Multivariate Data Analysis

1. Introduction

This project is based on Chapter 4, Multivariate Distributions, from Applied Multivariate Statistical Analysis (Sixth Edition) by Härdle, Simar and Fengler. Statistical analysis often requires tools to effectively handle and interpret multivariate data. In this project, we will explore the essential probability tools and techniques for understanding and analyzing multivariate data, such as the distributions, transformations and sampling methods.

2. Distribution and Density Function

A random vector $X = \left(X_1, X_2, \ldots, X_p\right)^T$ has a cumulative distribution function (cdf) defined as $F(x) = P(X \leq x) = P\left(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p\right)$. For continuous X, a nonnegative probability density function (pdf) f exists and satisfies $F(x) = \int_{-\infty}^{x} f(u)du$, where $\int_{-\infty}^{\infty} f(u)du = 1$. Note, the integral $\int_{-\infty}^{x} f(u)du$ refers to $\int_{-\infty}^{x_p} \ldots \int_{-\infty}^{x_1} f\left(u_1, \ldots, u_p\right)du_1 \ldots du_p$. For discrete X, the probability of events $\{X \in D\}$ is computed as $P(X \in D) = \sum_{j:c_j \epsilon D} P\left(X = c_j\right)$.

Let $X = (X_1, X_2)^T$ be partitioned into $X_1 \in \mathbb{R}^k$, $X_2 \in \mathbb{R}^{p-k}$ with joint cdf F, the marginal cdf of X_1 is $F_{X_1}(x_1) = P(X_1 \leq x_1) = F(x_{11}, \ldots, x_{1k}, \infty, \ldots, \infty)$, and the marginal pdf of X_1 can be obtained by $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2$. Different joint pdfs may have the same marginal pdfs. The conditional pdf of X_2 given $X_1 = x_1$ is given as $f(x_2|x_1) = \frac{f(x_1,x_2)}{f_{X_1}(x_1)}$.

X_1 and X_2 are independent if $f(x) = f(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ or equivalently, the conditional pdfs are equal to the marginal densities, i.e., $f(x_1|x_2) = f_{X_1}(x_1)$ and $f(x_2|x_1) = f_{X_2}(x_2)$. Independence means that knowing $X_2 = x_2$ does not change the probability assessments of X_1 and conversely.

Copula is a function which connects marginals to form joint cdfs. By Skylar's Theorem, any joint distribution F, with marginal distribution functions F_{X_1} and F_{X_2}, have a copula C such that $F(x_1, x_2) = C\left\{F_{X_1}(x_1), F_{X_2}(x_2)\right\}$ for every $x_1, x_2 \epsilon \mathbb{R}$. . If F_{X_1} and F_{X_2} are continuous, then C is unique, and furthermore X_1 and X_2 are independent if and only if $C_{X_1,X_2} = \Pi$, where $\Pi(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i$. Conversely, if C is a copula and F_{X_1} and F_{X_2} are distribution functions, then the function $F(x_1, x_2)$ as defined is a joint distribution function with marginals F_{X_1} and F_{X_2}.

3. Moments and Characteristic Functions

Expectation and Covariance Matrix

If X is a random vector with density $f(x)$, then the expectation of X is defined component-wise as

$$EX = \begin{pmatrix} EX_1 \\ \vdots \\ EX_p \end{pmatrix} = \int xf(x)dx = \begin{pmatrix} \int x_1 f(x)dx \\ \vdots \\ \int x_2 f(x)dx \end{pmatrix} = \mu,$$

Expectations are linear, i.e., $E(\alpha X + \beta Y) = \alpha EX + \beta EY$, and $E(AX) = AEX$ when applied matrix transformation with $A \in \mathbb{R}^{q \times p}$. For independent X and Y, $E(XY^T) = EXEY^T$. First-order moment of X is $E(X) = \mu$ and Second-order moment of X is $E(XX^T) = \left\{E\left(X_i X_j\right)\right\}$, for $i, j = 1, \ldots, p$.

The covariance matrix of X is $Var(X) = \Sigma = E(X - \mu)(X - \mu)^T$, and the cross-covariance matrix of $X \sim (\mu, \Sigma_{XX})$ and $Y \sim (v, \Sigma_{YY})$ is $\Sigma_{XY} = Cov(X, Y) = E(X - \mu)(Y - v)^T = E(XY^T) - EXEY^T$.

$Cov(X,Y) = 0$ when X and Y are independent. Covariance Σ is symmetric, i.e., $\Sigma_{XY} = \Sigma_{YX}{}^T$ and is positive semi-definite, i.e., $\Sigma \geq 0$. The diagonal entries of Σ is $\sigma_{X_iX_i} = Var(X_i)$, and the off-diagonal entries are $\sigma_{X_iX_j} = Cov(X_i, X_j)$. When $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$, the covariance $\Sigma_{ZZ} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}$.

Variances satisfy $Var(a^T X) = a^T Var(X) a$, and $Var(AX + b) = AVar(X)A^T$. Covariances also satisfy $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$ and $Cov(AX, BY) = ACov(X, Y)B^T$. We have $Var(X + Y) = Var(X) + Cov(X, Y) + Cov(Y, X) + Var(Y)$.

Conditional Expectations

Conditional expectation $E(X_2|X_1 = x_1) = \int x_2 f(x_2|x_1)dx_2$ represents the location parameter of the conditional pdf of X_2 given $X_1 = x_1$. Similarly, conditional variance $Var(X_2|X_1 = x_1)$ measures the dispersion of X_2 given $X_1 = x_1$, calculated as $E\left(X_2 X_2{}^T \middle| X_1 = x_1\right) - E(X_2|X_1)E\left(X_2{}^T \middle| X_1 = x_1\right)$. Conditional correlations (or partial correlations) quantify the relationship between components of X, conditioned on others, and value may differ from simple correlations depending on the conditioning. For example, $\rho_{X_2X_3|X_1=x_1} = \frac{Cov(X_2, X_3|X_1 = x_1)}{\sqrt{Var(X_2|X_1 = x_1)Var(X_3|X_1 = x_1)}}$.

Key properties include $E(X_2) = E\{E(X_2|X_1)\}$ and $Var(X_2) = \{Var(X_2|X_1)\} + Var\{E(X_2|X_1)\}$. $E(X_2|X_1)$ can be viewed as a regression function of $h(X_1)$, providing a conditional approximation of X_2 by a function of X_1 with an error term $U = X_2 - E(X_2|X_1)$. For $X_1 \in \mathbb{R}^k$, $X_2 \in \mathbb{R}^{p-k}$ and the error term U, $E(U) = 0$ and $E(X_2|X_1)$ is the best approximation of X_2 by a function $h(X_1)$, minimizing the mean squared error $MSE(h) = E[\{X_2 - h(X_1)\}^T\{X_2 - h(X_1)\}]$.

Characteristic Functions (cf)

Characteristic function of a random vector $X \in \mathbb{R}^p$ is $\varphi_X(t) = E\left(e^{it^TX}\right) = \int e^{it^TX} f(x)dx$, $t \in \mathbb{R}^p$ where $\boldsymbol{i^2} = -1$ is the complex unit. Key properties include $\varphi_X(0) = 1$ and $|\varphi_X(t)| \leq 1$. For $X \in \mathbb{R}^p$ and $t \in \mathbb{R}^p$, the marginal cf of X_j is $\varphi_{X_j}(t_j) = \varphi_X(0, \dots, t_j, \dots, 0)$. If $X_1, \dots, X_p$ are independent, $\varphi_X(t) = \varphi_{X_1}(t_1) \cdot \dots \cdot \varphi_{X_p}(t_p)$ for $t \in \mathbb{R}^p$ and if $t \in \mathbb{R}$, $\varphi_{X_1+\dots+X_p}(t) = \varphi_{X_1}(t) \cdot \dots \cdot \varphi_{X_p}(t)$.

If φ is absolutely integrable i.e., $\int_{-\infty}^{\infty} |\varphi_X(t)|\, dx < \infty$, then $f(x) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} e^{it^Tx} \varphi_X(t)\, dt$, which the pdf can be recovered. The cf can also recover all the cross-product moments of any order, i.e., $\forall j_k \geq 0, k = 1, \dots, p$ and for $t = (t_1, t_2, \dots, t_p)^T$ we have $E\left(X_1^{j_1} \cdot \dots \cdot X_p^{j_p}\right) = \frac{1}{i^{j_1+\dots+j_p}} \left[\frac{\partial \varphi_X(t)}{\partial t_1^{j_1} \dots \partial t_p^{j_p}}\right]_{t-0}$.
A variety of distributional characteristics can be computed from $\varphi_X(t)$. Deviations from normal covariance structures can be identified through cf deviations. Table 1 gives an overview of the cfs for a variety of distributions. By Cramer-Wold Theorem, the distribution of $X \in \mathbb{R}^p$ is completely determined by linear combination of all one-dimensional distributions of t^TX where $t \in \mathbb{R}^p$, through $\sum_{j=1}^{p} t_j X_j = t^T X$.

Table 1: Characteristic functions for some common distributions.

	pdf	cf
Uniform	$f(x) = \mathrm{I}(x \in [a, b])/(b - a)$	$\varphi_X(t) = (e^{\mathbf{i}bt} - e^{\mathbf{i}at})/(b - a)\mathbf{i}t$
$N_1(\mu, \sigma^2)$	$f(x) = (2\pi\sigma^2)^{-1/2}\exp\{-(x - \mu)^2/2\sigma^2\}$	$\varphi_X(t) = e^{\mathbf{i}\mu t - \sigma^2 t^2/2}$
$\chi^2(n)$	$f(x) = \mathrm{I}(x > 0)x^{n/2-1}e^{-x/2}/\{\Gamma(n/2)2^{n/2}\}$	$\varphi_X(t) = (1 - 2\mathbf{i}t)^{-n/2}$
$N_p(\mu, \Sigma)$	$f(x) = \|2\pi\Sigma\|^{-1/2}\exp\{-(x - \mu)^\top\Sigma(x - \mu)/2\}$	$\varphi_X(t) = e^{\mathbf{i}t^\top\mu - t^\top\Sigma t/2}$

Cumulant Functions

Moments, $m_k = \int x^k f(x)dx$, describes key distributional characteristics. The normal distribution in one dimension is completely characterized by its standard normal density $f = \varphi$, and the moment parameters are $\mu = m_1$ and $\sigma^2 = m_2 - {m_1}^2$. Cumulants (or semi-invariants), κ_j, does not change (for $j > 1$) under a shift transformation $(X \mapsto X + a)$, making it a natural parameter in dimension reduction methods. For a one-dimensional random variable X with density f and finite moments of order k, cumulants are derived from the cf as $\kappa_j = \frac{1}{i^j}\left[\frac{\partial^j \log\{\varphi_X(t)\}}{\partial t^j}\right]_{t=0}, j = 1, \ldots, k$. The relationship between the first k moments $m_1, \ldots, m_k$ and the cumulants are given by

$$\kappa_k = (-1)^{k-1}\begin{vmatrix} m_1 & 1 & \cdots & 0 \\ m_2 & \binom{1}{0}m_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ m_k & \binom{k-1}{0}m_{k-1} & \cdots & \binom{k-1}{k-2}m_1 \end{vmatrix}.$$

A relationship can be observed between the cumulants and the central moments, $\mu_k = E(X-\mu)^k$, where the first cumulant is the mean, and the second is the variance. The third and fourth cumulants relate to the skewness and kurtosis, which characterize the shape of a distribution.

4. Transformations

To find the pdf of a transformed random variable Y from X, we use the principle of a one-to-one transformations, $u: \mathbb{R}^p \to \mathbb{R}^p$. If X has pdf $f_X(x)$, then a transformed random vector Y, i.e., $X = u(Y)$, has pdf $f_Y(y) = abs(|J|) \cdot f_X\{u(y)\}$. J denotes the Jacobian $J = \left(\frac{\delta x_i}{\delta y_j}\right) = \left(\frac{\delta u_i(y)}{\delta y_j}\right)$ and $abs(|J|)$ is the absolute value of the determinant of this Jacobian.

For linear transformation where $Y = AX + b$ and A is non-singular, the inverse transformation is $X = A^{-1}(Y - b)$. Therefore, $J = A^{-1}$, and hence the pdf of X and Y are related via $f_Y(y) = abs(|A|^{-1}) \cdot f_X\{A^{-1}(Y-b)\}$.

5. The Multinormal Distribution

$f(x) = \left|2\pi\Sigma\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$ is the density of a p-dimensional multinormal distribution $N_p(\mu, \Sigma > 0)$. The random variable $X \sim N_p(\mu, \Sigma)$ can be transformed into a multivariate standard normal variable $Y \sim N_p(0, I_p)$ using Mahalanobis transformation $Y = \Sigma^{-\frac{1}{2}}(X - \mu)$, where the components of Y are independent $N(0,1)$ variables. Conversely, the inverse transformation $X = \Sigma^{\frac{1}{2}}Y + \mu$ transforms $Y \sim N_p(0, I_p)$ to $X \sim N_p(\mu, \Sigma)$. For linear transformation with $Y = AX + c$, where $c \in \mathbb{R}^p$, $X \sim N_p(\mu, \Sigma)$ and A_{pxp} is non-singular, then $Y \sim N_p(A\mu + c, A\Sigma A^T)$.

The density of $N_p(\mu, \Sigma)$ is constant on ellipsoids defined by $(x-\mu)^T\Sigma^{-1}(x-\mu) = d^2$. The half-length of the axes in the contour ellipsoid are $\sqrt{d^2/\lambda_i}$ where λ_i are the eigenvalues of Σ. If Σ is a diagonal, the rectangle circumscribing the contour ellipse has sides with length $2d\sigma_i$, and is naturally proportional to the standard deviations of X_i.

If $X \sim N_p(\mu, \Sigma)$, the variable $U = (X-\mu)^T\Sigma^{-1}(X-\mu)$ has χ_p^2 distribution. The cf of $X \sim N_p(\mu, \Sigma)$ is given by $\varphi_X(t) = \exp\left(\boldsymbol{i}\boldsymbol{t}^T\mu - \frac{1}{2}t^T\Sigma t\right)$. If $Y \sim N_p(0, I_p)$, then $\varphi_Y(t) = \varphi_{Y_1}(t_1) \cdot \ldots \cdot \varphi_{Y_p}(t_p)$.

When the covariance matrix Σ of $X \sim N_p(\mu, \Sigma)$ is singular (i.e., $rank(\Sigma) < p$), it defines a singular normal distribution with density $f(x) = \frac{(2\pi)^{-k/2}}{(\lambda_1 \dots \lambda_k)^{1/2}} \exp\{-\frac{1}{2}(x-\mu)^T \Sigma^- (x-\mu)$, where x lies on the hyperplane $N^T(x-\mu) = 0$. Here, $N \in \mathbb{R}^{p \times (p-k)}$ satisfies $N^T \Sigma = 0$ and $N^T N = I_k$. The matrix Σ^- is the G-inverse of Σ, and $\lambda_1, \dots, \lambda_k$ are the non-zero eigenvalues of Σ. If $Y \sim N_k(0, \Lambda_1)$ with $\Lambda_1 = diag(\lambda_1, \dots, \lambda_k)$, there exists an orthogonal matrix $B \in \mathbb{R}^{p \times k}$ with $B^T B = I_k$, such that $X = BY + \mu$, linking the singular distribution of X to a k-dimensional multivariate normal distribution.

The Gaussian (normal) copula is defined as $C_\rho(u, v) = \int_{-\infty}^{\Phi_1^{-1}(u)} \int_{-\infty}^{\Phi_2^{-1}(v)} f_\rho(x_1, x_2) dx_2 dx_1$, where f_ρ denotes the bivariate normal density function with correlation ρ for $n = 2$. The functions Φ_1 and Φ_2 are the standard normal cdfs of the marginals. If $\rho = 0$, the Gaussian copula becomes $C_0(u, v) = uv = \prod(u, v)$, which corresponds to the independent copula.

6. Sampling Distributions and Limit Theorems

A multivariate random variable X is observed, and under random sampling, the sample $\{x_i\}_{i=1}^n$ consists of realizations of iid random variables $X_1, \dots, X_n$, where each X_i is a p-variate random variable that replicates the population random variable X, and x_i is the ith observation in the sample. The purpose of statistical inference is to analyse characteristics such as mean and covariance matrix of the population variable X, typically using statistics as observable functions of the sample data. Sampling distribution of the statistics is derived to understand its relationship to the corresponding population parameter. If $X_1, \dots, X_n$ are iid with $X_i \sim N_p(\mu, \Sigma)$, then $\bar{x} \sim N_p\left(\mu, \frac{1}{n}\Sigma\right)$.

Sampling distributions in multivariate statistics can be difficult to derive, so approximations provided by limit theorems are often used. The Central Limit Theorem (CLT) states that When the sample size is large, even if the parent distribution is not normal, the sampling distribution approximates normal distribution. Specifically, if $X_1, \dots, X_n$ are iid with $X_i \sim (\mu, \Sigma)$, then the distribution of $\sqrt{n}(\bar{x} - \mu)$ is asymptotically $N_p(0, \Sigma)$ as $n \to \infty$.

The asymptotic normal distribution allows the construction of confidence intervals for unknown parameters. Confidence interval at level $1 - \alpha, \alpha \in (0,1)$, is an interval that covers the true parameter with probability $1 - \alpha$: $P\left(\theta \in \left[\hat{\theta}_l, \hat{\theta}_u\right]\right) = 1 - \alpha$, where θ is the unknown parameter and $\hat{\theta}_l$ and $\hat{\theta}_u$ are the lower and upper bounds, respectively. If $X_1, \dots, X_n$ are iid random variables with $X_i \sim (\mu, \sigma^2)$, and σ^2 is known, the CLT implies that $P\left(-u_{1-\frac{\alpha}{2}} \le \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \le u_{1-\frac{\alpha}{2}}\right) \to 1 - \alpha$, as $n \to \infty$, where $u_{1-\alpha/2}$ denotes the $(1 - \alpha/2)$-quantile of the standard normal distribution. Hence, the approximate $(1 - \alpha)$-confidence interval for μ is $\bar{x} \pm \frac{\hat{\sigma}}{\sqrt{n}} u_{1-\alpha/2}$. CLT will still hold if the sample variance $\Sigma\hat{}$, a consistent estimate for Σ, is used

If t is a statistic that is asymptotically normal, i.e., $\sqrt{n}(t - \mu) \to N_p(0, \Sigma)$, and $f = \left(f_1, \dots, f_q\right)^T$: $\mathbb{R}^p \to \mathbb{R}^q$ is differentiable at $\mu \in \mathbb{R}^p$, then $f(t)$ is also asymptotically normal with mean $f(\mu)$ and covariance $D^T \Sigma D$ where $D = \left(\frac{\partial f_j}{\partial t_i}\right)(t)|_{t=\mu}$ is the $(p \times q)$ matrix of partial derivatives. , i.e., $\sqrt{n}\{f(t) - f(\mu)\} \to N_q(0, D^T \Sigma D)$ for $n \to \infty$.

7. Bootstrap

When sample sizes are too small for CLT to provide accurate critical value approximations, the bootstrap method is an alternative to estimate parameters and approximate the distribution of interest. Given observations $x_1, \dots, x_n$ from a sample $X_1, \dots, X_n$, the empirical distribution function (edf) F_n is estimated as $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \le x)$. The edf is a step function which is constant between neighbouring data points. A bootstrap sample $x_1^*, \dots, x_n^*$, is drawn randomly with replacement from

F_n, ensuring each observation has a selection probability of 1/n, and preserves $E_{F_n}(X_i^*) = \bar{x}$ and $Var_{F_n}(X_i^*) = \hat{\sigma}^2$.

By CLT, the bootstrap distribution of $\frac{\sqrt{n}(\bar{x}^* - \bar{x})}{\hat{\sigma}^*}$ converges asymptotically to $N(0,1)$, where $\bar{x}^*$ and $\hat{\sigma}^*$ are the mean and standard deviation of the bootstrap sample. However, quantiles such as $u_{1-\alpha/2}$ from N(0,1), may not be accurate for small samples if $\frac{\sqrt{n}(\bar{x} - \mu)}{\hat{\sigma}}$ diverges from then normal limit. Thus, to address this, the bootstrap simulates many samples to estimate an empirical $u_{1-\alpha/2}^*$. The bootstrap confidence interval is constructed as $C_{1-\alpha}^* = \bar{x} \pm \frac{\hat{\sigma}}{\sqrt{n}} u_{1-\alpha/2}^*$, satisfying $P(\mu \in C_{1-\alpha}^*) \to 1 - \alpha$ as $n \to \infty$, with improved convergence rates (Hall 1992).

8. Heavy-Tailed Distributions

Heavy-tailed distribution has higher probability density in its tail compared to a normal distribution with the same mean and variance. This property makes them useful for modelling extreme events in various domains. In terms of kurtosis, heavy-tailed distributions are leptokurtic (kurtosis > 3), unlike mesokurtic distributions (kurtosis = 3) and platykurtic distributions (kurtosis < 3). The univariate properties lay the foundation for analyzing their multivariate counterparts and tail behaviours.

Generalized Hyperbolic Distribution

The Generalized Hyperbolic (GH) distribution, introduced by Barndorff-Nielson, is designed for modelling heavy-tailed data. Its log-density is hyperbolic, unlike the parabolic log-density of the normal distribution.

The pdf for $x \in \mathbb{R}$ is $f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \frac{\left(\sqrt{\alpha^2 - \beta^2}/\delta\right)^{\lambda}}{\sqrt{2\pi} K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \frac{K_{\lambda - 1/2}\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\}}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2 - \lambda}} e^{\beta(x-\mu)}$, where K_λ is a modified Bessel function of the third kind, and $\mu, \alpha, \beta, \delta, \lambda$ are parameters determining the location, scale and shape. The mean and variance for f_{GH} are $E(X) = \mu + \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}} \frac{K_{\lambda+1}(\delta\sqrt{\alpha^2 - \beta^2})}{K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})}$, and

$Var(X) = \delta^2 \left[\frac{K_{\lambda+1}(\delta\sqrt{\alpha^2 - \beta^2})}{\delta\sqrt{\alpha^2 - \beta^2} K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} + \frac{\beta^2}{\alpha^2 - \beta^2} \left[\frac{K_{\lambda+2}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} - \left\{\frac{K_{\lambda+1}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}{K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}\right\}^2\right]\right]$. With specific values of λ, we obtain different subclasses of GH. If $\lambda = 1$, GH simplifies to the Hyperbolic distribution (HYP) and if $\lambda = -1/2$, it simplifies to the Normal-Inverse Gaussian distribution (NIG).

The multivariate generalized hyperbolic distribution (GH_d) generalizes GH distribution to multiple dimension with location vector $\mu \in \mathbb{R}^d$, and covariance matrix $\Delta \in \mathbb{R}^{d \times d}$. Similarly, for $\lambda = \frac{d+1}{2}$, we obtain multivariate HYP distribution and for $\lambda = -1/2$, we obtain NIG distribution.

A second parameterizations (ζ, Π, Σ), introduced by Blæsild and Jensen (1981), can be used to find the mean and variance of $X \sim GH_d$. If partition X to X_1 and X_2 with r and k dimensions, respectively, then the distribution of X_1 is r-dimensional GH, and the conditional distribution of X_2 given X_1 is k-dimensional GH. If $Y = XA + B$ is a regular affine transformation of X, then Y is d-dimensional GH.

Student's T-Distribution

The Student's t-distribution, introduced by Gosset (1908), is followed by $X\sqrt{n}/Y$ with n degrees of freedom if $X \sim N(\mu, \Sigma)$ and $\frac{Y^2}{\sigma^2} \sim \chi_n^2$, and X and Y are independent. With n degrees of freedom, the pdf is $f_t(x; n) = \frac{\Gamma\left\{\frac{n+1}{2}\right\}}{\sqrt{n\pi}\Gamma(n/2)} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$, where Γ is the gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. The t-distribution ($n > 4$) has $\mu = 0, \sigma^2 = \frac{n}{n-2}$, $skewness = 0$ and $kurtosis = 3 + \frac{6}{n-4}$.

Student's t-distribution approaches the normal distribution as $n \to \infty$, $\lim_{n\to\infty} f_t(x;n) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. The t-distribution is widely used, but its flexibility of modelling is restricted because of the integer-valued tail index. In tail area, x is proportional to $|x|^{-(n+1)}$, hence t-distribution decays faster for higher n and have heavier tails for smaller n.

In multivariate case, if $X \sim N_p(\mu, \Sigma)$ and $Y \sim \chi_n^2$ are independent, and $X\sqrt{n/Y} = t - \mu$, then the pdf of t is given by $f_t(t; n, \Sigma, \mu) = \frac{\Gamma\{(n+p)/2\}}{\Gamma\left(\frac{n}{2}\right) n^{p/2} \pi^{p/2} |\Sigma|^{1/2} \{1 + \frac{1}{n}(t-\mu)^T \Sigma^{-1} (t-\mu)\}^{(n+p)/2}}$. The distribution of t is the non-central t-distribution with n degrees of freedom and the non-centrality parameter μ, Giri (1996).

Laplace Distribution

Laplace distribution (Laplace 1774), also known as double exponential distribution, is defined as the distribution of differences between two independent variates with identical exponential distributions. The pdf and cdf of a Laplace distribution with mean μ and scale parameter θ is $f_{Laplace}(x; \mu, \theta) = \frac{1}{2\theta} e^{-|x-\mu|/\theta}$ and the cdf is $F_{Laplace}(x; \mu, \theta) = \frac{1}{2}\left\{1 + sign(x - \mu)\left(1 - e^{-\frac{|x-\mu|}{\theta}}\right)\right\}$. The Laplace distribution has $\sigma^2 = 2\theta^2,\ skewness = 0, kurtosis = 6$. Standard Laplace distribution is obtained with mean 0 and $\theta = 1$.

Multivariate Laplace distribution can be generalized to higher dimension, with the pdf as $f_{MLaplace_d}(x; m, \Sigma) = \int_0^\infty g\left(z^{-\frac{1}{2}}x - z^{\frac{1}{2}}m\right) z^{-\frac{d}{2}} e^{-z} dz$, where g is pdf of $N_d(0, \Sigma)$, and the cdf as $F_{MLaplace_d}(x; m, \Sigma) = \int_0^\infty G\left(z^{-\frac{1}{2}}x - z^{\frac{1}{2}}m\right) e^{-z} dz$, where G is the cdf of $N_d(0, \Sigma)$. The pdf can also be described with $\lambda = (2 - d)/2$ and $K_\lambda(x)$, the modified Bessel function of the third kind. The multivariate Laplace distribution has mean and variance, $E(x) = m$ and $Cov(X) = \Sigma + mm^T$.

Cauchy Distribution

Cauchy distribution, with m and s as the location and scale parameter, respectively, has pdf and cdf of $f_{Cauchy}(x; m, s) = \frac{1}{s\pi} \frac{1}{1 + (\frac{x-m}{s})^2}$ and $F_{Cauchy}(x; m, s) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-m}{s}\right)$. Standard Cauchy distribution is obtained with $m = 0$ and $s = 1$. The mean, variance, skewness, and kurtosis of Cauchy distribution are all undefined, since its moment-generating function diverges. But it has mode and median both equal to the location parameter m.

Mixture Model

Mixture modelling models a statistical distribution as a weighted sum of different distributions. The mixture model can approximate any continuous density to arbitrary accuracy, provided that the number of component density functions is sufficiently large and the parameters of the model are chosen correctly. The pdf of a mixture distribution consists of L distributions is $f(x) = \sum_{l=1}^{L} w_l p_l(x)$ with the weights w_l to be $0 \le w_l \le 1$ and $\sum_{l=1}^{L} w_l = 1$, and for the pdf of the l'th component density $p_l(x)$ to be $\int p_l(x)dx = 1$. The mean, variance, skewness, and kurtosis of the mixture model are computed as the weighted averages of the respective moments of the individual component distributions. The most common approach involves mixture distribution having Gaussian components. Multivariate mixture model comprises multivariate distributions, e.g., pdf of a multivariate Gaussian distribution $f(x) = \sum_{l=1}^{L} \frac{w_l}{|2\pi\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x - \mu_l)^T \Sigma^{-1} (x - \mu_l)\}$.

Comparison

GH distributions have an exponential decaying speed. Among GH($\lambda = 0.5$), GH($\lambda = 1.5$), NIG and HYP, GH($\lambda = 1.5$) has the lowest decay, while NIG has the fastest decays,. Chen et al. (2008), compared NIG, Laplace, Cauchy and Gaussian distribution, all with mean and variance set to 0 and 1 respectively. Cauchy distribution has the lowest peak and fattest tails, followed by NIG distribution which decays quickly but has the highest peak,

9. Copulae

Correlation is suitable for measuring dependency only in elliptical or spherical distributions, such as the multivariate normal distribution. Hence, Copulae, introduced in earlier sections, are functions that link multivariate distributions functions to their one-dimensional marginals, allowing F_{X_i} to be modelled separately from their dependence structure and then coupled to form F_X, making them essential in quantitative finance (Härdle et al. 2009). We first consider the two-dimensional case and later extend to the d-dimensional case. Let U_1 and U_2 be two sets in $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ and $F: U_1 \times U_2 \to \overline{\mathbb{R}}$.

Conceptual Foundation

The F-volume of a rectangle $B = [x_1, x_2] \times [y_1, y_2] \subset U_1 \times U_2$ is given by $V_F(B) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$, which measures how the function F accumulates over B. F is a two-increasing function if $V_F(B) \geq 0$ for any rectangle B, ensuring F adheres to the monotonicity required for joint distributions. If $F(x, minU_2) = 0\ \forall x \epsilon U_1$, and $F(minU_1, y) = 0\ \forall y \epsilon U_2$, then F is grounded ensuring F starts from zero at the boundaries. F has margins given by $F(x) = F(x, maxU_2)\ \forall x \epsilon U_1$, and $F(y) = F(maxU_1, y) \forall y \epsilon U_2$. A cdf is a function from $\overline{\mathbb{R}}^2 \mapsto [0,1]$ which is grounded, two-increasing and satisfies $F(\infty, \infty) = 1$.

A 2-dimensional copula C is a multivariate distribution defined on the unit-square $[0,1]^2$ such that

- For every $u \in [0,1]$: $C(0, u) = C(u, 0) = 0$, i.e., C is grounded.
- For every $(u_1, u_2), (v_1, v_2) \epsilon [0,1] x [0,1]$ with $u_1 \leq v_1$ and $u_2 \leq v_2$: $C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0$, i.e., C is two-increasing
- For every $u \in [0,1]$: $C(u, 1) = C(1, u) = u$, i.e., uniform marginals.

Informally, a copula is a joint distribution function defined on the unit square $[0,1]^2$ which has uniform marginals. That means that if $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ are univariate distribution functions, then $C\{F_{X_1}(x_1), F_{X_2}(x_2)\}$ is a 2-dimensional distribution function with marginals $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$.

Key Theorems

In Section 2, Skylar Therorem was introduced to shows the relationship between copulas and joint cdf. The dependence structure described by copula is invariant under increasing and continuous transformations of the marginal distributions.If (X_1, X_2) have copula C, and g_1, g_2 are continuously increasing functions, then $(g_1\ (X_1), g_2(X_2))$ has the same copula C.

For copula C, $|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$, meaning every copula C is uniformly continuous on its domain. The partial derivatives of a copula, e.g., $\frac{\partial C(u.v)}{\partial v}$, exists for almost all $u \in [0,1] = I$. For such u and v, $\frac{\partial C(u.v)}{\partial v} \in I$. The partial derivates are defined and non-increasing almost everywhere on I.

Multivariate Copulas

Extending the concepts to the d-dimensional case, for a random variable in $\mathbb{R}^d$. Let $U_1, U_2, \ldots, U_d$ be non-empty sets in $\overline{\mathbb{R}}$ and $F: U_1 \times U_2 \ldots \times U_d \to \overline{\mathbb{R}}$. For $a, b \in \mathbb{R}^d$ with $a \leq b$, let $B = [a, b] = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_d, b_d] \subset U_1 \times U_2 \ldots \times U_d$ be the d-box with vertices $c \in \mathbb{R}^d$, where each c_k is either equal to a_k or b_k.

The F-volume of a d-box B is defined as $V_F(B) = \sum_{k=1}^{d} sign(c_k)F(c_k)$, where $signa(c_k) = 1$ if $c_k = a_k$ for even k and $signa(c_k) = -1$ if $c_k = a_k$ for odd k. F is d-increasing if all d-boxes B with vertices in $U_1 \times U_2 \ldots \times U_d$ holds $V_F(B) \geq 0$. F is grounded if $F(x) = 0$ for all $x \in U_1 \times U_2 \ldots \times U_d$ such that $x_k = minU_k$ for at least one k.

A d-dimensional copula (d-copula) is a function C defined on the unit d-cube I^d such that

- For every $u \in I^d$: $C(u) = 0$, if at least one coordinate of u is zero, i.e., C is grounded.
- For every $a, b \in I^d$ with $a \leq b$: $V_C([a, b]) \geq 0$, i.e., C is two-increasing
- For every $u \in I^d$: $C(u) = u_k$, if all coordinated of u are 1 except u_k , i.e., uniform marginals.

The Skylar's Theorem is also extended to the d-dimensional case.

Gaussian and Gumbel-Hougaard Copulas

The Gumbel-Hougaard copula reduces to the product copula, i.e., $C_\theta^{GH}(u, v) = uv$ at $\theta = 1$, and approaches the Fréchet-Hoeffding upper bound $M(u, v) = \min(u, v)$ as $\theta \to \infty$. The Fréchet-Hoeffding inequality states that for any copula, $W(u, v) = max(u + v - 1, 0) \leq C(u, v) \leq M(u, v)$. Gaussian copula, introduced in Section 5, and Gumbel-Hougaard copula can both be generalized to d-dimensional. The Gaussian copula can generate joint symmetric dependence, but is not possible to model a tail dependence, while Gumbel-Hougaard can generate an upper tail dependence.

10. Implementation

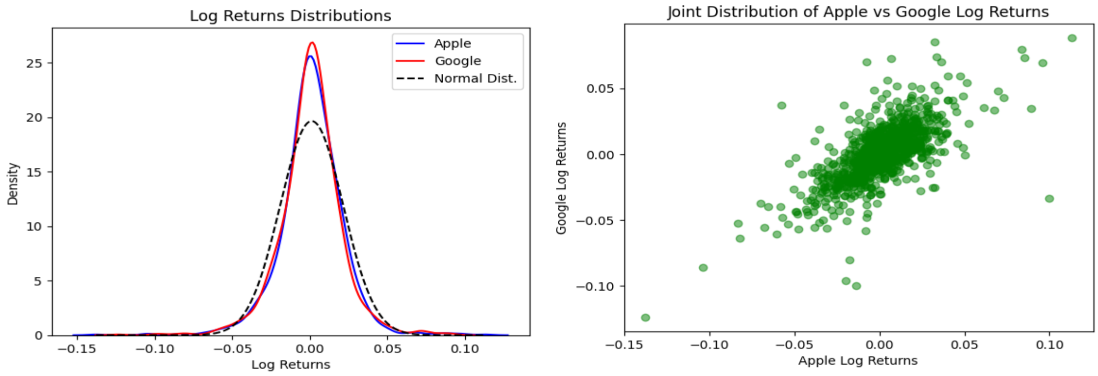

Fig. 1

We analyse the adjusted closing prices of Apple (AAPL) and Google (GOOGL) from January 1st 2019 to January 1st 2024, focusing on their daily log returns. The log return distributions of Apple and Google appear similar, as shown in Fig. 1 (left). Both distributions have mean close to 0 and standard deviation approximately 0.02, indicating no significant daily directional bias and have similar levels of volatility. The kurtosis is 5.4 for Apple and 4.1 for Google, both greater than 3, implying heavier tails and higher peaks compared to the normal distribution, as reflected in Fig. 1 (left).

Fig. 1 (right) shows the joint distribution of Apple and Google log returns, highlighting the dependence structure between the two assets, motivating the use of copula models to explore the relationship further, especially in the tails of the distribution where extreme events may occur.

We transform the log returns onto a uniform scale to fit Gaussian Copula and Gumbel-Hougaard Copula models. The Gaussian Copula has a fitted correlation parameter of 0.68, indicating moderate linear dependence between the log returns of Apple and Google, the upper tail dependence value is calculated as -0.83, reflecting weak upper tail dependence in extreme scenarios. The Gumbel Copula has fitted θ parameter of 1.96 which indicates the model captures strong upper tail dependence, where extreme positive events for a stock are more likely to coincide with similar events for the other.

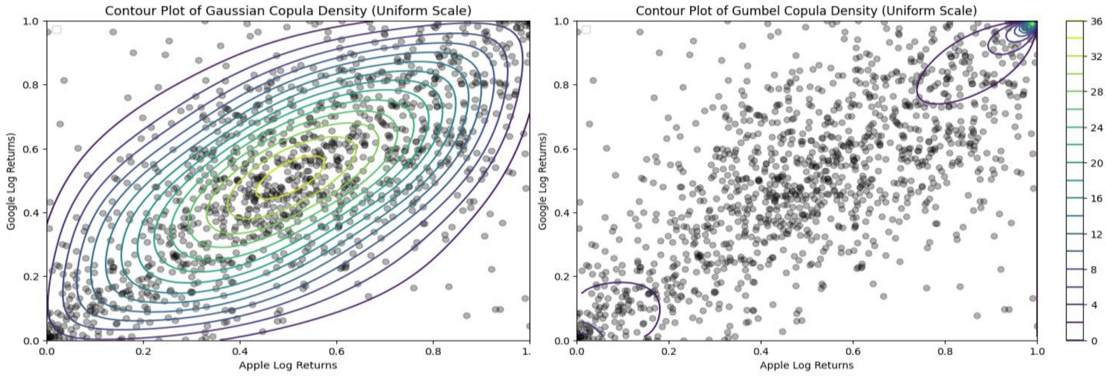

Fig. 2

Fig. 2 shows the contour plot of Gaussian Copula (left) and Gumbel Copula (right) on the uniform scaled-data. The Gaussian Copula shows elliptical contours, representing symmetrical dependence and limited tail behaviour. The Gumbel Copula contours are more focused at the corners, capturing tail dependence (extreme events) more effectively.

The models are evaluated using Akaike Information Criterion (AIC), where the AIC for Gaussian Copula and Gumbel Copula are 6087 and -705 respectively. The significantly lower AIC of the Gumbel Copula indicates a much better fit to the data, especially since the marginal distributions exhibit heavy tails. This highlights the Gumbel Copula's strength in modelling extreme events and tail dependence compared to the Gaussian Copula.

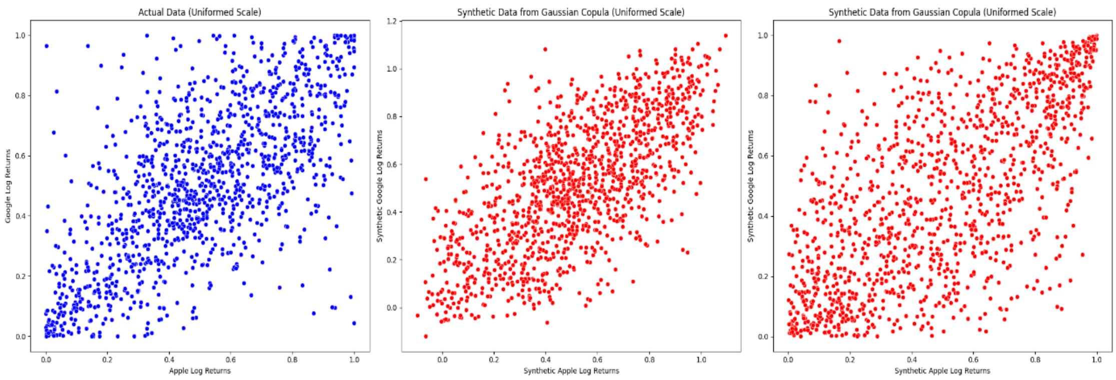

Fig. 3

In Fig, 3, synthetic data generated from the copula models is compared with the observed data (left), on a uniform scale. The Gumbel Copula (right) better replicates the characteristics of the actual data, especially in the tails, while the Gaussian Copula (middle) struggles to capture the extreme behaviour. This further confirms the Gumbel Copula's suitability for modelling heavy-tailed distributions and tail dependence in financial time series.

11. Conclusion

In this project, we have learned the fundamental tools in multivariate analysis, including various distributions, sampling methods, as well as copula. Gaussian Copula and Gumbel-Hougaard Copula were applied to the Apple and Google stock data to capture dependencies between variables. The analysis highlighted Gumbel Copula’s strength in capturing tail dependence, a critical feature for modelling tail-heavy data which are commonly observed in real-world stock markets.

References

A. Sklar, Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231 (1959)

N.C. Giri, Multivariate Statistical Analysis. (Marcel Dekker, New York, 1996)

P. Blæsild, J.L. Jensen, Multivariate distributions of hyperbolic type, in Statistical Distributions in Scientific Work – Proceedings of the NATO Advanced Study Institute held at the Università degli studi di Trieste, vol. 4 (1981), pp. 45–66

P. Embrechts, A. McNeil, D. Straumann, Correlation and Dependence in Risk Management: Properties and Pitfalls. Preprint ETH Zürich, 1999

P. Hall, The Bootstrap and Edgeworth Expansion. Statistical Series. (Springer, New York, 1992)

P.-S. Laplace, Mémoire sur la Probabilité des Causes par les événements. Savants étranges 6, 621–656 (1774)

R.B. Nelsen, An Introduction to Copulas (Springer, New York, 1999)

W. Härdle, L. Simar, & M. Fengler, Applied Multivariate Statistical Analysis, 6th ed., 107-170 (Springer, Cham, 2024)

W. Härdle, N. Hautsch, L. Overbeck, Applied Quantitative Finance, 2nd edn. (Springer, Heidel-berg, 2009)

W. S. Gosset, The probable error of a mean. Biometrika 6, 1–25 (1908)

Y. Chen, W. Härdle, S.-O. Jeong, Nonparametric risk management with generalized hyperbolic distributions. J. Am. Stat. Assoc. 103, 910–923 (2008)