 1 \\ &= 4 \int_{\mathbb{R}^2} F_{X_1,X_2}(x_1, x_2)\, \mathrm{d}F_{X_1,X_2}(x_1, x_2) - 1 \\ &= 4 \int_{[0,1]^2} C(u_1, u_2)\, \mathrm{d}C(u_1, u_2) - 1,\end{aligned}$$

which proves (3.1).

For Spearman's rho, $U_i = F_i(X_i) \sim \mathrm{Unif}[0, 1]$, so $\mathrm{Var}(U_i) = 1/12$. Hoeffding's covariance identity,

applied on $[0,1]^2$, gives

$$\operatorname{Cov}(U_1, U_2) = \int_0^1 \int_0^1 \big(C(u_1, u_2) - u_1 u_2\big)\, \mathrm{d}u_1\, \mathrm{d}u_2.$$

Dividing by $1/12$ proves (3.2). □

Proposition 3.5 If both marginals are continuous, Kendall's tau and Spearman's rho satisfy the following properties:

1. Both measures take values in $[-1, 1]$.
2. If X_1 and X_2 are independent, then $\rho_\tau(X_1, X_2) = \rho_S(X_1, X_2) = 0$. The reverse implication fails in general.
3. If (X_1, X_2) is comonotonic, then $\rho_\tau(X_1, X_2) = \rho_S(X_1, X_2) = 1$.
4. If (X_1, X_2) is countermonotonic, then $\rho_\tau(X_1, X_2) = \rho_S(X_1, X_2) = -1$.

Proof. The first statement follows immediately from the probabilistic definition of ρ_τ and the fact that ρ_S is a Pearson correlation. Under independence, the copula is $C(u_1, u_2) = u_1 u_2$, so (3.1) gives

$$\rho_\tau = 4 \int_0^1 \int_0^1 u_1 u_2 \, \mathrm{d}u_1\, \mathrm{d}u_2 - 1 = 0,$$

and $F_1(X_1)$ and $F_2(X_2)$ are independent, so $\rho_S = 0$.

For a comonotonic pair, $C(u_1, u_2) = \min\{u_1, u_2\}$. Hence

$$\begin{aligned}
\rho_S &= 12 \int_0^1 \int_0^1 (\min\{u_1, u_2\} - u_1 u_2)\, \mathrm{d}u_1\, \mathrm{d}u_2 \\
&= 12 \left\{ \int_0^1 \int_0^{u_2} u_1(1 - u_2)\, \mathrm{d}u_1\, \mathrm{d}u_2 + \int_0^1 \int_0^{u_1} u_2(1 - u_1)\, \mathrm{d}u_2\, \mathrm{d}u_1 \right\} \\
&= 12 \left(\frac{1}{6} - \frac{1}{8} + \frac{1}{6} - \frac{1}{8} \right) = 1.
\end{aligned}$$

Moreover, two independent copies of a comonotonic pair are almost surely concordant, so $\rho_\tau = 1$. The countermonotonic case follows analogously: the two copies are almost surely discordant, and the copula C^{L} substituted into (3.2) gives $\rho_S = -1$. □

Tail Dependence

Definition 3.6 Let (X_1, X_2) have continuous marginal distribution functions F_1, F_2. Provided the limits exist, the coefficients of upper and lower tail dependence are, respectively,

$$\lambda_{\mathrm{U}} = \lim_{q \uparrow 1} \mathbb{P}(X_2 > F_2^{-1}(q) \mid X_1 > F_1^{-1}(q)), \tag{3.3}$$

$$\lambda_{\mathrm{L}} = \lim_{q \downarrow 0} \mathbb{P}(X_2 \leq F_2^{-1}(q) \mid X_1 \leq F_1^{-1}(q)). \tag{3.4}$$

Both coefficients lie in $[0, 1]$. The pair has upper tail dependence if $\lambda_{\mathrm{U}} > 0$; if $\lambda_{\mathrm{U}} = 0$, it is asymptotically independent in the upper tail. The analogous terminology applies to the lower tail.

The numerator events in the two conditional probabilities are located in Figure 21. As $q \downarrow 0$, the lower tail square contracts toward the origin; as $q \uparrow 1$, the upper tail square contracts toward $(1, 1)$.

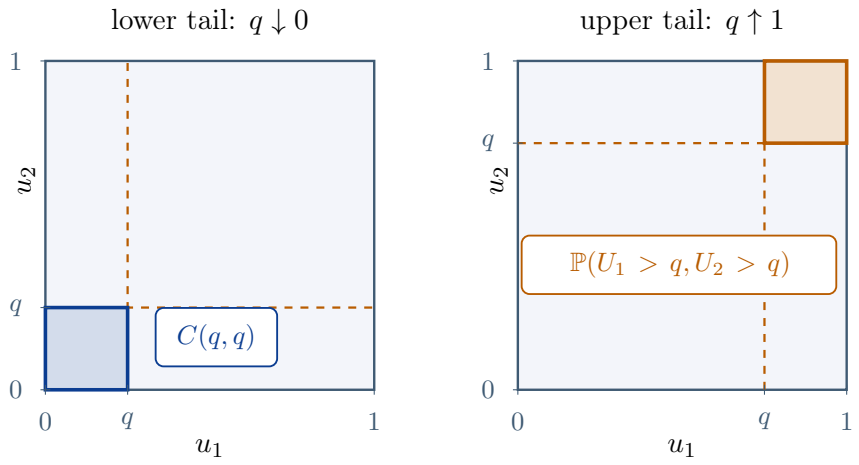

FIGURE 21

Proposition 3.7 If F_1 and F_2 are continuous, then

$$\lambda_{\mathrm{L}} = \lim_{q \downarrow 0} \frac{C(q, q)}{q}.$$

Proof. By (3.4) and Sklar's theorem,

$$\begin{aligned}\lambda_{\mathrm{L}} &= \lim_{q\downarrow 0} \frac{\mathbb{P}(X_1 \le F_1^{-1}(q), X_2 \le F_2^{-1}(q))}{\mathbb{P}(X_1 \le F_1^{-1}(q))} \\ &= \lim_{q\downarrow 0} \frac{C(F_1(F_1^{-1}(q)), F_2(F_2^{-1}(q)))}{F_1(F_1^{-1}(q))} \\ &= \lim_{q\downarrow 0} \frac{C(q,q)}{q}.\end{aligned}$$

□

Survival Copulas

For a random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, its **joint survival function** is

$$\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(X_1 > x_1, \ldots, X_n > x_n),$$

which is not generally equal to $1 - F_{\boldsymbol{X}}(\boldsymbol{x})$. The marginal survival functions are

$$\overline{F}_i(x) = \mathbb{P}(X_i > x) = 1 - F_i(x).$$

A survival version of Sklar's theorem yields a copula $\widehat{C}$, called the **survival copula**, such that

$$\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \widehat{C}(\overline{F}_1(x_1), \ldots, \overline{F}_n(x_n)) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n.$$

Proposition 3.8 For a bivariate copula C and its survival copula $\widehat{C}$,

$$\widehat{C}(1-u_1, 1-u_2) = 1 - u_1 - u_2 + C(u_1, u_2) \quad \text{for all } (u_1, u_2) \in [0,1]^2.$$

Proof. Let (U_1, U_2) have copula C. Since each marginal is standard uniform, $\overline{F}_i(u_i) = 1 - u_i$. Inclusion–exclusion gives

$$\begin{aligned}\widehat{C}(1-u_1, 1-u_2) &= \mathbb{P}(U_1 > u_1, U_2 > u_2) \\ &= 1 - \mathbb{P}(U_1 \le u_1) - \mathbb{P}(U_2 \le u_2) + \mathbb{P}(U_1 \le u_1, U_2 \le u_2) \\ &= 1 - u_1 - u_2 + C(u_1, u_2).\end{aligned}$$

The four regions involved are shown in Figure 22.

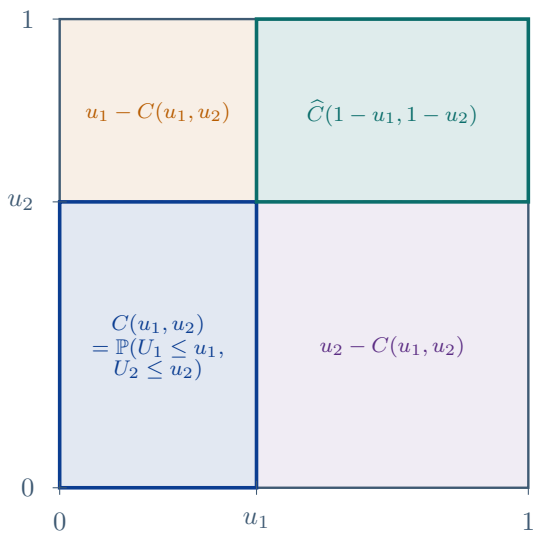

FIGURE 22

□

Proposition 3.9 If F_1 and F_2 are continuous, then

$$\lambda_{\mathrm{U}} = \lim_{q \uparrow 1} \frac{\widehat{C}(1-q, 1-q)}{1-q} = \lim_{q \uparrow 1} \frac{1 - 2q + C(q, q)}{1-q}.$$

Proof. Beginning with (3.3) and applying the survival version of Sklar's theorem gives

$$\begin{aligned}
\lambda_{\mathrm{U}} &= \lim_{q \uparrow 1} \frac{\mathbb{P}(X_1 > F_1^{-1}(q), X_2 > F_2^{-1}(q))}{\mathbb{P}(X_1 > F_1^{-1}(q))} \\
&= \lim_{q \uparrow 1} \frac{\widehat{C}(1 - F_1(F_1^{-1}(q)), 1 - F_2(F_2^{-1}(q)))}{1-q} \\
&= \lim_{q \uparrow 1} \frac{\widehat{C}(1-q, 1-q)}{1-q}.
\end{aligned}$$

The second equality follows from Proposition 3.8. □

Lecture 4 — Tuesday, March 30, 2021

4. Risk Measures and Expected Shortfall

Risk Measures

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and let $\mathcal{X}$ be a space of real-valued random variables on it.

Definition 4.1 A **risk measure** is a function

$$\rho\colon \mathcal{X} \longrightarrow \overline{\mathbb{R}}.$$

Remark 4.2 The loss convention is used throughout: $X > 0$ represents a loss, whereas $X < 0$ represents a gain.

A risk measure orders positions by severity: $\rho(Y) \geq \rho(X)$ indicates that Y is considered at least as risky as X. Applications include insurance premium calculation, regulatory capital requirements, and the modelling of risk averse preferences. The last application is motivated in part by observed probability distortions: individuals may overweight extremely rare favourable events, such as lottery wins, while underweighting extremely rare adverse events, such as natural catastrophes. Distortion risk measures model this behaviour explicitly.

Desirable Properties

The following properties are commonly imposed on a risk measure ρ:

1. It is **law invariant** if $X \overset{d}{=} Y$ implies $\rho(X) = \rho(Y)$.
2. It is **monotone** if $X \leq Y$ almost surely implies $\rho(X) \leq \rho(Y)$.
3. It is **translation invariant** if $\rho(X + c) = \rho(X) + c$ for every $c \in \mathbb{R}$.
4. It is **convex** if
$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$$
for every $\lambda \in [0, 1]$. This property rewards diversification.

5. It is **subadditive** if $\rho(X+Y) \leq \rho(X) + \rho(Y)$.
6. It is **positively homogeneous** if $\rho(\lambda X) = \lambda\rho(X)$ for every $\lambda \geq 0$.
7. It is **comonotonic additive** if $\rho(X+Y) = \rho(X) + \rho(Y)$ whenever (X, Y) is comonotonic.

Remark 4.3 Translation invariance supports the interpretation of a risk measure as a capital requirement. The amount $\rho(X)$ is the capital that must be set aside against the loss X to make the position acceptable to a risk manager or regulator. Positions satisfying $\rho(X) \leq 0$ require no additional capital.

If $\rho(0) = 0$ and ρ is translation invariant, then a sure gain $c < 0$ satisfies

$$\rho(c) = \rho(0 + c) = \rho(0) + c = c < 0.$$

After the capital amount $\rho(X)$ is set aside against a loss X, the adjusted loss is $\widetilde{X} = X - \rho(X)$ and

$$\rho(\widetilde{X}) = \rho(X - \rho(X)) = \rho(X) - \rho(X) = 0,$$

so the adjusted position is acceptable.

Notation 4.4 All risk measures considered in this topic are law invariant, so $\rho(X)$ may also be written as $\rho(F_X)$.

Law invariance makes statistical estimation possible because observations identify the distribution of X, rather than the particular random variable on its underlying probability space. Given observations $x_1, \ldots, x_M$, the empirical distribution function is

$$\widehat{F}_X(t) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{\{x_i \leq t\}}, \quad t \in \mathbb{R}.$$

Without law invariance, the value of $\rho(X)$ need not be recoverable from this distributional information alone.

Proposition 4.5 A subadditive risk measure satisfies

$$\rho(nX) \leq n\rho(X) \quad \text{for every } n \in \mathbb{N}.$$

Proof.

$$\rho(nX) = \rho(X + (n-1)X) \leq \rho(X) + \rho((n-1)X) \leq \cdots \leq n\rho(X).$$

□

Proposition 4.6 A positive homogeneous risk measure is convex if and only if it is subadditive.

Proof. *Forward implication.* Assume that ρ is positively homogeneous and convex. Then

$$\rho(X+Y) = 2\rho\left(\frac{1}{2}X + \frac{1}{2}Y\right) \leq \rho(X) + \rho(Y).$$

Reverse implication. Assume that ρ is positively homogeneous and subadditive. For every $\lambda \in [0,1]$,

$$\rho(\lambda X + (1-\lambda)Y) \leq \rho(\lambda X) + \rho((1-\lambda)Y) = \lambda\rho(X) + (1-\lambda)\rho(Y).$$

□

Definition 4.7 A risk measure is **coherent** if it is monotone, translation invariant, subadditive, and positively homogeneous. It is a **convex risk measure** if it is monotone, translation invariant, and convex.

Remark 4.8 Neither definition includes law invariance as an axiom.

Remark 4.9 The axioms for coherent risk measures were proposed as desirable principles, but they are not universally accepted. Positive homogeneity, for example, may be unrealistic at very large scales. Concentration and liquidity effects can instead lead to

$$\rho(\lambda X) > \lambda\rho(X)$$

for large λ: the position λX concentrates exposure in a single risky asset and may be costly to liquidate.

Convex Risk Measures

Subadditivity implies $\rho(nX) \leq n\rho(X)$ for every positive integer n. To allow for concentration and liquidity effects, convex risk measures replace subadditivity and positive homogeneity by the weaker diversification principle of convexity.

Proposition 4.10 On $\mathcal{L}^1$, the expectation $\rho(X) = \mathbb{E}[X]$ is a law invariant coherent risk measure.

Proof. Law invariance follows because

$$\rho(X) = \int_{\mathbb{R}} x \, \mathrm{d}F_X(x)$$

depends only on F_X. Translation invariance follows from

$$\rho(X + c) = \mathbb{E}[X + c] = \mathbb{E}[X] + c = \rho(X) + c$$

for $c \in \mathbb{R}$. For $\lambda \geq 0$, positive homogeneity follows from

$$\rho(\lambda X) = \mathbb{E}[\lambda X] = \lambda \mathbb{E}[X] = \lambda \rho(X).$$

Additivity, and hence subadditivity, follows from

$$\rho(X + Y) = \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = \rho(X) + \rho(Y).$$

Finally, if $X \leq Y$ almost surely, monotonicity of expectation gives

$$\rho(X) = \mathbb{E}[X] \leq \mathbb{E}[Y] = \rho(Y).$$

Thus expectation is law invariant and coherent. □

Definition 4.11 The **Value at Risk** of X at level $\alpha \in (0, 1)$ is

$$\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha) = \inf\{y \in \mathbb{R} : \mathbb{P}(X \leq y) \geq \alpha\}.$$

Proposition 4.12 The risk measure VaR_α is law invariant, monotone, translation invariant, positively homogeneous, and comonotonic additive. It is generally neither subadditive nor convex.

Proof. Law invariance follows directly from Definition 4.11. If $X \leq Y$ almost surely, then

$$F_X(x) = \mathbb{P}(X \leq x) \geq \mathbb{P}(Y \leq x) = F_Y(x) \quad \text{for all } x \in \mathbb{R}.$$

Consequently,

$$F_X^{-1}(\alpha) = \mathrm{VaR}_\alpha(X) \le F_Y^{-1}(\alpha) = \mathrm{VaR}_\alpha(Y),$$

as illustrated in Figure 23. Thus VaR_α is monotone.

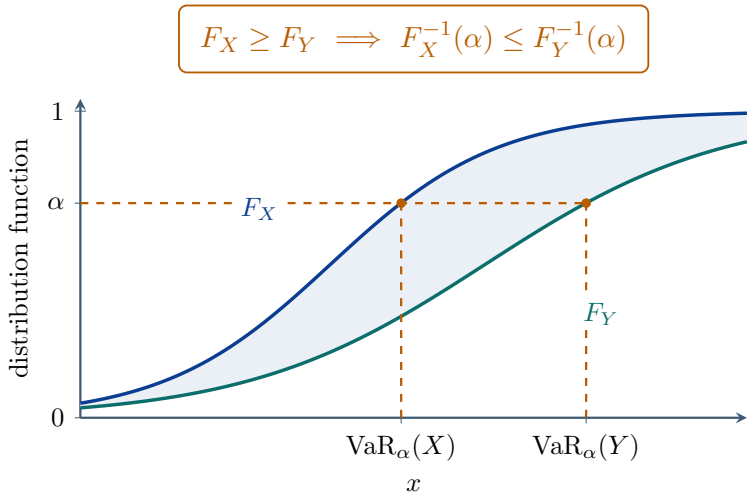

FIGURE 23

For $m \in \mathbb{R}$, translation invariance follows from

$$\begin{aligned}\mathrm{VaR}_\alpha(X+m) &= \inf\{y \in \mathbb{R} : \mathbb{P}(X+m \le y) \ge \alpha\} \\ &= \inf\{z+m : z \in \mathbb{R},\ \mathbb{P}(X \le z) \ge \alpha\} \\ &= \mathrm{VaR}_\alpha(X) + m.\end{aligned}$$

For $\lambda > 0$, the same change of variables gives

$$\begin{aligned}\mathrm{VaR}_\alpha(\lambda X) &= \inf\{y \in \mathbb{R} : \mathbb{P}(\lambda X \le y) \ge \alpha\} \\ &= \lambda \inf\{z \in \mathbb{R} : \mathbb{P}(X \le z) \ge \alpha\} \\ &= \lambda\, \mathrm{VaR}_\alpha(X).\end{aligned}$$

The case $\lambda = 0$ is immediate. □

Proposition 4.13 Value at Risk is not subadditive and is not convex in general.

Example 4.14 Let $X_1, \ldots, X_{100}$ be independent and identically distributed, with

$$X_i = \begin{cases} 100, & \text{with probability } 0.02, \\ -2, & \text{with probability } 0.98. \end{cases}$$

The distribution function is shown in Figure 24.

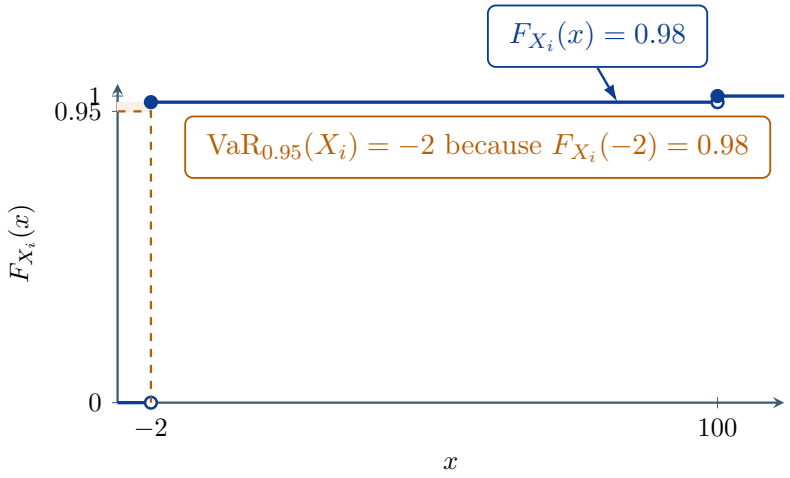

FIGURE 24

Consider the diversified portfolio P_1 and the concentrated portfolio P_2 given by

$$P_1 = \sum_{i=1}^{100} X_i \qquad \text{and} \qquad P_2 = 100X_1.$$

Positive homogeneity gives

$$\text{VaR}_{0.95}(P_2) = 100\,\text{VaR}_{0.95}(X_1) = 100(-2) = -200.$$

To find the risk of P_1, let $Y_1, \ldots, Y_{100}$ be independent Bernoulli random variables with success probability 0.02. Then

$$X_i = 100Y_i - 2(1 - Y_i) = 102Y_i - 2,$$
$$P_1 = 102 \sum_{i=1}^{100} Y_i - 200.$$

Because $\sum_{i=1}^{100} Y_i \sim \text{Bin}(100, 0.02)$,

$$\text{VaR}_{0.95}\left(\sum_{i=1}^{100} Y_i\right) = 5.$$

Therefore,

$$\begin{aligned}\text{VaR}_{0.95}(P_1) &= 102\,\text{VaR}_{0.95}\left(\sum_{i=1}^{100} Y_i\right) - 200 \\ &= 310 > -200 = \sum_{i=1}^{100} \text{VaR}_{0.95}(X_i).\end{aligned}$$

This violates subadditivity. Since $\text{VaR}_{0.95}$ is positively homogeneous, Proposition 4.6 shows that it cannot be convex either.

Proposition 4.15 If $g\colon \mathbb{R} \to \mathbb{R}$ is nondecreasing and continuous, then

$$F^{-1}_{g(X)}(u) = g(F^{-1}_X(u)) \quad \text{for all } u \in (0,1).$$

Proposition 4.16 Value at Risk is comonotonic additive.

Proof. Let X and Y be comonotonic. For a common $U \sim \text{Unif}[0,1]$,

$$(X, Y) \overset{d}{=} (F^{-1}_X(U), F^{-1}_Y(U)).$$

The sum $q(u) = F^{-1}_X(u) + F^{-1}_Y(u)$ is nondecreasing and left-continuous, so it is the quantile function of $q(U)$. Consequently,

$$\begin{aligned}\text{VaR}_\alpha(X+Y) &= F^{-1}_X(\alpha) + F^{-1}_Y(\alpha) \\ &= \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y).\end{aligned}$$

□

Expected Shortfall

Definition 4.17 The **Expected Shortfall** of X at level $\alpha \in [0,1)$ is

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(u)\,\mathrm{d}u = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(X)\,\mathrm{d}u.$$

Unlike Value at Risk, Expected Shortfall averages the entire upper tail beyond the level α, as depicted in Figure 25.

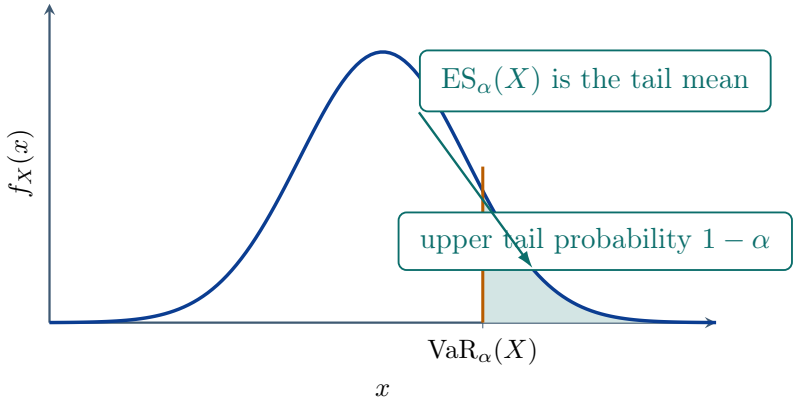

FIGURE 25

Proposition 4.18 On $\mathcal{L}^1$, Expected Shortfall is law invariant, coherent, and comonotonic additive.

The subadditivity of Expected Shortfall follows most transparently from the following optimization representation.

Theorem 4.19 Let $\alpha \in (0,1)$, and write $(x)_+ = \max\{x, 0\}$. Then

$$\mathrm{ES}_\alpha(X) = \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\mathbb{E}[(X - \mathrm{VaR}_\alpha(X))_+], \tag{4.1}$$

$$\mathrm{ES}_\alpha(X) = \inf_{t\in\mathbb{R}}\left\{t + \frac{1}{1-\alpha}\mathbb{E}[(X-t)_+]\right\}. \tag{4.2}$$

If F_X is continuous and strictly increasing, the infimum in (4.2) is attained at $t = \mathrm{VaR}_\alpha(X)$.

Proof. For (4.1), let $U \sim \mathrm{Unif}[0,1]$. Since F_X^{-1} is nondecreasing and $F_X^{-1}(U) \overset{d}{=} X$,

$$\begin{aligned}\mathrm{ES}_\alpha(X) &= F_X^{-1}(\alpha) + \frac{1}{1-\alpha}\int_\alpha^1 (F_X^{-1}(u) - F_X^{-1}(\alpha))\,\mathrm{d}u \\ &= \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\mathbb{E}[(F_X^{-1}(U) - \mathrm{VaR}_\alpha(X))_+] \\ &= \mathrm{VaR}_\alpha(X) + \frac{1}{1-\alpha}\mathbb{E}[(X - \mathrm{VaR}_\alpha(X))_+].\end{aligned}$$

For the optimization formula, assume for simplicity that F_X is continuous and strictly increasing, and define

$$\Gamma(t) = t + \frac{1}{1-\alpha}\mathbb{E}[(X-t)_+].$$

The pointwise form of the positive part function is illustrated in Figure 26.

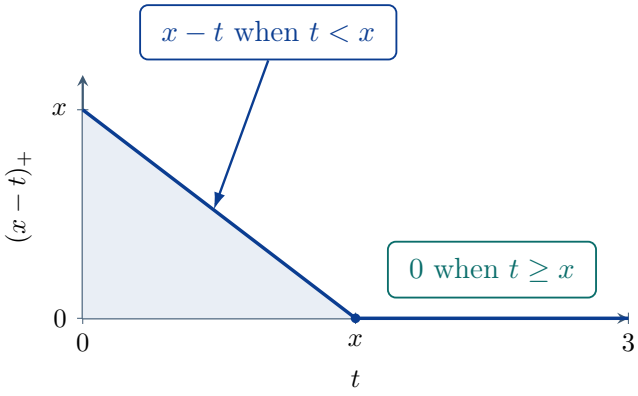

FIGURE 26

Differentiation yields

$$\Gamma'(t) = 1 - \frac{\mathbb{P}(X > t)}{1-\alpha} = 1 - \frac{1 - F_X(t)}{1-\alpha}.$$

Thus $\Gamma'(t) = 0$ precisely when $F_X(t) = \alpha$, or $t = F_X^{-1}(\alpha)$. The derivative is negative to the left of this point and positive to the right, so this point attains the infimum. Substitution into Γ and (4.1) proves (4.2). The general formula follows by the corresponding subgradient argument. □

Proof of Proposition 4.18. Law invariance is immediate from Definition 4.17. Monotonicity, translation invariance, and positive homogeneity follow by integrating the corresponding Value at Risk properties over $u \in [\alpha, 1]$.

If X and Y are comonotonic, Proposition 4.16 gives

$$\begin{aligned}\mathrm{ES}_\alpha(X+Y) &= \frac{1}{1-\alpha}\int_\alpha^1 \big(\mathrm{VaR}_u(X)+\mathrm{VaR}_u(Y)\big)\,\mathrm{d}u \\ &= \mathrm{ES}_\alpha(X)+\mathrm{ES}_\alpha(Y).\end{aligned}$$

Finally, let $t = \mathrm{VaR}_\alpha(X)+\mathrm{VaR}_\alpha(Y)$. Since $(x+y)_+ \le x_+ + y_+$, (4.1) and (4.2) yield

$$\begin{aligned}\mathrm{ES}_\alpha(X)+\mathrm{ES}_\alpha(Y) &\ge t + \frac{1}{1-\alpha}\mathbb{E}[(X+Y-t)_+] \\ &\ge \inf_{s\in\mathbb{R}}\left\{ s + \frac{1}{1-\alpha}\mathbb{E}[(X+Y-s)_+]\right\} \\ &= \mathrm{ES}_\alpha(X+Y).\end{aligned}$$

Therefore Expected Shortfall is subadditive and hence coherent. □

Representation Theorems

Assume throughout this section that $(\Omega, \mathcal{A}, \mathbb{P})$ is atomless. Let $\mathcal{L}^\infty$ denote the space of essentially bounded random variables: $X \in \mathcal{L}^\infty$ if there exists $M < \infty$ such that $|X| \le M$ almost surely.

Definition 4.20 A risk measure ρ satisfies the **Fatou property** if every uniformly bounded sequence $(X_n)_{n\in\mathbb{N}} \subseteq \mathcal{L}^\infty$ that converges in probability to X satisfies

$$\rho(X) \le \liminf_{n\to\infty} \rho(X_n).$$

Example 4.21 Expected Shortfall satisfies the Fatou property.

Proposition 4.22 Law invariant convex risk measures on $\mathcal{L}^\infty$ satisfy the Fatou property.

Corollary 4.23 Law invariant coherent risk measures satisfy the Fatou property.

Proposition 4.24 Suppose that ρ is comonotonic additive and that the map $\lambda \mapsto \rho(\lambda X)$ is continuous on $[0,\infty)$ for every X. Then ρ is positively homogeneous.

Proof. Comonotonic additivity applied to $0 + 0$ gives $\rho(0) = 2\rho(0)$, so $\rho(0) = 0$. Since X is comonotonic with itself, repeated addition gives

$$\rho(nX) = n\rho(X) \quad \text{for every } n \in \mathbb{N}.$$

Consequently, for $m, p, q \in \mathbb{N}$,

$$\rho\left(\frac{1}{m}X\right) = \frac{1}{m}\rho(X),$$
$$\rho\left(\frac{p}{q}X\right) = \frac{p}{q}\rho(X).$$

The assumed continuity extends the identity from nonnegative rational scalars to all nonnegative real scalars. □

Corollary 4.25 Finite-valued convex risk measures that are comonotonic additive are coherent.

Proof. For each X, convexity and finite-valuedness make $\lambda \mapsto \rho(\lambda X)$ a finite convex function on $\mathbb{R}$, and hence a continuous function. By Proposition 4.24, ρ is positively homogeneous, so the defining properties of a convex risk measure imply coherence. □

Theorem 4.26 A risk measure ρ on $\mathcal{L}^\infty$ is coherent and satisfies the Fatou property if and only if there exists a nonempty convex set $\mathcal{Q}$ of probability measures absolutely continuous with respect to $\mathbb{P}$, closed in the weak topology $\sigma(\mathcal{L}^1, \mathcal{L}^\infty)$ through their densities, such that

$$\rho(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X] \quad \text{for every } X \in \mathcal{L}^\infty.$$

Example 4.27 Expected Shortfall has the dual representation

$$\mathrm{ES}_\alpha(X) = \sup_{\mathbb{Q} \in \mathcal{Q}_\alpha} \mathbb{E}^{\mathbb{Q}}[X],$$

where

$$\mathcal{Q}_\alpha = \left\{ \mathbb{Q} \ll \mathbb{P} : \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \leq \frac{1}{1-\alpha} \quad \mathbb{P}\text{-almost surely} \right\}.$$

Theorem 4.28 A convex risk measure ρ on $\mathcal{L}^\infty$ satisfies $\rho(0) = 0$ and has the Fatou property if and only if it admits the representation

$$\rho(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \left\{ \mathbb{E}^{\mathbb{Q}}[X] - \alpha\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right) \right\} \quad \text{for every } X \in \mathcal{L}^\infty,$$

where α is a penalty function and

$$\mathcal{Q} = \left\{ \mathbb{Q} \ll \mathbb{P} : \alpha\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right) < \infty \right\}.$$

Remark 4.29 The penalty term measures how costly it is to depart from the reference probability measure $\mathbb{P}$.

The **Kusuoka representation** states that a risk measure ρ is law invariant and coherent if and only if

$$\rho(X) = \sup_{m \in \mathcal{M}} \int_0^1 \mathrm{ES}_\alpha(X)\, m(\mathrm{d}\alpha) \quad \text{for every } X \in \mathcal{L}^\infty,$$

where $\mathcal{M}$ is a suitable set of probability measures on $[0,1]$. Here one either restricts the representing measures to have no mass at 1, or defines $\mathrm{ES}_1(X) = \operatorname{ess\,sup} X$.

Remark 4.30 There is an analogous Kusuoka representation for law invariant convex risk measures; like Theorem 4.28, it includes a penalty term.

If comonotonic additivity is imposed as well, the supremum reduces to a single mixture: ρ is law invariant, comonotonic additive, and coherent if and only if there exists a probability measure m on $[0,1]$ such that

$$\rho(X) = \int_0^1 \mathrm{ES}_\beta(X)\, m(\mathrm{d}\beta) \quad \text{for every } X \in \mathcal{L}^\infty.$$

Remark 4.31 The Kusuoka representations express law invariant risk measures through mixtures of Expected Shortfall. Expected Shortfall is therefore a fundamental building block for law invariant coherent risk measurement.

Distortion Risk Measures

Definition 4.32 A **distortion function** is a nondecreasing function $g\colon [0,1] \to [0,1]$ satisfying $g(0) = 0$ and $g(1) = 1$. Its associated **distortion risk measure** is the Choquet integral

$$\rho_g(X) = -\int_{-\infty}^{0} \big[1 - g(1 - F_X(x))\big]\,\mathrm{d}x + \int_{0}^{\infty} g(1 - F_X(x))\,\mathrm{d}x.$$

Example 4.33 For the identity distortion $g(u) = u$, one recovers expectation:

$$\rho_g(X) = \mathbb{E}[X].$$

Proof. Substitution of $g(u) = u$ into Definition 4.32 gives the tail integral representation

$$\rho_g(X) = -\int_{-\infty}^{0} F_X(x)\,\mathrm{d}x + \int_{0}^{\infty} (1 - F_X(x))\,\mathrm{d}x.$$

Integration by parts on the two half-lines yields

$$\rho_g(X) = \int_{-\infty}^{0} x\,\mathrm{d}F_X(x) + \int_{0}^{\infty} x\,\mathrm{d}F_X(x) = \int_{\mathbb{R}} x\,\mathrm{d}F_X(x) = \mathbb{E}[X].$$

□

Proposition 4.34 If g is absolutely continuous, define, up to changes on a Lebesgue-null set,

$$\gamma(u) = g'(1-u), \quad u \in [0,1],$$

where the derivative exists almost everywhere. Then

$$\rho_g(X) = \int_0^1 \gamma(u) F_X^{-1}(u)\,\mathrm{d}u = \int_0^1 \gamma(u)\,\mathrm{VaR}_u(X)\,\mathrm{d}u.$$

The function γ is called the **distortion weight function**.

Notation 4.35 The notation ρ_γ denotes the distortion risk measure with weight function

γ:

$$\rho_\gamma(X) = \int_0^1 \gamma(u) F_X^{-1}(u)\,\mathrm{d}u.$$

Proof. Integration by parts in the Choquet integral, followed by the layer-cake representation of the quantile function, gives

$$\rho_g(X) = \int_0^1 F_X^{-1}(u)\gamma(u)\,\mathrm{d}u.$$

The boundary terms vanish whenever the positive and negative parts in Definition 4.32 are finite. This quantile argument remains valid when F_X has atoms; the tempting substitution $u = F_X(x)$ in a Lebesgue–Stieltjes integral would not, because $F_X(X)$ need not then be uniform. □

Proposition 4.36 For an absolutely continuous distortion function g and its weight function γ,

$$\gamma(u) = g'(1-u) \quad \text{for almost every } u, \tag{4.3}$$

$$g(u) = \int_{1-u}^1 \gamma(s)\,\mathrm{d}s. \tag{4.4}$$

Proof. Identity (4.3) is the definition. For (4.4), substitute $t = 1 - s$ to obtain

$$\int_{1-u}^1 \gamma(s)\,\mathrm{d}s = \int_{1-u}^1 g'(1-s)\,\mathrm{d}s = \int_0^u g'(t)\,\mathrm{d}t = g(u) - g(0) = g(u).$$

□

Proposition 4.37 An integrable function $\gamma\colon [0,1] \to \mathbb{R}$, identified up to equality almost everywhere, is a distortion weight function if and only if the following two conditions hold:

1. The function γ is nonnegative on $[0,1]$.
2. The function γ is normalized, meaning that

$$\int_0^1 \gamma(u)\,\mathrm{d}u = 1.$$

Proof. *Forward implication.* If g is a distortion function and $\gamma(u) = g'(1-u)$, then $\gamma \geq 0$ because g is nondecreasing. By (4.4),

$$\int_0^1 \gamma(s)\,\mathrm{d}s = g(1) = 1.$$

Reverse implication. Suppose that γ satisfies (1) and (2), and define

$$g(u) = \int_{1-u}^1 \gamma(s)\,\mathrm{d}s.$$

Then g maps $[0,1]$ into $[0,1]$. If $u_1 \leq u_2$, then

$$g(u_2) - g(u_1) = \int_{1-u_2}^{1-u_1} \gamma(s)\,\mathrm{d}s \geq 0,$$

as illustrated in Figure 27. Hence g is nondecreasing.

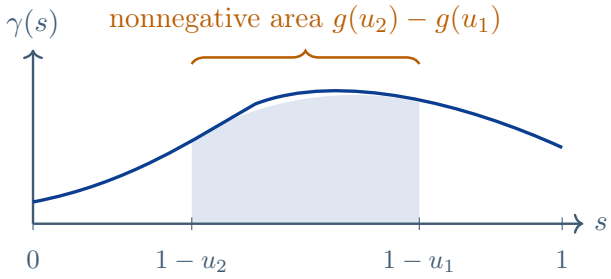

FIGURE 27

Finally,

$$g(0) = \int_1^1 \gamma(s)\,\mathrm{d}s = 0 \qquad \text{and} \qquad g(1) = \int_0^1 \gamma(s)\,\mathrm{d}s = 1.$$

Thus g is a distortion function. □

Thus a distortion weight function is a probability density on $[0,1]$.

Proposition 4.38 Every distortion risk measure ρ_γ has the following properties:

1. It is monotone.
2. It is positively homogeneous.

3. It is translation invariant.
4. It is comonotonic additive.
5. It is law invariant.

Proof. Each property follows by integrating the corresponding property of VaR_u against the nonnegative normalized weight $\gamma(u)\,\mathrm{d}u$. □

Definition 4.39 A distortion risk measure is a **spectral risk measure** if its weight function γ is nondecreasing, or equivalently, if its distortion function g is concave.

Remark 4.40 The equivalence between the two conditions follows directly from (4.3).

Proof under twice differentiability. From (4.4),

$$g'(u) = \gamma(1-u) \qquad \text{and} \qquad g''(u) = -\gamma'(1-u).$$

Forward implication. If g is concave, then $g'' \leq 0$, and hence $\gamma' \geq 0$. Therefore, γ is nondecreasing.

Reverse implication. If γ is nondecreasing, then $\gamma' \geq 0$, and hence $g'' \leq 0$. Therefore, g is concave. □

Theorem 4.41 If ρ_γ is a spectral risk measure with $\gamma \in \mathcal{L}^2[0,1]$, then

$$\rho_\gamma(X) = \int_0^1 \mathrm{ES}_p(X)\,\mu(\mathrm{d}p),$$

where μ is a probability measure on $[0,1)$ satisfying, for almost every s,

$$\gamma(s) = \int_0^s \frac{1}{1-p}\,\mu(\mathrm{d}p).$$

Proof. Fubini's theorem, applied to the region $0 \leq p \leq u \leq 1$, gives

$$\begin{aligned}\rho_\gamma(X) &= \int_0^1 F_X^{-1}(u) \int_0^u \frac{1}{1-p}\,\mu(\mathrm{d}p)\,\mathrm{d}u \\ &= \int_0^1 \frac{1}{1-p} \int_p^1 F_X^{-1}(u)\,\mathrm{d}u\,\mu(\mathrm{d}p) \\ &= \int_0^1 \mathrm{ES}_p(X)\,\mu(\mathrm{d}p).\end{aligned}$$

□

Theorem 4.42 Every spectral risk measure is coherent.

Proof. Only subadditivity remains to be checked. By Theorem 4.41 and the subadditivity of Expected Shortfall,

$$\rho_\gamma(X+Y) = \int_0^1 \mathrm{ES}_p(X+Y)\,\mu(\mathrm{d}p) \leq \int_0^1 \big(\mathrm{ES}_p(X) + \mathrm{ES}_p(Y)\big)\,\mu(\mathrm{d}p) = \rho_\gamma(X) + \rho_\gamma(Y).$$

□

Theorem 4.43 The class of law invariant, comonotonic additive, coherent risk measures coincides with the class of spectral risk measures.

The following are important examples of distortion risk measures. In the first example, the weight is a probability measure rather than an ordinary density; this is the natural Stieltjes-measure extension of the preceding absolutely continuous notation.

1. For Value at Risk at level α, the distortion is $g(x) = \mathbb{1}_{\{x \geq 1-\alpha\}}$, and the weighting measure is the Dirac measure at α.
2. For Range Value at Risk at levels $0 \leq \alpha < \beta \leq 1$,

$$\begin{aligned}g(x) &= \min\left\{\max\left\{\frac{x+\beta-1}{\beta-\alpha}, 0\right\}, 1\right\}, \\ \gamma(u) &= \frac{1}{\beta-\alpha}\mathbb{1}_{\{u \in (\alpha,\beta]\}}, \\ \mathrm{RVaR}_{\alpha,\beta}(X) &= \frac{1}{\beta-\alpha}\int_\alpha^\beta F_X^{-1}(u)\,\mathrm{d}u.\end{aligned}$$

Range Value at Risk is not coherent in general.

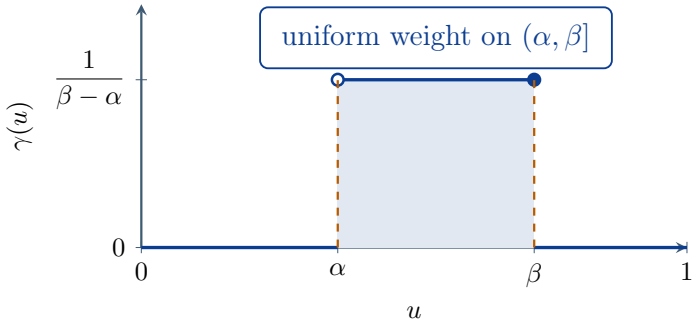

FIGURE 28

As Figure 28 shows, Range Value at Risk assigns uniform weight to the quantile levels in $(\alpha, \beta]$ and zero weight elsewhere.

3. For Expected Shortfall at level α,

$$g(x) = \min\left\{\frac{x}{1-\alpha}, 1\right\} \qquad \text{and} \qquad \gamma(u) = \frac{1}{1-\alpha}\mathbb{1}_{\{u\in(\alpha,1]\}}.$$

4. For the dual power distortion with $\beta > 0$,

$$g(x) = 1 - (1-x)^{\beta} \qquad \text{and} \qquad \gamma(u) = \beta u^{\beta-1}.$$

5. For the Wang transform with $q_0 \in (0,1)$,

$$g(x) = \Phi(\Phi^{-1}(x) + \Phi^{-1}(q_0)),$$
$$\gamma(u) = \frac{\phi(\Phi^{-1}(1-u) + \Phi^{-1}(q_0))}{\phi(\Phi^{-1}(1-u))},$$

where Φ and ϕ are the standard normal distribution and density functions.

Proposition 4.44 For integrable X, Value at Risk, Range Value at Risk, and Expected Shortfall are related by

$$\lim_{\beta\uparrow 1} \mathrm{RVaR}_{\alpha,\beta}(X) = \mathrm{ES}_{\alpha}(X),$$
$$\lim_{\alpha'\uparrow\alpha} \mathrm{RVaR}_{\alpha',\alpha}(X) = \mathrm{VaR}_{\alpha}(X).$$

In general, the second limit is not equal to $\lim_{\alpha'\downarrow\alpha} \mathrm{RVaR}_{\alpha,\alpha'}(X)$.

Lecture 5 — Tuesday, April 06, 2021

5. Dependence Uncertainty and Risk Bounds

Risk Versus Uncertainty

Frank Knight distinguished risk from uncertainty in 1921:

Risk. The outcome is unknown, but its probabilities can be measured accurately. The model is a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with known probability measure $\mathbb{P}$.

Uncertainty. There is insufficient information to assign accurate probabilities. One begins with a measurable space $(\Omega, \mathcal{A})$, while the probability measure, or equivalently the relevant distributions, is not fully known.

Risk Measurement Under Uncertainty

On a measurable space $(\Omega, \mathcal{A})$, a real-valued random variable is a measurable function $X : \Omega \to \mathbb{R}$. A probability measure $\mathbb{P}$ on this space satisfies $\mathbb{P}(A) \in [0,1]$ for every $A \in \mathcal{A}$,

$$\mathbb{P}(\Omega) = 1 \qquad \text{and} \qquad \mathbb{P}(\varnothing) = 0,$$

and is countably additive: if $(A_i)_{i \in \mathcal{I}}$ is a countable family of pairwise disjoint events, then

$$\mathbb{P}\left(\bigcup_{i \in \mathcal{I}} A_i\right) = \sum_{i \in \mathcal{I}} \mathbb{P}(A_i).$$

Let $\mathcal{P}$ be the set of all probability measures on $(\Omega, \mathcal{A})$.

Definition 5.1 A distributional **uncertainty set**, or **ambiguity set**, is a collection $\mathcal{M}$ of distributions compatible with the available information. The best-case and worst-case values of a risk measure are

$$\inf_{F_X \in \mathcal{M}} \rho(X) \qquad \text{and} \qquad \sup_{F_X \in \mathcal{M}} \rho(X).$$

For risk aggregation under uncertainty, the corresponding bounds are

$$\inf_{F_{\boldsymbol{X}} \in \mathcal{M}} \rho\left(\sum_{i=1}^{n} X_i\right) \qquad \text{and} \qquad \sup_{F_{\boldsymbol{X}} \in \mathcal{M}} \rho\left(\sum_{i=1}^{n} X_i\right),$$

where $\mathcal{M}$ is a collection of n-dimensional distributions.

Example 5.2 A basic example fixes the marginal distributions of $\boldsymbol{X} = (X_1, \ldots, X_n)$ but allows its copula to vary.

Stochastic Ordering

Assume throughout this section that $(\Omega, \mathcal{A}, \mathbb{P})$ is atomless and that all expectations under consideration exist.

Definition 5.3 The following stochastic orders compare random variables through test functions:

1. The relation $X \preccurlyeq_{\text{st}} Y$, called **first-order stochastic dominance**, holds if
$$\mathbb{E}[h(X)] \leq \mathbb{E}[h(Y)]$$
for every increasing function h.
2. The relation $X \preccurlyeq_{\text{icx}} Y$, called **increasing convex order**, holds if the same inequality holds for every increasing convex function h.
3. The relation $X \preccurlyeq_{\text{cx}} Y$, called **convex order**, holds if the same inequality holds for every convex function h.

Proposition 5.4 The following conditions are equivalent:

1. The relation $X \preccurlyeq_{\text{st}} Y$ holds.
2. The inequality $F_X(x) \geq F_Y(x)$ holds for all $x \in \mathbb{R}$.
3. The inequality $F_X^{-1}(u) \leq F_Y^{-1}(u)$ holds for all $u \in (0, 1)$.
4. There exist a probability space $(\Omega', \mathcal{A}', \mathbb{P}')$ and random variables X', Y' on it such that $X' \overset{d}{=} X$, $Y' \overset{d}{=} Y$, and $X' \leq Y'$ almost surely.

To see directly that (3) implies (4), take $U \sim \text{Unif}[0,1]$ and define $X' = F_X^{-1}(U)$ and $Y' = F_Y^{-1}(U)$. Then

$$X'(\omega) = F_X^{-1}(U(\omega)) \leq F_Y^{-1}(U(\omega)) = Y'(\omega)$$

for almost every ω.

Remark 5.5 Increasing convex order is also called second-order stochastic dominance or stop-loss order, denoted by $\preccurlyeq_{\text{SL}}$. In particular, $X \preccurlyeq_{\text{icx}} Y$ holds if and only if

$$\mathbb{E}[(X-d)_+] \leq \mathbb{E}[(Y-d)_+] \quad \text{for all } d \in \mathbb{R}.$$

The quantity $\mathbb{E}[(X-d)_+]$ is called the stop-loss premium in actuarial science.

Proposition 5.6 If $X \preccurlyeq_{\text{icx}} Y$, then there exists a random variable Z such that

$$X \preccurlyeq_{\text{st}} Z \preccurlyeq_{\text{cx}} Y.$$

Proposition 5.7 For integrable X and Y, the following conditions are equivalent:

1. The relation $X \preccurlyeq_{\text{cx}} Y$ holds.
2. The relation $X \preccurlyeq_{\text{SL}} Y$, equivalently $X \preccurlyeq_{\text{icx}} Y$, holds, and $\mathbb{E}[X] = \mathbb{E}[Y]$.
3. There exist versions $X' \overset{d}{=} X$ and $Y' \overset{d}{=} Y$ on a common probability space such that
$$X' = \mathbb{E}[Y' \mid X'].$$
4. $$\int_\alpha^1 F_X^{-1}(u)\,\mathrm{d}u \leq \int_\alpha^1 F_Y^{-1}(u)\,\mathrm{d}u$$
for every $\alpha \in (0,1)$, and $\mathbb{E}[X] = \mathbb{E}[Y]$.

Remark 5.8 Applying convex order to the affine functions $h(x) = x$ and $h(x) = -x$ gives both $\mathbb{E}[X] \leq \mathbb{E}[Y]$ and $-\mathbb{E}[X] \leq -\mathbb{E}[Y]$, hence equality of the means. Moreover, conditional expectation reduces variability in convex order:

$$\mathbb{E}[Y' \mid X'] \preccurlyeq_{\text{cx}} Y'.$$

Theorem 5.9 If ρ is monotone and law invariant, then

$$X \preccurlyeq_{\text{st}} Y \Longrightarrow \rho(X) \leq \rho(Y).$$

Proof. By (4), there are versions $X' \overset{d}{=} X$ and $Y' \overset{d}{=} Y$ such that $X' \leq Y'$ almost surely. Law invariance and monotonicity give

$$\rho(X) = \rho(X') \leq \rho(Y') = \rho(Y).$$

□

Theorem 5.10 If ρ is convex, law invariant, and satisfies the Fatou property, then

$$X \preccurlyeq_{\text{cx}} Y \Longrightarrow \rho(X) \leq \rho(Y).$$

Remark 5.11 Only convexity of the functional is needed here; ρ need not satisfy all axioms of a convex risk measure.

Proof. By (3), we may construct versions X', Y' such that $X' = \mathbb{E}[Y' \mid X']$. On an atomless probability space, this conditional expectation can be approximated in probability by random variables Z_n that are finite convex combinations of copies of Y' having the same distribution as Y'. Convexity and law invariance therefore give

$$\rho(Z_n) \leq \rho(Y') = \rho(Y) \quad \text{for every } n.$$

The approximating sequence may be chosen uniformly bounded when $X, Y \in \mathcal{L}^\infty$. The Fatou property then yields

$$\rho(X) = \rho(X') \leq \liminf_{n\to\infty} \rho(Z_n) \leq \rho(Y).$$

□

Theorem 5.12 If ρ is law invariant and coherent, then

$$X \preccurlyeq_{\text{icx}} Y \Longrightarrow \rho(X) \leq \rho(Y).$$

Proof. There exists Z such that $X \preccurlyeq_{\text{st}} Z \preccurlyeq_{\text{cx}} Y$. A coherent risk measure is monotone and convex, and a law invariant coherent risk measure has the Fatou property. Therefore, Theorems 5.9 and 5.10 give

$$\rho(X) \leq \rho(Z) \leq \rho(Y).$$

Equivalently, the result follows from the Kusuoka representation because $\text{ES}_\beta(X) \leq \text{ES}_\beta(Y)$ for every $\beta \in (0,1)$. □

Proposition 5.13 For random variables X, Y, $X \preccurlyeq_{\text{st}} Y$ holds if and only if

$$\text{VaR}_\alpha(X) \leq \text{VaR}_\alpha(Y) \quad \text{for every } \alpha \in (0,1).$$

Theorem 5.14 For integrable random variables X, Y, $X \preccurlyeq_{\text{icx}} Y$ holds if and only if

$$\text{ES}_\alpha(X) \leq \text{ES}_\alpha(Y) \quad \text{for every } \alpha \in (0,1).$$

Proof. *Forward implication.* Expected Shortfall is law invariant and coherent, so Theorem 5.12 gives $\text{ES}_\alpha(X) \leq \text{ES}_\alpha(Y)$ for every α.

Reverse implication. The assumed inequalities are equivalent to

$$\int_\alpha^1 F_X^{-1}(u)\,\mathrm{d}u \leq \int_\alpha^1 F_Y^{-1}(u)\,\mathrm{d}u \quad \text{for every } \alpha \in (0,1).$$

Letting $\alpha \downarrow 0$ also gives $\mathbb{E}[X] \leq \mathbb{E}[Y]$. These integrated quantile inequalities characterize increasing convex order. □

Theorem 5.15 For random variables X, Y, $X \preccurlyeq_{\text{st}} Y$ holds if and only if

$$\rho_\gamma(X) \leq \rho_\gamma(Y) \quad \text{for every distortion weight function } \gamma.$$

Proof. *Forward implication.* Every distortion risk measure is law invariant and monotone, so the result follows from Theorem 5.9.

Reverse implication. The assumed ordering says

$$\int_0^1 F_X^{-1}(u)\gamma(u)\,\mathrm{d}u \leq \int_0^1 F_Y^{-1}(u)\gamma(u)\,\mathrm{d}u$$

for every probability density γ on $[0,1]$. Concentrating such densities on arbitrarily small intervals implies $F_X^{-1}(u) \le F_Y^{-1}(u)$ at every continuity point, and hence everywhere by left continuity. The conclusion follows from Proposition 5.4. □

Theorem 5.16 For random variables X, Y, $X \preccurlyeq_{\text{icx}} Y$ holds if and only if

$$\rho_\gamma(X) \le \rho_\gamma(Y) \quad \text{for every nondecreasing distortion weight } \gamma.$$

Remark 5.17 Thus increasing convex order is equivalent to the ordering of all spectral risk measures, or equivalently all concave distortion risk measures.

Proof. *Forward implication.* Spectral risk measures are law invariant and coherent, so Theorem 5.12 applies.

Reverse implication. For each $\beta \in (0,1)$, the weight

$$\gamma_\beta(u) = \frac{1}{1-\beta}\mathbb{1}_{\{u>\beta\}}$$

is nondecreasing and satisfies $\rho_{\gamma_\beta} = \text{ES}_\beta$. The assumed ordering therefore gives $\text{ES}_\beta(X) \le \text{ES}_\beta(Y)$ for every β, and Theorem 5.14 completes the proof. □

Complete Dependence Uncertainty

Suppose that the marginal distributions of $X_1, \dots, X_n$ are known but their copula is not. The worst-case aggregate risk is

$$\sup_{\text{copulas of } \boldsymbol{X}} \rho\left(\sum_{i=1}^n X_i\right).$$

Theorem 5.18 Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be integrable, let Y be any random variable, and let $U \sim \text{Unif}[0,1]$. Then

$$\sum_{i=1}^n \mathbb{E}[X_i \mid Y] \preccurlyeq_{\text{cx}} \sum_{i=1}^n X_i \preccurlyeq_{\text{cx}} \sum_{i=1}^n F_{X_i}^{-1}(U).$$

Remark 5.19 The bounds in Theorem 5.18 have the following interpretations:

1. The vector $(F_{X_1}^{-1}(U), \ldots, F_{X_n}^{-1}(U))$ is the comonotonic version of $\boldsymbol{X}$. It has the same marginals and is often denoted by $(X_1^{\mathrm{c}}, \ldots, X_n^{\mathrm{c}})$.
2. Among all sums with the prescribed marginals, the comonotonic sum is largest in convex order.
3. The conditional expectations $\mathbb{E}[X_i \mid Y]$ need not have the same marginal distributions as X_i.

Theorem 5.20 Every law invariant, comonotonic additive risk measure that preserves increasing convex order is subadditive.

Proof. Let $(X^{\mathrm{c}}, Y^{\mathrm{c}})$ be a comonotonic version of (X, Y). By Theorem 5.18,

$$X + Y \preccurlyeq_{\mathrm{cx}} X^{\mathrm{c}} + Y^{\mathrm{c}},$$

and hence also $X + Y \preccurlyeq_{\mathrm{icx}} X^{\mathrm{c}} + Y^{\mathrm{c}}$. Therefore,

$$\rho(X + Y) \leq \rho(X^{\mathrm{c}} + Y^{\mathrm{c}}) = \rho(X^{\mathrm{c}}) + \rho(Y^{\mathrm{c}}) = \rho(X) + \rho(Y).$$

□

Theorem 5.21 If ρ is a law invariant convex risk measure, then

$$\rho\left(\sum_{i=1}^{n} X_i\right) \leq \rho\left(\sum_{i=1}^{n} X_i^{\mathrm{c}}\right),$$

where $\boldsymbol{X}^{\mathrm{c}}$ is the comonotonic version of $\boldsymbol{X}$.

Proof. The convex order inequality follows from Theorem 5.18, and ρ preserves convex order by Theorem 5.10. □

Remark 5.22

1. The comonotonic sum has the largest risk under any law invariant convex risk measure.
2. It is straightforward to construct an empirical comonotonic sum. For $n = 2$, reorder the second column according to the ranks of the first:

$$\begin{pmatrix} x_1^{(1)} & x_2^{(1)} \\ x_1^{(2)} & x_2^{(2)} \\ \vdots & \vdots \\ x_1^{(M)} & x_2^{(M)} \end{pmatrix} \xrightarrow{\substack{\text{reorder } x_2^{(i)} \\ \text{according to} \\ \text{the rank} \\ \text{of } x_1^{(i)}}} \begin{pmatrix} x_1^{(1)} & x_2^{s(1)} \\ x_1^{(2)} & x_2^{s(2)} \\ \vdots & \vdots \\ x_1^{(M)} & x_2^{s(M)} \end{pmatrix}$$

Notation 5.23 Permuting a data column changes only the empirical copula; it does not change the empirical marginal distribution.

Alternatively, both columns may be sorted in the same order, although sorting only one column is numerically more efficient.

Example 5.24 For example, reordering the second column according to the ranks in the first gives

$$\begin{pmatrix} 1 & 5 \\ 3 & 1 \\ 2 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 \\ 3 & 5 \\ 2 & 3 \end{pmatrix}.$$

Proposition 5.25 For a law invariant coherent risk measure,

$$\rho\left(\sum_{i=1}^{n} X_i\right) \leq \sum_{i=1}^{n} \rho(X_i) = \sum_{i=1}^{n} \rho(X_i^{\mathrm{c}}),$$

and the final expression depends only on the marginals.

Proposition 5.26 If ρ_γ is a spectral risk measure, then

$$\rho_\gamma\left(\sum_{i=1}^{n} X_i\right) \leq \sum_{i=1}^{n} \rho_\gamma(X_i) = \sum_{i=1}^{n} \rho_\gamma(X_i^{\mathrm{c}}).$$

Proof. By Theorem 5.18 and Theorem 5.16,

$$\rho_\gamma\left(\sum_{i=1}^n X_i\right) \le \rho_\gamma\left(\sum_{i=1}^n X_i^{\mathrm{c}}\right) = \sum_{i=1}^n \rho_\gamma(X_i^{\mathrm{c}}) = \sum_{i=1}^n \rho_\gamma(X_i).$$

□

The same conclusion also follows immediately from the coherence of spectral risk measures.

Remark 5.27 For law invariant convex risk measures, the comonotonic copula attains the worst-case aggregate risk. For coherent and spectral risk measures, this maximizing dependence structure need not be unique; comonotonicity is one maximizer.

Bounds on VaR

Definition 5.28 For distribution functions $F_1, \dots, F_n$, the **Fréchet class** is

$$\mathcal{F}(F_1, \dots, F_n) = \{F_{\boldsymbol{X}} : X_i \sim F_i \text{ for every } i = 1, \dots, n\}.$$

When the marginals are clear, this class is denoted simply by $\mathcal{F}$.

For prescribed marginals, the Value at Risk bounds are

$$\inf_{F_{\boldsymbol{X}} \in \mathcal{F}} \mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i\right) \qquad \text{and} \qquad \sup_{F_{\boldsymbol{X}} \in \mathcal{F}} \mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i\right).$$

The optimization is therefore over all possible copulas. Although comonotonic additivity gives

$$\mathrm{VaR}_\alpha\left(\sum_{i=1}^n X_i^{\mathrm{c}}\right) = \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i),$$

the comonotonic copula need not maximize Value at Risk. In some examples, even independence produces a larger value.

We write an overline for a supremum over the Fréchet class and an underline for an infimum; thus $\overline{\mathrm{VaR}^+}_\alpha$ and $\underline{\mathrm{VaR}}_\alpha$ denote the worst-case right Value at Risk and best-case Value at Risk, respectively, for the fixed marginals under discussion.

Definition 5.29 For $\alpha \in [0,1)$, the **right quantile**, or **right Value at Risk**, is

$$\mathrm{VaR}_\alpha^+(X) = F_X^{+,-1}(\alpha) = \inf\{y \in \mathbb{R} : F_X(y) > \alpha\}.$$

Remark 5.30 If F_X is strictly increasing, then $\mathrm{VaR}_\alpha^+(X) = \mathrm{VaR}_\alpha(X)$. In general, the two quantiles may differ across a flat part of F_X, as shown in Figure 29.

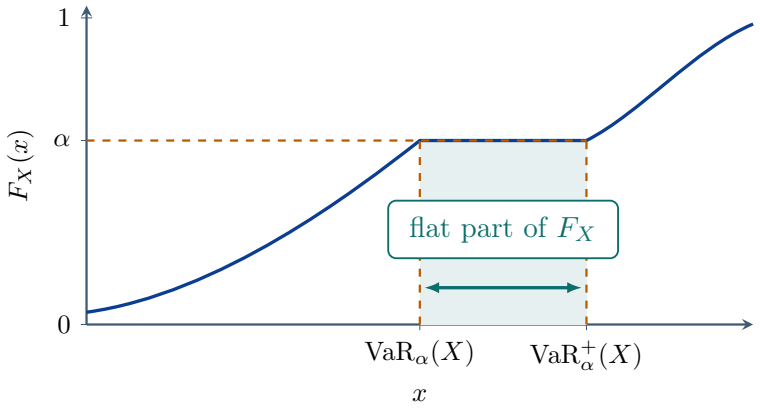

FIGURE 29

By definition,

$$\mathrm{VaR}_\alpha(X) \le \mathrm{VaR}_\alpha^+(X) \quad \text{for every } X \text{ and } \alpha \in (0,1).$$

Definition 5.31 The **left Expected Shortfall** of X at level $\alpha \in (0,1]$ is

$$\mathrm{LES}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha F_X^{-1}(u)\,\mathrm{d}u.$$

Theorem 5.32 Let $X_1, \dots, X_n$ be integrable, and set

$$S = \sum_{i=1}^n X_i \qquad \text{and} \qquad S^{\mathrm{c}} = \sum_{i=1}^n X_i^{\mathrm{c}}.$$

Then

$$A := \sum_{i=1}^{n} \mathrm{LES}_\alpha(X_i) = \mathrm{LES}_\alpha(S^{\mathrm{c}}) \leq \mathrm{VaR}_\alpha(S) \leq \mathrm{VaR}_\alpha^+(S)$$
$$\leq \mathrm{ES}_\alpha(S) \leq \mathrm{ES}_\alpha(S^{\mathrm{c}}) = \sum_{i=1}^{n} \mathrm{ES}_\alpha(X_i) =: B. \tag{5.1}$$

Thus A and B bound the possible Value at Risk, right Value at Risk, and Expected Shortfall of the aggregate.

Proof of the upper bounds. For every random variable Z, $\mathrm{VaR}_\alpha(Z) \leq \mathrm{VaR}_\alpha^+(Z) \leq \mathrm{ES}_\alpha(Z)$. Furthermore, Expected Shortfall is increasing with respect to convex order. By Theorem 5.18,

$$S \preccurlyeq_{\mathrm{cx}} S^{\mathrm{c}},$$

so

$$\mathrm{VaR}_\alpha^+(S) \leq \mathrm{ES}_\alpha(S) \leq \mathrm{ES}_\alpha(S^{\mathrm{c}}) = \sum_{i=1}^{n} \mathrm{ES}_\alpha(X_i) = B.$$

The lower bounds follow by the analogous lower tail argument. □

Remark 5.33

1. The upper bound for ES_α is sharp and is attained by the comonotonic sum.
2. The upper bound B for VaR_α^+ is sharp. It is attained by an aggregate S^* if and only if its quantile function is constant and equal to B on $[\alpha, 1]$.

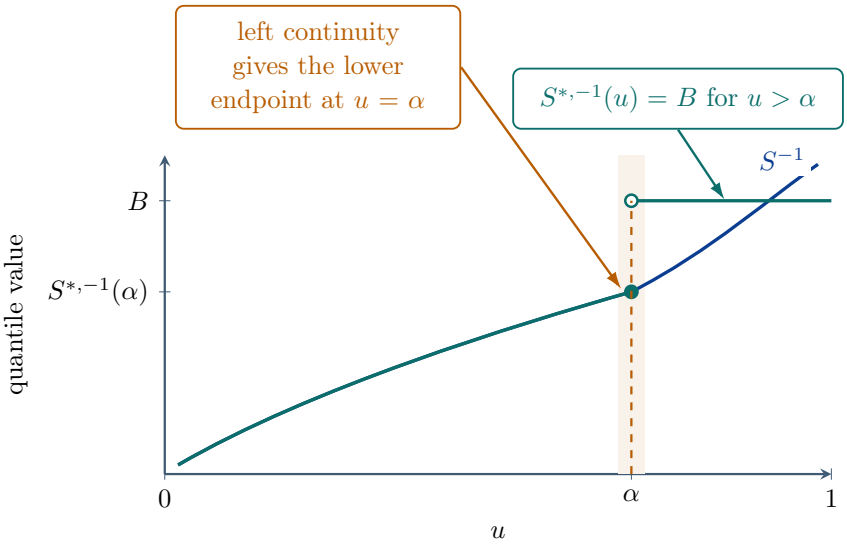

FIGURE 30

As Figure 30 illustrates, left continuity prevents the displayed nondegenerate extremal construction from attaining the same upper bound B for VaR_α. Degenerate cases in which the lower and upper bounds coincide are excluded from this nonattainment statement.

A symmetric argument shows that $\mathrm{VaR}_\alpha(S)$ is bounded below by A. The bound is attained by S^* if and only if $(S^*)^{-1}$ is constant and equal to A on $[0, \alpha)$. Outside degenerate cases, the lower bound for VaR_α^+ is not attained.

1. The copula attaining the upper bound B for VaR_α^+ uses negative dependence in the upper part of the aggregate, corresponding to $u \in [\alpha, 1]$. This contrasts with the comonotonic copula that attains the upper bound for ES_α.
2. The density schematic in Figure 31 represents the corresponding concentration of upper tail mass at the bound B.

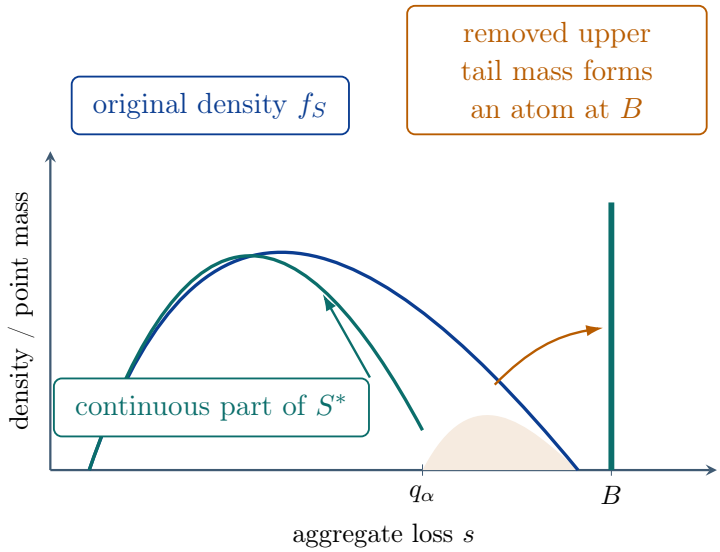

FIGURE 31

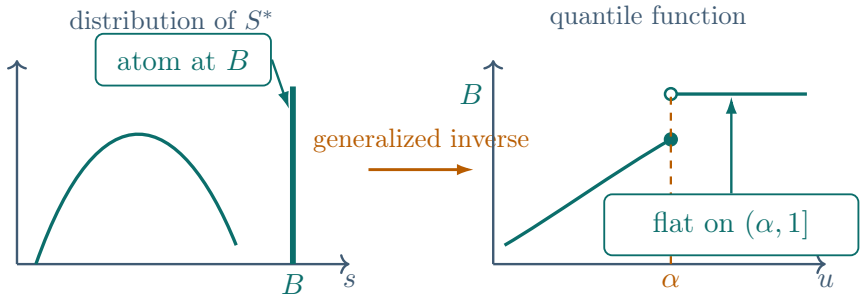

FIGURE 32

As shown in Figure 32, the point mass represented by the spike corresponds to the flat portion of the quantile function.

Definition 5.34 Distribution functions $F_1, \ldots, F_n$ are **jointly mixable** if there exists a random vector $\boldsymbol{X}^* = (X_1^*, \ldots, X_n^*)$ with $X_i^* \sim F_i$ such that

$$\sum_{i=1}^{n} X_i^* = c \quad \mathbb{P}\text{-almost surely}$$

for some constant $c \in \mathbb{R}$.

Example 5.35

1. If $F_1 = F_2$ is the standard uniform distribution, take $(U_1^*, U_2^*) = (U, 1-U)$ for $U \sim \text{Unif}[0,1]$. Then $U_1^* + U_2^* = 1$ almost surely.
2. Suppose $F_1 = F_2 = F$ is symmetric about 0, so
$$F^{-1}(u) = -F^{-1}(1-u) \quad \text{for almost every } u \in (0,1).$$
The choice $X_1^* = F^{-1}(U)$ and $X_2^* = F^{-1}(1-U)$ gives $X_1^* + X_2^* = 0$ almost surely. The standard normal distribution is one example.
3. In dimension two, joint mixability means precisely that we can construct versions satisfying $X_2^* = k - X_1^*$ almost surely for a constant k.

Remark 5.36

1. Joint mixability is a strong dependence condition, and determining which families of marginals are jointly mixable is difficult when $n > 2$.
2. Joint mixability is a form of extremal negative dependence. In dimension two it coincides with a countermonotonic construction whose sum is constant; in higher dimensions it need not reduce to pairwise countermonotonicity.
3. The upper bound for VaR_α^+ is attainable when the upper tail restrictions of the marginals are jointly mixable.

Definition 5.37 Let F be a distribution function.

1. A function $f\colon [0,1] \to \mathbb{R}$ is a **rearrangement of F** if $f(U) \overset{d}{=} F^{-1}(U)$ for $U \sim \text{Unif}[0,1]$.
2. A function $f\colon [a,b] \to \mathbb{R}$ is a **rearrangement of F on $[a,b]$** if $f(V) \overset{d}{=} F^{-1}(V)$ for $V \sim \text{Unif}[a,b]$.

On an atomless auxiliary space, any coupling may be represented as $X_i = f_i(U)$ for rearrangements f_i and a common $U \sim \text{Unif}[0,1]$. By randomizing within ties, the upper $1-\alpha$ probability block

may be parameterized by $U \in [\alpha, 1]$. If the aggregate

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

has no atom at $\mathrm{VaR}_\alpha(S)$, this gives, up to null sets,

$$\{S > \mathrm{VaR}_\alpha(S)\} = \{U \in (\alpha, 1]\}.$$

With an atom at the quantile, only the required fraction of that atom belongs to the randomized tail block; the entire event $\{S \geq \mathrm{VaR}_\alpha(S)\}$ can have probability larger than $1 - \alpha$.

Theorem 5.38 Let $X_i = f_i(U)$ and $S = \sum_{i=1}^{n} X_i$.

1. The worst-case right Value at Risk $\overline{\mathrm{VaR}^+}_\alpha(S)$ is attained if and only if
   (a) Each f_i is a rearrangement of F_{X_i} on $[\alpha, 1]$.
   (b) The components are jointly mixable on $\{U \geq \alpha\}$, meaning that, for some $c \in \mathbb{R}$,
   $$\sum_{i=1}^{n} f_i(u) = c \quad \text{for almost every } u \in [\alpha, 1].$$
2. The best-case Value at Risk $\underline{\mathrm{VaR}}_\alpha(S)$ is attained if and only if
   (a) Each f_i is a rearrangement of F_{X_i} on $[0, \alpha)$.
   (b) The components are jointly mixable on $\{U < \alpha\}$, meaning that $\sum_{i=1}^{n} f_i(u) = c$ for almost every $u \in [0, \alpha)$ and some $c \in \mathbb{R}$.

Remark 5.39 For the upper bound, there is no restriction on dependence over $\{U < \alpha\}$; only the upper tail rearrangements in Theorem 5.38 matter. Although exact joint mixability is strong, weaker asymptotic forms of mixability hold under mild assumptions as $n \to \infty$.

Theorem 5.40 Let the marginals be integrable, and let F_i^α be the upper tail restriction of F_i, that is, the distribution of $F_i^{-1}(V)$ for $V \sim \mathrm{Unif}[\alpha, 1]$. Write $\overline{\mathrm{VaR}^+}_\alpha(F_1, \ldots, F_n)$ for the supremum of the right Value at Risk over all couplings with those marginals.

1. The tail reduction identity is

$$\overline{\mathrm{VaR}^+}_\alpha(F_1,\dots,F_n) = \sup_{Z_i\sim F_i^\alpha} \mathrm{VaR}_0^+\left(\sum_{i=1}^n Z_i\right),$$

where $\mathrm{VaR}_0^+(Z) = \operatorname{ess\,inf} Z$.

2. Suppose that $X_i^\alpha, Y_i^\alpha \sim F_i^\alpha$ and

$$\sum_{i=1}^n Y_i^\alpha \preccurlyeq_{\mathrm{cx}} \sum_{i=1}^n X_i^\alpha.$$

Then

$$\begin{aligned}\mathrm{VaR}_0^+\left(\sum_{i=1}^n X_i^\alpha\right) &\le \mathrm{VaR}_0^+\left(\sum_{i=1}^n Y_i^\alpha\right)\\ &\le \mathbb{E}\left[\sum_{i=1}^n Y_i^\alpha\right] = B,\end{aligned}$$

where B is the upper bound in (5.1). Equality in the final inequality holds if and only if $Y_1^\alpha,\dots,Y_n^\alpha$ form a joint mix.

Remark 5.41 The worst-case right Value at Risk is obtained by making the upper tail aggregate as small as possible in convex order. This is the opposite of the worst-case Expected Shortfall construction, which uses the comonotonic copula and maximizes convex order.

Example 5.42 For two upper tail marginals, the countermonotonic coupling

$$X_1^\alpha = F_1^{-1}(V) \qquad \text{and} \qquad X_2^\alpha = F_2^{-1}(1+\alpha-V), \quad V\sim \mathrm{Unif}[\alpha,1],$$

is the smallest coupling in convex order. The upper bound for VaR_α^+ is attained precisely when $X_1^\alpha + X_2^\alpha$ is constant almost surely.

For example, take $U\sim\mathrm{Unif}[0,1]$ and set

$$\begin{aligned}X_1 &= F_1^{-1}(U),\\ X_2 &= F_2^{-1}(U)\mathbb{1}_{\{U<\alpha\}} + F_2^{-1}(1+\alpha-U)\mathbb{1}_{\{U\ge\alpha\}}.\end{aligned}$$

The pair is comonotonic on $\{U < \alpha\}$ and countermonotonic on $\{U \geq \alpha\}$, as shown in Figure 33.

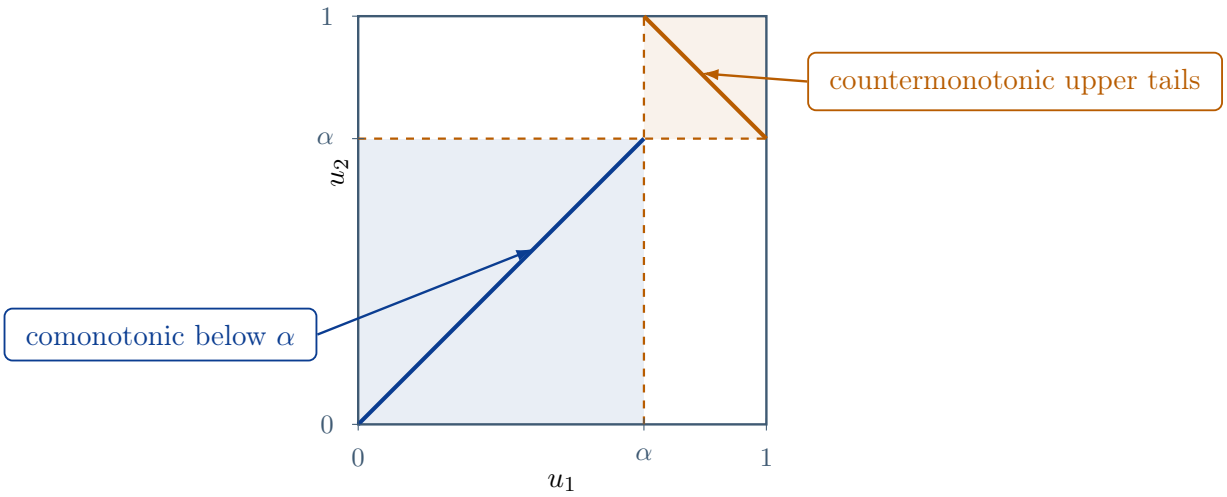

FIGURE 33

Rearrangement Algorithm

The **rearrangement algorithm** heuristically approximates

$$\overline{\mathrm{VaR}^{+}}_{\alpha}\left(\sum_{i=1}^{n} X_i\right)$$

from observations of $\boldsymbol{X} = (X_1, \ldots, X_n)$. By Theorem 5.38, the ideal arrangement makes the sum of variables having the upper-tail laws $F_1^{\alpha}, \ldots, F_n^{\alpha}$ as nearly constant as possible. This is not the same as summing thresholded variables $X_j \mathbb{1}_{\{X_j > F_j^{-1}(\alpha)\}}$, especially when the marginals have atoms. The algorithm reduces the empirical variance of the retained row sums; exact variance zero is generally unattainable.

1. Arrange M observations in the data matrix

$$\mathbf{X} = (x_{ij}) = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{Mn} \end{pmatrix},$$

where column j contains observations of X_j.

2. Assume that $\alpha = k/M$ for some $k \in \{0, \dots, M-1\}$. For example, $\alpha = 0.95$ and $M = 100$ give $k = 95$.

3. Sort each column in ascending order, writing its order statistics as $x^{c}_{1j} \le \cdots \le x^{c}_{Mj}$, and retain the $(M-k) \times n$ upper tail matrix

$$\overline{\mathbf{X}} = \begin{pmatrix} x^{c}_{k+1,1} & \cdots & x^{c}_{k+1,n} \\ \vdots & \ddots & \vdots \\ x^{c}_{M1} & \cdots & x^{c}_{Mn} \end{pmatrix}.$$

These are the largest $M(1-\alpha)$ observations in each column. The extraction is illustrated in Figure 34.

(a) The separately sorted columns have the form

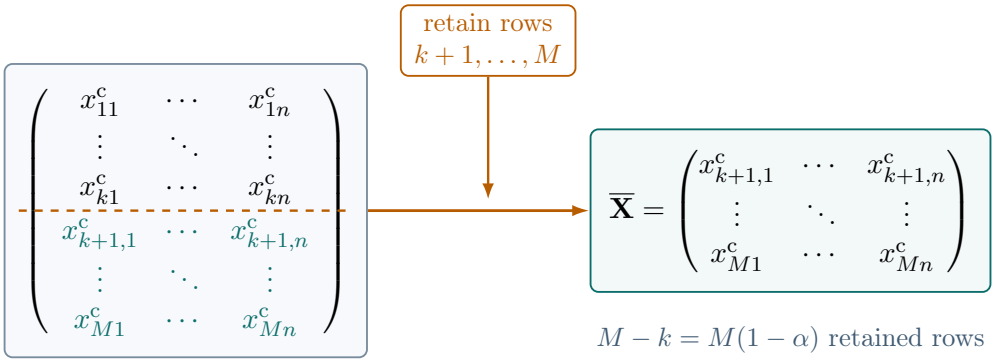

FIGURE 34

(b) The retained rows correspond to the upper $1-\alpha$ fraction of each marginal sample.

4. For each $j \in \{1, \dots, n\}$, rearrange column j in the opposite order from the row sums of all other columns.

5. Cycle through the columns until the row sums meet a chosen convergence tolerance. Denote the resulting matrix by $\mathbf{X}^* = (x^*_{rj})$, for $r \in \{k+1, \dots, M\}$ and $j \in \{1, \dots, n\}$. The procedure can stop at a local solution of the discrete maximin problem, so different initial permutations and finer tail grids should be compared in numerical work.

6. $$\overline{\mathrm{VaR}^{+}}_{\alpha}\left(\sum_{i=1}^{n} X_i\right) \simeq \min_{k+1 \le r \le M} \sum_{j=1}^{n} x^*_{rj}.$$

Thus the approximation is the smallest converged row sum.

The algorithm for the lower bound is analogous: apply the rearrangement procedure to the $M\alpha$ smallest values in every column and use the largest resulting row sum,

$$\underline{\mathrm{VaR}}_\alpha\left(\sum_{i=1}^n X_i\right) \simeq \max_{1\le r\le k}\sum_{j=1}^n x^*_{rj}.$$

Figures 35 and 36 show one complete pass through the columns and the subsequent converged pass, respectively.

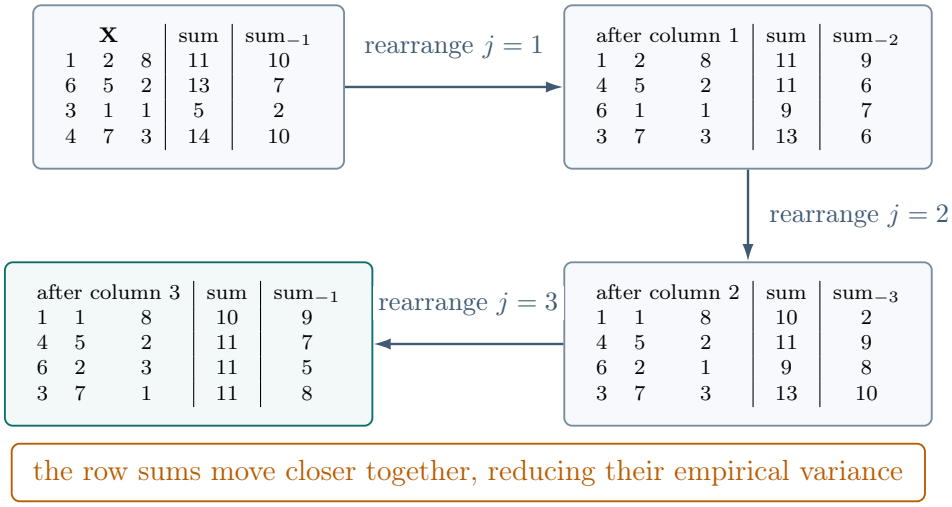

FIGURE 35

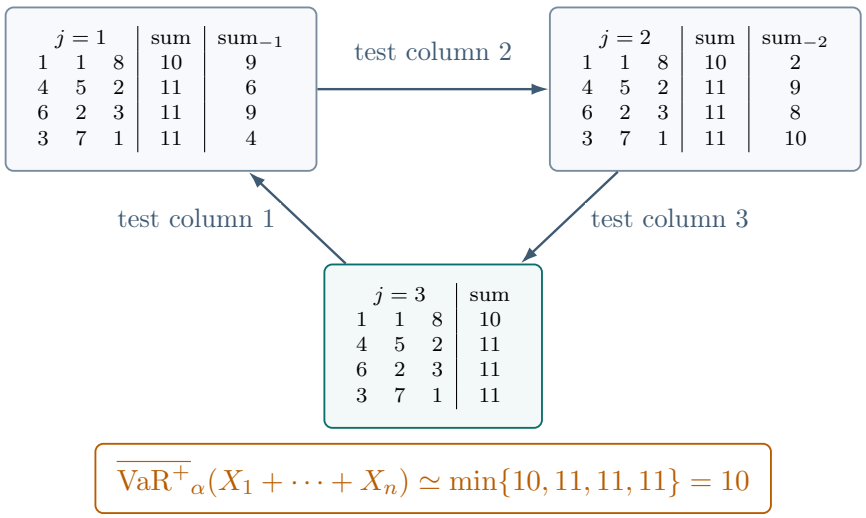

FIGURE 36

Bounds with Moment Information

Risk bounds based only on the marginals can be extremely wide. One way to improve them is to restrict the copula of $\boldsymbol{X} = (X_1, \dots, X_n)$. Another is to study the univariate distribution of the aggregate

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

and impose moment information on S. The same method applies to a nonlinear aggregate $S = g(X_1, \dots, X_n)$.

Definition 5.43 For $\mu \in \mathbb{R}$ and $\sigma > 0$, let

$$\mathcal{M}(\mu, \sigma^2) = \left\{ F : \begin{array}{l} F \text{ is a univariate distribution function,} \\ \displaystyle\int_{\mathbb{R}} x \, \mathrm{d}F(x) = \mu, \\ \displaystyle\int_{\mathbb{R}} (x-\mu)^2 \, \mathrm{d}F(x) = \sigma^2 \end{array} \right\}.$$

Theorem 5.44 For $0 < \alpha < \beta < 1$,

$$\begin{aligned} \sup_{F_S \in \mathcal{M}(\mu,\sigma^2)} \mathrm{ES}_\alpha(S) &= \sup_{F_S \in \mathcal{M}(\mu,\sigma^2)} \mathrm{RVaR}_{\alpha,\beta}(S) \\ &= \sup_{F_S \in \mathcal{M}(\mu,\sigma^2)} \mathrm{VaR}_\alpha(S) \\ &= \sup_{F_S \in \mathcal{M}(\mu,\sigma^2)} \mathrm{VaR}_\alpha^+(S) \\ &= \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}}. \end{aligned}$$

Remark 5.45 The definitions give

$$\begin{aligned} \lim_{\beta \uparrow 1} \mathrm{RVaR}_{\alpha,\beta}(S) &= \mathrm{ES}_\alpha(S), \\ \lim_{\alpha' \uparrow \alpha} \mathrm{RVaR}_{\alpha',\alpha}(S) &= \mathrm{VaR}_\alpha(S), \\ \lim_{\alpha' \downarrow \alpha} \mathrm{RVaR}_{\alpha,\alpha'}(S) &= \mathrm{VaR}_\alpha^+(S). \end{aligned}$$

These limiting relations help explain why all four worst-case values in Theorem 5.44 coincide.

Theorem 5.46 The worst-case values of VaR_α^+, $\mathrm{RVaR}_{\alpha,\beta}$, and ES_α in Theorem 5.44 are attained by the two-point distribution with quantile function

$$G^{-1}(u) = \begin{cases} \mu - \sigma\sqrt{\dfrac{1-\alpha}{\alpha}}, & 0 < u \le \alpha, \\ \mu + \sigma\sqrt{\dfrac{\alpha}{1-\alpha}}, & \alpha < u \le 1. \end{cases}$$

The worst-case value of VaR_α is not attained.

Remark 5.47 The left- and right-continuous quantile values of the extremal distribution are compared in Figure 37.

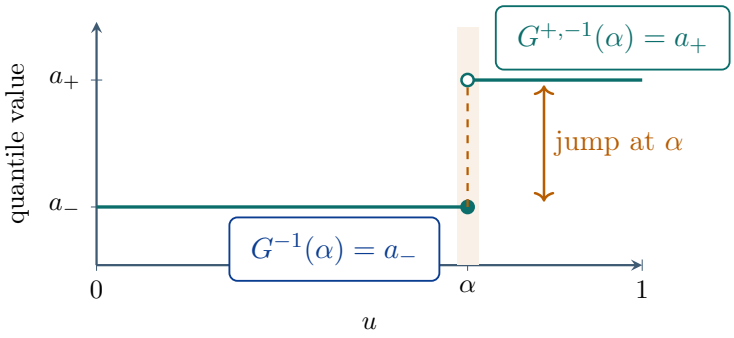

FIGURE 37

$$G^{-1}(\alpha) = \mu - \sigma\sqrt{\frac{1-\alpha}{\alpha}},$$
$$G^{+,-1}(\alpha) = \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}.$$

Indeed,

$$\mathrm{VaR}_\alpha(G) = G^{-1}(\alpha) = \mu - \sigma\sqrt{\frac{1-\alpha}{\alpha}},$$
$$\mathrm{VaR}_\alpha^+(G) = G^{+,-1}(\alpha) = \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}.$$

$$\mathrm{RVaR}_{\alpha,\beta}(G) = \frac{1}{\beta-\alpha}\int_\alpha^\beta G^{-1}(u)\,\mathrm{d}u = \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}},$$
$$\mathrm{ES}_\alpha(G) = \frac{1}{1-\alpha}\int_\alpha^1 G^{-1}(u)\,\mathrm{d}u = \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}.$$

Theorem 5.48

$$\inf_{F_S\in\mathcal{M}(\mu,\sigma^2)} \mathrm{ES}_\alpha(S) = \mu,$$
$$\inf_{F_S\in\mathcal{M}(\mu,\sigma^2)} \mathrm{RVaR}_{\alpha,\beta}(S) = \mu - \sigma\sqrt{\frac{1-\beta}{\beta}},$$
$$\inf_{F_S\in\mathcal{M}(\mu,\sigma^2)} \mathrm{VaR}_\alpha(S) = \mu - \sigma\sqrt{\frac{1-\alpha}{\alpha}}.$$

Because $\sigma > 0$, the lower Expected Shortfall bound μ is an infimum rather than a minimum.

Theorem 5.49 If ρ_γ is a spectral risk measure, then

$$\begin{aligned}\sup_{F_S\in\mathcal{M}(\mu,\sigma^2)} \rho_\gamma(S) &= \mu + \sigma\sqrt{\int_0^1 (\gamma(s)-1)^2\,\mathrm{d}s} \\ &= \mu + \sigma\sqrt{\mathrm{Var}(\gamma(U))}, \quad U\sim\mathrm{Unif}[0,1],\end{aligned}$$

whereas

$$\inf_{F_S\in\mathcal{M}(\mu,\sigma^2)} \rho_\gamma(S) = \mu.$$

If γ is nonconstant, the upper bound is attained by the quantile function

$$G^{-1}(u) = \mu + \sigma\frac{\gamma(u)-1}{\sqrt{\int_0^1(\gamma(s)-1)^2\,\mathrm{d}s}}, \quad u\in[0,1].$$

If γ is nonconstant, the lower bound is not attained. If $\gamma = 1$ almost everywhere, then $\rho_\gamma(S) = \mathbb{E}[S] = \mu$ for every distribution in $\mathcal{M}(\mu,\sigma^2)$, so both bounds equal μ and are attained throughout the ambiguity set. Without the square-integrability assumption, the upper supremum need not be finite.

Theorem 5.50 Let ρ_γ be a distortion risk measure with $\gamma \in \mathcal{L}^2[0,1]$. If $\gamma^\uparrow$ and $\gamma^\downarrow$ are nonconstant, then

$$\sup_{F_S \in \mathcal{M}(\mu,\sigma^2)} \rho_\gamma(S) = \mu + \sigma \sqrt{\int_0^1 (\gamma^\uparrow(u) - 1)^2 \, \mathrm{d}u},$$

$$\inf_{F_S \in \mathcal{M}(\mu,\sigma^2)} \rho_\gamma(S) = \mu - \sigma \sqrt{\int_0^1 (\gamma^\downarrow(u) - 1)^2 \, \mathrm{d}u}.$$

If the relevant projection is constant, the corresponding bound equals μ.

Definition 5.51 For $h \in \mathcal{L}^2[0,1]$, its isotone projections onto the closed convex cones of nondecreasing and nonincreasing functions are the almost-everywhere unique minimizers defined by

$$h^\uparrow \in \underset{\substack{k \in \mathcal{L}^2[0,1] \\ k \text{ nondecreasing and left-continuous}}}{\arg\inf} \int_0^1 (h(s) - k(s))^2 \, \mathrm{d}s,$$

$$h^\downarrow \in \underset{\substack{k \in \mathcal{L}^2[0,1] \\ k \text{ nonincreasing and left-continuous}}}{\arg\inf} \int_0^1 (h(s) - k(s))^2 \, \mathrm{d}s.$$

Theorem 5.52

1. If $\gamma^\uparrow$ is nonconstant, the worst-case quantile function is

$$G^{-1}(u) = \mu + \frac{\sigma(\gamma^\uparrow(u) - 1)}{\sqrt{\int_0^1 (\gamma^\uparrow(s) - 1)^2 \, \mathrm{d}s}}.$$

2. If $\gamma^\downarrow$ is nonconstant, the best-case quantile function is

$$G^{-1}(u) = \mu - \frac{\sigma(\gamma^\downarrow(u) - 1)}{\sqrt{\int_0^1 (\gamma^\downarrow(s) - 1)^2 \, \mathrm{d}s}}.$$

A constant projection gives the corresponding bound μ. It is unattained unless $\gamma = 1$ almost everywhere; in that exceptional case every distribution attains both bounds.

Appendix: Lecture and Tutorial Index

1. Lecture 1 — Tuesday, March 09, 2021
2. Lecture 2 — Tuesday, March 16, 2021
3. Lecture 3 — Tuesday, March 23, 2021
4. Lecture 4 — Tuesday, March 30, 2021
5. Lecture 5 — Tuesday, April 06, 2021