Stochastic Machine Learning
Chapter 03 - Time series and LSTM

Thorsten Schmidt

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

SS 2024

Multivariate dependencies

We mainly follow Schmidt (2007), while there is of course a rich literature on the topic.

An explanatory example

- Consider two rvs X_1 and X_2 which shall give us two numbers out of $\{1, 2, \ldots, 6\}$.
- Assume that X_1 is communicated to us and we may enter a bet on X_2.
- The question is, how much information can be gained from the observation of X_1, or what is the dependence ?
  1. **independence:** the knowledge of X_1 gives us no information about X_2.
  2. **comontonicity:** with X_1 we have full information on X_2.
  3. A quite different answer will be given if X_1 is always the number of the smaller throw and X_2 the larger one. Then we have a strict monotonic relation between these two, namely $X_1 \leq X_2$. In the case where $X_1 = 6$ we also know X_2. If $X_1 = 5$, we would guess that X_2 is either 5 or 6, both with a chance of 50%, and so on.

A deeper look

- The marginals in cases 1. and 2. are the same, $\mathbb{P}(X_i = k) = 1/6$.
- The cdf in case 1 is simply

$$P(X_1 \leq x_1, X_2 \leq x_2) = F(x_1) \cdot F(x_2).$$

- In the third case we obtain (assume $x_1 \leq x_2$)

$$\begin{aligned} P(X_1 \leq x_1, X_2 \leq x_2) &= P(Z_1 \leq x_1, Z_2 \leq x_2) + P(Z_2 \leq x_1, Z_1 \leq x_2) \\ &\quad - P(Z_1 \leq x_1, Z_2 \leq x_1) \\ &= 2F(x_1)F(x_2) - F(x_1)^2, \end{aligned}$$

  and - in general -

$$= 2F(\min\{x_1, x_2\})F(x_2) - F(\min\{x_1, x_2\})^2.$$

Hence, to obtain a full description of X_1 and X_2 together we used *two* ingredients: the marginals and the type of interrelation, for example independence. The question is if this kind of separation between marginals and dependence can also be realized in a more general framework. Luckily the answer is yes, and the right concept for this is copulas.

It was Hoeffding's idea (already in the 1940s) to study multivariate distributions under "arbitrary changes of scale", and although he did not introduce copulas directly, his work contributed many interesting results (eg, see Fisher (1995)).

- Copulas help in the understanding of **dependence** at a deeper level;
- They show us potential pitfalls of approaches to dependence that focus only on correlation;
- They allow us to define useful **alternative dependence measures**;
- They express dependence on a **quantile scale**, which is natural in QRM;
- They facilitate a **bottom-up approach** to multivariate model building;
- They are easily simulated and thus lend themselves to **Monte Carlo risk studies**.

What is a Copula?

Definition

A copula is a multivariate distribution function with standard uniform margins.

Equivalently, a copula if any function $C : [0,1]^d \to [0,1]$ satisfying the following properties:

1. $C(u_1, \dots, u_d)$ is increasing in each component u_i.
2. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ for all $i \in \{1, \dots, d\}$, $u_i \in [0,1]$.
3. For all $(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0,1]^d$ with $a_i \leq b_i$ we have:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \dots, u_{di_d}) \geq 0,$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \dots, d\}$.

Probability and Quantile Transforms

Definition (Quantile function)

The **generalized inverse** of the cdf F is

$$F^{-1}(t) := \inf\{x \in \mathbb{R} : F(x) \geq t\}$$

for any $t \in (0,1)$. $x_\alpha := F^{-1}(\alpha)$ is the α-**Quantile** of F.

Lemma 1: probability transform

Let X be a random variable with **continuous** distribution function F. Then

$$F(X) \sim U(0,1) \qquad \text{(standard uniform).}$$

$$P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}F(u)) = u, \quad \forall u \in (0,1).$$

Lemma 2: quantile transform

Let U be uniform and F the distribution function of **any** rv X. Then

$$F^{-1}(U) \stackrel{\mathscr{L}}{=} X$$

so that $P(F^{-1}(U) \leq x) = F(x)$.

These facts are the key to all statistical simulation and essential in dealing with copulas.

Sklar's Theorem

Theorem (Sklar)

Let F be a joint distribution function with margins $F_1, \dots, F_d$.

- *There exists a copula C such that for all $x_1, \dots, x_d$ in $[-\infty, \infty]$*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

- *If the margins are **continuous** then C is unique; otherwise C is uniquely determined on Ran$F_1 \times$ Ran$F_2 \dots \times$ RanF_d.*
- ***Conversely**, if C is a copula and $F_1, \dots, F_d$ are univariate distribution functions, then F defined above is a multivariate df with margins $F_1, \dots, F_d$.*

Sklar's Theorem: Proof in Continuous Case

Henceforth, **unless explicitly stated**, vectors $\mathbf{X}$ will be assumed to have **continuous** marginal distributions. In this case:

$$\begin{aligned} F(x_1, \ldots, x_d) &= P\left(X_1 \leq x_1, \ldots, X_d \leq x_d\right) \\ &= P\left(F_1(X_1) \leq F_1(x_1), \ldots, F_d(X_d) \leq F_d(x_d)\right) \\ &= C(F_1(x_1), \ldots, F_d(x_d)). \end{aligned}$$

The unique copula C can be calculated from $F, F_1, \ldots, F_d$ using

$$C(u_1, \ldots, u_d) = F\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right).$$

Copulas and Dependence Structures

Sklar's theorem shows how a unique copula C describes the **dependence structure** of the multivariate df of a random vector $\mathbf{X}$. This motivates a further definition.

Definition: Copula of $\mathbf{X}$

The copula of $(X_1, \ldots, X_d)$ is the df C of $(F_1(X_1), \ldots, F_d(X_d))$.

Invariance

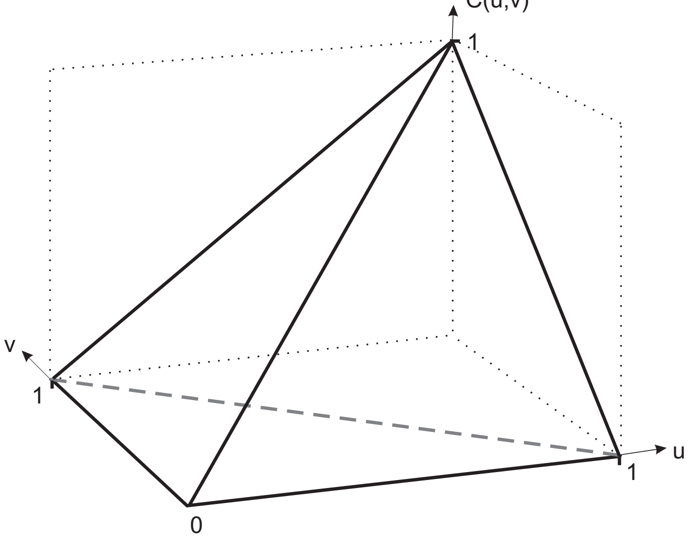
C is invariant under **strictly increasing** transformations of the marginal distributions. If $T_1, \ldots, T_d$ are strictly increasing, then $(T_1(X_1), \ldots, T_d(X_d))$ has the same copula as $(X_1, \ldots, X_d)$.

The Fréchet Bounds

For every copula $C(u_1, \ldots, u_d)$ we have the important bounds

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \leq C(\mathbf{u}) \leq \min\left\{u_1, \ldots, u_d\right\}. \quad (1)$$

- The upper bound is the df of $(U, \ldots, U)$ and the copula of the random vector $\mathbf{X}$ where $X_i \stackrel{\text{a.s.}}{=} T_i(X_1)$ for increasing functions $T_2, \ldots, T_d$. It represents **perfect positive dependence** or **comonotonicity**.
- The lower bound is only a copula when $d = 2$. It is the df of the vector $(U, 1 - U)$ and the copula of (X_1, X_2) where $X_2 \stackrel{\text{a.s.}}{=} T(X_1)$ for T decreasing. It represents **perfect negative dependence** or **countermonotonicity**.
- The copula representing **independence** is $C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$.



According to the Fréchet-Hoeffding bounds every copula has to lie inside of the pyramid shown in the graph. The surface given by the bottom and back side of the pyramid (the lower bound) is the countermonotonicity-copula $C(u, v) = \max\{u + v - 1, 0\}$, while the front side is the upper bound, $C(u, v) = \min(u, v)$.

Examples of Implicit Copulas

Probably the most famous example is the Gaussian copula. It is also at the root of the financial crisis (See "The formula that killed Wall street" at https://www.sps.ed.ac.uk/sites/default/files/assets/pdf/Formula12.pdf by Donald MacKenzie and Taylor Spears).
The principle was borrowed from actuarial sciences - but from the current viewpoint, simple statistical principles were overlooked: do we have confidence bounds of the parameters? Does the model fit the data? How is the time-evoluation of the model?

Gaussian Copula

$$C_P^{\mathsf{Ga}}(\mathbf{u}) = \mathbf{\Phi}_P\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)\right),$$

where Φ denotes the standard univariate normal df, $\mathbf{\Phi}_P$ denotes the joint df of $\mathbf{X} \sim N_d(\mathbf{0}, P)$ and P is a correlation matrix. Write C_ρ^{Ga} when $d = 2$.
$P = I_d$ gives independence and $P = J_d$ gives comonotonicity.

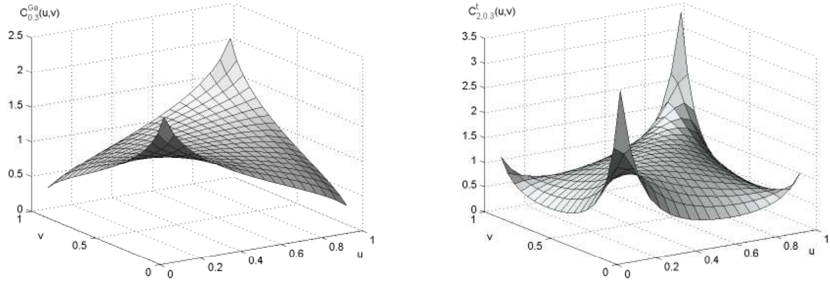
Copula densities

- For illustration of the copula a density is much more useful than the cdf. Lets shortly also look at this concept.
- If the copula is sufficiently differentiable the **copula density** can be computed:

$$c(\boldsymbol{u}) := \frac{\partial^d\, C(u_1, \ldots, u_d)}{\partial u_1 \,\cdots\, \partial u_d}.$$

- If the copula is given by $C(\boldsymbol{u}) = F\big(F_1(^{\leftarrow}u_1), \ldots, F_d(^{\leftarrow}u_d)\big)$, then

$$c(\boldsymbol{u}) = \frac{f\big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\big)}{f_1(F_1^{-1}(u_1)) \,\cdots\, f_d(F_d^{-1}(u_d))}.$$



The copula densities of a Gaussian copula (left) and a Student t-copula (right). Both copulas have correlation coefficient $\rho = 0.3$ and the t-copula has 2 degrees of freedom.

Simulating Copulas

How to simulate the Gaussian copula $C_P^{\mathbf{Ga}}$

- Simulate $\mathbf{X} \sim N_d(\mathbf{0}, P)$
- Set $\mathbf{U} = (\Phi(X_1), \ldots, \Phi(X_d))'$ (probability transformation)

an the t copula $C_{\nu,P}^{\mathbf{t}}$

- Simulate $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$
- Set $\mathbf{U} = (t_\nu(X_1), \ldots, t_\nu(X_d))'$ (probability transformation)
  t_ν is df of univariate t distribution.

Simulation of Archimedean copulas is less obvious, but also turns out to be fairly simple in the majority of cases.

Factor Copulas

- Now we look at a factor structure and study the related copula
- Consider an m-dimensional copula C and assume that $U \sim C$
- Suppose that there is a $p < m$-dimensional random vector V, such that conditional on V, $U_1, \ldots, U_m$ are independent. In that case the copula is called a **factor copula**.
- We can compute the cdf as follows:

$$\begin{aligned} F(x_1, \ldots, x_m) &= C\left(F_1(x_1), \ldots, F_m(x_m)\right) \\ &= P(U_1 \leq F_1(x_1), \ldots, U_m \leq F_m(x_m)) \\ &= E\left[\prod_{i=1}^{m} P\left(U_i \leq F_i(x_i) \mid V\right)\right] \end{aligned}$$

  based on the **conditional distribution function** of $U_1, \ldots, U_m$.
- Lets go for an example

Example: One-Factor Gauss Copula

- Let $X_i = \sqrt{\rho_i}\, V + \sqrt{1-\rho_i}\, \epsilon_i$, $\rho_i \in (0,1)$ and $V, (\epsilon_i)_{1 \le i \le m}$ iid standard normal rvs
- then $X \sim \mathcal{N}_m(0, \Sigma)$
- Σ has as entries the covariances $\sigma_{ij} = \sqrt{\rho_i \rho_j}$
- et $U_i = \Phi(X_i)$, so that $U \sim C_\Sigma^{\mathbf{Ga}}$ (the Gaussian copula with covariance matrix Σ)
- Now we compute the distribution function: first, observe that with $d_i = \Phi^{-1}(F_i(x_i))$,

$$\begin{aligned} P\left(U_i \le F_i(x_i) \mid V = v\right) &= P\left(\Phi(X_i) \le F_i(x_i) \mid V = v\right) \\ &= P\left(X_i \le \Phi^{-1}(F_i(x_i)) \mid V = v\right) \\ &= P\left(\epsilon_i \le \frac{d_i - \sqrt{\rho_i}\, v}{\sqrt{1-\rho_i}} \mid V = v\right) \\ &= \Phi\left(\frac{d_i - \sqrt{\rho_i}\, v}{\sqrt{1-\rho_i}}\right) \end{aligned}$$

- Inserting this into the above representation, we obtain the multivariate distribution function

$$F(x_1, \ldots, x_m) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \prod_{i=1}^{m} \Phi\left(\frac{d_i - \sqrt{\rho_i} v}{\sqrt{1-\rho_i}}\right) e^{-v^2/2} dv\,.$$

- The **exchangeable special case** is obtained when the correlations are all equal. Then the formula simplifies further and we have only one parameter - ρ.

Simulation of factor copula models

- Note that we have a mixture

$$F(x_1, \dots, x_m) = E\left[\prod_{i=1}^{m} P\left(U_i \leq F_i(x_i) \mid V\right) \right]$$
$$= \int \prod_{i=1}^{m} F_i(x_i, v) f_V(v) dv.$$

This allows a straightforward simulation

1. Simulate a realization of V.
2. Simulate independent rvs U_i with df $F_i(\cdot, v)$.

Estimating - Rank Correlation

Working on ranks is a classical way in doing non-parametric statistics. Note that working on ranks is like working on the copula directly, so ideally suited for estimating the copula. Let us look on the most popular measures.

Spearman's rho

$$\begin{aligned}\rho_S(X_1, X_2) &= \rho(F_1(X_1), F_2(X_2)) = \rho(\text{copula}) \\ \rho_S(X_1, X_2) &= 12 \int_0^1 \int_0^1 \{C(u_1, u_2) - u_1 u_2\} du_1 du_2.\end{aligned}$$

Kendall's tau

Take an independent copy of (X_1, X_2) denoted $(\widetilde{X}_1, \widetilde{X}_2)$.

$$\begin{aligned}\rho_\tau(X_1, X_2) &= 2P\left((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0\right) - 1 \\ \rho_\tau(X_1, X_2) &= 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.\end{aligned}$$

Properties of Rank Correlation

The following statements are true for Spearman's rho (ρ_S) or Kendall's tau (ρ_τ), but not for Pearson's linear correlation (ρ).

- ρ_S depends only on copula of (X_1, X_2).
- ρ_S is invariant under strictly increasing transformations of the random variables.
- $\rho_S(X_1, X_2) = 1 \iff X_1, X_2$ comonotonic.
- $\rho_S(X_1, X_2) = -1 \iff X_1, X_2$ countermonotonic.

Sample Rank Correlations

Consider iid bivariate data $\{(X_{1,1}, X_{1,2}), \ldots, (X_{n,1}, X_{n,2})\}$. The standard estimator of $\rho_\tau(X_1, X_2)$ is

$$\frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \mathsf{sgn}\left[(X_{i,1} - X_{j,1})(X_{i,2} - X_{j,2})\right],$$

and the estimator of $\rho_S(X_1, X_2)$ is

$$\frac{12}{n(n^2-1)} \sum_{i=1}^{n} \left(\mathsf{rank}(X_{i,1}) - \frac{n+1}{2}\right)\left(\mathsf{rank}(X_{i,2}) - \frac{n+1}{2}\right).$$

Fitting Copulas to Data

- We have data vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with identical distribution function F.
- We write $\mathbf{X}_t = (X_{t,1}, \ldots, X_{t,d})'$ for an individual data vector and $\mathbf{X} = (X_1, \ldots, X_d)'$ for a generic random vector with df F.
- We assume further that this df F has continuous margins $F_1, \ldots, F_d$ and thus
- by Sklar's theorem a unique representation $F(x) = C(F_1(x_1), \ldots, F_d(x_d))$.

This part is devoted to the problem of estimating the parameters $\boldsymbol{\theta}$ of a parametric copula $C_{\boldsymbol{\theta}}$. The main method we consider is **maximum likelihood estimation**, but we first outline a simpler **method-of-moments procedure** using sample rank correlation estimates; this method has the advantage that marginal distributions do not need to be estimated so that inference about the copula is margin-free.

Method-of-Moments Using Rank Correlation

Recall the standard estimators of Kendall's rank correlation and Spearman's rank correlation. We will use the notation R^τ and R^S to denote matrices of pairwise estimates. These can be shown to be positive semi-definite.

- Calibrating Gauss copula with Spearman's rho

Suppose we assume a meta-Gaussian model for $\mathbf{X}$ with copula C_P^{Ga} and we wish to estimate the correlation matrix P. It follows from Theorem 5.36 in book that

$$\rho_S(X_i, X_j) = \frac{6}{\pi} \arcsin \frac{\rho_{ij}}{2} \approx \rho_{ij},$$

where the final approximation is very accurate. This suggests we estimate P by the matrix of pairwise Spearman's rank coefficients R^S.

Calibrating t Copula with Kendall's tau

Suppose we assume a meta t model for $\mathbf{X}$ with copula $C^{\mathbf{t}}_{\nu,P}$ and we wish to estimate the correlation matrix P. The theoretical relationship between Spearman's rho and P is not known in this case, but a relationship between Kendall's tau and P is known.
It follows from Proposition 5.37 in book that

$$\rho_\tau(X_i, X_j) = \frac{2}{\pi} \arcsin \rho_{ij},$$

so that a possible estimator of P is the matrix R^* with components given by $r^*_{ij} = \sin(\pi r^\tau_{ij}/2)$ This may not be positive definite, in which case R^* can be transformed by the eigenvalue method given in Algorithm 5.55 to obtain a positive definite matrix that is **close** to R^*.
The remaining parameter ν of the copula could then be estimated by maximum likelihood.

Maximum Likelihood Method

To estimate the copula by ML we require a so-called **pseudo-sample** of observations from the copula. To construct such a sample we are required to estimate marginal distributions. This can be done with

1. parametric models $\widehat{F}_1, \dots, \widehat{F}_d$,
2. a form of the empirical distribution function such as
   $\widehat{F}_j(x) = \frac{1}{n+1} \sum_{i=1}^{n} 1_{\{X_{i,j} \le x\}}, \quad j = 1, \dots, d,$

The second method, known as pseudo-maximum likelihood, means that we essentially work with the **ranks** of the original data, standardized to lie on the copula scale. For statistical properties see Genest and Rivest (1993).

Estimating the Copula

We form the pseudo-sample

$$\widehat{\mathbf{U}}_i = \left(\widehat{U}_{i,1}, \ldots, \widehat{U}_{i,d}\right)' = \left(\widehat{F}_1(X_{i,1}), \ldots, \widehat{F}_d(X_{i,d})\right)', \quad i = 1, \ldots, n.$$

and fit parametric copula C by maximum likelihood.
The copula density is

$$c(u_1, \ldots, u_d; \boldsymbol{\theta}) = \frac{\partial}{\partial u_1} \cdots \frac{\partial}{\partial u_d} C(u_1, \ldots, u_d; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ denotes the unknown parameters. The **log-likelihood** is

$$l(\boldsymbol{\theta}; \widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n) = \sum_{i=1}^{n} \log c(\widehat{U}_{i,1}, \ldots, \widehat{U}_{i,d}; \boldsymbol{\theta}).$$

Independence of vector observations assumed for simplicity.

BMW-Siemens Example

Copula	ρ, β	ν	std.error(s)	log-likelihood
Gauss	0.70		0.0098	610.39
t	0.70	4.89	0.0122,0.73	**649.25**
Gumbel	1.90		0.0363	584.46
Clayton	1.42		0.0541	527.46

Summary

- We shortly dived into multivariate dependencies
- Copulas are a general concept to capture all multivariate dependencies
- If we can find a factor structure, we have a simple dependencies - but this is only a very special case of the picture.
- Now we study the application to multivariate time series.

Risk and Risk measures

- Now we start an important discussion, which is central in the regulation of banks: the measurement of **risk**
- It has become an own strand of research - the study of risk measures.
- Here we only shortly touch upon this topic. For more details or literature we refer to Föllmer and Schied (2011).

Loss Distributions

Risks are represented by **random variables** mapping unforeseen future states of the world into values representing **profits and losses**.

The risks which interest us are **aggregate** risks. In general we consider a **portfolio** which might be

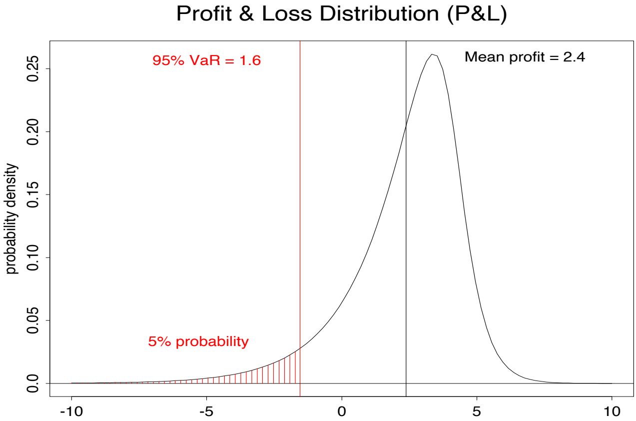
- a collection of **stocks and bonds**;
- a book of **derivatives**;
- a collection of risky **loans**;
- a financial institution's **overall position** in risky assets.

Portfolio Values and Losses

- Consider a portfolio and let V_t denote its **value** at time t; we assume this random variable is **observable** at time t.
- Suppose we look at risk from perspective of time t and we consider the time period $[t, t+1]$. The value V_{t+1} at the end of the time period is unknown to us.
- The distribution of $(V_{t+1} - V_t)$ is known as the profit-and-loss or **P&L distribution**. We denote the **loss** by $L_{t+1} = -(V_{t+1} - V_t)$. By this convention, losses will be positive numbers and profits negative.

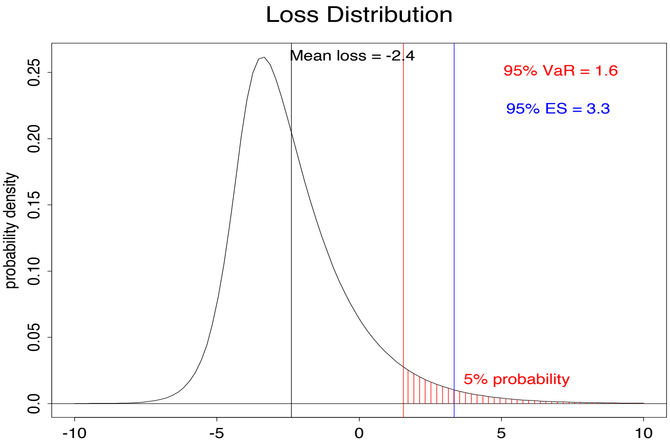
We refer to the distribution of L_{t+1} as the **loss distribution**.

An example



- Profits and losses. Now we turn it around and obtain ...

An example



- Profits and losses. Now we turn it around and obtain ...

Risk Measures

Risk measures

Risk measures attempt to quantify the riskiness of a portfolio. The most popular risk measures like value-at-risk **VaR** describe the right tail of the loss distribution of L_{t+1} (or the left tail of the P&L).

- Think of a risk measure as a method of quantifying the amount of capital which is needed to make a financial position acceptable. This can be made mathematically precise (soon) and will lead to **coherent** and **convex** risk measures.
- We will also meet risk measures in the context of **fairness of AI** - here it is used as a deviation of a positive random variable from zero. It will be one of our tasks to critically review this approach and probably suggest improved alternatives !
- So first - what is a risk measure?

Probability and Quantile Transforms

- Denote the distribution function of the loss $L := L_{t+1}$ by F_L so that $P(L \leq x) = F_L(x)$.
- Recall: The (generalized) **inverse** of the cdf F is

$$F^{-1}(t) := \inf\{x \in \mathbb{R} : F(x) \geq t\}$$

for any $t \in (0, 1)$. $q_\alpha := F^{-1}(\alpha)$ is the α-**Quantile** of F.

VaR and Expected Shortfall

Let $0 < \alpha < 1$ (typically $\alpha = 0.95,\ 0.99$). We use

- **Value at Risk** is defined as

$$\mathrm{VaR}_\alpha = q_\alpha(F_L) = F_L^{-1}(\alpha)\,, \tag{2}$$

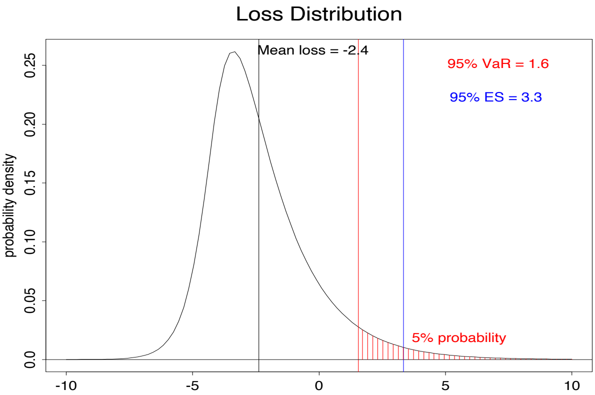
  where we use the notation $q_\alpha(F_L)$ or $q_\alpha(L)$ for a quantile of the distribution of L and F_L^{-1} for the (generalized) inverse of F_L.

- Provided $E(|L|) < \infty$ **expected shortfall** is defined as

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)du. \tag{3}$$

- Here - instead of fixing one level α, we average over all levels above α.

VaR in Visual Terms



Expected Shortfall

- For **continuous** loss distributions expected shortfall is the expected loss, given that the VaR is exceeded.

Lemma

Assume that L is continuous. Then, for any $\alpha \in (0,1)$ we have

$$\mathrm{ES}_\alpha = \frac{E\big[L \cdot \mathbb{1}_{L \geq q_\alpha(L)}\big]}{1-\alpha} = E[L \mid L \geq \mathrm{VaR}_\alpha]\,.$$

Proof(blackboard)

For a discontinuous loss df we have the more complicated expression

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\Big(E\big[L \cdot \mathbb{1}_{L \geq q_\alpha(L)}\big] + q_\alpha(1-\alpha-P(L \geq q_\alpha))\Big)\,.$$

Coherent Measures of Risk

- There are many possible measures of the risk in a portfolio such as VaR, ES or stress losses. To decide which are reasonable risk measures a systematic approach is called for.
- The authors in Artzner et al. (1999) achieved a list of properties which a risk measure should have - and called them **coherent**
- Thereafter a study of coherence started (value-at-risk is not always coherent, expected shortfall is, etc.)
- and further extensions have been proposed: for example, convex risk measures.

Purposes of Risk Measurement

Risk measures are used for the following purposes:

- Determination of **risk capital**. Risk measure gives amount of capital needed as a buffer against (unexpected) future losses to satisfy a regulator.
- **Management tool**. Risk measures are used in internal limit systems.
- **Insurance premia** can be viewed as measure of riskiness of insured claims.

Interpretation: Risk measure gives amount of capital that needs to be added to a position with loss L, so that the position becomes **acceptable** to an (internal/external) regulator.

The Axioms

A coherent risk measure is a real–valued function ϱ on some space of rv's (representing losses), s.t.

1. **Monotonicity**. For two rv's with $L_1 \geq L_2$ we have $\varrho(L_1) \geq \varrho(L_2)$.
2. **Subadditivity**. For any L_1, L_2 we have $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$.

   This is the most debated property. Necessary for following reasons:

   - Reflects idea that risk can be reduced by **diversification** and that "a merger creates no extra risk".
   - Makes **decentralized** risk management possible.
   - If a regulator uses a non-subadditive risk measure, a financial institution could reduce risk capital by splitting into subsidiaries.
3. **Positive homogeneity**. For $\lambda \geq 0$ we have that $\varrho(\lambda L) = \lambda \varrho(L)$. If there is no diversification we should have equality in subadditivity axiom.
4. **Translation invariance**. For any $a \in \mathbb{R}$ we have that $\varrho(L + a) = \varrho(L) + a$.

Remarks:

- VaR is in general not coherent. ES (as we have defined it) is coherent.
- Non-subadditivity of VaR is relevant in presence of **skewed** loss distributions (credit-risk management, derivative books), or if traders **optimize against VaR**.
- Many recent papers study **convex** risk measures. Here instead of Subadditivity and Positive Homogeneity one has convexity:

$$\rho(\lambda L_1 + (1-\lambda)L_2) \leq \lambda\rho(L_1) + (1-\lambda)\rho(L_2)$$

  for all $\lambda \in [0,1]$.

Non-Coherence of VaR: an Example

Consider portfolio of 50 defaultable bonds with independent defaults. Default probability identical and equal to 2%. Current price of bonds equal to 95, face value equal to 100.

1. Portfolio A: buy 100 units of bond 1; current value is $V_0 = 9500$.
2. Portfolio B : buy 2 units of each bond; current value is $V_0 = 9500$.

Common sense. Portfolio B is less risky (better diversified) than Portfolio A. This is wrong if we measure risk with VaR !

Loss of each bond equals

$$L_i := 95 - 100(1 - Y_i) = 100Y_i - 5,$$

where $Y_i = 1$ if default occurs, $Y_i = 0$ else. Y_i are iid Bernoulli(0.02).

Non-Coherence of VaR: an Example II

$$L_i := 95 - 100(1 - Y_i) = 100Y_i - 5.$$

Portfolio A: $L = 100L_1$ and hence

$$\mathrm{VaR}_{0.95}(L) = 100\,\mathrm{VaR}_{0.95}(L_1) = -500,$$

i.e. we may take 500 out of portfolio and still satisfy regulator.

Portfolio B: $L = \sum_{i=1}^{50} 2L_i = 200\sum_{i=1}^{50} Y_i - 500$, and hence

$$\mathrm{VaR}_\alpha(L) = 200\,q_\alpha\left(\sum_{i=1}^{50} Y_i\right) - 500.$$

Inspection shows that $q_{0.95}(\sum_{i=1}^{50} Y_i) = 3$, so that $\mathrm{VaR}_{0.95}(L) = 100$, i.e. **extra capital is needed** to hold the portfolio.

This is directly linked to non-coherence of VaR.

Examples

Further, we want to compute some examples and learn how to elaborate with this context. We focus on the easiest case - when losses are normal.

- Assume that L is normally distributed with mean μ and variance σ^2.
- The value-at-risk is easily computed:

$$\mathrm{VaR}_\alpha = q_\alpha(L) = \Phi^{-1}\Big(\frac{\alpha - \mu}{\sigma}\Big).$$

Lemma

For normally distributed losses we have that

$$\mathrm{ES}_\alpha = \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

Proof (blackboard)

Summary

- We have met risk measures and their two important classes: coherent and convex risk measures.
- We also saw the two important special cases value-at-risk and expected shortfall
- and showed that expected shortfall is an average over the value-at-risk when distribution functions are continuous (inspiring spectral risk measures).

Artzner, Philippe et al. (1999). „Coherent measures of risk“. In: **Mathematical finance** 9.3, pp. 203–228.

Föllmer, Hans and Alexander Schied (2011). **Stochastic finance: an introduction in discrete time**. Walter de Gruyter.

Genest, C. and L. Rivest (1993). „Statistical inference procedures for bivariate archimedean copulas“. In: 88, pp. 1034–1043.

Schmidt, Thorsten (2007). „Coping with copulas“. In: **Copulas-From theory to application in finance** 3, pp. 1–34.