Copula and multivariate dependencies

Eric Marsden

<eric.marsden@risk-engineering.org>

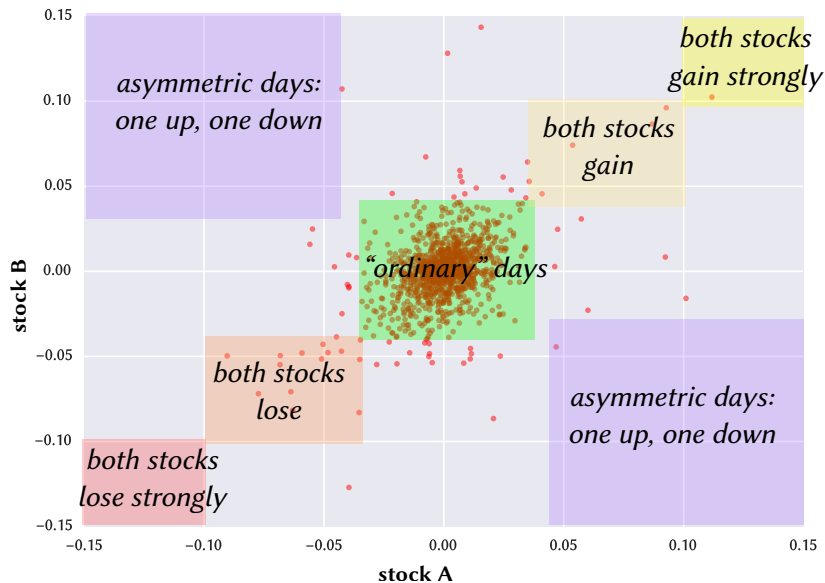**Warmup**. Before reading this material, we suggest you consult the following associated slides:

▷ Slides on *Modelling correlations with Python*

▷ Slides on *Estimating Value at Risk*

Available from risk-engineering.org & slideshare.net

Dependencies and risk: stock portfolios

Dependencies and risk: life insurance

▷ Correlation of deaths for life insurance companies
- marginal distributions: probabilities of time until death for each spouse
- joint distribution shows the probability of spouses dying in close succession

▷ Aim (actuarial studies): estimate the conditional probability when one spouse dies, that the succeeding spouse will die shortly afterwards

▷ Common risk factors:
- common disaster (fatal accidents involving both spouses)
- common lifestyle
- "broken-heart syndrome"

RISK
ENGINEERING

Dependencies and risk: bank loans

▷ Correlation of default: what is the likelihood that if one company defaults, another will default soon after?

- $r = 0$: default events are independent

- perfect positive correlation (r=1): if one company defaults, the other will automatically do the same

- perfect negative correlation: if one company defaults, the other will certainly not
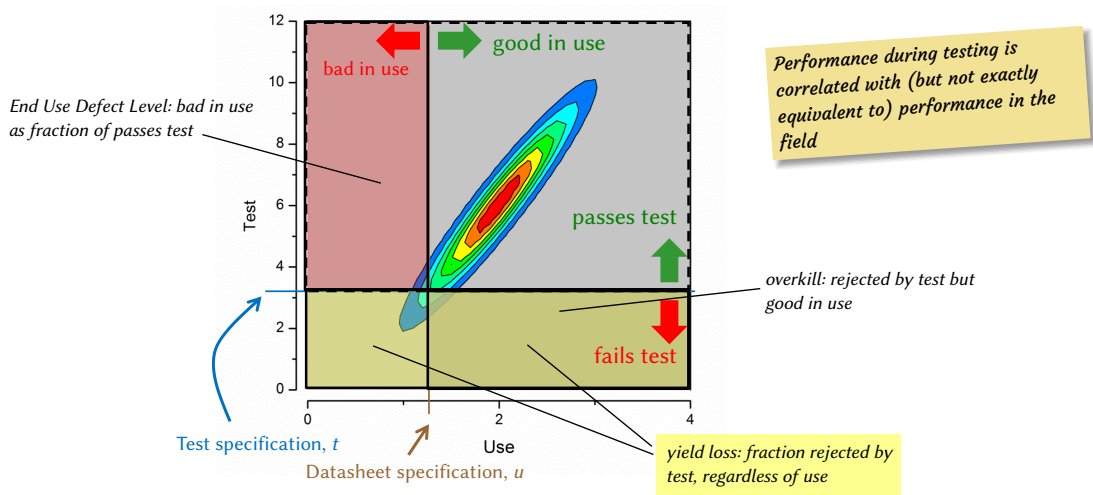

borrower j

j pays, k defaults

both j & k pay

k pays, j defaults

both j & k default

borrower k
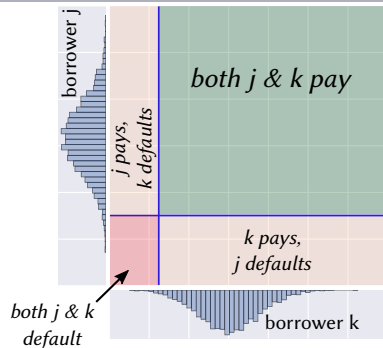
Example of correlated defaults

A bank lends money to two companies: a dairy farm and a dairy. The farm has a 10% chance of going bust and the dairy a 5% chance. But if the farm does go under, the chances that the dairy will follow will quickly rise above 5% if the farm was its main milk supplier.

Poor estimation of these correlations led to failure of LTCM hedge fund (USA, 1998)

RISK ENGINEERING

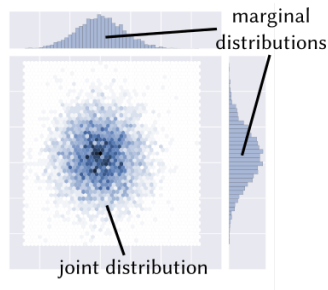# Dependencies and risk: testing in semiconductor manufacturing

End Use Defect Level: bad in use as fraction of passes test

bad in use

good in use

passes test

overkill: rejected by test but good in use

fails test

Performance during testing is correlated with (but not exactly equivalent to) performance in the field

Test specification, t

Datasheet specification, u

yield loss: fraction rejected by test, regardless of use

Source: *Copula Methods in Manufacturing Test: A DRAM Case Study*, C. Glenn Shirley & W. Robert Daasch

Dependencies and risk: other applications

Modelling dependencies is an important and widespread issue in risk analysis:

▷ Civil engineering: reliability analysis of highway bridges

▷ Insurance industry: estimating exposure to *systemic risks*
  • a hurricane causes deaths, property damage, vehicle damage, business interruption…

▷ Medicine: failure of paired organs

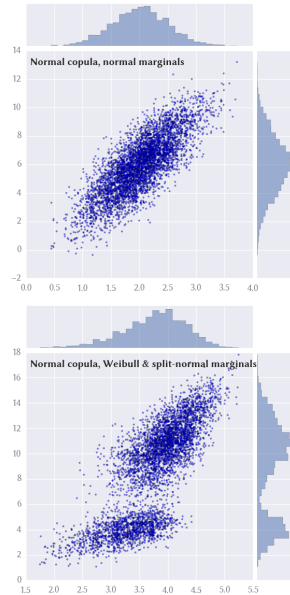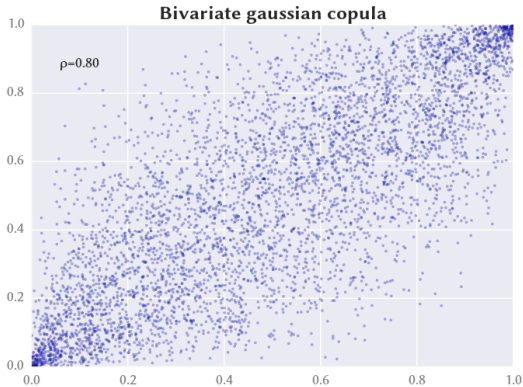▷ Note: often the dependency is more complicated than a simple linear correlation…

RISK
ENGINEERING

Copula

marginal distributions

joint distribution

▷ Latin word that means "to fasten or fit"

▷ A **bridge** between marginal distributions and a joint distribution

- dependency between stocks (*e.g.* CAC40 & DAX)

- dependency between defaults on loans

- dependency between annual peak of a river and volume (hydrology)

▷ Widely used in quantitative finance & insurance

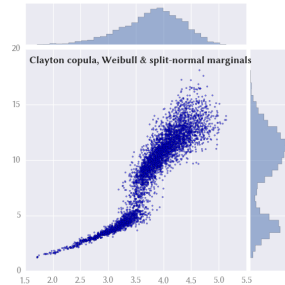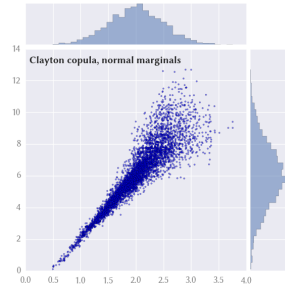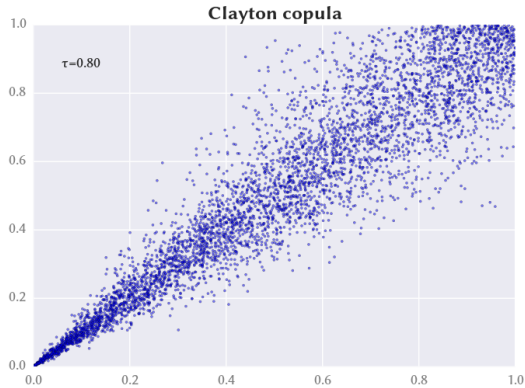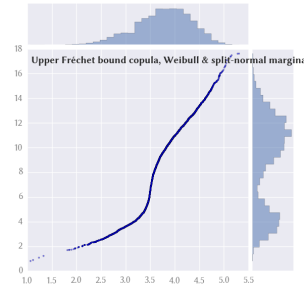▷ Let's look at plots of a few 2D copula functions to try to visualize their impact

R!SK
ENGINEERING

Same copula, different marginals

Bivariate gaussian copula

ρ=0.80

Normal copula, normal marginals

Normal copula, Weibull & split-normal marginals

RISK ENGINEERING

Another copula

Clayton copula

τ=0.80

Clayton copula, normal marginals

Clayton copula, Weibull & split-normal marginals

Copula representing perfect correlation

Upper Fréchet-Hoeffding bound copula

Upper Fréchet bound copula, normal marginals

Upper Fréchet bound copula, Weibull & split-normal marginals

Copula representing perfect negative correlation

Fréchet lower bound copula

Lower Fréchet bound copula, normal marginals

Lower Fréchet bound copula, Weibull & split-normal marginals
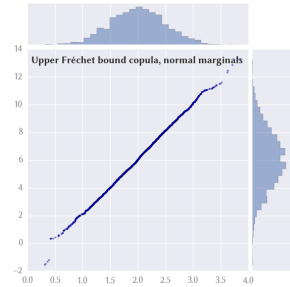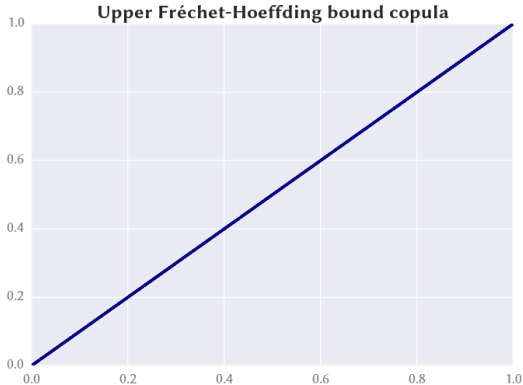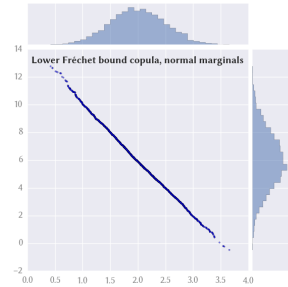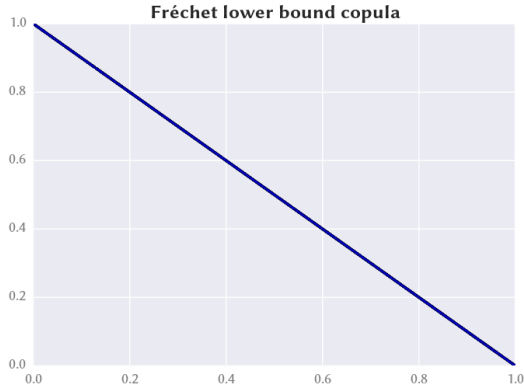
Copula representing independence

Independence copula

Independence copula, normal marginals


Independence copula, Weibull & split-normal marginals
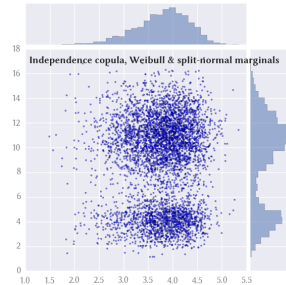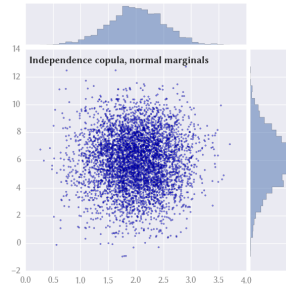
Copula: definitions

▷ Copula functions are a tool to separate the specification of marginal distributions and the dependence structure

- unlike most multivariate statistics, allow combination of different marginals

▷ Say two risks A and B have joint probability $H(X, Y)$ and marginal probabilities F_X and F_Y

- $H(X, Y) = C(F_X, F_Y)$
- C is a copula function

▷ Characteristics of a copula:

- $C(1, 1) = 1$
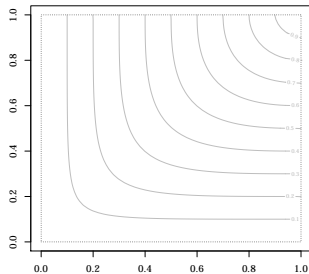- $C(x, 0) = 0$
- $C(0, y) = 0$
- $C(x, 1) = x$
- $C(1, y) = y$
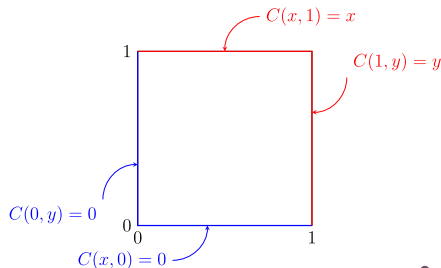
Gaussian copula, dimension 2, rho=0.8
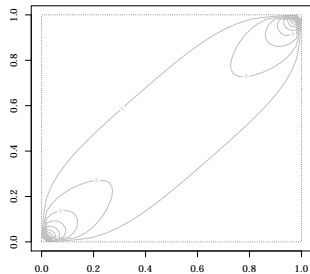
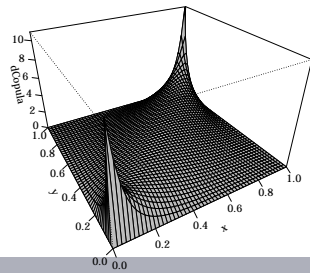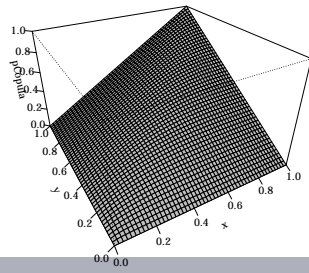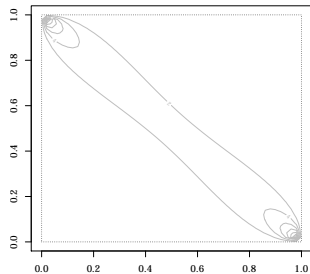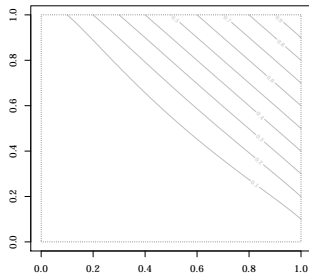

PDF

CDF

PDF

CDF

RISK ENGINEERING
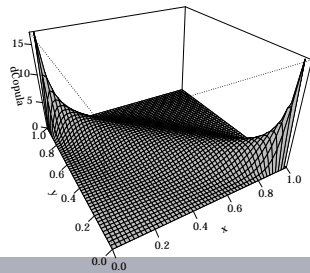
Gaussian copula, dimension 2, rho=-0.9
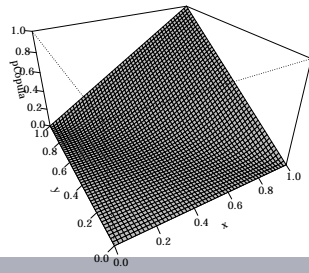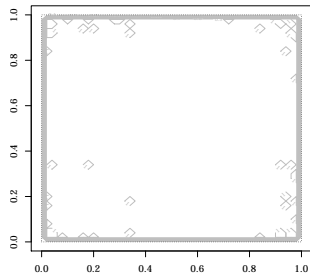
Gaussian copula, dimension 2, rho=0
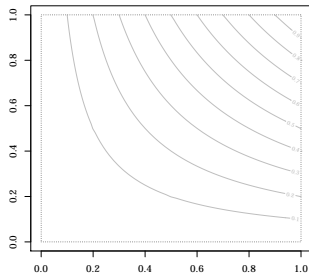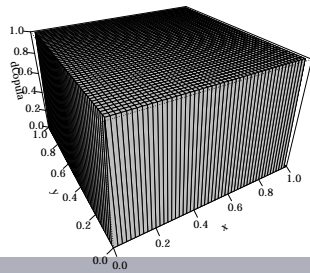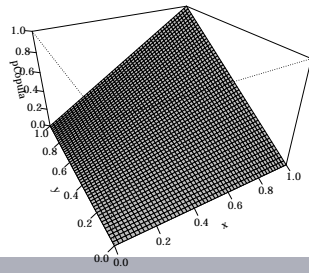
Gaussian copula, dimension 3

RISK
ENGINEERING
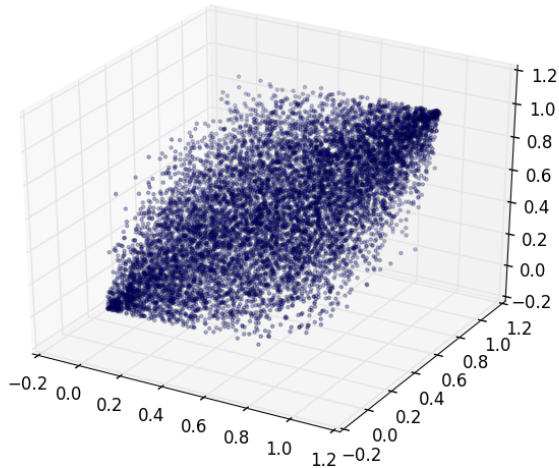
Student t copula, dimension 2, rho=0.8
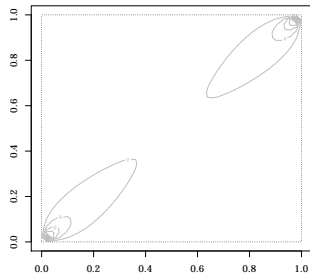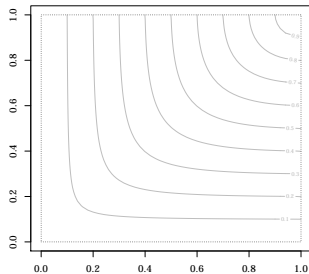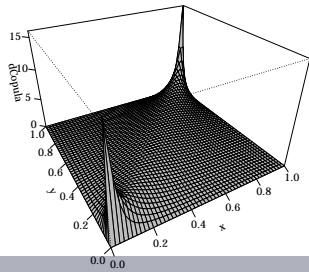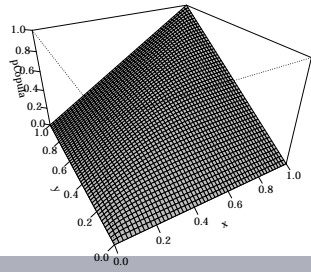
Gumbel copula, dimension 2, rho=0.8
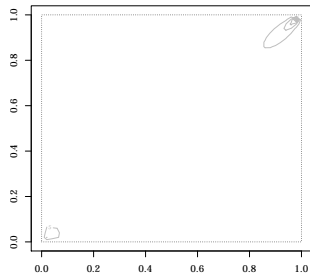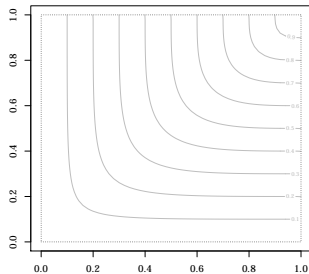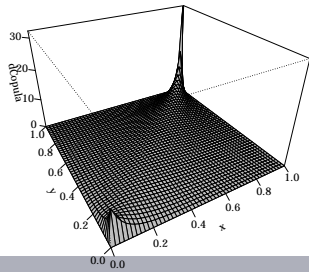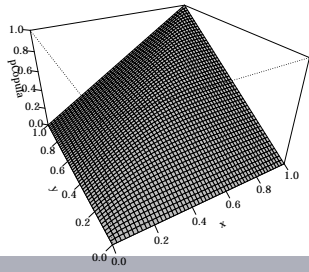
Clayton copula, dimension 2, rho=0.8
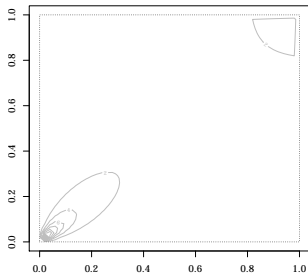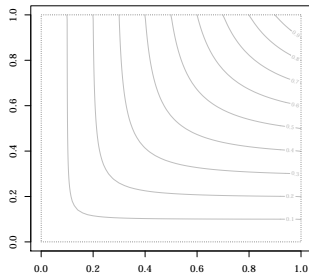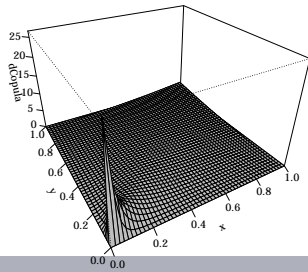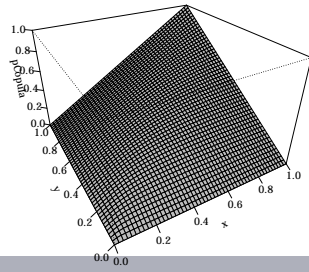
Independence copula, $C(u, v) = u \times v$

PDF

CDF

PDF

CDF

RISK ENGINEERING
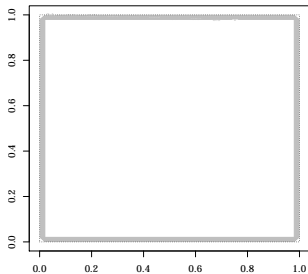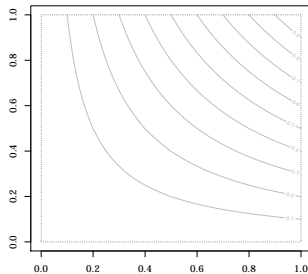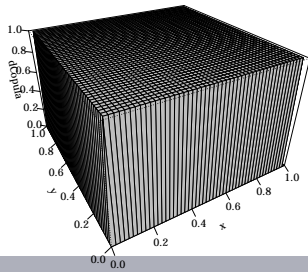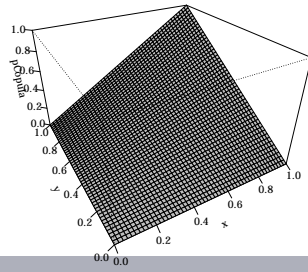
Copula representing perfect positive dependence

$C(u, v) = \min(u, v)$

Upper Fréchet-Hoeffding bound copula
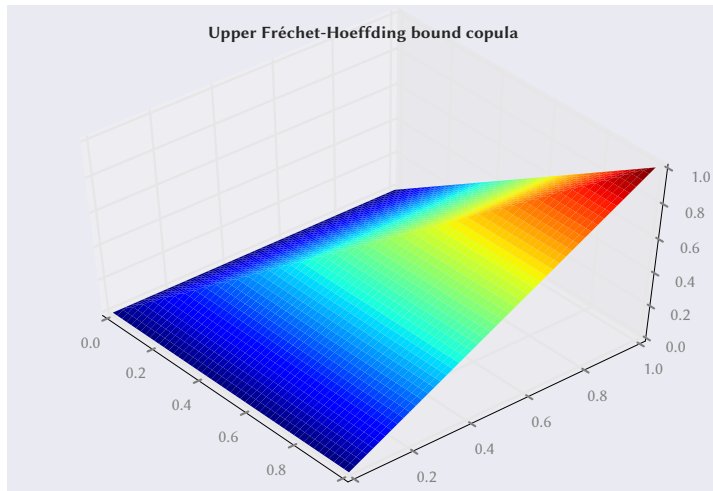
Copula representing perfect negative dependence

$C(u, v) = \max(u + v - 1, 0)$
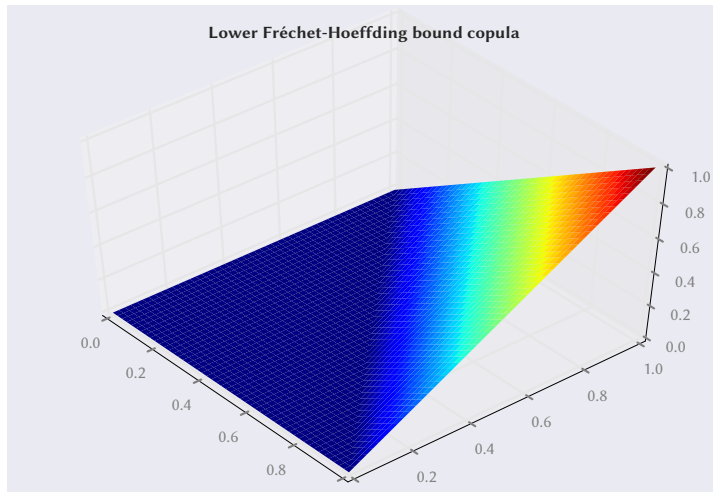
Tail dependence

▷ Risk management is concerned with the **tail of the distribution** of losses

▷ Large losses in a portfolio are often caused by simultaneous large moves in several components

▷ One interesting aspect of any copula is the probability it gives to **simultaneous extremes in several dimensions**

▷ The *lower tail dependence* of X_i and X_j is defined as

$$\lambda_l = \lim_{u \to 0} \Pr[X_i \leq F_i^{-1}(u) | X_j \leq F_j^{-1}(u)]$$

▷ Only depends on the copula, and is $\lim_{u \to 0} \frac{1}{u} C_{i,j}(u, u)$

▷ Tail dependence is symmetric
  • tail dependence of X_i and X_j is the same as that of X_j and X_i

▷ *Upper tail dependence* λ_u is defined similarly

RISK
ENGINEERING

Mathematical recap: joint probability distribution

▷ Joint probability distributions are defined in the form below:

$$f(x, y) = \Pr(X = x, Y = y)$$

representing the probability that events x and y occur at the same time

▷ The Cumulative Distribution Function (CDF) for a joint probability distribution is given by:

$$F_{XY}(x, y) = \Pr(X \leq x, Y \leq y)$$

▷ Note: examples here are bivariate, but principles are valid for multivariate distributions

RISK
ENGINEERING

Joint distribution — discrete case

	b_1	b_2	b_3	...	b_k
a_1	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	...	$p_{1,k}$
a_2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	...	$p_{2,k}$
...	...	...	...	...	...
...	...	...	...	...	...
a_m	$p_{m,1}$	$p_{m,2}$	$p_{m,3}$	...	$p_{m,k}$

$\Pr(A = a_i \text{ and } B = a_j) = p_{i,j}$

RISK ENGINEERING

Joint distribution — continuous case

▷ $f_{XY} : R^n \to R$

▷ $f_{XY} \geq 0 \quad \forall v \in R^n$

▷ $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) = 1$

▷ $\Pr(v \in B) = \iint\limits_B f_{XY}$

▷ $\Pr(X \leq x) = F_X(x) = \int_{-\infty}^{x} F_X(z) dz$

▷ $\Pr(Y \leq y) = F_Y(y) = \int_{-\infty}^{y} F_Y(z) dz$

▷ $\Pr(X \leq x, Y \leq y) = F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(w, z) dw \, dz$
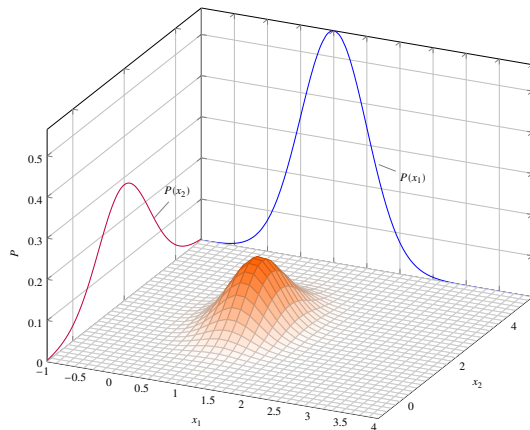
Bivariate gaussian distribution

Marginal distributions — discrete case

	b_1	b_2	b_3	...	b_k
a_1	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	...	$p_{1,k}$
a_2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	...	$p_{2,k}$
...	...	...	...	...	...
...	...	...	...	...	...
a_m	$p_{m,1}$	$p_{m,2}$	$p_{m,3}$	...	$p_{m,k}$

$$p_X(x) = \Pr(X = x) = \sum_y p(x, y)$$

RISK
ENGINEERING

Marginal distributions — discrete case

	b_1	b_2	b_3	...	b_k
a_1	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	...	$p_{1,k}$
a_2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	...	$p_{2,k}$
...	...	...	...	...	...
...	...	...	...	...	...
a_m	$p_{m,1}$	$p_{m,2}$	$p_{m,3}$	...	$p_{m,k}$

$$p_X(x) = \Pr(X = x) = \sum_y p(x, y)$$

$$p_Y(y) = \Pr(Y = y) = \sum_x p(x, y)$$

RISK ENGINEERING

Marginal distributions — continuous case

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

RISK ENGINEERING

Sklar's theorem

For every joint probability distribution F_{XY} there is a copula C such that:

copula function

$$F_{XY}(x, y) = C(F_X(x), F_Y(y))$$

marginal distribution of X

marginal distribution of Y
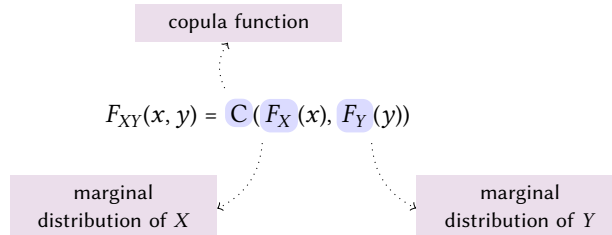
If F_{XY} is continuous then C is unique.

RISK ENGINEERING

Copula: summary

▷ Copula captures the *dependency* between the random variables

▷ Marginals capture individual distributions

▷ Sklar's theorem "glues" them together

▷ "Shape" and degree of joint tail dependence is a copula property
   • are independent of the marginals

Understanding the copula function

space of our "real variables"

Y

X
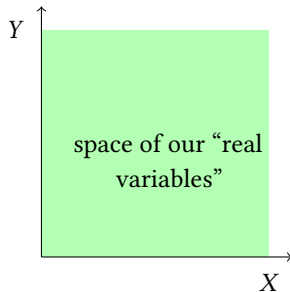
Understanding the copula function

$F_{XY}(x, y) = t$ is the probability that $X < x$ and $Y < y$

joint distribution function

$F_{XY}(x, y) = t$
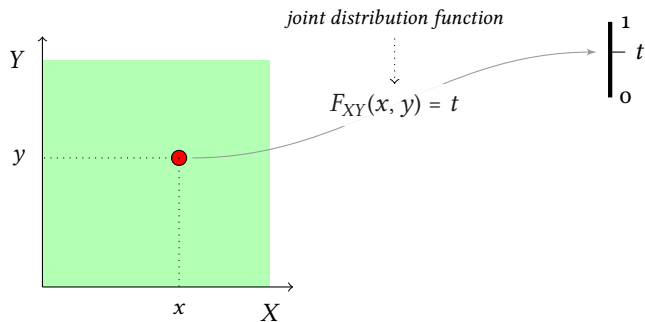
Understanding the copula function
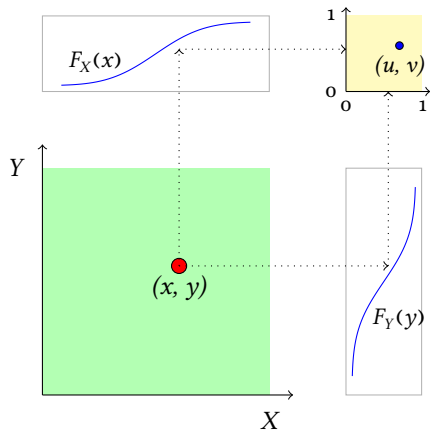
We can use the marginal CDFs to map from (x, y) to a point (u, v) on the unit square.

Understanding the copula function
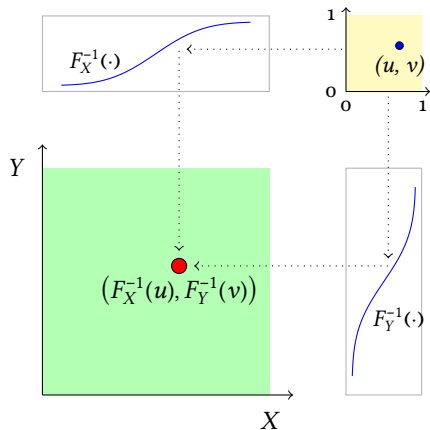
We can use the marginal inverse CDFs to map from (u, v) to $(F_X^{-1}(u), F_Y^{-1}(v))$.

Understanding the copula function

Then map from $(F_X^{-1}(u), F_Y^{-1}(v))$ to $[0, 1]$ using the joint distribution function.

joint distribution function

$F_{XY}(\cdot, \cdot)$

$(F_X^{-1}(u), F_Y^{-1}(v))$

Y

X

$$1$$
$$t = F_{XY}\left(F_X^{-1}(u), F_Y^{-1}(v)\right)$$
$$0$$

RISK ENGINEERING
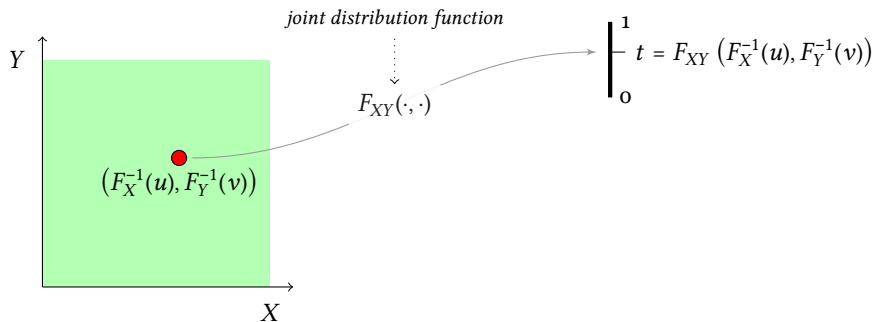
Understanding the copula function

$$C(u, v) = F_{XY}\,(F_X^{-1}(u), F_Y^{-1}(v))$$
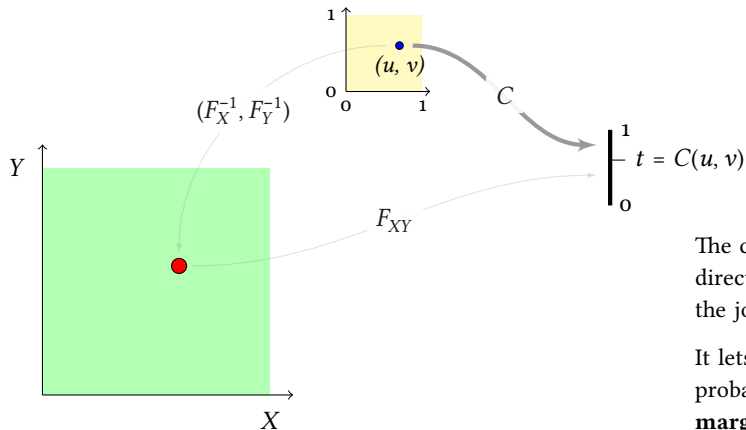
The copula function lets us map directly from the unit square to the joint distribution.

It lets us express the joint probability **as a function of the marginal distributions**.

$$F_{XY}(x, y) = C\,(F_X(x), F_Y(y))$$

Multivariate simulation...

We generate random points on $[0, 1]^2$ using the copula function random generator.
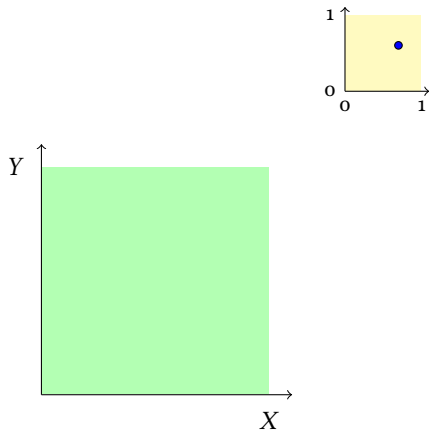
Multivariate simulation...
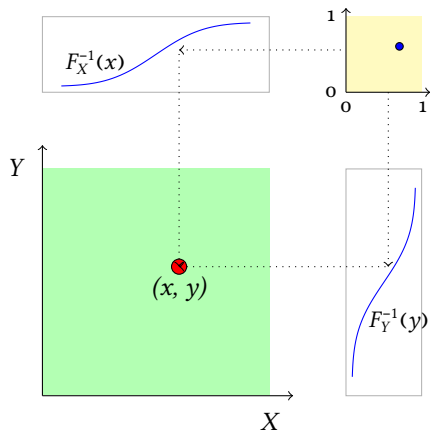
We generate random points on $[0, 1]^2$ using the copula function random generator.

We use the inverse CDFs to generate red points in the space of our real variables.

The joint distribution of the red points has marginals F_X and F_Y, with the required dependency structure.

Simulating dependent random vectors

Various situations in practice where we might wish to simulate dependent random vectors $(X_1, ..., X_n)^t$:

▷ **finance**: simulate the future development of the values of assets in a portfolio, where we know these assets to be dependent in some way

▷ **insurance**: "multi-line products" where payouts are triggered by the occurrence of losses in one or more dependent business lines, and wish to simulate typical losses

▷ **environmental modelling**: measures such as wind speed, temperature and atmospheric pressure are correlated

RISK
ENGINEERING

Simulation of correlated stock returns

(Coming back to our estimation of VaR of a CAC40/DAX stock portfolio)

CAC/DAX returns, dependency simulated with t copula

t copula with σ=0.941300, df=2.869400
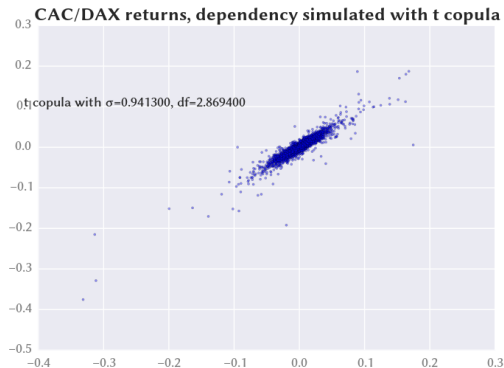
Simulation using the following parameters:

▷ CAC: student-t distribution with tμ = 0.000505, tσ = 0.008974, df = 2.768865

▷ DAX: student-t distribution with tμ = 0.000864, tσ = 0.008783, df = 2.730707

▷ dependency: t copula with ρ=0.9413, df = 2.8694

Assumption: the correlation structure does not change with time
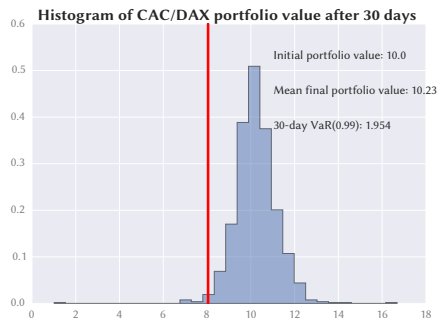
RISK ENGINEERING

VaR of a CAC-DAX portfolio

- ▷ 10 M€ portfolio, equally weighted between CAC and DAX indexes
  - CAC: daily returns with Student's t with $t\mu = 0.000505$, $t\sigma = 0.008974$, df = 2.768865
  - DAX: daily returns with Student's t with $t\mu = 0.000864$, $t\sigma = 0.008783$, df = 2.730707
  - Dependency between CAC and DAX indexes modelled using a Student t copula with $\rho = 0.9413$, df=2.8694

- ▷ Monte Carlo simulation of portfolio returns

- ▷ 30-day VaR(0.99) is 1.95 M€

Histogram of CAC/DAX portfolio value after 30 days

Initial portfolio value: 10.0

Mean final portfolio value: 10.23

30-day VaR(0.99): 1.954

Download the associated Python notebook at `risk-engineering.org`
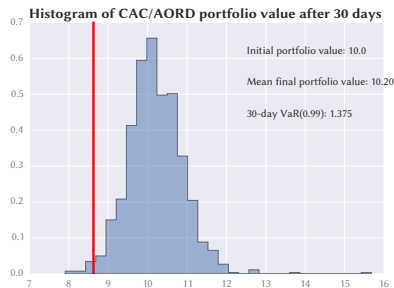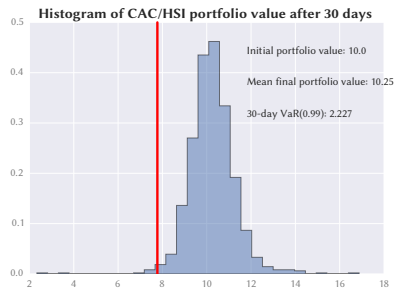
RISK ENGINEERING

VaR of a CAC-AORD portfolio

▷ 10 M€ portfolio, equally weighted between CAC and AORD indexes

- CAC: daily returns with Student's t with $t\mu$ = 0.000505, $t\sigma$ = 0.008974, df = 2.768865

- AORD: daily returns with Student's t with $t\mu$ = 0.0007309, $t\sigma$ = 0.0082751, df = 3.1973

- Dependency between CAC and AORD indexes modelled using a Student t copula with ρ=0.3101, df=2.795

▷ Monte Carlo simulation of portfolio returns

▷ 30-day VaR(0.99) is 1.37 M€

▷ **Lower than for CAC-DAX portfolio** because of lower degree of dependency!

Histogram of CAC/AORD portfolio value after 30 days

Initial portfolio value: 10.0

Mean final portfolio value: 10.20

30-day VaR(0.99): 1.375

RISK ENGINEERING

VaR of a CAC–HSI portfolio

▷ 10 M€ portfolio, equally weighted between CAC and HSI indexes

- CAC: daily returns with Student's t with tμ = 0.000505, tσ = 0.008974, df = 2.768865

- HSI: daily returns with Student's t with tμ = 0.000988, tσ = 0.01032, df = 2.269018

- Dependency between CAC and DAX indexes modelled using a Student t copula with ρ=0.35695, df=2.542247

▷ Monte Carlo simulation of portfolio returns

▷ 30-day VaR(0.99) is 2.23 M€

▷ **Higher than for CAC-DAX portfolio** despite lower degree of dependency (why?)

Histogram of CAC/HSI portfolio value after 30 days

Initial portfolio value: 10.0

Mean final portfolio value: 10.25

30-day VaR(0.99): 2.227

Further reading

▷ Teaching material by Prof. Paul Embrecht from ETH Zurich, an important researcher in the use of copula techniques in finance: `qrmtutorial.org`

▷ Article *'The Formula That Killed Wall Street'? The Gaussian Copula and the Material Cultures of Modelling* by Donald MacKenzie and Taylor Spears

▷ Slides on extreme and correlated risks by Prof. Arthur Charpentier: `perso.univ-rennes1.fr/arthur.charpentier/slides-edf-2.pdf`

For more free course materials on risk engineering,
visit `https://risk-engineering.org/`

RISK
ENGINEERING

Feedback welcome!

Was some of the content unclear? Which parts of the lecture were most useful to you? Your comments to feedback@risk-engineering.org (email) or @LearnRiskEng (Twitter) will help us to improve these course materials. Thanks!

@LearnRiskEng

fb.me/RiskEngineering

google.com/+RiskengineeringOrgCourseware

For more free course materials on risk engineering, visit https://risk-engineering.org/

RISK ENGINEERING