An introduction to pair-copula constructions

Ramiro Ruiz Cárdenas

1º Workshop do projeto PROCAD:

Seguro Agrícola: Modelagem Estatística e Precificação

UFMG - Belo Horizonte, Novembro 25-27 de 2009

Multivariate Distributions

Consider n random variables $X = (X_1, ..., X_n)$ with

- joint density $f(x_1, ..., x_n)$ and marginal densities $f_i(x_i)$, $i = 1, ..., n$
- joint cdf $F(x_1, ..., x_n)$ and marginal cdf's $F_i(x_i)$, $i = 1, ..., n$
- $f(. \mid .)$ denote corresponding conditional densities

and consider the factorization

$$\begin{aligned} f(x_1, ..., x_n) &= f(x_n \mid x_1, ..., x_{n-1}).f(x_1, ..., x_{n-1}) \\ &= \left[\prod_{t=2}^{n} f(x_t \mid x_1, ..., x_{t-1})\right].f_1(x_1) \end{aligned}$$

Copula

A copula is a multivariate distribution on $[0,1]^n$ with uniformly distributed marginals.

- copula cdf $C(u_1, ..., u_n)$
- copula density $c(u_1, ..., u_n)$

Using Sklar's Theorem (1959) we have for absolutely continuous bivariate distributions with continuous marginal cdf's

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)).f_1(x_1).f_2(x_2)$$
$$f(x_1 \mid x_2) = c_{12}\left(F_1(x_1), F_2(x_2)\right).f_1(x_1)$$

for some bivariate copula density $c_{12}(.)$.

Pair-copula constructions (PCC)

- Multivariate data can be modelled using a cascade of pair-copulae, acting on two variables at a time.
- The basic idea is to decompose an arbitrary distribution function into simple bivariate building blocks and stitch them together appropriately.
- These bivariate blocks are two-dimensional copulas and we have a large selection to choose from.

The two dimensional case

For the base case in two dimensions we can easily see that

$$f(x_1, x_2) = c_{12}(F_1(x1), F_2(x_2)).f_1(x_1).f_2(x_2)$$

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

$$f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = c_{12}(F_1(x_1), F_2(x_2)).f_2(x_2)$$

The three dimensional case

- Any three-dimensional density function can be written in the form

$$f(x_1, x_2, x_3) = f_1(x_1).f_{2|1}(x_2|x_1).f_{3|1,2}(x_3|x_1, x_2)$$

- we can write $f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)).f_2(x_2)$

- conditioning in X_2, we have that

$$f_{3|1,2}(x_3|x_1, x_2) = c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)).f_{3|2}(x_3|x_2)$$

- This yields the full decomposition

$$\begin{aligned} f(x_1, x_2, x_3) =& f_1(x_1). \\ & c_{12}(F_1(x_1), F_2(x_2))f_2(x_2). \\ & c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)).c_{23}(F_2(x_2).F_3(x_3)).f_3(x_3) \end{aligned}$$

The four dimensional case

- For a four-dimensional density we start with

$$f(x_1,x_2,x_3,x_4) = f_1(x_1).f_{2|1}(x_2|x_1).f_{3|1,2}(x_3|x_1,x_2).f_{4|1,2,3}(x_4|x_1,x_2,x_3)$$

- and rewrite it in terms of six pair-copulas and the four marginal densities $f_i(x_i)$ for $i = 1,2,3,4$:

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = & f_1(x_1). \\ & c_{12}(F_1(x_1),F_2(x_2)).f_2(x_2). \\ & c_{23|1}(F_{2|1}(x_2|x_1),F_{3|1}(x_3|x_1)).c_{13}(F_1(x_1),F_3(x_3)).f_3(x_3). \\ & c_{34|12}(F_{3|12}(x_3|x_1,x_2),F_{4|12}(x_4|x_1,x_2)). \\ & c_{24|1}(F_{2|1}(x_2|x_1),F_{4|1}(x_4|x_1)). \\ & c_{14}(F_1(x_1),F_4(x_4)).f_4(x_4) \end{aligned}$$

Pair-copula constructions (PCC)

- For distinct $i, j, i_1, ..., i_k$ with $i < j$ and $i_1 < ... < i_k$ let

$$c_{i,j|i_1,...,i_k} := c_{i,j|i_1,...,i_k}(F(x_i \mid x_{i_1}, ..., x_{i_k}), (F(x_j \mid x_{i_1}, ..., x_{i_k}))$$

- Reexpress $f(x_t \mid x_1, ..., x_{t-1})$ as

$$\begin{aligned} f(x_t \mid x_1, ..., x_{t-1}) &= c_{1,t|2,...,t-1} . f(x_t \mid x_1, ..., x_{t-2}) \\ &= \left[\prod_{s=1}^{t-2} c_{s,t|s+1...,t-1}\right] . c_{(t-1),t} \times f_t(x_t) \end{aligned}$$

- Using (1) and $s = i$, $t = i + j$ it follows that

$$\begin{aligned} f(x_1, ..., x_n) &= \left[\prod_{t=2}^{n}\prod_{s=1}^{t-2} c_{s,t|s+1,...,t-1}\right] \cdot \left[\prod_{t=2}^{n} c_{(t-1),t}\right]\left[\prod_{k=1}^{n} f_k(x_k)\right] \\ &= \left[\prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{i,(i+j)|(i+1),...,(i+j-1)}\right] \cdot \left[\prod_{k=1}^{n} f_k(x_k)\right] \end{aligned}$$

Marginal conditional distributions

- many of the pair-copulas need to be evaluated at a conditional distribution of the form $F(x|\mathbf{v})$, where **v** denotes a vector of variables.
- The calculation of these conditional distributions is also recursive.
- Let $\mathbf{v}_{-j}$ denote the vector **v** but excluding the jth component v_j. For every j,

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v_{-j}}}(F(x|\mathbf{v_{-j}}), F(v_j|\mathbf{v_{-j}}))}{\partial F(v_j|\mathbf{v_{-j}})}$$

  where $C_{x,v_j|\mathbf{v_{-j}}}$ is a bivariate copula function.
- For the special case where **v** has only one component we have

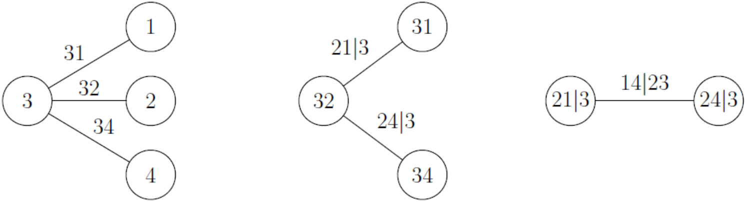
$$F(x|v) = \frac{\partial C_{x,v}(F_x(X), F_v(V))}{\partial F_v(V)}$$

Pair-Copula Constructions and Vines

- The above decomposition is called a pair-copula construction (PCC).
- The decomposition is not unique. That is, for high-dimensional distributions there are many possible pair-copula constructions.
- Bedford and Cooke (2002) introduced a graphical model called regular vine that help us organize a subset of all possible decompositions.
- The class of regular vines is large and embraces a large number of possible PCC's. Two special cases are:
  - D-Vine
  - Canonical Vine

  Both consists of sequences of trees that show us how to write a joint density function into pair-copulas and marginal densities

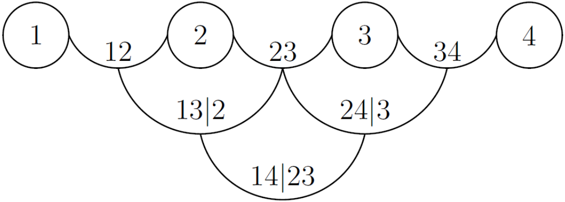
Canonical Vine Representation



$$
\begin{aligned}
f(x_1,x_2,x_3,x_4) = & f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_4) \\
& c_{31}(F_3(x_3),F_1(x_1))c_{32}(F_3(x_3),F_2(x_2))c_{34}(F_3(x_3),F_4(x_4)) \\
& c_{21|3}(F_{2|3}(x_2|x_3),F_{1|3}(x_1|x_3))c_{24|3}(F_{2|3}(x_2|x_3),F_{4|3}(x_4|x_3)) \\
& c_{14|23}(F_{1|23}(x_1|x_2,x_3),F_{4|23}(x_4|x_2,x_3))
\end{aligned}
$$

The intuition behind canonical vines is that one variable plays a key role in the dependency structure and so everyone is linked to it.

D-Vine Representation



$$
\begin{aligned}
f(x_1, x_2, x_3, x_4) = & f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \\
& c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) c_{34}(F_3(x_3), F_4(x_4)) \\
& c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{24|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \\
& c_{14|23}(F_{1|23}(x_1|x_2, x_3), F_{4|23}(x_4|x_2, x_3))
\end{aligned}
$$

Estimating the Pair-Copula Decomposition

The canonical or D-vine constructions decompose an *n*-dimensional multivariate density function into two main components.

- the product of each of the marginal density functions.
- the product of the density functions of n(n-1)/2 bivariate copulas.

To estimate the parameters of either construction we need to

1. decide which family to use for each pair-copula and
2. estimate all necessary parameters simultaneously

chi-plots (Fischer and Switzer, 1985)

- A chi-plot is a graphical method to help us extract information about the dependence between two random variables.
- The essence of the chi-plot is to compare the empirical bivariate distribution against the null hypothesis of independence at each point in the scatterplot.
- To construct this plot from a set of points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ we calculate three empirical distribution functions: the bivariate distribution H and the two marginal distributions F and G.
- For each point (x_i, y_i) let H_i be the proportion of points below and to the left of (x_i, y_i). Also let F_i and G_i be the proportion of points to the left and below of the point (x_i, y_i), respectively.

chi-plots (Fischer and Switzer (1985)

Each point (χ_i, λ_i) of the χ-plot is then defined by

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}} \tag{24}$$

and

$$\lambda_i = 4S_i \max\left\{\left(F_i - \frac{1}{2}\right)^2, \left(G_i - \frac{1}{2}\right)^2\right\}, \tag{25}$$

where

$$S_i = \text{sign}\left\{\left(F_i - \frac{1}{2}\right)\left(G_i - \frac{1}{2}\right)\right\}. \tag{26}$$

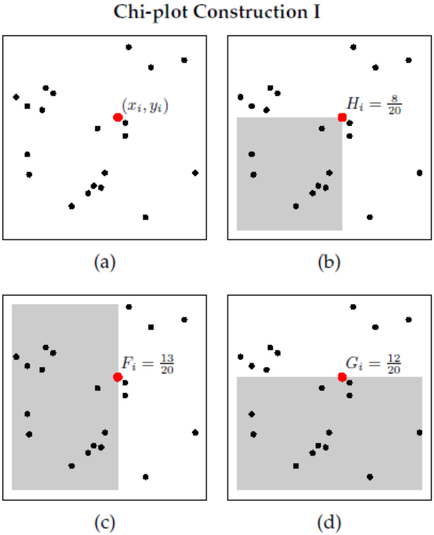
The formal definitions for H, F and G are

$$H_i = \frac{1}{n-1}\sum_{j \neq i} I(x_j \leq x_i, y_j \leq y_i), \tag{27}$$

$$F_i = \frac{1}{n-1}\sum_{j \neq i} I(x_j \leq x_i), \tag{28}$$

$$G_i = \frac{1}{n-1}\sum_{j \neq i} I(y_j \leq y_i), \tag{29}$$

chi-plots (Fischer and Switzer (1985)



Some References

Aas, K., Czado, C. Frigessi, A. and Bakken, H. (2009) Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, **44**, 182–198.

Bedford, T. and Cooke, R.M. (2002) Vines - a new graphical model for dependent random variables. *Annals of Statistics* **30**, 1031–1068.

Fisher, N. I., and Switzer, P. (1985) Chi-plots for assessing dependence. *Biometrika*, **72**, 253–265.

Min, A. and Czado, C. (2008) Bayesian inference for multivariate copulas using pair-copula constructions. Technical Report, Center for Mathematical Sciences, Munich University of Technology.

Schirmacher, D. and Schirmacher, E. (2008) Multivariate Dependence Modeling Using Pair-Copulas. Liberty Mutual Group. available at: `http://64.49.242.50/library/monographs/other-monographs/2008/april/mono-2008-m-as08-1-schirmacher.pdf`

Genest, C. Rémillard, B. and Beaudoin, D. (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**, 199–213.