Copulas, Sklar's Theorem, and Distributional Transform

Matvei Kreinin, BSc.

Moscow Institute of Physics and Technology

September 24, 2024

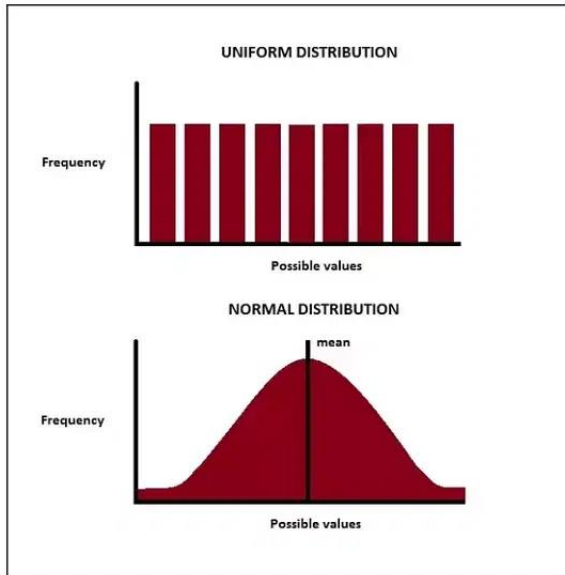# Table of contents

Motivation

Problem statement

Definition
Consider a random vector $(X_1, X_2, ..., X_d)$. Suppose its marginals are continuous, i.e. the marginal CDFs $F_i(x) = \mathbb{P}(X_i \leq x)$ are continuous functions. By applying the probability integral transform to each component, the random vector

$$(U_1, U_2, ..., U_d) = (F_1(X_1), F_2(X_2), ..., F_d(X_d))$$

has the marginals thar are uniformly distributed on the interval $[0, 1]$. The copula of $(X_1, X_2, ..., X_d)$ is defined at the join cumulative distribution function of $(U_1, U_2, ..., U_d)$:

$$C(u_1, u_2, ..., u_d) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2, ..., U_d \leq u_d)$$

Problem statement

Definition
The reverse of these steps can be used to generate pseudo-random samples from general classes of multivariate probability distributions. That is, given a procedure to generate a sample

$$(X_1, X_2, ..., X_d) = (F_1^{-1}(U_1), F_2^{-1}(U_2), ..., F_d^{-1}(U_d))$$

The generalized inverses F_i^{-1} are unproblematic almost surely, since the F_i were assumed to be continuous. Furthermore, the above formula for the copula function can be rewritten as:

$$C(u_1, u_2, ..., u_d) = \mathbb{P}(X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), ..., X_d \leq F_d^{-1}(u_d))$$

Theorem

Definition
In probabilistic terms, $C : [0,1]^d \to [0,1]$ is a d-dimensional copula if C is a joint cumulative distribution function of a d-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals.

Theorem
Let $F \in \mathcal{F}(F_1, ..., F_n)$ be an $n-$dimensional distribution function with marginals $F_1, ..., F_n$. Then there exists a copula $C \in \mathcal{F}(\mathcal{U}, ..., \mathcal{U})$ with uniforms marginals such that

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n))$$

Applications

- Generative Models and Variational Autoencoders (VAEs) Traditional VAEs typically assume independent latent variables with Gaussian priors. By applying copulas based on Sklar's Theorem, more complex dependencies among latent variables can be captured, allowing for more flexible and realistic data generation.

- Time Series Modeling and Sequential Data (Financial, Weather, IoT) In time series analysis, especially in applications like finance or weather forecasting, the relationship between variables can exhibit complex dependencies beyond what can be captured using simple correlation. Copulas can capture tail dependencies and non-linear relationships between variables, making them valuable for multivariate time series prediction.

Links

- ▶ Deriving the joint probability of default of two entities with the Gaussian Copula, book, p. 77
- ▶ Copulas, Sklar's Theorem, and Distributional Transform, article
- ▶ Copula Short Course