COPULA MODELS IN LIFE INSURANCE

Jana Špirková, Vladimír Hiadlovský

ELEVENTH INTERNATIONAL CONFERENCE
ON FUZZY SET THEORY AND APPLICATIONS
2 February 2012

Groupe Consultatif Actuarial Europeen (European Actuarial Consultative Group)

Core Syllabus for Actuarial Training in Europe

- basic probability theory
- random variables and related concepts
- correlation and regression analysis
- simulation methods

Actuaries strive

- to understand stochastic outcomes of financial security systems
- to estimate of joint life mortality and multidecrement models

Copula

- a copula is a function that links univariate marginals to their full multivariate distribution
- copulas are useful for examining the dependence structure of multivariate random variables

Advantages of copula models

- their relative mathematical simlicity
- posibility to built a variety of dependence structures based on existing parametric or non-parametric models of the marginal distributions

The range of copulas applications

- civil engineering- reliability of analysis of highway bridges
- climate and weather related research
- analysis of extremas in financial assets and returns
- failure of paired organs in health science
- human mortality in insurance (actuarial science)
    - mortalities of spouses
    - mortalities of parents and children
    - twins (identical or nonidentical)

Actual science

- two lives are subject to failure
  - a joint life insurance
  - annuity (pension) insurance

Risk factors

- common disaster (fatal accidents involving both spouses)
- common lifestyle
- broken-heart syndrome

Basic properties of copulas

- p uniform (on the unit interval) random variables $u_1, u_2, \ldots, u_p$
- we do not assume that $u_1, u_2, \ldots, u_p$ are independent; yet they may be related
- this relationship is described through their joint distribution function

$$C(u_1, u_2, \ldots, u_p) = P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_p \leq u_p)$$

- $$C(F_1(x_1), F_2(x_2), \ldots, F_p(x_p)) = F(x_1, x_2, \ldots, x_p)$$

The copula construction does not constrain the choice of marginal distribution

Example 1

Suppose we are considering modeling male and female lifetimes for a joint-life annuity product.
Then, with $p = 2$, we might choose the Gompertz disribution to represent mortality at the older ages, yet with different parameters to reflect gender differencies in mortality.

Example 2

We consider a bivariate outcome associated with the loss and the expense associated with administering a property and casualty claim.
We could elect to use

- a lognormal distribution for expenses,
- a longer tail distribution (such as Pareto) for losses associated with the claim.

A copula function

Definition

A copula function is defined as a binary function $C: [0,1]^2 \to [0,1]$, which satisfies the following three properties:

1. $C(u,0) = C(0,u) = 0$ for any $u \in [0,1]$,
2. $C(u,1) = C(1,u) = u$ for any $u \in [0,1]$,
3. for all $0 \le u_1 \le u_2 \le 1$ and $0 \le v_1 \le v_2 \le 1$
   $C([u_1,v_1] \times [u_2,v_2]) =$
   $C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \ge 0$

$u = F_1(x_1)$, $v = F_2(x_2)$ are univariate distribution functions

$$C(u,v) = H\left(F_1^{-1}(u), F_2^{-1}(v)\right)$$

Shemyakin, Youn, 2011

Gaussian copula may be represented as

$$C^G(u, v; \rho) = \Phi_\rho^2\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$$

$$C^G(u, v; \rho) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left(-\frac{x_1^2 - \rho x_1 x_2 + x_2^2}{2 \cdot (1-\rho^2)}\right) dx_1 dx_2$$

$\Phi(x)$ is the standard normal distribution function.
$\Phi_\rho^2(x_1, x_2)$ is the bivariate normal distribution function with correlation ρ between the marginals.

Shemyakin, Youn, 2011

Student t-copula may be represented as

$$C^t(u,v;\rho,\nu) = t^2_{\rho,\nu}\left(t_\nu^{-1}(u), t_\nu^{-1}(v)\right)$$

$$C^t(u,v;\rho,\nu) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \frac{1}{\left(1+\frac{x_1^2-\rho x_1 x_2 + x_2^2}{\nu\cdot(1-\rho^2)}\right)} dx_2 dx_1$$

$t_\nu(x)$ is t-distribution function with ν degrees of freedom.
$t^2_{\rho,\nu}(x_1,x_2)$ is the bivariate t-distribution function with correlation ρ between the marginals.

It is also possible to use marginal survival function instead of marginal distribution functions

If the arguments of the copula function are univariate survival functions

$$S_1(x_1) = P(X_1 > x_1) \qquad S_2(x_2) = P(X_2 > x_2)$$

The coplula function is a legitimate joint (bivariate) survival function with marginals S_1, S_2

$$C(S_1, S_2) = S(x_1, x_2) = S(X_1 > x_1, X_2 > x_2)$$

Exponential distribution

Consider a claims random variable X that, given a risk classification parameter γ, can be modeled as an exponential distribution

$$P\left(X \leq x|\gamma\right) = 1 - exp(-\gamma x).$$

If γ has a Gamma distribution, then the marginal distribution of X is Pareto.

$$F(x) = P(X \leq x) = 1 - \left(1 + \frac{x}{\lambda}\right)^{-\alpha}$$

the joint distribution

$$F(x_1,x_2) = 1 - P(X_1 > x_1) - P(X_2 > x_2) + P\left(X_1 > x_1, X_2 > x_2\right)$$

$$= 1 - \left(1 + \frac{x_1}{\lambda}\right)^{-\alpha} - \left(1 + \frac{x_2}{\lambda}\right)^{-\alpha} + \left[1 + \frac{x_1 + x_2}{\lambda}\right]^{-\alpha}$$

$$= F_1(x_1) + F_2(x_2) - 1 + \left[(1 - F_1(x_1))^{-\frac{1}{\alpha}} + (1 - F_2(x_2))^{-\frac{1}{\alpha}} - 1\right]^{-\alpha}$$

This yields the copula function

$$C(u_1,u_2) = u_1 + u_2 - 1 + \left[(1-u_1)^{-\frac{1}{\alpha}} + (1-u_2)^{-\frac{1}{\alpha}} - 1\right]^{-\alpha}$$

A family of copulas

The two main methods for specifying a family of copulas are:

- the Archimedean approach,
- the compounding approach.

The Archimedean copula

Suppose that $\Phi : [0, \infty] \rightarrow [0, 1]$ is a strictly decreasing convex function such that $\Phi(0) = 1$.
Then an Archimedean copula may be generated as

$$C(u, v; \alpha) = \Phi\left(\Phi^{-1}(u) + \Phi^{-1}(v)\right) ; \; u, v \in]0, 1],$$

α is the parameter of association.

Frank's copula

generated by

$$\Phi^{-1}(u) = -log\left[\frac{exp(-\alpha u) - 1}{exp(-\alpha) - 1}\right]$$

with association $\alpha \neq 0$, is given by

$$C_F(u, v; \alpha) = -\frac{1}{\alpha}\left[1 + \frac{(exp(-\alpha u) - 1) \cdot (exp(-\alpha v) - 1)}{exp(-\alpha) - 1}\right].$$

Clayton's copula

generated by

$$\Phi^{-1}(u) = \frac{1 - u^{\alpha}}{\alpha u^{\alpha}}$$

with association $\alpha > 0$, is given by

$$C_C(u, v; \alpha) = \left(u^{-\alpha} + v^{-\alpha} - 1\right)^{\frac{-1}{\alpha}}$$

Stable (Gumbel-Hougaard) copula

generated by

$$\Phi^{-1}(u) = (-logu)^{\alpha}$$

with association $\alpha \geq 1$, can be represented as

$$C_{GH}(u, v; \alpha) = exp\left\{-\left[(-logu)^{\alpha} + (-logv)^{\alpha}\right]^{1/\alpha}\right\}$$

Joint survival analysis

Assume

- for lives L_j let X_j be “lifelenth” random variables (age at death),
- associated pairs of lives L_1 and L_2 are observed during a certain limited period of time T,
- an observation begins at entry age a_1, a_2;
- a_1, a_2 represent effect of “left trancation”,
- $a_1 + T$, $a_2 + T$ represent “right censoring”.

In a sample $y = (y_1, y_2, \ldots, y_n)$ each $i - th$ observation y_i of an associated pair of lives (L_{i1}, L_{i2}) may be represented as a vector

$$y_i = (a_{i1}, a_{i2}, t_{i1}, t_{i2}, c_{i1}, c_{i2}),$$

- t_{ij} is the termination time for life L_{ij}
- c_{ij} is the censoring indicator

$$c_{ij} = \begin{cases} 0 & t_{ij} = T \text{ (censoring)} \\ 1 & t_{ij} < T \text{ (no censoring)} \end{cases}$$

One of the important tasks is to estimate the future lifelenth probabilities for given entry ages:

- the joint first-life survival function

$$p_{FL}(t; a_1, a_2) =$$

$$= P\left(min\left\{X_1 - a_1,\, X_2 - a_2\right\} > t \middle| min\left\{X_1 - a_1,\, X_2 - a_2\right\} > 0\right)$$

- the joint last-survivor function

$$p_{LS}(t; a_1, a_2) =$$

$$= P\left(max\left\{X_1 - a_1,\, X_2 - a_2\right\} > t \middle| min\left\{X_1 - a_1,\, X_2 - a_2\right\} > 0\right)$$

Survival functions

$$S_1(t_1) = P(X_1 > t_1) \qquad S_2(t_2) = P(X_2 > t_2)$$

The copula function is a legitimate joint (bivariate) survival function with marginals S_1, S_2

$$C(S_1, S_2) = S(t_1, t_2) = S(X_1 > t_1, X_2 > t_2)$$

A very natural approach to the estimation of $p_{FL}(t; a_1, a_2)$ and $p_{LS}(t; a_1, a_2)$ is to first estimate the bivariate survival function $S(t_1, t_2)$ and then use the formulas

$$p_{FL}(t; a_1, a_2) = \frac{S(a_1 + t, a_2 + t)}{S(a_1, a_2)}$$

$$p_{LS}(t; a_1, a_2) = \frac{S(a_1, a_2 + t) + S(a_1 + t, a_2) - S(a_1 + t, a_2 + t)}{S(a_1, a_2)}$$

The choice of two-parameter Weibull distribution

for survival function

$$S(t_j) = P(X_j > t_j) = exp\left\{-\left(\frac{t}{\beta_j}\right)^{\gamma_j}\right\};\ t \geq 0$$

Gumbel-Hougaard copula for the model of association

$$C_{GH}(u, v; \alpha) = exp\left\{-\left[(-log\, u)^{\alpha} + (-log\, v)^{\alpha}\right]^{1/\alpha}\right\}$$

$$S(t_1, t_2) = C_{GH}(S(t_1; \beta_1, \gamma_1), S(t_2; \beta_2, \gamma_2); \alpha)$$

$$S(t_1, t_2) = exp\left\{-\left[\left(\frac{t}{\beta_1}\right)^{\alpha\gamma_1} + \left(\frac{t}{\beta_2}\right)^{\alpha\gamma_2}\right]^{1/\alpha}\right\}$$