Copula
Applied Stochastic Processes (FIN 514)

Instructor: Jaehyuk Choi

Peking University HSBC Business School, Shenzhen, China

2018-19 Module 1 (Fall 2018)

Joint probability distribution

- Two random variables X_1 and X_2 have PDF $f_1(x)$ and $f_2(x)$, and CDF $F_1(x)$ and $F_2(x)$ respectively.

$$F_1(x) = \text{Prob}(X_1 \le x)$$
$$F_2(x) = \text{Prob}(X_2 \le x)$$

- However, the knowledge of the PDFs and CDFs of individual RVs does not tell us how the two RVs are related. We still need to define the joint PDF and CDF:

$$F_{1,2}(x_1, x_2) = \text{Prob}(X_1 \le x_1 \text{ and } X_2 \le x_2)$$

- Note that the definition of $F_{1,2}(x_1, x_2)$ is not related to those of $F_1(x)$ and $F_2(x)$. In two extremes, X_1 and X_2 can be independent or completely correlated, often characterized by the correlation coefficient ρ.

Multivariate normal distribution

- The PDF of multivariate normal variable $\boldsymbol{x}$ (vector) with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ (matrix) is given as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

- For the independent standard normals ($\Sigma = I(\det\Sigma = 1)$ and $\mu = 0$),

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = n(z_1)\cdots n(z_n) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(z_1^2 + \cdots + z_n^2)\right)$$

- The bivariate case (n=2) is more explicit (see wikipedia).
- In practice, the joint PDF is not often used (remind how we generate correlated normal RNs)
- There are only handful distributions whose joint CDF is known for a given covariance: normal, Student's t, etc.

Joint distribution via copula

- The joint CDF $F_{1,2}(x_1, x_2)$ is *not completely independent* from the individual RNs. For one thing, the function domain has to be same as that of RN: $[0, 1]$ for uniform, $(-\infty, \infty)$ for normal, etc.
- For the CDFs $F_1(x)$ and $F_2(x)$, we know that $F_1(X_1)$ and $F_2(X_2)$ are uniform RNs. (In the same way we generate RNs $X_1 = F_1^{-1}(U)$.)
- So can implicitly define the joint CDF via a copula function $C : [0, 1]^2 \to [0, 1]$,

$$C(u_1, u_2) = F_{1,2}(x_1 = F_1^{-1}(u_1), x_2 = F_2^{-1}(u_2))$$
$$C(u_1 = F_1(x_1), u_2 = F_2(x_2)) = F_{1,2}(x_1, x_2)$$

  Defining either $F_{1,2}(x_1, x_2)$ or $C(u_1, u_2)$ is equivalent.
- The function C can be understood as a joint CDF on uniform RNs.

Copula – Mathematical definition

Now we generalize to n-dimensional case: $C : [0,1]^n \to [0,1]$. Because C is a joint CDF function, it should satisfy:

- $C(u_1, \cdots, u_{k-1}, 0, u_{k+1}, \cdots, u_n) = 0$
- $C(1, \cdots, 1, u_k, 1, \cdots, 1) = u_k$
- The probability on any hypercube is always non-negative.
  - For $n = 1$, it means $C(u_1^a) \le C(u_1^b)$ if $u_1^a \le u_1^b$.
  - For $n = 2$, the probability over $[u_1^a, u_1^b] \times [u_2^a, u_2^b]$ should be non-negative:

$$0 \le C(u_1^b, u_2^b) - C(u_1^a, u_2^b) - C(u_1^b, u_2^a) + C(u_1^a, u_2^a)$$

  - If $C(u_1, \cdots, u_n)$ is a continuous function, the PDF is non-negative:

$$0 \le c(u_1, \cdots, u_n) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} C(u_1, \cdots, u_n)$$

Copula – Mix and match

- Copula $C(\cdots)$ is a way of defining joint distribution not bounded by the original RVs.
- There are only a few well-known distribution from which $C(\cdots)$ is defined from the multivariate joint distribution, $F_{\boldsymbol{X}}(\cdots)$
- For the distributions difficult to define joint distributions, we borrow the copulas from those well-known joint distributions and apply to the original distributions.

Copula - Families

- Gaussian Copula:

$$C_{\boldsymbol{R}}(u_1, \cdots, u_n) = N_n(N^{-1}(u_1), \cdots, N^{-1}(u_n))$$

  where N_n is the multi-variate cumulative normal distribution with correlation matrix $\boldsymbol{R}$.

- Independent Copula:

$$C(u_1, \cdots, u_n) = u_1\, u_2 \quad \text{or} \quad c(u_1, \cdots, u_n) = 1$$

- Completely dependent Copula:

$$C(u_1, \cdots, u_n) = \min(u_1, \cdots, u_n)$$

- Others: see the copula families in wikipedia.

RN generation from Gaussian copula

- Imagine CDFs $F_1(x_1)$ and $F_2(x_2)$ are given and we want to generate joint RNs, e.g., (X_1, X_2) in order to evaluate an expectation, $E\left[g(X_1, X_2)\right]$ (e.g., a price of a derivative).
- As long as we generate joint uniform RNs (U_1, U_2), we can transform them to $(X_1, X_2) = (F_1^{-1}(U_1), F_2^{-1}(U_2))$.
- We borrow Gaussian variables to generate (U_1, U_2):
  - Generate pairs of independent normal RNs: $(Z_1,\ Z_2)$.
  - Correlate the normal RNs: $(Z_1', Z_2') = (Z_1,\ \rho Z_1 + \sqrt{1-\rho^2} Z_2)$
  - Generate the joint uniform RNs: $(U_1, U_2) = (N(Z_1'),\ N(Z_2'))$
  - Generate the original RNs: $(X_1, X_2) = (F_1^{-1}(U_1),\ F_2^{-1}(U_2))$
- Finally Monte-Carlo method is applied as

$$E\left[g(X_1, X_2)\right] = \frac{1}{N} \sum_{k=1}^{N} g(X_1^{(k)}, X_2^{(k)})$$

A case: spread option under SV models

- We want to price a spread option, i.e., $E(S_{1T} - S_{2T} - K)^+$ using MC.
- If two stocks S_1 and S_2 follow GBMs, we know how to correlate them (HW3) since the the distribution is transformed from normal RVs. However, GBMs are not right due to the volatility smile.
- If the stocks follow SV models, creating a joint distribution is not easy. So we use copula.
- We first build discrete CDFs for $S_1(T)$ and $S_2(T)$ from the call prices at the series of strikes, $K_j = S_0 + j\Delta K$ for $j = 0, \pm 1, \pm 2, \cdots$.

$$F(K_j) = -\frac{\partial}{\partial K} C(K) \approx \frac{C(K_{j-1}) - C(K_{j+1})}{2\Delta K}$$

- The discrete inverse CDF is the interpolation from the inversed pairs, $(F(K_j), K_j)$.

A case: Collateralized debt obligation (CDO)

- A COD is a bond backed by a pool of loans.
- Naturally the joint distribution of the default of the underlying loans are important. So Gaussian copula is used as a standard way of pricing CDOs.
- While the underlying loans are sub-primes (below investment grade `BBB-`), the super-senior *tranche* of CDO got `AAA` credit rating as the pools were considered *diversified*. The correlation was typically estimated from historical data.
- In financial crisis, however, the correlation across all assets significantly increased: when a bond defaults, the others do so. So the pool is not really diversified.
- The use of copula is criticized as one reason behind the financial criss in 2008–2009. Copula in general can not capture the dynamic changes of the correlation over time.