Copulas

MOU Lili

December, 2014

Outline

Preliminary

Probability density function (pdf) f

$$f(x) = \frac{P(x - \epsilon \leq X \leq x + \epsilon)}{2\epsilon}$$

Cumulative distribution function (cdf) F

$$F(x) = P(X \leq x)$$

Relationship between pdf and cdf

$$f(x) = \frac{\partial F(x)}{\partial x}$$

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Introduction

From Marginal to Joint Probability

We have random variables $\boldsymbol{X} = (X_1, X_2, \cdots, X_d)$

If we know the marginal distribution $F_1(x_1), \cdots, F_d(x_d)$,
Say $X_i \sim U[0, 1]$

- What is the joint distribution?
- Is the joint distribution unique?

No, because we do not know the dependencies among the variables.

Basic Idea

Life will be easier, if ...

we have some effective way of modeling dependencies between variables. (Copula)

Copula + Marginal distribution $\Longleftrightarrow$ Joint distribution

Formal Definition

Copulas

Let $\boldsymbol{U} = (U_1, \cdots, U_d)$ be $d-$dimensional random variables,
with marginal distribution $U_i \sim U[0,1]$

Define a **copula** function as a joint cdf on $\boldsymbol{U}$

$$C(u_1, \cdots, u_d) = P(U_1 \leq u_1, \cdots, U_d \leq u_d)$$

Essentially, a copula is defined as a joint distribution on a unit hyper-cube.

Theorem

For a multivariate distribution $F(x_1, \cdots, x_n)$, with marginal distribution $F_i(x_i)$,

$\exists$ a copula C s.t.

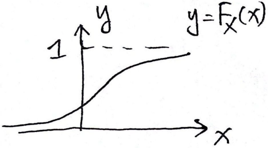
$$C(F_1(x_1), \cdots, F_d(x_d)) = F(x_1, \cdots, x_d)$$

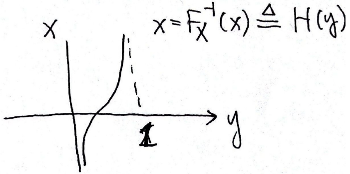
Furthermore, if the marginal distributions are continuous, the copula is unique.

1. To prove $F(X) \sim U[0,1]$,

Let $y = F_X(x)$ and $x = F_X^{-1}(y) \triangleq H(y)$

$$f_Y(y) = |H'(y)| f_X[H(y)] = \frac{dH(y)}{dy} f_X[H(y)] = \frac{dx}{dF_X(x)} f_X(x) = 1$$





2. By the definition of copula $C(u_1, \cdots, u_d) = P(U_1 \le u_1, \cdots, U_d \le u_d)$, we have

$$C(F_1(x_1), \cdots, F_d(x_d)) = P(F_1(X_1) \le F_1(x_1), \cdots, F_d(X_d) \le F_d(X_d))$$

$$= P(X_1 \le x_1, \cdots, X_d \le x_d) = F_{\mathbf{X}}(x_1, \cdots, x_d)$$

■

Joint Probability Density Function

Through differentiation of

$$C(F_1(x_1), \cdots, F_d(x_d)) = F(x_1, \cdots, x_d)$$

We obtain

$$f(x_1, \cdots, x_n) = c(F_1(x_1), \cdots, F_d(x_d)) \prod_{i=1}^{d} f_i(x_i)$$

where $c(u_1, \cdots, u_d) = \dfrac{\partial^d C(u_1, \cdots, u_d)}{\partial u_1 \cdots \partial u_d}$

Importance of Sklar's Theorem

Instead of specifying directly an odd joint distribution on $X_1, \cdots, X_d$, we

- Model the marginal distribution, which is easy, and
- Specify a coplua C, modeling nontrivial dependencies between variables.

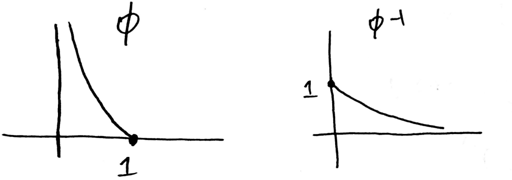
Sklar's Theorem guarantees: Copula(Marginal)=Joint

Copula Functions

Archimedean Copula Family

$$C_\phi(u_1, \cdots u_d) = \phi^{-1}\left(\sum_{i=1}^{d} \phi(u_i)\right)$$

Constraints on Φ



Verify the marginal distribution

$$F(u_1) = C_\phi(u_1, +\infty, \cdots, +\infty) = \phi^{-1}\phi(u_1) = u_1$$
$$\Rightarrow U_i \sim U[0,1]$$

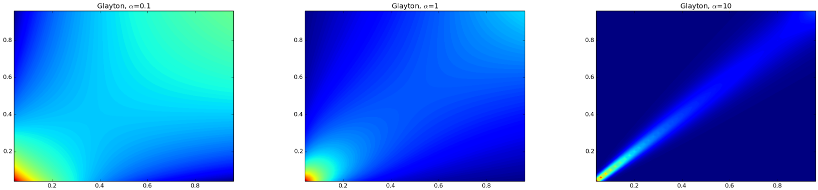
Examples

- Glayton, $\phi(x) = x^{-\alpha} - 1$, $\alpha > 0$,
- Gumbel, $(-\log(x))^{\alpha}$, $\alpha > 1$,
- Frank, $\phi(x) = -\log\left(\frac{\exp(-\alpha x) - 1}{\exp(-\alpha) - 1}\right)$, $\alpha \in (-\infty, +\infty), \alpha \neq 0$

The parameters are learned with training examples.

Specially, if $\alpha = 1$ in Gumbel copula, then variables are independent.

$\phi(x) = x^{-\alpha} - 1,\ \alpha > 0$



"Correct" Copula

Substitute $u_i = F_i(x_i)$ into

$$C(F_1(x_1), \cdots, F_d(x_d)) = F(x_1, \cdots, x_d)$$

We obtain

$$C(u_i, \cdots, u_d) = F(F_1^{-1}(u_1), \cdots F_d^{-1}(u_d)$$

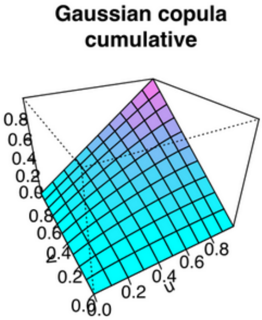
Note that for any valid multivariate cumulative function

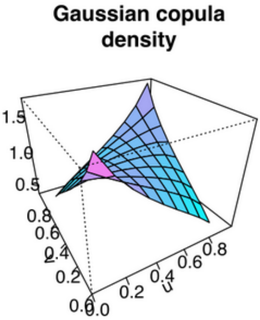
$$\begin{aligned} C(u_1, 1, \cdots, 1) &= F(F_1^{-1}(u_1), \infty, \cdots, \infty) \\ &= P(F_1^{-1}(U_1) < F_1^{-1}(u_1)) \\ &= P(U_1 < u_1) \\ &= u_1 \end{aligned}$$

$\Rightarrow C$ is valid regardless of F and F_i.

Let F_i to be standard norm distribution, and F to be norm with 0 mean and covariance matrix Σ.

Example $\rho = .4$





[Source: http://en.wikipedia.org/wiki/Copula_(probability_theory)]

Estimating the Parameters

Semi-Parametric Estimation

Log-likelihood

$$\ell(\alpha, \theta_1, \cdots, \theta_d) = \sum_{j=1}^{n} \log\left(c_\alpha \left(F_{\theta_1}(x_1^{(j)}), \cdots, F_{\theta_d}(x_d^{(j)}) \right) + \sum_{j=1}^{n} \sum_{i=1}^{d} \log(f_{\theta_1}(x_i^{(j)})) \right)$$

Inference for Margins (IFM), 2 two-stage procedure:

1. Estimate marginal distributions
2. Estimate copula parameters

$$\ell(\alpha) = \log\left(c_\alpha \left(\hat{F}_1(x_1^{(j)}), \cdots, \hat{F}_d(x_d^{(j)}) \right) \right)$$

The marginal distribution is estimated empirically.

$$\tilde{F}_i(y) = \tilde{P}(X_i \leq y) \approx \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}\{x_i^{(j)} \leq y\}$$

The log-likelihood

$$\ell(\alpha) = \log\left(c_\alpha\left(\tilde{F}_1(x_1^{(j)}), \cdots, \tilde{F}_d(x_d^{(j)})\right)\right)$$

Example

Teany Tiny Baby Example

- Bivariate normal distribution
- $\mu = (00)^T$
- $\Sigma = \begin{pmatrix} 1.25 & 0.43 \\ 0.43 & 1.75 \end{pmatrix}$
- 1000 training data samples
- Frank Archimedean copula
- semiparametric version of IFM

Results: Density

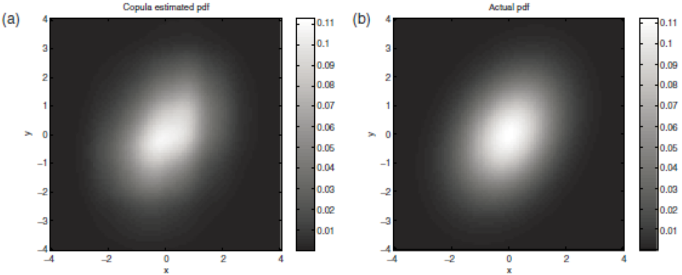
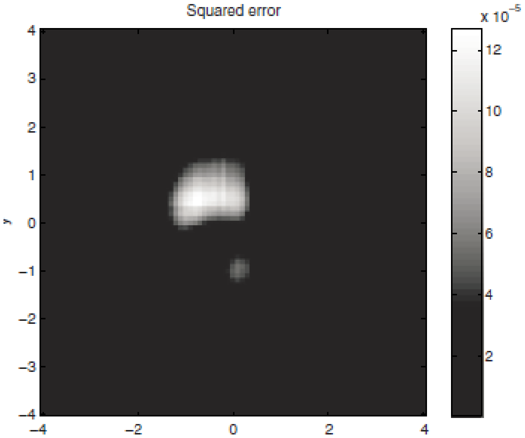


Figure 4.29 Copula estimate of the probability density function (a), and the actual probability density function (b).

Results: Squared Error



Results: Marginal Distribution

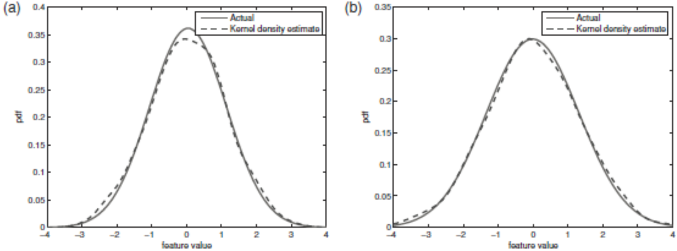


Figure 4.31 The marginal probability densities for the first (a) and second (b) features, for the actual densities (solid) and the kernel density estimates (dashed).

Conclusion and Discussion

Conclusion

- A copula is defined as a joint cdf on a unit hypercude with marginal distributions $U[0,1]$.
- Copulas link joint distribution with marginal distributions.

  $$\text{Copula (Marginals)} = \text{Joint}$$

- We can specify copulars explicitly, or via a (joint, marginals) pair
- Parameters are learned by MLE or its variations.

\+ Copulas provide a means of modeling non-trivial dependencies.

\- Choosing copulas is a $64,000,000 question. Inappropriate copulas, which fail to model true dependencies in data, may also cause considerable financial losses.

Q&A

Thanks for listening!