Project 1B:
Two-Dimensional Distributions, Marginals and Covariance Structure

LIU XIAOLONG ID: 201900170025
E-mail: 201900170025@mail.sdu.edu.cn

May 16, 2022

Abstract

In this Project, I will discuss some special properties about the two dimensional distributions and their marginals, introduce moment generating functions and characteristic functions which can be used to study distributions more deeply. One particular can we consider is the bivariate Cauchy distribution. Moreover, we introduce how to construct a bivariate distribution from two single, marginal, distributions.

Contents

1 Introduction

The first part of the Project is to introduce the basic definitions and properties. This part aiming to provide the preliminaries.

In the second part, I will write the moment generating functions (m.g.f.s) and characteristic functions (ch.f.s) of bivariate distributions and some important properties and theorems about them. As an application, I discuss a special distribution: bivariate Cauchy distribution.

In the final part, I will introduce the Farlie-Gumbel-Morgenstern (FGM) copula and derive some properties about it. A copula is a function that makes marginals F_X and F_Y to some joint distribution F. It was first introduced by Sklar. Now copulas and several parametric families of copulas have been widely used in statistics. One of the most popular parametric families is the FGM copula, whose properties were discussed by Farlie [1].

2 Random Variables, Distributions and Covariance

For now I will introduce some basic definitions of bivariate r.v.s and their distributions.

Definition 2.1. We denote Ω be the whole sample space and let $\mathscr{F}$ be the σ-algebra of all events, i.e. $\mathscr{F}$ satisfies

(i) $\Omega \in \mathscr{F}$;

(ii) If $A \in \mathscr{F}$, then $A^c = \Omega \backslash A \in \mathscr{F}$;

(iii) If $A_n \in \mathscr{F}, n = 1, 2, ...$, then $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$.

Let $\mathsf{P} : \mathscr{F} \to \mathbb{R}$ be a probability, i.e. a function on $\mathscr{F}$ satisfying

(a) $\mathsf{P}(A) \geq 0$ for all $A \in \mathscr{F}$;

(b) $\mathsf{P}(\Omega) = 1$;

(c) If $A_i \in \mathscr{F}$ for all $i = 1, 2, ...$ which are pairwise disjoint, then

$$\mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i), \quad (\sigma - \text{additivity}).$$

It is generally accepted to call the triple $(\Omega, \mathscr{F}, \mathsf{P})$ a probability space. We assume that all r.v.s considered below are defined in a given probability space $(\Omega, \mathscr{F}, \mathsf{P})$.

Definition 2.2. Consider the random vector $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega))$, its components are r.v.s. The function

$$F(x, y) = \mathsf{P}(\xi_1(\omega) \leq x, \xi_n(\omega) \leq y), \quad (x, y) \in \mathbb{R}^2$$

is called a **distribution function** (d.f.) of ξ. Its dimension is 2.

Suppose, there exists a non-negative function $p(x, y)$, $(x, y) \in \mathbb{R}^2$, such that

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(x, y) dx dy,$$

where $\int_{\mathbb{R}^2} p(x, y) dx dy = 1$; it is called a **probability density function**.

Definition 2.3 (Marginal)**.** We consider two-dimensional random vector (ξ, η).

(1)[General] Let its d.f. be F as defined above. Let

$$F_1(x) = \mathsf{P}(\xi \leq x) = F(x, +\infty) \text{ and } F_2(y) = \mathsf{P}(\eta \leq y) = F(+\infty, y).$$

These are called the **marginal d.f.s** of $F(x, y)$, also of the components ξ and η.

(2)[Discrete] Suppose that ξ takes values in the set $\{x_1, x_2, ..., x_n\}$ and η takes values in the set $\{y_1, y_2, ..., y_m\}$, and we know $\mathsf{P}(\xi = x_i, \eta = y_j) = p_{ij}$. Then $\mathsf{P}(\xi = x_i) := a_i, \mathsf{P}(\eta = y_j) := b_j$ are found as follows:

$$\sum_j p_{ij} = a_i, \quad \sum_i p_{ij} = b_j.$$

The two coefficients, $\{a_i\}$ and $\{b_j\}$ are the probability distributions of ξ and η, respectively.

(3)[Continuous] Let (ξ, η) have a probability density function $p(x, y)$. Then

$$F_1(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} p(u, y) du dy, \quad F_2(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} p(x, v) dx dv.$$

Hence the probability density functions of $F_1(x)$ and $F_2(y)$, also of ξ and η, are

$$p_1(x) = \int_{\mathbb{R}} p(x, y) dy, \quad p_2(y) = \int_{\mathbb{R}} p(x, y) dx.$$

They are called marginal densities functions of ξ and η, respectively.

Definition 2.4. The r.v.s ξ_1, ξ_2 are said to be **independent** if

$$\mathsf{P}(\xi_1(\omega) \leq x, \xi_2(\omega) \leq y) = \mathsf{P}(\xi_1(\omega) \leq x)\mathsf{P}(\xi_2(\omega) \leq y)$$

for all x, y from the range of values of ξ_1, ξ_2, respectively.

Definition 2.5 (Expectation)**.** In the general case, we will use a special integral, called Riemann-Stieltjes integral, to define it.

Let the d.f. of the random vector (ξ_1, ξ_2) be $F(x_1, x_2)$, $(x_1, x_2) \in \mathbb{R}^2$. The expectation of (ξ_1, ξ_2) is denoted by $(\mathsf{E}\xi_1, \mathsf{E}\xi_2)$, where

$$\mathsf{E}\xi_i = \int_{\mathbb{R}^2} x_i dF(x_1, x_2) = \int_{\mathbb{R}} x_i dF_i(x_i) < \infty$$

where F_i is the marginal d.f. of ξ_i, $i = 1, 2$. If the integral is not finite, then we say that the expectation of (ξ_1, ξ_2) does not exists.

So in the case of discrete random variables, $\mathsf{E}(\xi_1) = \sum_i x_i a_i$, $\mathsf{E}(\xi_2) = \sum_j x_j b_j$; in the continuous case, $\mathsf{E}(\xi_i) = \int_{\mathbb{R}} x p_i(x) dx$.

Remark 2.6. *More generally, we have the following formula for any measurable function $g(x_1, x_2)$, $(x_1, x_2) \in \mathbb{R}^2$, assuming the integral w.r.t. F exists:*

$$\mathsf{E}g(\xi_1, \xi_2) = \int_{\mathbb{R}^2} g(x_1, x_2) dF(x_1, x_2).$$

Definition 2.7 (Variance and Covariance)**.** Consider a random vector $\boldsymbol{\xi} = (\xi_1, \xi_2)$, we define its covariance matrix as

$$\mathsf{Cov}(\boldsymbol{\xi}) = \begin{pmatrix} \mathsf{Var}\xi_1 & \mathsf{Cov}(\xi_1, \xi_2) \\ \mathsf{Cov}(\xi_2, \xi_1) & \mathsf{Var}\xi_2 \end{pmatrix},$$

where $\mathsf{Var}(\xi_i) = \mathsf{E}[(\xi_i - \mathsf{E}\xi_i)^2]$ is the variance of ξ_i and $\mathsf{Cov}(\xi_i, \xi_j) = \mathsf{E}[(\xi_i - \mathsf{E}\xi_i)(\xi_j -$

$\mathsf{E}\xi_j)]$ is the covariance between ξ_i and ξ_j for $i, j = 1, 2$ if all of these are finite. So $\mathsf{Cov}(\xi_i, \xi_i) = \mathsf{Var}(\xi_i)$ and $\mathsf{Cov}(\boldsymbol{\xi}) = (\mathsf{Cov}(\xi_i, \xi_j))_{2\times 2}$.

Remark 2.8. *The matrix* $\mathsf{Cov}(\boldsymbol{\xi})$ *is non-negative definite: for all* $t_j \in \mathbb{R}$ *we have*

$$\sum_{j,k} \mathsf{Cov}(\xi_j, \xi_k) t_j t_k = \mathsf{E}\left(\sum_j t_j(\xi_j - \mathsf{E}\xi_j)\right)^2 \geq 0.$$

3 Moment Generating Functions and Characteristic Functions

Now we will introduce two tools, moment generating functions and characteristic functions, to analyse the distributions. For two random variables X and Y, the joint moment generating functions of them contains rich information about their joint probability distribution. (See [3])

Definition 3.1. For two r.v.s X and Y, the moment generating function (m.g.f.) is defined as

$$M_{X,Y}(t_1, t_2) = \mathsf{E}(\exp(t_1 X + t_2 Y))$$

where real vector (t_1, t_2) take values in a closed rectangle $I_1 \times I_2 \subset \mathbb{R}^2$ containsing the origin $(0, 0)$.

Hence we find that $M_X(t_1) = M_{X,Y}(t_1, 0)$ and $M_Y(t_2) = M_{X,Y}(0, t_2)$ are the marginal m.g.f.s of X and Y, repectively. Now we can use m.g.f.s to express properties and facts we are familiar with from our Probability course.

Theorem 3.2. *Consider two r.v.s X and Y with d.f. $F(x, y)$ and m.g.f. $M_{X,Y}(t_1, t_2)$. Then for all $r, s \in \mathbb{Z}_{\geq 0}$*

$$\mathsf{E}(X^r Y^s) = \frac{\partial^{r+s} M_{X,Y}}{\partial t_1^r \partial t_2^s}(0, 0)$$

and

$$\mathsf{Cov}(X^r, Y^s) = \frac{\partial^{r+s} M_{X,Y}}{\partial t_1^r \partial t_2^s}(0,0) - \frac{\partial^r M_X}{\partial t_1^r}(0)\frac{\partial^s M_Y}{\partial t_2^s}(0).$$

Proof. We have

$$\begin{aligned}
\frac{\partial^{r+s} M_{X,Y}}{\partial t_1^r \partial t_2^s}(t_1,t_2) &= \frac{\partial^{r+s}}{\partial t_1^r \partial t_2^s}\int_{\mathbb{R}^2} e^{t_1 x + t_2 y} dF(x,y) \\
&= \int_{\mathbb{R}^2} \frac{\partial^{r+s} e^{t_1 x + t_2 y}}{\partial t_1^r \partial t_2^s} dF(x,y) \\
&= \int_{\mathbb{R}^2} x^r y^s e^{t_1 x + t_2 y} dF(x,y).
\end{aligned}$$

For the second equality, the existence of the m.g.f. allows to change the differentiation and the integration by Fubini type theorem.

Hence

$$\frac{\partial^{r+s} M_{X,Y}}{\partial t_1^r \partial t_2^s}(0,0) = \int_{\mathbb{R}^2} x^r y^s dF(x,y) = \mathsf{E}(X^r Y^s).$$

Moreover,

$$\frac{\partial^{r+s} M_{X,Y}}{\partial t_1^r \partial t_2^s}(0,0) - \frac{\partial^r M_X}{\partial t_1^r}(0)\frac{\partial^s M_Y}{\partial t_2^s}(0) = \mathsf{E}(X^r Y^s) - \mathsf{E}(X^r)\mathsf{E}(Y^s)$$

which is exactly $\mathsf{Cov}(X^r, Y^s)$. □

Corollary 1. *If the m.g.f. $M_{X,Y}$ exists, then all moments of X and Y are finite.*

Theorem 3.3 (Uniqueness in terms of m.g.f.)**.** *Consider (X,Y) and (U,V), two r.v.s. Then $M_{X,Y} = M_{U,V}$ in some neighborhood of the origin if and only if (X,Y) and (U,V) have the same joint d.f.s.*

Proof. This proof relies on Laplace transforms. I refer to [2] for the entire proof. □

Here, we will use m.g.f.s to make conclusions about the independence of the components of the random vectors.

Theorem 3.4. *Consider two r.v.s X and Y with joint d.f. $F(x, y)$, the two marginal d.f.s $G(x)$ and $H(y)$ and the m.d.f. $M_{X,Y}(t_1, t_2)$. Then X and Y are independent if and only if*

$$M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$$

for all (t_1, t_2) in some neighborhood of the origin, $(0, 0)$ in $\mathbb{R}^2$.

Proof. If (X, Y) are independent by the definition of expectation, we have

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= \mathsf{E}(e^{t_1X+t_2Y}) = \mathsf{E}(e^{t_1X}e^{t_2Y}) \\ &= \mathsf{E}(e^{t_1X})\mathsf{E}(e^{t_2Y}) = M_X(t_1)M_Y(t_2). \end{aligned}$$

Conversely, if $M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$ for all (t_1, t_2) in some neighborhood of the origin, we claim that X and Y are independent. By the uniqueness of a d.f. from its m.g.f., the bivariate d.f. $F(x, y)$ is the unique distribution that corresponds to $M_{X,Y}$ and $G(x)H(y)$ is the unique distribution that corresponds to M_XM_Y. Hence $F(x, y) = G(x)H(y)$, $(x, y) \in \mathbb{R}^2$. □

Corollary 2. *Consider two r.v.s X and Y. Then X and Y are independent if and only if*

$$\mathsf{Cov}(e^{t_1X}, e^{t_2Y}) = 0$$

for all (t_1, t_2) in some neighborhood of origin.

Proof. Use the previous theorem and the following fact, assuming that all expectations are finite,

$$\mathsf{Cov}(e^{t_1X}, e^{t_2Y}) = \mathsf{E}(e^{t_1X+t_2Y}) - \mathsf{E}(e^{t_1X})\mathsf{E}(e^{t_2Y}) = M_{X,Y}(t_1, t_2) - M_X(t_1)M_Y(t_2),$$

then well done. □

Actually, we have the following expansion:

$$\mathsf{Cov}(e^{t_1 X}, e^{t_2 Y}) = \sum_{r,s=0}^{\infty} \frac{t_1^r t_2^s}{r!s!} \mathsf{Cov}(X^r, Y^s).$$

Hence we can show that if X, Y have bounded supports then $\mathsf{Cov}(X^r, Y^s) = 0$ for all $r, s > 0$ if and only if X and Y are independent. I have seen details in [3].

Example 1. *In the theorem if X and Y are arbitrary r.v.s which are independent, then we have*

$$M_{X+Y}(t) = M_{X,Y}(t,t) = M_X(t)M_Y(t).$$

But conversely, if X and Y satisfies $M_{X+Y}(t) = M_{X,Y}(t,t) = M_X(t)M_Y(t)$, then X and Y may not be independent.

Indeed, consider (X, Y) to be a two-dimensional random vector defined by the table:

X/Y	1	2	3	Total of X
1	$\frac{2}{18}$	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{1}{3}$
2	$\frac{3}{18}$	$\frac{2}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
3	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{2}{18}$	$\frac{1}{3}$
Total of Y	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Then the sum $Z = X + Y$ is given in the following table:

Z	2	3	4	5	6
P	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

We find separately the m.g.f. of each of X, Y and Z:

$$M_X(t) = \frac{1}{3}(e^t + e^{2t} + e^{3t}), M_Y(t) = \frac{1}{3}(e^t + e^{2t} + e^{3t}),$$
$$M_Z(t) = \frac{1}{9}(e^{2t} + 2e^{3t} + 3e^{4t} + 2e^{5t} + e^{6t}).$$

Now it is easy to see that $M_Z(t) = M_X(t)M_Y(t)$. Howvever the r.v.s X and Y are not independent. Because $\mathsf{P}(X = i, Y = j) \neq \mathsf{P}(X = i)\mathsf{P}(Y = j)$ for all $i \neq j$. For example, if $i = 1, j = 2$, we have $\mathsf{P}(X = i, Y = j) = \frac{1}{18}$, $\mathsf{P}(X = i) = \frac{1}{3}$, $\mathsf{P}(Y = j) = \frac{1}{3}$.

We already mentioned above that all moments exist if its m.g.f. exists. Here is an interesting question: Is there a r.v. X such that all moment $\mathsf{E}(X^k)$ are finite, however the m.g.f. does not exist? To answer this question, we consider univariate case. (See [5]) Consider a r.v. Z with density $f(x) = \frac{1}{2}\exp(-\sqrt{x})\mathbb{1}_{\mathbb{R}_{\geq 0}}$, then

$$\begin{aligned}
\mathsf{E}(Z^k) &= \int_0^\infty x^k \frac{1}{2}\exp(-\sqrt{x})dx = \frac{1}{2}\int_0^\infty x^k e^{-x^{1/2}} dx \\
&\overset{\sqrt{x}=t}{=\!=\!=} \frac{1}{2}\int_0^\infty t^{2k} e^{-t} 2t dt = \int_0^\infty t^{2k+1} e^{-t} dt \\
&= \Gamma(2k+2) = (2k+1)!.
\end{aligned}$$

By the definition of a m.g.f.,

$$M(z) = \frac{1}{2}\int_0^\infty \exp(zx - \sqrt{x})dx.$$

So, is this function finite for Z in a neighborhood of the origin? If $\varepsilon > 0$ is small enough then for every z with $0 < z < \varepsilon$ we have $zx - \sqrt{x} = \sqrt{x}(z\sqrt{x} - 1) \to \infty$ as $x \to \infty$. So $\frac{1}{2}\int_0^\infty \exp(zx - \sqrt{x})dx = \infty$, hence $M(z)$ does not exist.

In general it is useful to introduce a function which exists for any probability distribution. Now we introduce the second tool, characteristic functions (ch.f.).

Definition 3.5. Consider a random vector (X, Y) with d.f. $F(x, y)$. The characteristic function (ch.f) is denoted by $\psi_{X,Y}$ and defined as follows:

$$\psi_{X,Y}(t_1, t_2) = \mathsf{E}(e^{it_1X+it_2Y}) = \int_{\mathbb{R}^2} e^{i(t_1x+t_2y)}dF(x, y), \quad (t_1, t_2) \in \mathbb{R}^2.$$

Theorem 3.6. *Consider two r.v.s X and Y with joint d.f. $F(x, y)$ and ch.f. $\psi_{X,Y}$, then for all $r, s \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$*

$$\mathsf{E}(X^rY^s) = i^{-r-s}\frac{\partial^{r+s}\psi_{X,Y}}{\partial t_1^r \partial t_2^s}(0, 0).$$

Theorem 3.7 (Uniqueness of d.f.)**.** *Consider (X, Y) and (U, V), two random vectors. Then $\psi_{X,Y} = \psi_{U,V}$ in some neighborhood of the origin if and only if (X, Y) and (U, V) have the same joint d.f.s.*

Theorem 3.8. *Consider two r.v.s X and Y, their ch.f. is $\psi_{X,Y}(t_1, t_2)$. Then X and Y are independent if and only if*

$$\psi_{X,Y}(t_1, t_2) = \psi_X(t_1)\psi_Y(t_2)$$

for all (t_1, t_2) in some neighborhood of the origin.

Proof. All of these theorem are the same as in the case of m.g.f. above. $\square$

We can use ch.f.s and the uniqueness theorem for d.f.s, to derived interesting statements. Such is the following fact.

Example. Consider an arbitrary matrix $\boldsymbol{M} \in \mathcal{M}_{2\times 2}(\mathbb{R})$ and a bivarate normal random vector $\boldsymbol{\xi} \sim \mathcal{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. For convenience we introduce the notations $\boldsymbol{\mu}$ and $\boldsymbol{C}$, where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T, \boldsymbol{C} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. We use the matrix $\boldsymbol{M}$ and the new random vector $\boldsymbol{\eta} = \boldsymbol{M}\boldsymbol{\xi}$ as a linear transformation. Then $\boldsymbol{\eta} \sim \mathcal{N}_2(\boldsymbol{M}\boldsymbol{\mu}, \boldsymbol{M}\boldsymbol{C}\boldsymbol{M}^T)$. Here $\mathcal{N}_2$ denotes 2-dimensional normal distribution.

Proof. Consider the ch.f. of $\boldsymbol{\eta}$. For any real vector $\boldsymbol{t} \in \mathbb{R}^2$, we have

$$\begin{aligned}\psi_{\boldsymbol{\eta}}(\boldsymbol{t}) &= \mathsf{E}e^{i\boldsymbol{t}^T\boldsymbol{M}\boldsymbol{\xi}} = \mathsf{E}e^{i(\boldsymbol{M}^T\boldsymbol{t})^T\boldsymbol{\xi}} \\ &= \exp\left(i(\boldsymbol{M}\boldsymbol{\mu})^T\boldsymbol{t} - \frac{1}{2}\boldsymbol{t}^T(\boldsymbol{M}\boldsymbol{C}\boldsymbol{M}^T)\boldsymbol{t}\right).\end{aligned}$$

The explicit form of $\psi_{\boldsymbol{\eta}}(\boldsymbol{t})$ allows to conclude that $\boldsymbol{\eta} = \boldsymbol{M}\boldsymbol{\xi} \sim \mathcal{N}_2(\boldsymbol{M}\boldsymbol{\mu}, \boldsymbol{M}\boldsymbol{C}\boldsymbol{M}^T)$. We tell this by words: A linear transformation of a normal random vector is normal. And we saw how exactly the parameters change. □

4 Bivariate Cauchy Distribution

Recall first that a r.v. ξ with values in $\mathbb{R}$ has Cauchy distribution, and we write $\boldsymbol{\xi} \sim \mathcal{C}_1$, if ξ is continuous and has density

$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Definition 4.1. We say that the random vectors (X, Y) follows bivariate Cauchy distribution, $(X, Y) \sim \mathcal{C}_2$, if its joint density is

$$p(x, y) = \frac{1}{2\pi}\frac{1}{(1+x^2+y^2)^{3/2}}, \quad x, y \in \mathbb{R}.$$

Proposition 4.2 (Marginals)**.** *If* $(X, Y) \sim \mathcal{C}_2$*, then* $X \sim \mathcal{C}_1$ *and* $Y \sim \mathcal{C}_1$*, i.e. the marginal densities are*

$$p_X(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}, \quad p_Y(y) = \frac{1}{\pi(1+y^2)}, y \in \mathbb{R}.$$

Proof. First, we need a specific indefinite integral:

$$\begin{aligned}
\int \frac{1}{(1+x^2+y^2)^{3/2}}dy &\xlongequal{y=\sqrt{1+x^2}\tan u} \int \frac{\sqrt{1+x^2}}{(1+x^2+(1+x^2)\tan^2 u)^{3/2}\cos^2 u}du \\
&= \int \frac{\sqrt{1+x^2}}{(1+x^2)^{3/2}(1+\tan^2 u)^{3/2}\cos^2 u}du \\
&= \frac{1}{1+x^2}\int \frac{\cos^3 u}{\cos^2 u}du = \frac{1}{1+x^2}\int \cos u du \\
&= \frac{y}{(1+x^2)\sqrt{1+x^2+y^2}} + C.
\end{aligned}$$

Second, use the Newton-Leibniz formula:

$$\int_{\mathbb{R}} \frac{1}{(1+x^2+y^2)^{3/2}}dy = \left.\frac{y}{(1+x^2)\sqrt{1+x^2+y^2}}\right|_{-\infty}^{\infty} = \frac{2}{1+x^2}.$$

Hence

$$p_X(x) = \int_{\mathbb{R}} p(x,y)dy = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}.$$

Well done. □

Another observation is that the expectation of a r.v. $X \sim \mathcal{C}_1$ does not exist! Indeed, by definition we have

$$\mathsf{E}(X) = \int_{\mathbb{R}} x p_X(x)dx = \int_{\mathbb{R}} \frac{x}{\pi(1+x^2)}dx,$$

which does not exist! As a consequence, higher order moments $\mathsf{E}(|X|^r)$ also do not exist for any $r > 1$, hence no variance.

Corollary 3. *This also tells us the m.g.f. of Cauchy distribution does not exists!*

However, even if the m.g.f. does not exists, the ch.f. always exists, and in many case it can be found explicitly. Here is the such a case.

Theorem 4.3. *If the random vectors* $(X, Y) \sim \mathcal{C}_2$, *then its ch.f. is*

$$\psi_{X,Y}(t_1, t_2) = \exp\left(-\sqrt{t_1^2 + t_2^2}\right), \quad (t_1, t_2) \in \mathbb{R}^2.$$

Proof. Just need to calculate $\frac{1}{2\pi}\int_{\mathbb{R}^2} e^{i(t_1 x + t_2 y)}(1 + x^2 + y^2)^{-3/2} dxdy$.

First it is easy to see that for any integrable function g, we have

$$\int_{-a}^{a} g(x)dx = \int_{-a}^{a} \frac{g(x) + g(-x)}{2} dx$$

where $a \in \mathbb{R}$ or $a = \infty$. Then, by using this and the known fact,

$$\cos u = \frac{1}{2}(e^{iu} + e^{-iu}),$$

we obtain:

$$\begin{aligned}
\psi_{X,Y}(t_1, t_2) &= \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{e^{i(t_1 x + t_2 y)}}{(1 + x^2 + y^2)^{3/2}} dxdy \\
&= \frac{1}{8\pi}\int_{\mathbb{R}^2} \frac{(e^{it_1 x} + e^{-it_1 x})(e^{it_2 y} + e^{-it_2 y})}{(1 + x^2 + y^2)^{3/2}} dxdy \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{\cos(t_1 x)\cos(t_2 y)}{(1 + x^2 + y^2)^{3/2}} dxdy \\
&= \frac{2}{\pi}\int_0^\infty \int_0^\infty \frac{\cos(|t_1|x)\cos(|t_2|y)}{(1 + x^2 + y^2)^{3/2}} dxdy \\
&= \exp\left(-\sqrt{t_1^2 + t_2^2}\right).
\end{aligned}$$

The last step is a tricky and complicated integral and we can see details of this and more general results in [7]. □

5 Farlie-Gumbel-Morgenstern (FGM) Family of Distributions

We will discribe how from two 1-dimensional distributions to construct a bivariate distribution and study its properties. To mention that FGM copulas are investigated in many publications e.g., in [4] and [6].

We start with introducing the general copula with the most basic definitions and some fundamental results.

Definition 5.1. Let S_1 and S_2 be nonempty subsets of $\mathbb{R}$ and let H be a function defined on $S_1 \times S_2 \subset \mathbb{R}^2$. For any rectangle $B = [x_1, x_2] \times [y_1, y_2]$ with all of whose vertices are in $S_1 \times S_2$, we define the H-volume of B is

$$V_H(B) = H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1).$$

A such function H is called 2-increasing if $V_H(B) \geq 0$ for all rectangles B whose vertices lie in $S_1 \times S_2$.

General fact: Suppose a r.v. $\xi \sim F$. Define a new r.v. $\eta = F(\xi)$. Clearly, η take values in $[0,1]$, and moreover, η is continuous uniform on $[0,1]$, its density is $f_\eta = \mathbb{1}_{[0,1]}$.

Definition 5.2. A two-dimensional copula is a function $C : [0,1] \times [0,1] \to [0,1]$ with the following properties

(i) For any $u, v \in [0,1]$

$$C(u,0) = C(0,v) = 0 \text{ and } C(u,1) = u, \quad C(1,v) = v.$$

(ii) C is 2-increasing in the square $[0,1]^2$.

Now we introduce the statement of the following fundamental theorem.

Theorem 5.3 (Sklar's Theorem)**.** *Let $H(x, y)$, $(x, y) \in \mathbb{R}^2$ be the 2-dimensional d.f. with marginal d.f.s F and G. Then there exists a copula C such that for all $x, y \in \mathbb{R}$,*

$$H(x, y) = C(F(x), G(y)).$$

If F and G are continuous, then C is unique. Conversely, if C is a copula function and F and G are d.f.s, then the function H defined by $H(-, -) = C(F(-), G(-))$ is a 2-dimensional d.f. with marginals F and G.

Now we back to FGM copula. We start with two d.f.s, say F_1 and F_2: each is defined on $(-\infty, \infty)$, values in $[0, 1]$, non-decreasing, right-continuous. Now, for any number α, $\alpha \in [-1, 1]$, we define the function

$$G(x, y) = F_1(x)F_2(y)(1 + \alpha(1 - F_1(x))(1 - F_2(y))), x, y \in \mathbb{R}.$$

We can check that G is a 2-dimensional d.f. Indeed, we just need to verify that $G(x, y)$ satisfies (i) $\lim_{x \to -\infty} G(x, y) = \lim_{y \to -\infty} G(x, y) = 0$, (ii) $\lim_{x \to \infty, y \to \infty} G(x, y) = 1$,(iii) $G(x, y)$ is right-continuous with x and y, (iv) for any $a < b, c < d$, one has

$$G(b, d) + G(a, c) \geq G(a, d) + G(b, c).$$

Actually (i),(ii),(iii) are trivial by the definition of F_1 and F_2. To see (iv), consider $u_1 < u_2$ and $v_1 < v_2$, then

$$\begin{aligned}
&G(u_2, v_2) - G(u_1, v_2) - G(u_2, v_1) + G(u_1, v_1) \\
&= (F_1(u_2) - F_1(u_1))(F_2(v_2) - F_2(v_1)) \\
&\times (1 + \alpha(1 - F_1(u_1) - F_1(u_2))(1 - F_2(v_1) - F_2(v_2))).
\end{aligned}$$

Hence it is non-negative when $\alpha \in [-1, 1]$, since $0 \leq F_i \leq 1$. Thus we conclude that by Sklar's theorem there is a random vector, say (X, Y), such that this G is its joint d.f.

Moreover, since $G(x, \infty) = F_1(x)$ and $G(\infty, y) = F_2(y)$, we see that the marginals of (X, Y) are F_1 and F_2 for any $\alpha \in [-1, 1]$. So this is an infinite family with the same marginals. As a consequence, the marginal distributions do not determine uniquely the 2-dimensional distribution.

Assume $X \sim F_1$ and $Y \sim F_2$ are continuous with densities $f_1(x), f_2(y)$. Then the 2-dimensional probability density function of (X, Y) is

$$g(x, y) = \frac{\partial^2 G(x, y)}{\partial x \partial y} = f_1(x) f_2(y)(1 + \alpha(2F_1(x) - 1)(2F_2(y) - 1)).$$

It is useful to know the correlation between X and Y. First we find:

$$\begin{aligned}
&\mathsf{Cov}(X, Y) = \int_{\mathbb{R}^2} (x - \mathsf{E}(X))(y - \mathsf{E}(Y)) g(x, y) dx dy \\
&= \left(\int_{\mathbb{R}} (x - \mathsf{E}(X)) f_1(x) dx \right) \left(\int_{\mathbb{R}} (y - \mathsf{E}(Y)) f_2(y) dy \right) \\
&+ \alpha \left(\int_{\mathbb{R}} (x - \mathsf{E}(X)) f_1(x)(2F_1(x) - 1) dx \right) \left(\int_{\mathbb{R}} (y - \mathsf{E}(Y)) f_2(y)(2F_2(y) - 1) dx \right) \\
&= \alpha \left(\int_{\mathbb{R}} (x - \mathsf{E}(X)) f_1(x)(2F_1(x) - 1) dx \right) \left(\int_{\mathbb{R}} (y - \mathsf{E}(Y)) f_2(y)(2F_2(y) - 1) dx \right)
\end{aligned}$$

Also we can find the correlation coefficient:

$$\rho(X, Y) = \frac{\mathsf{Cov}(X, Y)}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}} = \frac{\alpha \times D}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}},$$

where

$$D = \left(\int_{\mathbb{R}} (x - \mathsf{E}(X)) f_1(x)(2F_1(x) - 1) dx \right) \left(\int_{\mathbb{R}} (y - \mathsf{E}(Y)) f_2(y)(2F_2(y) - 1) dx \right).$$

It is of great interest to find the upper bound of the correlation coefficient. Since we have the decomposition $D = D_1 D_2$ where

$$D_1 = \left(\int_{\mathbb{R}} (x - \mathsf{E}(X)) f_1(x)(2F_1(x) - 1) dx \right)$$

and

$$D_2 = \left(\int_{\mathbb{R}} (y - \mathsf{E}(Y)) f_2(y)(2F_2(y) - 1) dx \right),$$

we just need to estimate one of them. Actually, we have the following chain of relations:

$$\begin{aligned}
D_1^2 &= \left(\int_{\mathbb{R}} (x - \mathsf{E}(X))(2F(x) - 1) f(x) dx \right)^2 \\
&= \left(\int_{\mathbb{R}} \left((x - \mathsf{E}(X)) \sqrt{f(x)} \right) \left((2F(x) - 1) \sqrt{f(x)} \right) dx \right)^2 \\
&\leq^{(1)} \left(\int_{\mathbb{R}} (x - \mathsf{E}(X))^2 f(x) dx \right) \left(\int_{\mathbb{R}} (2F(x) - 1)^2 f(x) dx \right) = \frac{\mathsf{Var}(X)}{3},
\end{aligned}$$

where (1) is Cauchy-Swartz inequality and the last step is because

$$\int_{\mathbb{R}} (2F(z) - 1)^2 f(z) dz = \int_0^1 (2u - 1)^2 du = \frac{1}{3}.$$

Hence $D_1 D_2 \leq \frac{\mathsf{Var}(X)}{3}$ and

$$\rho(X, Y) = \frac{\alpha D_1 D_2}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}} \leq \frac{\alpha}{3},$$

well done.

6 Conclusions

For now after we introduce the basic definitions, we have described the properties of moment generating functions and characteristic functions, the bivariate Cauchy distribution and some properties about the FGM copula.

7 Symbols and Notations

Table 1: Symbols

Symbol	meaning	Symbol	meaning
P	Probability Measure	p	Probability Density Function
E	Expectation	Var	Variance
Cov	Covariance	ρ	Correlation Coefficient
$\mathcal{C}$	Cauchy Distribution	$\mathcal{N}$	Normal Distribution

References

[1] D. J. G. Farlie. *The performance of some correlation coefficients for a general bivariate distribution*. Biometrika, 47(3/4):307-323, 1960.

[2] William Feller. *An Introduction to Probability Theory and its Applications, vol 2, 2nd ed*. Wiley, New York, 1971.

[3] Yongmiao Hong. *Probability and Statistics for Economists*. World Scientific, Singapore, 2018.

[4] Roger B Nelsen. *An Introduction to Copulas*. Springer Science & Business Media, 2007.

[5] Jordan Stoyanov. *Counterexamples in Probability, 3rd ed.* Dover Publications Inc., 2014.

[6] Tsutomu T Takeuchi. *Constructing a bivariate distribution function with given marginals and correlation: application to the galaxy luminosity function.* Monthly Notices of the Royal Astronomical Society, 406(3):1830-1840, 2010.

[7] Ferguson, T. S. *A representation of the symmetric bivariate Cauchy distribution.* The Annals of Mathematical Statistics, 33(4), 1256-1266, 1962.