# 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 $\{1, 2, \dots, 6\}$.
- 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 $\alpha$-**Quantile** of $F$.

## Lemma 1: probability transform

Let $X$ be a random variable with **continuous** distribution function $F$.
Then $F(X) \sim U(0,1)$ (standard uniform).

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

## Lemma 2: quantile transform

Let $U$ be uniform and $F$ the distribution function of **any** rv $X$.
Then $F^{-1}(U) \overset{\mathscr{L}}{=} X$ so that $P(F^{\leftarrow}(U) \leq x) = F(x)$.

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 $F_1, \ldots, F_d$.*

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

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

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

# 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

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

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

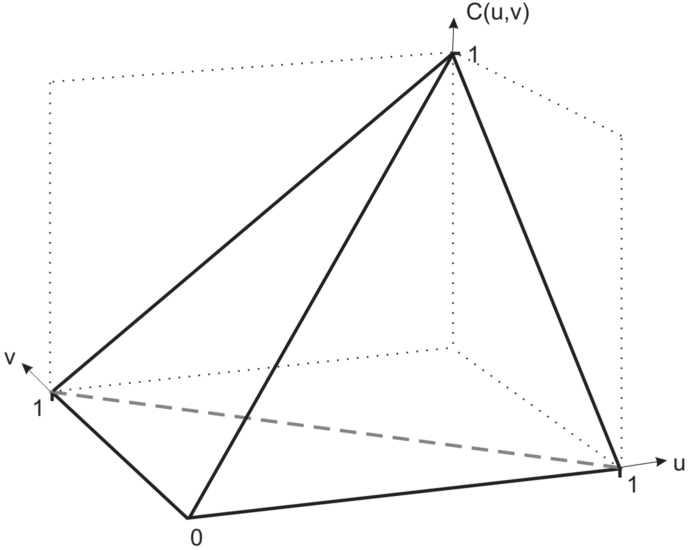
$C$ is invariant under **strictly increasing** transformations of the marginal distributions. If $T_1, \ldots, T_d$ are strictly increasing, then $(T_1(X_1), \ldots, T_d(X_d))$ has the same copula as $(X_1, \ldots, X_d)$.

# The Fréchet Bounds

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

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

- The upper bound is the df of $(U, \ldots, U)$ and the copula of the random vector $\mathbf{X}$ where $X_i \overset{\text{a.s.}}{=} T_i(X_1)$ 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 $X_2 \overset{\text{a.s.}}{=} T(X_1)$ 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 $\Phi$ 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 $C_\rho^{\mathsf{Ga}}$ when $d = 2$.
$P = I_d$ gives independence and $P = J_d$ gives comonotonicity.

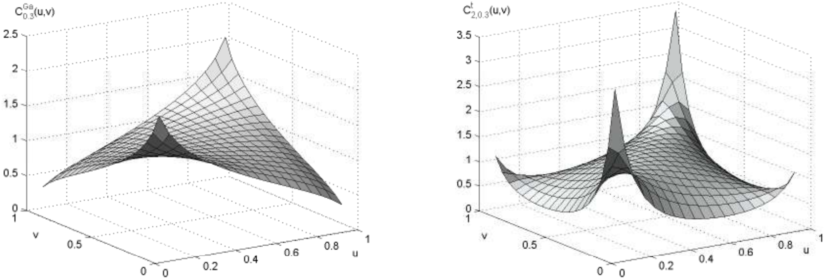
# Copula densities

- 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 $C_P^{\textsf{Ga}}$

- 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 $C_{\nu,P}^{\textsf{t}}$

- Simulate $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$
- Set $\mathbf{U} = (t_\nu(X_1), \ldots, t_\nu(X_d))'$ (probability transformation)
  $t_\nu$ 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.

# 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 ($\rho_S$) or Kendall's tau ($\rho_\tau$), but not for Pearson's linear correlation ($\rho$).

- $\rho_S$ depends only on copula of $(X_1, X_2)$.
- $\rho_S$ 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 module 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 $R^\tau$ 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 $C_P^{\mathsf{Ga}}$ 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 $\nu$ 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 $\widehat{F}_1, \ldots, \widehat{F}_d$,
2. a form of the empirical distribution function such as
   $\widehat{F}_j(x) = \frac{1}{n+1} \sum_{i=1}^{n} 1_{\{X_{i,j} \leq x\}}, \quad j = 1, \ldots, d$,
3. empirical df with EVT tail model.

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 | $\rho, \beta$ | $\nu$ | 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 |

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.