Kernel Smoothing Technique and Copula for Dependency Structure

Author: Tianyi Zhang
Version: 1.0
Date: February 6, 2020

1 Kernel Density Estimation

Kernel density estimation (KDE) is a non-parametric way to estimate the distribution function of a random variable. Practically, KDE uses an integral form of the transformed data at each sample point with a kernel probability function

$$\hat{f}(x_0) = \frac{1}{nh} \sum_i K(\frac{x_i - x_0}{h}), \tag{1}$$

where n is the number of observations, h is a smoothing parameter bandwidth which controls the neighborhood around x_0, K is the kernel probability function satisfying the following properties

- $K(u) \geq 0$,
- $K(u) = K(-u)$,
- $\int K(u)du = 1$,
- $\int u^2 K(u)du > 0$.

The kernel function is designed to control the weights given to observations $\{x_i\}$ at point x_0. In our micro service, we support four kernel functions with the following forms:

- (normal) $K(u) = \sqrt{\frac{1}{2\pi}} \exp(-\frac{u^2}{2})$
- (box) $K(u) = \begin{cases} \frac{1}{2}, & \text{if } u \in [-1, 1], \\ 0, & \text{otherwise.} \end{cases}$
- (epanechinikov) $K(u) = \begin{cases} 1 + u, & \text{if } u \in [-1, 0], \\ 1 - u, & \text{if } u \in [0, 1], \\ 0, & \text{otherwise.} \end{cases}$
- (triangle) $K(u) = \begin{cases} \frac{3}{4}(1 - u^2), & \text{if } u \in [-1, 1], \\ 0, & \text{otherwise.} \end{cases}$

KDE has the advantage that it takes in data with essentially irregular or non-smoothing histogram and return a smoothed version of the distribution function, even with non-smooth kernel functions. The mathematical interpretation is that the integral operator can produce functions with higher regularity than the integrand. Moreover, the quality of the estimator depends less on the choice of the kernel function K than the bandwidth selection h[10].

Once kernel function is selected, the bandwidth should be estimated so that the kernel density and empirical density is sufficiently close. Certain measures can be introduced in the aspect, for example, mean squared errors (MSE), mean integrated squared error (MISE) and asymptotic mean integrated squared error (AMISE). Below is the mathematical form of the mentioned measures

$$\begin{aligned} MSE(f, \hat{f}_h) &= E((f(x) - \hat{f}_h(x))^2), & (2) \\ MISE(f, \hat{f}_h) &= E(\int (f(x) - f_h\hat{(}x))^2 dx), & (3) \\ AMISE(f, \hat{f}_h) &= \frac{\int K(y)^2 dy}{nh} + \frac{\sigma_K^4 h^4 \int (f''(x))^2 dx}{4}, & (4) \end{aligned}$$

where $\sigma_K^2 = \int x^2 K(x) dx$ is the variance of the kernel.

Optimal bandwidth is associated with the curvature of unknown density, therefore we tend to assign a priori (usually normal) density and calculate its curvature first. Sometimes this approach is called "plug-in" selectors. Here are some practical rules of thumb:

- Silverman's reference bandwidth [11]: $h_{AMISE} = 0.9An^{-1/5}$, where A equals minsample std, sample interquantile range (IQR)/1.34 ;
- Normal reference rule [1, 9]: in this approach, $h = \hat{\sigma}(\frac{4}{3n})^{1/5}$, where $\hat{\sigma}$ is sample standard deviation or more robust estimate, for example, IQR or inverse standard normal $\phi^{-1}(\frac{3}{4}) * MAD(x)$ (adopted by MATLAB and the micro service).

Once bandwidth is determined, KDE will follow the next steps to evaluate pdf or cdf:

1. Calculate kernel estimates of individual points, then use equally weighting scheme to evaluate kernel function values over an evenly spaced region of a specified interval;
2. Guard against rounding off errors to secure positive pdf and $[0, 1]$ cdf estimates.

With cdf and pdf, Newton's method allows us to calculate inverse cdf, i.e., to find x that meets $cdf(x) = y$ for a given y. The algorithm iteratively calls cdf and pdf (as first order derivative of cdf) to find the root, or recursively $x_{k+1} = x_k - \frac{cdf(x_k)}{\max\{pdf(x_k), \varepsilon\}}$. Since we have a vector input of the icdf, the stopping criterion is componentwisely applied in the sense that the component which satisfies the stopping criterion will no longer be updated. Note here we introduce ϵ to refine potential constant region of the cdf to avoid invalid division. Since it requires more function evaluations, icdf is more computationally expensive than pdf and cdf.

2 General Copula Description

Definition 2.1 (Copula)**.** *A copula is a function $C : [0,1]^d \to [0,1]$ that maps marginal cumulative distributions to joint cumulative distribution. That is, C must satisfy the following conditions:*

- $C(u_1, \cdots, u_{i-1}, 0, u_{i+1}, \cdots, u_d) = 0,$
- $C(1, \cdots, 1, u_i, 1, \cdots, 1) = u_i,$
- $\int_B dC(u) \geq 0$ *for* $B = \prod_i [x_i, y_i] \subset [0,1]^d$.

Theorem 2.1 (Existence)**.** *Let $(X_1, \cdots, X_d)$ be a vector of random variables with joint distribution $F(x_1, \cdots, x_d) := \mathbb{P}[X_1 \leq x_1, \cdots, X_d \leq x_d]$ and each margin distribution $F_i(x_i) := \mathbb{P}[X_i \leq x_i]$. There exists a copula C, such that $F(x_1, \cdots, x_d) = C(F_1(x_1), \cdots, F_d(x_d))$.*

Theorem 2.2 (Boundedness)**.** *The copula function $C : [0,1]^d \to [0,1]$ must satisfy the following bounds:*

$$\max\{1 - d + \sum_i u_i, 0\} \leq C(u_1, \cdots, u_d) \leq \min\{u_1 \cdots, u_d\}. \tag{5}$$

3 EllipticalCopula

3.1 Gaussian Copula

First, we consider a d-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ. For simplicity, we use standardization to get $\mu = \mathbf{0}$ and Σ has all ones on the diagonal. Let Φ denotes the cdf of a Gaussian distribution, we can formulate the Gaussian copula as

$$\begin{aligned} C_\Sigma(\mathbf{u}) &= \Phi\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \cdots, \Phi^{-1}(u_d)\right) \qquad (6) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det\Sigma}} \exp(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x})d\mathbf{x}. \end{aligned}$$

By maximization of the log-likelihood function, we can fit the Gaussian copula structure to given data. Let R be the correlation matrix, the log likelihood function has the form

$$\ln c_\Sigma(\mathbf{u}) = -\frac{1}{2}\ln(\det\Sigma) - \frac{1}{2}\mathbf{y}^T(\Sigma^{-1} - I)\mathbf{y} - \frac{d}{2}\ln(2\pi),$$

where $\mathbf{y} := \Phi^{-1}(u) = \left(\Phi^{-1}(u_1), \cdots \Phi^{-1}(u_d)\right)^t$. The maximum log likelihood framework gives the problem

$$\ln c_{\Sigma^*}(\mathbf{u}) = \max_\Sigma \ln c_\Sigma(\mathbf{u}),$$

where Σ^* denotes the maximum likelihood estimate (MLE) for Σ. The simulation of a Gaussian copula is straightforward, given that Σ^* is known.

As in the framework proposed in [6], Gaussian copula is widely used in company asset models prior to the financial crisis in 2007-2008, in which the company asset X_i is decomposed to a systematic risk component Y and an idiosyncratic risk component Z_i by $X_i = \sqrt{\rho}Y + \sqrt{1-\rho}Z_i$. We define a default of company i if its asset is under a threshold x_i, i.e., $x_i = \Phi^{-1}(F_i(\tau_i))$, where τ_i denotes its default time and F_i the marginal distribution of default events, modeled by a jump process:

$$F_i(t) = \mathbb{P}\left[\tau \leq t\right] = 1 - e^{-\int_{t_0}^{t} \lambda_i(s)ds}.$$

The marginal distribution F_i and the market quotes for collateralized debt obligation (CDO) can be used to estimate the implied correlation ρ for the CDO tranche. However, in times of financial crisis, asset in a single tranche moves extremely highly coherent, which the Gaussian copula model alone cannot capture and account for the asset asset under risk within a tolerance level.

3.2 Student t Copula

It is well known that financial data is skewed and with heavy tail. Accordingly, it is necessary to introduce distributions with extra degree of freedom (DOF) to explain the skewness and tail performance, and the student t copula is widely applied due to such considerations.

First we revisit the construction of t-distribution,. Let $X_1, X_2, \cdots, X_d$ be a set of random variables drawn from the standard normal distribution $\mathcal{N}(0,1)$. The chi-squared distribution of DOF d is constructed by letting $SS(d) := \sum_{i=1}^{d} X_i^2 \sim \chi^2(d)$. Then the student t-distribution of DOF d is constructed by letting $T_i(d) := \frac{X_i}{\sqrt{\frac{SS(d)}{d}}}$. Note that this is a discrete version, as d can only take values in natural number set. For a continuous version, one can replace the chi-squared distribution by gamma, since $\Gamma(\nu, 2)$ has exactly the same distribution as $\chi(\nu)$. Thus, we can reach the one-dimensional pdf of t-distribution $p_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\phi\nu}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$, where we omit the second argument in gamma function for simplicity.

In the multi-dimensional cases, the density function for student t-distribution $\mathbf{u} = (u_1, \cdots, u_n)$ has the form

$$c_{\nu,\Sigma}(\mathbf{u}) = \frac{\Gamma(\frac{\nu+n}{2})}{(\pi\nu)^{\frac{n}{2}}\Gamma(\frac{\nu}{2})}\left(1+\frac{\mathbf{u}^T\Sigma\mathbf{u}}{\nu}\right)^{-\frac{\nu+n}{2}}. \tag{7}$$

Standard approach to estimate ν can be done via maximizing the log of the likelihood function. First, without loss of generality, we can assume P to be the identity matrix and put the derivative of the log likelihood function to zero to get

$$\frac{1}{2}\sum_i \ln(1+\frac{u_i^2}{\nu}) = I(\nu) + \frac{\nu+1}{2}\sum_i \frac{u_i^2}{\nu+u_i^2} \tag{8}$$

where $I(\nu) := \frac{I_1(\nu)}{\Gamma^n(\frac{\nu+1}{2})(\pi\nu)^{\frac{\nu}{2}}\Gamma^n(\frac{\nu}{2})}$, $\Psi(u) := \frac{\partial \log\Gamma(u)}{\partial u}$ and $I_1(\nu) := \frac{\frac{n}{2}\Gamma^n(\frac{\nu+1}{2})\Psi(\frac{\nu+1}{2})}{(\pi\nu)^{\frac{\nu}{2}}\Gamma^n(\frac{\nu}{2})} - \frac{\Gamma^n(\frac{\nu+1}{2})}{2(\pi\nu)^{\frac{\nu}{2}}\Gamma^n(\frac{\nu}{2})\frac{n}{\nu}} - \frac{\Gamma^n(\frac{\nu+1}{2})}{2(\pi\nu)^{\frac{\nu}{2}}\Gamma^n(\frac{\nu}{2})n\Psi(\frac{\nu}{2})}$.

It is hard to evaluate the log likelihood function (8), plus, the optimization problem is a constrained conic programming.

$$\max_{\nu\geq 0, \Sigma\in\mathcal{D}} \ln c_{\nu,\Sigma}(\mathbf{u}), \tag{9}$$

where $\mathcal{D}$ denotes the set of matrices that is all one on diagonal and SSPD. Thus, an approximation form is proposed for the likelihood function. It is clear that we can rewrite the copula as $\hat{c}_{\nu,\Sigma}(\mathbf{u}) = \frac{1}{\sqrt{\det\Sigma}}\frac{c^d_{\nu,\Sigma}(\mathbf{u})}{\prod_{i=1}^d c^1_{\nu,1}(u_i)}$, where we use the superscription to emphasize the impact of dimensionality. Then, the log likelihood function is

$$\begin{aligned} \ln \hat{c}_{\nu,\Sigma}(\mathbf{u}) \quad = \quad & \frac{n(\nu+1)}{2}\sum_i \log(1+\frac{u_i^2}{\nu}) - \frac{\nu+n}{2}\sum_i \log(1+\frac{w_i^2}{\nu}) \\ & +N\log\Gamma(\frac{\nu+n}{2}) + (n-1)N\log\Gamma(\frac{\nu}{2}) \\ & -nN\log\Gamma(\frac{\nu+1}{2}) - \frac{1}{2}\mathrm{Tr}(\log L), \end{aligned} \tag{10}$$

where $\Sigma = LL^T$ is the Cholesky decomposition of Σ and $L\mathbf{w} = \mathbf{u}$.

The simulation of a sample $\mathbf{u}$ generated by t copula is due to its definition. Given ν, Σ with $\Sigma = LL^T$, we implement below

- Generate a normal random sample, $\xi \sim \mathcal{N}(0,1)^{N\times n}$ and implement linear transformation by $\tilde{\xi} := L\xi$;
- Generate a gamma random sample $\zeta \sim \Gamma(\nu, 2)^N$ and scale by $\tilde{\zeta} := \sqrt{\frac{\zeta}{\nu}}$;
- Compute a t random sample given by $u := \frac{\tilde{\xi}}{\tilde{\zeta}}$.

3.3 Calibration and Optimization Technique

The maximum likelihood estimator framework gives an optimization problem $\max_{\nu,\Sigma} \ln \hat{c}_{\nu,\Sigma}(\mathbf{u})$. Since ν is unbounded we first apply an interval search to find the bounds for ν.

Data: Initialize P, upper limit K, initial interval $[L, U]$;
Preset $B_{\text{old}} = B_{\text{new}} = U$, evaluate lle $z_{\text{old}} := -\ln \hat{c}_{B_{\text{new}},\Sigma}(\mathbf{u})$;
while $abs(B) <= K$ **do**
  Update $B_{\text{new}} = 2B_{\text{new}}$, evaluate lle $z_{\text{new}} := -\ln \hat{c}_{B_{\text{new}},\Sigma}(\mathbf{u})$;
  if $z_{new} > z_{old}$ **then**
    Update upper bound $U = B_{\text{new}}$;
    break;
  else
    Update lower bound $L = B_{\text{old}}$;
  end
  Update $B_{\text{old}} = B_{\text{new}}$ and $z_{\text{old}} = z_{\text{new}}$;
end

Algorithm 1: interval search

Once we set the search interval $[L, U]$, we employee a bounded minimization algorithm [2, 3]. In each iteration, we update the following parameters.

- x: the current minimizer, or the most recent evaluation,
- w: the next lowest value of objective,
- v: previous value of w,
- u: the last evaluation point,
- a and b: the feasible interval and m the midpoint.

Data: Initialize $v = w = x = \frac{\sqrt{5}-1}{2}L + \frac{3-\sqrt{5}}{2}U$ and $m = \frac{L+U}{2}$;
  Update errors $e = 0$, $e_1 = \tau_1|x| + \tau_2$ and $e_2 = 2e_1$;
while $|x - m| + \frac{L-U}{2} > e_2$ **do**
  if $|e| > e_1$ **then**
    Apply a parabolic interpolation;
  end
  if *golden section needed* **then**
    Define e to be the largest distance between x and bounds and $d = c$;
  end
  Update parameters
end

Algorithm 2: bounded optimization

Once ν is fixed, $\Sigma$is simply updated by $\Sigma = \frac{\nu+d}{n\nu}\mathbf{s}\mathbf{s}^T$ for $s_i = \frac{w_i}{1+\frac{1}{\nu}\sum_j w_j^2}$ with $L\mathbf{w} = \mathbf{u}$.

The algorithm above has the following convergence result.

Proposition 3.1. *If the function* $\log L(\nu, P; \mathbf{x})$ *is* C^2 *near an minimum* ν *with* $f''(\nu) > 0$*, then the following algorithm has superlinear convergence.*

4 Archimedean Copula

Definition 4.1. *An Archimedean generator is a nonincreasing, continuous function* $\psi : [0,\infty] \to [0,1]$ *satisfying* $\psi(0) = 1$, $\psi(\infty) = 0$ *and strictly decreasing on* $[0, \inf_t\{\psi(t) = 0\}]$. *A corresponding Archimedean copula is given by* $C(\mathbf{u}) = \psi(\sum_{i=1}^d \psi^{-1}(u_i))$ *for* $\mathbf{u} \in [0,1]^d$.

In our micro service, we support five types of Archimedean copulas, with the generator and copula function given below.

Name	Generator $\psi_\theta(t)$	Copula $C_\theta(u,v)$
AMH	$\frac{1-\theta}{e^t-\theta}$	$\frac{uv}{1-\theta(1-u)(1-v)}$
Clayton	$(1+t\theta)^{-\frac{1}{\theta}}$	$\max\{u^{-\theta}+v^{-\theta}-1,0\}^{-\frac{1}{\theta}}$
Frank	$-\frac{1}{\theta}\log(1+e^{-t}(e^{-\theta}-1))$	$-\frac{1}{\theta}\log(1+\frac{(e^{-u\theta}-1)(e^{-v\theta}-1)}{e^{-\theta}-1})$
Gumbel	$\exp(-t^{\frac{1}{\theta}})$	$\exp(-((-\log u)^\theta+(-\log v)^\theta)^{\frac{1}{\theta}})$
Joe	$1-(1-e^{-t})^{\frac{1}{\theta}}$	$1-\left[(1-u)^\theta+(1-v)^\theta+(1-u)^\theta(1-v)^\theta\right]^{\frac{1}{\theta}}$

Table 1: Archimedean copulas and generators

Here we use the subscription θ to emphasize the impact of parameter of the generator and the distribution function, in the sense that the calibration process to find the optimal choice of θ so that the likelihood function is maximized. To better describe the copula distribution function, we denote $t_\theta(\mathbf{u}) = \sum_{i=1}^d \psi_\theta^{-1}(u_i)$ so that $C_\theta(\mathbf{u}) = \psi(t_\theta(\mathbf{u}))$.

4.1 Calibration

The calibration framework is based on maximum likelihood. We brief the calculation for two dimension cases below.

For AMH copula, we have

$$c_\theta(u,v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u,v) = \frac{1-\theta}{[1-\theta(1-u)(1-v)]^2}, \quad -1 \le \theta \le 1.$$

For Clayton copula, we have

$$c_\theta(u,v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u,v) = (\theta+1)u^{-\theta-1}v^{-\theta-1}(u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}-2}, \quad \theta > 0.$$

For Frank copula, we have

$$c_\theta(u,v) = \frac{\theta(1-e^{-\theta})}{e^{\theta(u+v)\left[1-e^{-\theta}-(1-e^{-\theta u})(1-e^{-\theta v})\right]}}, \quad \theta \ne 0.$$

For Gumbel copula, we have

$$c_\theta(u,v) = -\ln C_\theta(u,v)(1-\theta-\ln C_\theta(u,v))\frac{\ln(uv)^{\theta-1}}{uv}, \quad \theta \ge 1.$$

For Joe copula, we have

$$c_\theta(u,v) = \tilde{u}^{\theta-1}\tilde{v}^{\theta-1}(\tilde{u}^\theta + \tilde{v}^\theta + \tilde{u}^\theta\tilde{v}^\theta)^{\frac{1}{\theta}-2}\left[(1+\tilde{u}^\theta)(1+\tilde{v}^\theta) - \theta\right], \quad \theta \geq 1,$$

where $\tilde{u} = 1-u$ and $\tilde{v} = 1-v$.

For higher dimensions, we refer to the result given in [4, 5]. The AMH copula has the density function

$$c_\theta(\mathbf{u}) = \frac{(1-\theta)^{d+1}}{\theta^2}\frac{h_\theta^A(\mathbf{u})}{\prod_{j=1}^d u_j^2}\mathrm{Li}_d(h_\theta^A(\mathbf{u})), \quad 0 \leq \theta < 1,$$

where $\mathrm{Li}_s(z) = \sum_{k\geq 1}\frac{z^k}{k^s}$ denotes the polylogarithm of order s at z and $h_\theta^A = \theta\prod_{j=1}^d \frac{u_j}{1-\theta(1-u_j)}$.

The Clayton copula has the density function

$$c_\theta(\mathbf{u}) = \prod_{k=0}^{d-1}(\theta k+1)(\prod_{j=1}^{d} u_j)^{-(1+\theta)}(1+t_\theta(\mathbf{u}))^{-d+\frac{1}{\theta}}, \quad \theta > 0.$$

The Frank copula has the density function

$$c_\theta(\mathbf{u}) = (\frac{\theta}{1-e^{-\theta}})^{d-1}\mathrm{Li}_{d-1}(h_\theta^F(\mathbf{u}))\frac{\exp\left(-\theta\sum_{j=1}^d u_j\right)}{h_\theta^F(\mathbf{u})}, \quad \theta > 0,$$

where $h_\theta^F = (1-e^{-\theta})^{1-d}\prod_{j=1}^d(1-e^{-\theta u_j})$.

The Gumbel copula has the density function

$$c_\theta(\mathbf{u}) = \theta^d \exp(-t_\theta(\mathbf{u})^\alpha)\frac{\prod_{j=1}^d(-\ln u_j)^{\theta-1}}{t_\theta(\mathbf{u})^d\prod_{j=1}^d u_j}P^G_{d,\frac{1}{\theta}}(t_\theta(\mathbf{u})^{\frac{1}{\theta}}), \quad \theta \geq 1,$$

where $P^G_{d,\alpha}(x) = \sum_{k=0}^{d-1} a^G_{d,k}(\alpha)x^k$ is the polyexpansion at x with the coefficient $a^G_{d,k}(\alpha) = (-1)^{d-k}\sum_{j=k}^d \alpha^j s(d,j)S(j,k)$, $s(i,j)$ and $S(i,j)$ the Stirling numbers of first and second kind, respectively.

The Joe copula has the density function

$$c_\theta(\mathbf{u}) = \theta^{d-1}\frac{\prod_{j=1}^d(1-u_j)^{\theta-1}}{(1-h_\theta^J(\mathbf{u}))^{1-\frac{1}{\theta}}}P^J_{d,\frac{1}{\theta}}(\frac{h_\theta^J(\mathbf{u})}{1-h_\theta^J(\mathbf{u})}), \quad \theta \geq 1,$$

where $P^J_{d,\alpha}(x) = \sum_{k=0}^{d-1} a^J_{d,k}(\alpha)x^k$ is the polyexpansion at x with the coefficient $a^J_{d,k}(\alpha) = \frac{\Gamma(k+1-\alpha)}{\Gamma(1-\alpha)}S(d,k+1)$ and $h_\theta^J = \prod_{j=1}^d(1-(1-u_j)^\theta)$.

We make several remarks below to describe the precise evaluation of the likelihood function in our implementation.

Remark. *The polylogarithm function is convergent on the unit disc* $\{z : |z| \leq 1\}$ *only if* $s > 1$*.This means that the AMH density function is valid only for* $\theta \in [0, 1)$ *for* $d > 2$*, and the Frank density function is not convergent for* $d = 2$ *when* θ *approaches* ∞*. For this reason the domain of* θ *differs for two and higher dimension cases, and we should separately implement for* $d = 2$ *and* $d \neq 2$*.*

Remark. *In our implementation, the polylogarithm function is evaluated by iteration until the stopping criterion* $\frac{z^k}{k^s} < \varepsilon$ *is met with* ε *a pre-defined error tolerance.*

Remark. *The Stirling numbers have the currence relation:*

$$\begin{aligned} s(n+1, k) &= s(n, k-1) - ns(n, k), \\ S(n+1, k) &= S(n, k-1) + kS(n, k), \end{aligned}$$

with the end cases $s(0, 0) = S(0, 0) = 1$ *and* $s(n, 0) = s(0, n) = S(n, 0) = S(n, 0) = 0$ *for* $n \geq 1$*. During our implementation, we pre-build the matrices in the copula classes based on the recurrence relation instead of the combinatoric representation due to efficiency consideration.*

4.2 Simulation

To simulate samples from an Archimedean copula, we use Marshall-Olkin algorithm:

1. Sample $\zeta \sim F = \mathcal{LS}^{-1}(\psi)$, where $\mathcal{LS}$ is the Laplace-Stieltjes transform;
2. Sample $\xi_i \sim U([0, 1])$ for $1 \leq i \leq d$;
3. Return $\mu_i = \psi(-\frac{\log \xi_i}{\zeta})$ for $1 \leq i \leq d$.

To reason the algorithm above, since $\xi_i = e^{-\zeta \psi^{-1}(\mu_i)}$, and ξ_i is independent, we know

$$\begin{aligned} \mathbb{P}\left[\mu_1 \leq u_1, \cdots, \mu_d \leq u_d\right] &= \mathbb{P}\left[\xi_1 \leq e^{-\zeta\psi^{-1}(u_1)}, \cdots, \xi_d \leq e^{-\zeta\psi^{-1}(u_d)}\right] \\ &= e^{-\zeta \sum_{j=1}^d \psi^{-1}(u_j)}. \end{aligned}$$

We also note that $\psi(x) = \mathbb{E}\left[e^{-\zeta x}\right]$ since $\zeta \sim \mathcal{LS}^{-1}(\psi)$. Accordingly, we have $C(\mathbf{u}) = \psi(\sum_{j=1}^d \psi^{-1}(u_i))$.

Now we focus on the problem how to generate numbers from F, and in particular, we are interested in AMH/Clayton/Frank/Gumbel/Joe type Archimedeans. For Clayton and Gumbel, the distribution F is in close form [4], with $\mathcal{LS}^{-1}(\psi) = \Gamma(\frac{1}{\theta}, 1)$ being the gamma disribution $\Gamma(\alpha, \beta)$ for

Clayton and $\mathcal{LS}^{-1}(\psi) = S(\frac{1}{\theta}, 1, (\cos\frac{\pi}{2\theta})^{\theta}, 0; 1)$ being the α-stable disribution $S(\alpha, \beta, \gamma, \delta; 1)$ of parametrization 1 for Gumbel. To sample from gamma distribution is straightforward for Clayton, however, to sample from the stable distribution $\zeta \sim S(\alpha, \beta, \gamma, \delta; 1)$, we first sample from standardized stable distribution of parametrization 0, i.e., $\vartheta \sim S(\alpha, \beta; 0)$, which can be achieved by the following [7]

- Sample $\upsilon \sim U(\left[-\frac{\pi}{2}, \frac{\pi}{2}\right])$ and $\omega \sim \mathrm{Exp}(1)$, i.e., the exponential distribution with $\lambda = 1$;

- Transform using

$$\vartheta = \begin{cases} (1+\varsigma^2)^{\frac{1}{2\alpha}} \frac{\sin(\alpha(\nu+\eta))}{(\cos\nu)^{\frac{1}{\alpha}}} \left(\frac{\cos(\nu-\alpha(\nu+\eta))}{\omega}\right)^{\frac{1-\alpha}{\alpha}}, & \alpha \neq 1, \\ \frac{1}{\eta}\left[(\frac{\pi}{2}+\beta\nu)\tan\nu - \beta\ln(\frac{\frac{\pi}{2}\omega\cos\nu}{\frac{\pi}{2}+\beta\nu})\right], & \alpha = 1, \end{cases}$$

  where $\varsigma = -\beta\tan\frac{\alpha\pi}{2}$ and $\eta = \begin{cases} \frac{1}{\alpha}\arctan(-\varsigma), & \alpha \neq 1, \\ \frac{\pi}{2}, & \alpha = 1. \end{cases}$

Once ϑ is generated we can sample ζ by applying the transformation [8]

$$\zeta = \begin{cases} \gamma\vartheta + \delta + \beta\gamma\tan\frac{\alpha\pi}{2}, & \alpha \neq 1, \\ \gamma\vartheta + \delta + \beta\gamma\frac{2}{\pi}\ln\gamma, & \alpha = 1. \end{cases}$$

Note that in the real calculation, whenθ is close to 1, ζ can be to small and causing overflow in the sampling setting. Due to this concern, we will generate a sample from uniform distribution when $\theta < 1 + \varepsilon$.

On the other hand, for AMH/Frank/Joe, F is a discrete distribution represented by a series of functions, and we need the following result.

Proposition 4.1. *Given an Archimedean generator ψ with $F = \mathcal{LS}^{-1}(\psi)$ and $G = \sum_{k\geq 1} y_k \mathbb{1}_{[x_k,\infty)}$ with $\sum_{k\geq 1} y_k = 1$ and $y_k > 0$, then $F = G$ in the sense of L^∞ if and only if $\psi(t) = \sum_{k\geq 1} y_k e^{-x_k t}$ for $t \in [0, \infty]$.*

Proof. To see the only if part, we implement Laplace Stieltjes transform on F and get

$$\mathcal{LS}(F) = \mathcal{LS}(G) = \int e^{-tx} \sum_{k\geq 1} y_k \delta_{x_k}(x) dx = \sum_{k\geq 1} y_k e^{-x_k t}.$$

The if part follows directly from the uniqueness theorem of inverse Laplace-Stieltjes theorem. ∎

With this theorem, we know that the discrete distribution has a constant probability $\sum_{1\leq i\leq k} y_i$ supported on $[x_k, x_{k+1})$.

Now consider the specific Archimedeans. For AMH copula, we use geometric expansion on

$$\begin{aligned}\psi(t) &= \frac{1-\theta}{e^t-\theta} = (1-\theta)e^{-t}\frac{1}{1-\theta e^{-t}} \\ &= (1-\theta)e^{-t}\sum_{k\geq 0}\theta^k e^{-kt} = (1-\theta)\sum_{k\geq 1}\theta^{k-1}e^{-kt}. \qquad (11)\end{aligned}$$

Thus, we have $x_k = k$ and $y_k = (1-\theta)\theta^{k-1}$.

For Frank copula, we use Taylor expansion and get

$$\begin{aligned}\psi(t) &= -\frac{1}{\theta}\log(1-(1-e^{-\theta})e^{-t}) \\ &= \frac{1}{\theta}\sum_{k\geq 1}\frac{1}{k}(1-e^{-\theta})^k e^{-kt} \qquad (12)\end{aligned}$$

Thus, we have $x_k = k$ and $y_k = \frac{(1-e^{-\theta})^k}{k\theta}$.

For Joe copula, we use binomial expansion on

$$\begin{aligned}\psi(t) &= 1-(1-e^{-t})^{\frac{1}{\theta}} = 1-\sum_{k\geq 0}(-1)^k\binom{\frac{1}{\theta}}{k}e^{-kt} \\ &= \sum_{k\geq 1}(-1)^{k+1}\binom{\frac{1}{\theta}}{k}e^{-kt}. \qquad (13)\end{aligned}$$

Thus, we have $x_k = k$ and $y_k = (-1)^{k+1}\binom{\frac{1}{\theta}}{k}$.

Using the fact that $\sum_{k\geq 1} y_k \mathbb{1}_{[x_k,\infty)} = \sum_{k\geq 1} p_k \mathbb{1}_{[x_k,x_{k+1})}$ for $p_k = \sum_{i\leq k} y_i$, we can formulate a finite approximation to the distribution when k satisfies $1-\sum_{i\leq k} y_k < \varepsilon$ with given error tolerance $\varepsilon > 0$ sufficiently small.

We discuss the sampling performance of a given distribution generated by Marshall-Olkin. The property of the Archimedean generator assumes that $\lim_{t\to 0}\psi(t) = 1$. Since $t = -\frac{\log\xi}{\zeta}$, so either ζ is sufficiently large or ξ is sufficiently close to 1, we will generate numbers that are close to 1. For AMH copula, the infinite sum condition gives $k = \lceil\frac{\log\varepsilon}{\log\theta}\rceil$, where $\lceil\cdot\rceil$ denotes the ceiling operator. And $p_k = 1-\theta^k$ implies that p_k converges exponentially, aka. x_k will have a small support when $|\theta|$ close to 0. On the other hand, when $|\theta|$ is close to 1, x_k will have a larger support with significantly increasing p_k, and thus, the probability for a large ζ sample is significantly higher, leading to a simulation that is highly concentrated on the upper tail. For Joe copula, the convergence of infinite sum is only polynomial, and thus, it will have a much larger support, especially in the case when θ is close to 1, leading to the simulated distribution highly concentrated on the upper tail.

For Frank copula, note that the domain of θ is different between $d = 2$ $(\theta \neq 0)$ and $d \geq 3$ $(\theta > 0)$. Moreover, the Taylor expansion is not convergent in the case of $\theta < 0$. In fact, it is convergent if and only if $-1 < 1 - e^{-\theta} \leq 1$. Thus, we use an alternative algorithm [12] in the case for $d = 2$ and $\theta < 0$.

- Sample $\xi, \eta \sim U([0,1])$;
- Return $\mu_1 = \xi$ and $\mu_2 = -\frac{1}{\theta} \ln(1 + \frac{\eta(1-\exp(-\theta))}{\eta(\exp(-\theta\xi)-1)-\exp(\theta\xi)})$.

References

[1] A. W. Bowman and A. Azzalini. *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, volume 18. OUP Oxford, 1997.

[2] R. P. Brent. *Algorithms for minimization without derivatives.* Courier Corporation, 2013.

[3] G. E. Forsythe, M. A. Malcolm, and C. B. Moler. *Computer methods for mathematical computations*, volume 259. Prentice-hall Englewood Cliffs, NJ, 1977.

[4] M. Hofert. Efficiently sampling nested archimedean copulas. *Computational Statistics & Data Analysis*, 55(1):57–70, 2011.

[5] M. Hofert, M. Mächler, and A. J. Mcneil. Likelihood inference for archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 110:133–150, 2012.

[6] D. X. Li. On default correlation: A copula function approach. *Available at SSRN 187289*, 1999.

[7] A. Misiorek and R. Weron. Heavy-tailed distributions in var calculations. In *Handbook of Computational Statistics*, pages 1025–1059. Springer, 2012.

[8] J. Nolan. *Stable distributions: models for heavy-tailed data.* Birkhauser New York, 2003.

[9] D. W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.

[10] S. J. Sheather. Density estimation. *Statistical science*, pages 588–597, 2004.

[11] B. W. Silverman. *Density estimation for statistics and data analysis.* Routledge, 2018.

[12] F. Wu, E. Valdez, and M. Sherris. Simulating from exchangeable archimedean copulas. *Communications in Statistics—Simulation and Computation®*, 36(5):1019–1034, 2007.