USING COPULAS
An introduction for practitioners

DANIEL BERG

DnBNOR Asset Management

Norwegian ASTIN society. Oslo - November 2008

DnBNOR Kapitalforvaltning ASA

- Stor bredde og dybde i forvalterkompetanse, 100 analytikere og porteføljeforvaltere
- Bredt produktspekter - og gode dokumenterte forvaltningsresultater.
- Gode systemer for risikostyring og - kontroll. Store volumer - kostnadseffektiv forvaltning
- Kontinuerlig prosess med produktutvikling og -forbedring.
- Ca 300 aarsverk. Ca NOK 500 milliarder til forvaltning.
  - Personkunder
  - Institusjonelle investorer
    - Ca 500 kunder i Norge og Sverige
    - Viktigste kundesegmenter: pensjonskasser, kommuner, bedrifter, organisasjoner/stiftelser
    - Høy raadgivningskompetanse
  - Strategiske mandat
    - Vital
    - Skandia Liv
    - Skandia Fonder

Outline

- Introduction
- Popular copula families
- Simulation
- Parameter estimation
- Model selection
- Model evaluation
- Examples (uranium, river flow/temp, precipitation, ...)
- Extensions
- Summary

Introduction

Motivation

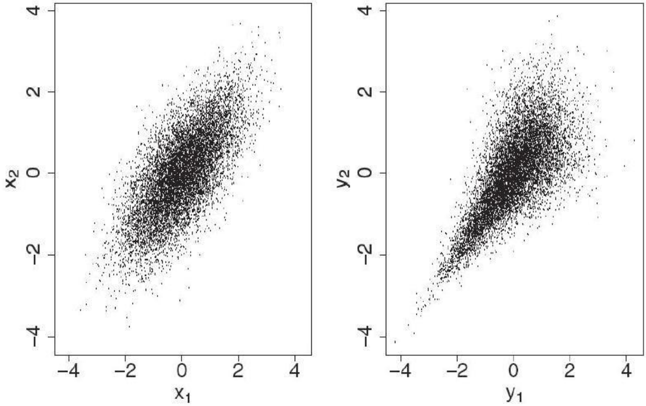


Figure: Simulations from two models, both with standard normal margins and correlation 0.7.

Introduction

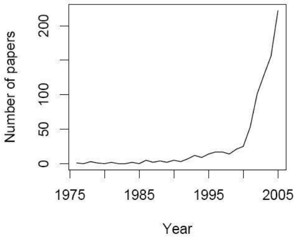
Motivation

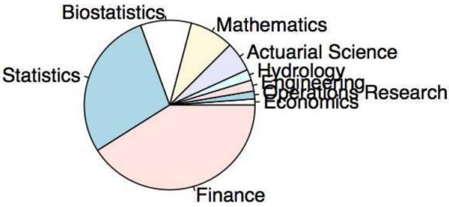
- For several applications empirical evidence has proved multinormal distribution inadequate for several reasons:
- Empirical marginal distributions are e.g. skewed and heavy-tailed
- Possibilities of extreme co-movements, in contrast to the multinormal distribution

Introduction

Brief historical background

- 1940: Hoeffding studies properties of multivariate distributions
- 1959: The word copula appears for the first time (Sklar, 1959)
- 1999: Introduced to financial applications (Embrechts et al., 1999)
- 2008: Widely used in insurance, finance, energy, hydrology, survival analysis, etc.





Introduction

Definition & theorem

Definition (Copula)

A d-dimensional copula is a multivariate distribution function $\mathcal{C}$ with standard uniform marginal distributions.

Theorem (Sklar, 1959)

Let H be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

Introduction

Useful results

- A general d-dimensional density h can be expressed, for some copula density c, as

$$h(x_1, \ldots, x_d) = c\{F_1(x_1), \ldots, F_d(x_d)\} f_1(x_1) \cdots f_d(x_d).$$

- Non-parametric estimate for $F_i(x_i)$ commonly used to transform original margins into standard uniform:

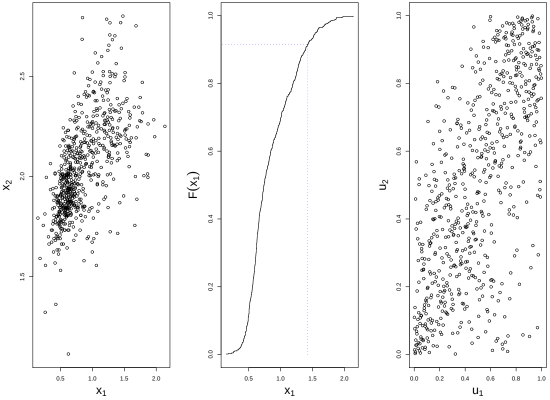
$$z_{ji} = \widehat{F}_i(x_{ji}) = \frac{R_{ji}}{n+1},$$

  where R_{ji} is the rank of x_{ji} amongst $x_{1i}, \ldots, x_{ni}$.
- z_{ji} commonly referred to as *pseudo-observations* and models based on non-parametric margins and parametric copulas are referred to as *semi-parametric* copulas
- Convenient to use empirical copula C_n as a consistent estimator of C (Deheuvels, 1979):

$$C_n(\boldsymbol{u}) = \frac{1}{n+1} \sum_{j=1}^{n} I\{z_{j1} \leq u_1, \ldots, z_{jd} \leq u_d\}$$

Introduction

Useful results



Introduction

Attractive features

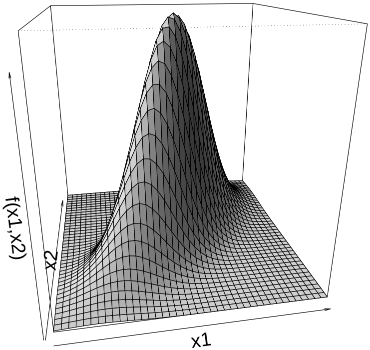
- The copula contains all the information about the dependence between random variables
- Copulas provide an alternative and often more useful representation of multivariate distribution functions compared to traditional approaches such as multivariate normality
- Most traditional representations of dependence are based on the linear correlation coefficient - restricted to multivariate elliptical distributions. Copula representations of dependence are free of such limitations.
- Copulas enable us to model marginal distributions and the dependence structure separately
- Copulas provide greater modeling flexibility, given a copula we can obtain many multivariate distributions by selecting different margins
- Any multivariate distribution can serve as a copula
- A copula is invariant under strictly increasing transformations
- Most traditional measures of dependence are measures of pairwise dependence. Copulas measure the dependence between all d random variables

Popular copula families

- Independence copula: $C_\Pi(u,v) = uv$
- Gaussian copula: $\mathcal{C}_\rho(uv) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\} \mathrm{d}x\mathrm{d}y$
- Student copula:
  $\mathcal{C}_{\rho,\nu}(u,v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{1 + \frac{x^2-2\rho xy+y^2}{\nu(1-\rho^2)}\right\}^{-(\nu+2)/2} \mathrm{d}x\mathrm{d}y$
- Archimedean copulas: $C_\theta(u,v) = \phi^{-1}\{\phi(u)+\phi(v)\}$ where ϕ is the copula generator.
  - Clayton copula: $C_\delta(u,v) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta}$
  - Gumbel copula: $C_\eta(u,v) = \exp\left[-((-\ln u)^\theta + (-\ln v)^\theta)^{1/\theta}\right]$

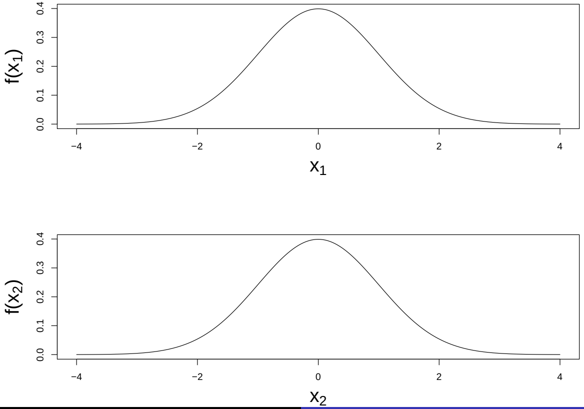
Popular copula families

Take the bivariate std. normal prob. density...



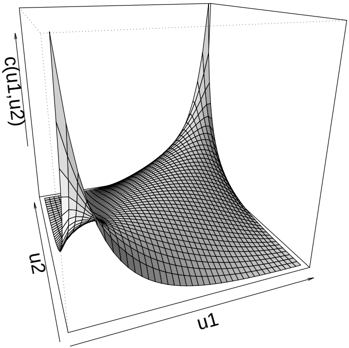
Popular copula families

...divide by std. normal marginal densities...



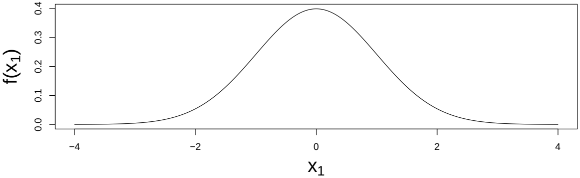
Popular copula families

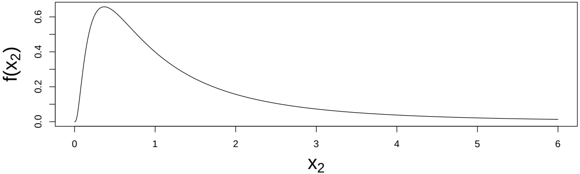
...and we obtain the normal copula density.



Popular copula families

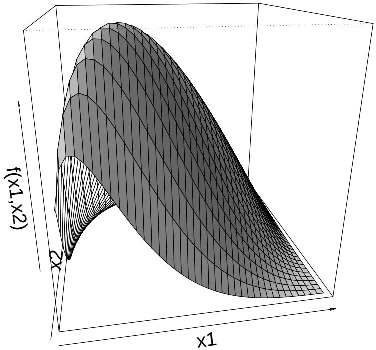
Now multiply the normal copula density with these...





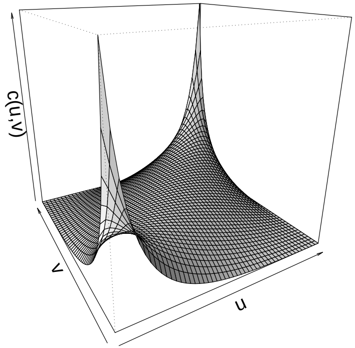
Popular copula families

...and we obtain...



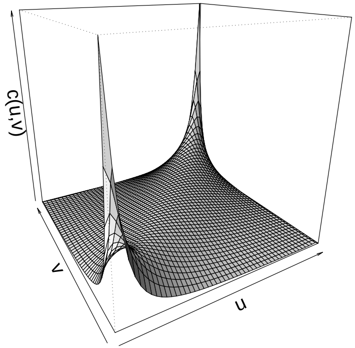
Popular copula families

Gaussian copula



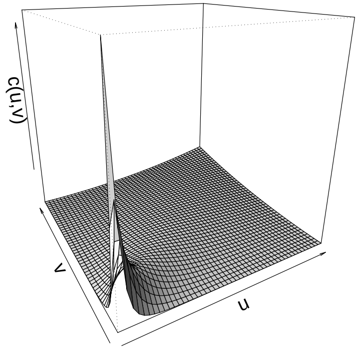
Popular copula families

Student copula



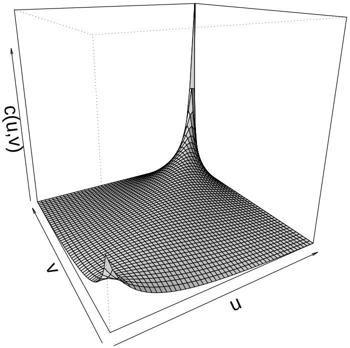
Popular copula families

Clayton copula

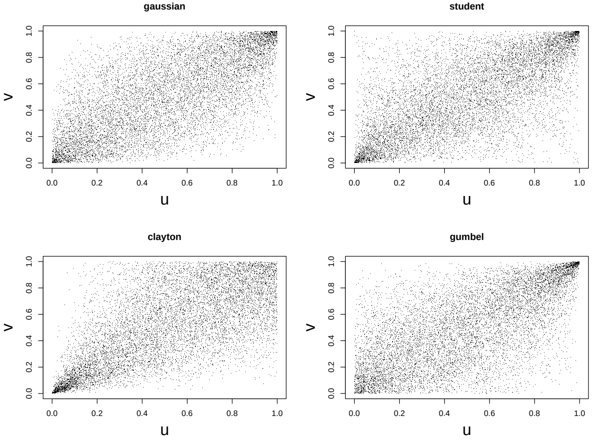


Popular copula families

Gumbel copula



Popular copula families



Simulation

Elliptical copulas

- Gaussian copula:
  - Simulate $\boldsymbol{x} \sim \mathcal{N}_{\boldsymbol{d}}(\boldsymbol{0}, \boldsymbol{R})$
  - Set $\boldsymbol{u} = \Phi(\boldsymbol{x})$
- Student copula:
  - Simulate $\boldsymbol{x} \sim t_{\boldsymbol{d}}(\boldsymbol{0}, \boldsymbol{R}, \nu)$
  - Set $\boldsymbol{u} = t_{\nu}(\boldsymbol{x})$

Simulation

Archimedean copulas

- Clayton copula:
    - Note that the inverse of the generator is equal to the Laplace transform of a Gamma variate $x \sim Ga(1/\delta, 1)$
    - Simulate a gamma variate $x \sim Ga(1/\delta, 1)$
    - Simulate d iid $U(0,1)$ variables $v_1, \ldots, v_d$
    - Return $\boldsymbol{u} = \left((1 - \frac{\log v_1}{x})^{-1/\delta}, \ldots, (1 - \frac{\log v_d}{x})^{-1/\delta}\right)$
- Gumbel copula:
    - Note that the inverse of the generator function is equal to the Laplace transform of a positive stable variate $x \sim St(1/\theta, 1, \gamma, 0)$, where $\gamma = \left(\cos\left(\frac{\pi}{2\theta}\right)\right)^{\theta}$ and $\theta > 1$
    - Simulate a positive stable variate $x \sim St(1/\theta, 1, \gamma, 0)$
    - Simulate d iid $U(0,1)$ variables $v_1, \ldots, v_d$
    - Return $\boldsymbol{u} = \left(\exp\left(-\left(-\frac{\log v_1}{x}\right)^{1/\theta}\right), \ldots, \exp\left(-\left(-\frac{\log v_d}{x}\right)^{1/\theta}\right)\right)$

Simulation

General algorithm

- In general we could apply the conditional marginal cdf's:

$$C_{i|1,\ldots,i-1}(u_i|u_1,\ldots,u_{i-1}) = \frac{\partial^{i-1}C(u_1,\ldots,u_i)}{\partial u_1\cdots\partial u_{i-1}} \Big/ \frac{\partial^{i-1}C(u_1,\ldots,u_{i-1})}{\partial u_1\cdots\partial u_{i-1}}.$$

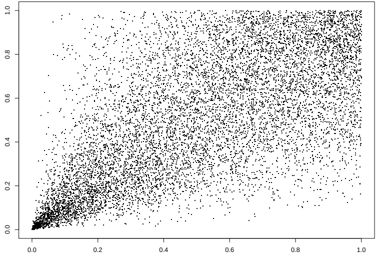
- Simulate a rv u_1 from $U(0,1)$,
- Simulate a rv u_2 from $C_{2|1}(\cdot|u_1)$,

  ...
- Simulate a rv u_d from $C_{d|1,\ldots,d-1}(\cdot|u_1,\ldots,u_{d-1})$.
- Generally means simulating a rv V_i from $U(0,1)$ from which $u_i = C^{-1}_{i|1,\ldots,i-1}(V_i|u_1,\ldots,u_{i-1})$ can be obtained, if necessary by numerical root finding.

Simulation

R example

```
> library(copulaGOF)
> u=SimulateCopulae(n=10000,d=2,construction=list(type=‘‘opc’’,copula=‘‘gumbel’’),param=2)
> plot(u,pch=".")
```



Parameter estimation

Model: $C_\theta(u_1, \ldots, u_d), \quad \theta \in \Theta, \quad \dim(\theta) \geq 1$
Data: $\boldsymbol{x_j} = (x_{j1}, \ldots, x_{jd}), \quad j = 1, \ldots, n$

- Method-of-moments
- Maximum likelihood
- Posterior density

Parameter estimation

Method-of-moments

- Moment m related to θ by one-to-one function $g_{\boldsymbol{m}}$: $m = g_{\boldsymbol{m}}(\theta; C)$
- If $\widehat{m}$ is a consistent estimator for m then $\widehat{\theta} = g_{\boldsymbol{m}}^{-1}(\widehat{m}; C)$ is a consistent estimator for θ
- In most cases of interest, as $n \to \infty$:

$$\sqrt{n}(\widehat{\theta} - \theta) \sim \mathcal{N}(0, \sigma^2(C_\theta))$$

- Examples: Spearman's rho, Kendall's tau

$$\widehat{\theta}_{\rho_{\boldsymbol{S}}} = g_{\rho_{\boldsymbol{S}}}^{-1}(\widehat{\rho_{\boldsymbol{S}}}; C), \quad \widehat{\theta}_{\tau} = g_{\tau}^{-1}(\widehat{\tau}; C)$$

$$\rho_{\boldsymbol{S}}(X, Y) = 12 \int_0^1 \int_0^1 C(u, v)\mathrm{d}u\mathrm{d}v - 3$$

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v)\mathrm{d}C(u, v) - 1$$

Parameter estimation

Maximum likelihood

- In classical statistics, ML estimation is usually more efficient than method-of-moments
- Adaptation needed since inference is based on ranks $\Rightarrow$ maximum pseudo-likelihood (Oakes, 1994; Genest et al., 1995; Shih and Louis, 1995)
- Maximize rank based log-likelihood

$$\widehat{\theta} = \arg\max_{\theta} \left[\frac{1}{n} \sum_{j=1}^{n} \log c_{\theta} \left\{ \widehat{F}_1(x_{j1}), \ldots, \widehat{F}_d(x_{jd}) \right\} \right]$$

- Requires density c_{θ} and usually numerical maximization
- Genest et al. (1995) show consistency and that as $n \to \infty$:

$$\sqrt{n}(\widehat{\theta} - \theta) \sim \mathcal{N}(0, \sigma^2(C_{\theta}))$$

- Inefficient in general (Genest and Werker, 2002), efficient at independence (Genest et al., 1995) and semi-parametrically efficient for the Gaussian copula (Klaassen and Wellner, 1997).

Parameter estimation

Posterior density

- While frequentist methods assume there is no prior knowledge about the parameter, Bayesian parameter estimation incorporates prior knowledge.
- Output is the entire probability density of the parameter and not only a point estimate

$$P\{\theta|\mathbf{x}\} = \frac{L\{\mathbf{x}|\theta\} \cdot \pi\{\theta\}}{\int_\Theta L\{\mathbf{x}|\theta\} \cdot \pi\{\theta\} \, \mathrm{d}\theta}$$

- $P\{\theta|\mathbf{x}\}$ is the posterior density of θ given the data $\mathbf{x}$ while $L\{\mathbf{x}|\theta\}$ is the likelihood and $\pi\{\theta\}$ is the prior density of θ.
- Applied to copulas:

$$P\{\theta|\mathbf{x}\} \propto L\{\mathbf{x}|\theta\} \cdot \pi\{\theta\}$$

$$L\{\mathbf{x}|\theta\} = \prod_{j=1}^{n} \left[c_\theta \left\{ F_1(x_{j1}|\theta), \dots, F_d(x_{jd}|\theta)|\theta \right\} \cdot \prod_{i=1}^{d} f_i(x_{ji}|\theta) \right]$$

Parameter estimation

R example

```
> library(copulaGOF)
> u=SimulateCopulae(n=10000,d=2,construction=list(type=‘‘opc’’,copula=‘‘gumbel’’),param=2)
> theta=EstimateCopulaParameter(u,construction=list(type=‘‘opc’’,copula=‘‘gumbel’’))
> theta
  $t
  [1] 1.980583
  $loglik
  [1] 3575.683
```

Model selection

Models: $C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}(u_{\mathbf{1}},\ldots,u_{\boldsymbol{d}}),\quad k=1,\ldots,K,\quad \theta_{\boldsymbol{k}}\in\Theta_{\boldsymbol{k}},\quad \dim(\theta_{\boldsymbol{k}})\geq 1$
Data: $\boldsymbol{x_j}=(x_{\boldsymbol{j1}},\ldots,x_{\boldsymbol{jd}}),\quad j=1,\ldots,\boldsymbol{n}$

- Akaike information criterion
- Pseudo-likelihood ratio tests
- Bayes factor

Model selection

Akaike information criterion

$$AIC(C_{k,\theta_k}) = -2\sum_{j=1}^{n} \log c_{k,\widehat{\theta}_k}\left\{\widehat{F}_1(x_{j1}),\ldots,\widehat{F}_d(x_{jd})\right\} + 2p_k, \quad p_k = \dim(\theta_k)$$

- Choose model with smallest AIC value
- Kullback-Leibler (KL) distance: Measure of closeness from true density $c_0(\cdot)$ to parametric density $c_\theta(\cdot)$
- ML estimator $\widehat{\theta}$ tends a.s. to the minimizer θ_0 of the KL distance from true model to approximate, parametric model
- AIC searches for model with smallest estimated KL distance
- AIC assumes true model is in class of considered models. If comparing non-nested models then p_k is no longer $\dim(\theta_k)$ and the formula above becomes inaccurate.
- Takeuchi information criterion (TIC) is a robustified version of AIC that deals with this issue.
- Suffers from working with pseudo-observations? Practical consequences?

Model selection

Pseudo-likelihood ratio tests

- Take into account randomness of the *AIC*; ensures that no model under consideration performs significantly better than selected model
- Does not require the considered models to include the true model. Hence allows for the comparison of non-nested models
- Compares each model to a benchmark model and chooses the model that is closest to the true model in terms of the KL distance

$$\widehat{T}_{kb} = \max_{1 \le k \le K; k \ne b} \left[\sqrt{\frac{n}{\hat{\sigma}_{kk}}} \left\{ \widehat{LR}_{\widehat{\theta}_k, \widehat{\theta}_b} \left(\widehat{F}_1, \ldots, \widehat{F}_d \right) + \frac{p_b - p_k}{n} \right\} G(\hat{\sigma}_{kk}), 0 \right]$$

$$\widehat{LR}_{\widehat{\theta}_k, \widehat{\theta}_b} \left(\widehat{F}_1, \ldots, \widehat{F}_d \right) = \frac{1}{n} \sum_{j=1}^{n} \log \left[\frac{c_{k, \widehat{\theta}_k} \left\{ \widehat{F}_1(x_{j1}), \ldots, \widehat{F}_d(x_{jd}) \right\}}{c_{b, \widehat{\theta}_b} \left\{ \widehat{F}_1(x_{j1}), \ldots, \widehat{F}_d(x_{jd}) \right\}} \right]$$

- Chen and Fan (2005) bootstrap to obtain p-value estimate for hypothesis that none of considered models are significantly better than benchmark model b. If hypothesis is not rejected then choose benchmark model.
- Results show consistency with *AIC*

Model selection

Bayes factor

- Idea: Compute posterior probability of copula model

$$
\begin{aligned}
P\{C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}|\boldsymbol{x}\} &\propto L\{\boldsymbol{x}|C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}\} \cdot \pi\{C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}\} \\
&= \pi\{C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}\} \cdot \int_{\Theta_{\boldsymbol{k}}} L_{\boldsymbol{k}}\{\boldsymbol{x}|\theta_{\boldsymbol{k}}\} \cdot \pi\{\theta_{\boldsymbol{k}}\} \mathrm{d}\theta_{\boldsymbol{k}}
\end{aligned}
$$

- $P\{C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}|\boldsymbol{x}\}$ is the posterior density of model k, $L_{\boldsymbol{k}}\{\boldsymbol{x}|\theta_{\boldsymbol{k}}\}$ is the likelihood under copula model k, $\pi\{C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}\}$ the prior on the copula model and $\pi\{\theta_{\boldsymbol{k}}\}$ the prior of $\theta_{\boldsymbol{k}}$.
- Bayes factor:

$$
B_{\boldsymbol{km}} = \frac{P\{C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}|\boldsymbol{x}\}/\pi\{C_{\boldsymbol{k},\theta_{\boldsymbol{k}}}\}}{P\{C_{\boldsymbol{m},\theta_{\boldsymbol{m}}}|\boldsymbol{x}\}/\pi\{C_{\boldsymbol{m},\theta_{\boldsymbol{m}}}\}} = \frac{\int_{\Theta_{\boldsymbol{k}}} L_{\boldsymbol{k}}\{\boldsymbol{x}|\theta_{\boldsymbol{k}}\} \cdot \pi\{\theta_{\boldsymbol{k}}\} \mathrm{d}\theta_{\boldsymbol{k}}}{\int_{\Theta_{\boldsymbol{m}}} L_{\boldsymbol{m}}\{\boldsymbol{x}|\theta_{\boldsymbol{m}}\} \cdot \pi\{\theta_{\boldsymbol{m}}\} \mathrm{d}\theta_{\boldsymbol{m}}}
$$

Model selection

Bayes factor

- Does not require preliminary estimation of $\theta_{\boldsymbol{k}}$
- Bayesian analogue of likelihood ratio test
- Prior and posterior information are combined in a ratio that provides evidence in favour of one model versus another
- Nested models not required
- Compared models should have the same dependent variable
- Huard et al. (2006) apply this methodology to copula selection. They have flat priors for parameter and copula and simply choose the copula with the highest posterior probability.

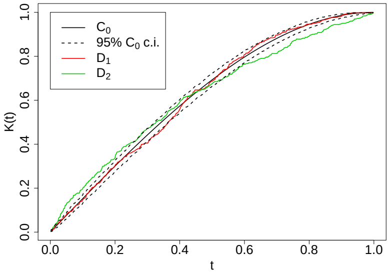
Model evaluation

- Given our "best" model - how good is it?
- Informal graphical diagnostics
- Goodness-of-fit tests

Model evaluation

Informal graphical diagnostics

- Compare pseudo-observations with random sample from model
- Compare, graphically some empirical estimate of model with parametric model, e.g. $\widehat{K}$ vs. $K_{\widehat{\theta}}$ where $K(t) = P(C(u_{\mathbf{1}}, \ldots, u_{\boldsymbol{d}}) \leq t)$
- Confidence/Credibility intervals



Model evaluation

Goodness-of-fit (gof) tests

- We wish to test the hypotheses

$$\mathcal{H}_0 : C \in \mathcal{F} = \{C_\theta; \theta \in \Theta\} \text{ vs. } \mathcal{H}_1 : C \notin \mathcal{F} = \{C_\theta; \theta \in \Theta\}$$

- Some proposed gof processes:

$$\begin{aligned} \mathcal{C}_n &= \sqrt{n}\left\{\widehat{C} - C_{\widehat{\theta}}\right\} \\ \mathcal{K}_n &= \sqrt{n}\left\{\widehat{K} - K_{\widehat{\theta}}\right\}, \quad K(t) = P(C(\boldsymbol{u}) \leq t) \\ \mathcal{S}_n &= \sqrt{n}\left\{\widehat{\theta}_{\rho_S} - \widehat{\theta}_\tau\right\} \end{aligned}$$

- Example: Cramér-von Mises statistic for $\mathcal{C}_n$:

$$\mathcal{V}_n = \int_0^1 \cdots \int_0^1 \{\mathcal{C}_n(x_1, \ldots, x_d)\}^2 \mathrm{d}x_1 \cdots \mathrm{d}x_d$$

- Null distribution of statistic depends on parameter - parametric bootstrap procedure to obtain proper p-value estimate

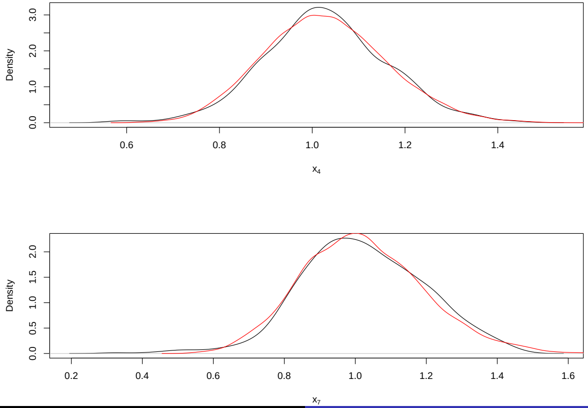
Examples

Uranium data

```
> x=read.table("/home/daniel/Jobb/Presentasjoner/ASTIN/example/uranium/COOK_Data.txt")
> n=nrow(x)
> z=EmpiricalMarginals(x)
> a1 = (mean(x[,4])/sd(x[,4]))^2
> a2 = (mean(x[,7])/sd(x[,7]))^2
> s1 = var(x[,4])/mean(x[,4])
> s2 = var(x[,7])/mean(x[,7])
> marg4.1 = rgamma(10000,shape=a1,scale=s1)
> marg4.2 = rgamma(10000,shape=a2,scale=s2)
> par(mfrow=c(2,1))
> plot(density(x[,4]),main="",xlab=expression(x[4]))
> lines(density(marg4.1),col=2)
> plot(density(x[,7]),main="",xlab=expression(x[7]))
> lines(density(marg4.2),col=2)
```

Examples

Urianium data



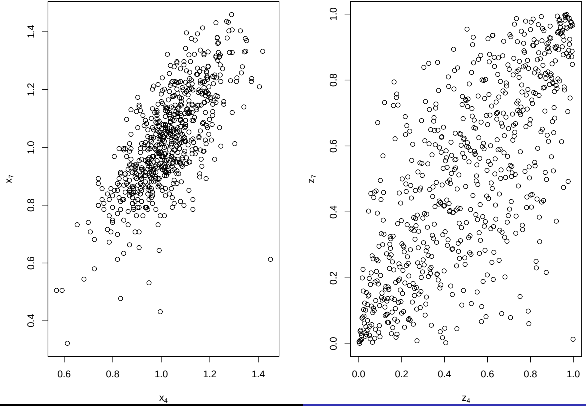
Examples

Uranium data

```
> par(mfrow=c(1,2))
> plot(x[,c(4,7)],xlab="x4",ylab="x7")
> plot(z[,c(4,7)],xlab="z4",ylab="z7")
```

Examples

Uranium data



Examples

Urianium data

```
> theta=EstimateCopulaParameter(z[,c(4,7)],construction=list(type="ell",copula="student")); theta
  $t
  [1] 0.7394778
  $nu
  [1] 8.102415
  $loglik
  [1] 254.3866
> theta=EstimateCopulaParameter(z[,c(4,7)],construction=list(type="opc",copula="clayton")); theta
  $t
  [1] 1.481104
  $loglik
  [1] 198.8813
> theta=EstimateCopulaParameter(z[,c(4,7)],construction=list(type="opc",copula="gumbel")); theta
  $t
  [1] 1.977482
  $loglik
  [1] 227.8120
```

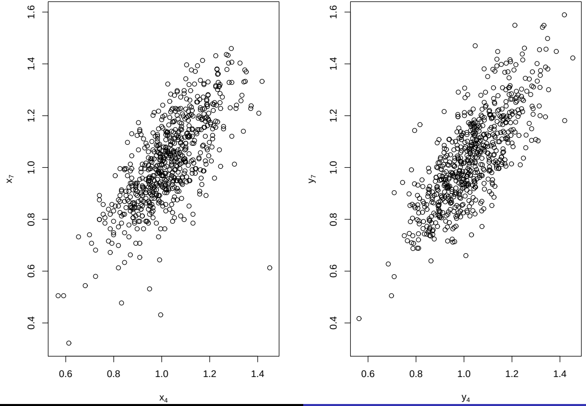
Examples

Urianium data

```
> theta=EstimateCopulaParameter(z[,c(4,7)],construction=list(type="ell",copula="student"))
> u = SimulateCopulae(n=nrow(x),d=2,construction=list(type="ell",copula="student"),param=theta)
> y = u
> y[,1]=qgamma(u[,1],shape=a1,scale=s1)
> y[,2]=qgamma(u[,2],shape=a2,scale=s2)
> par(mfrow=c(1,2))
> plot(x[,c(4,7)],xlab="x4",ylab="x7")
> plot(y,xlab="y4",ylab="y7")
```

Examples

Uranium data



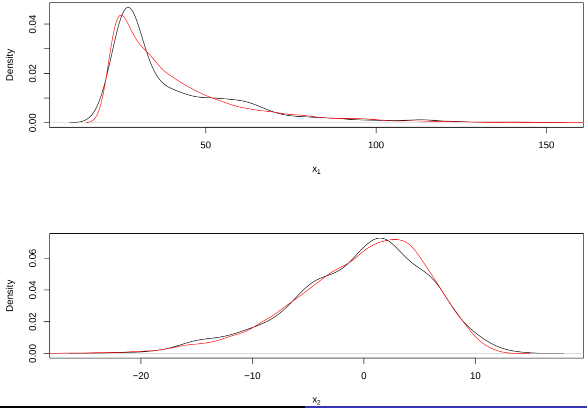
Examples

Icelandic river flow/temp data

```
> library(tseries); data(ice.river)
> x=ice.river[,c(2,4)]
> z=EmpiricalMarginals(x,ties=1)
> par(mfrow=c(2,1))
> plot(density(x[,1]))
> plot(density(x[,2]))
```

Examples

Icelandic river flow/temp data



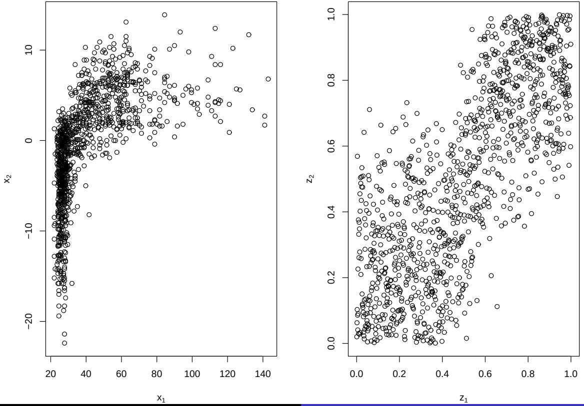
Examples

Icelandic river flow/temp data

```
> plot(x[,1],x[,2],xlab=expression(x[1]),ylab=expression(x[2]))
> plot(z,xlab=expression(z[1]),ylab=expression(z[2]))
```

Examples

Icelandic river flow/temp data



Examples

Icelandic river flow/temp data

```
> theta=EstimateCopulaParameter(z,construction=list(type="ell",copula="student")); theta
  $t
  [1] 0.757387
  $nu
  [1] 300
  $loglik
  [1] 342.0470
> theta=EstimateCopulaParameter(z,construction=list(type="opc",copula="clayton")); theta
  $t
  [1] 1.094927
  $loglik
  [1] 225.4335
> theta=EstimateCopulaParameter(z,construction=list(type="opc",copula="gumbel")); theta
  $t
  [1] 1.846864
  $loglik
  [1] 314.2530
> theta=EstimateCopulaParameter(z,construction=list(type="opc",copula="frank")); theta
  $t
  [1] 6.650217
  $loglik
  [1] 429.5826
```

Examples

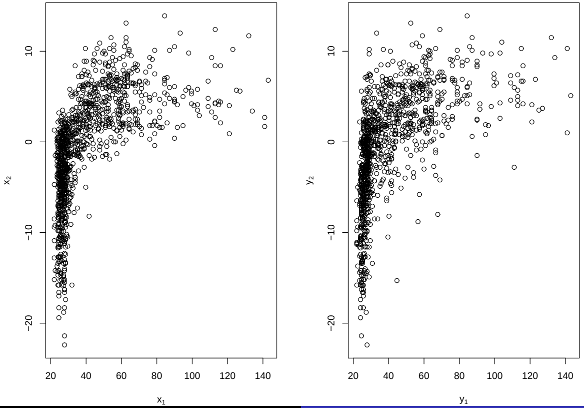
Icelandic river flow/temp data

```
> theta=EstimateCopulaParameter(z,construction=list(type="opc",copula="frank"))
> u = SimulateCopulae(n=nrow(x),d=2,construction=list(type="opc",copula="frank"),param=theta)
> y = u
> y[,1] = sort(x[,1])[rank(y[,1])] # Use empirical marginals to transform back.
> y[,2] = sort(x[,2])[rank(y[,2])]

> par(mfrow=c(1,2))
> plot(x[,1],x[,2],xlab=expression(x[1]),ylab=expression(x[2]),xlim=c(min(min(x[,1],min(y[,1]))),max(max(x[,1]),max(y[,1]))),yl
> plot(y,xlab=expression(y[1]),ylab=expression(y[2]),xlim=c(min(min(x[,1],min(y[,1]))),max(max(x[,1]),max(y[,1]))),ylim=c(min(m
```

Examples

Icelandic river flow/temp data



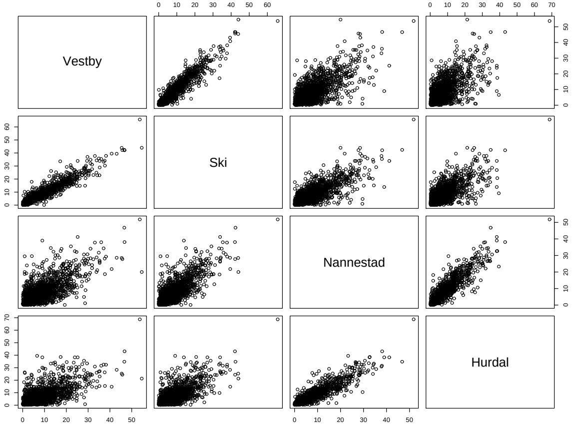
Examples

Precipitation data

```
> x.rain = as.data.frame(as.matrix(read.table("/home/daniel/Jobb/Presentasjoner/ASTIN/example/precipitation/regnData.txt")))
> names(x.rain)=c("Vestby","Ski","Nannestad","Hurdal")
> z.rain = as.data.frame(EmpiricalMarginals(x.rain,ties=1))
> names(z.rain)=c("Vestby","Ski","Nannestad","Hurdal")
> pairs(x.rain)
> pairs(z.rain)
```

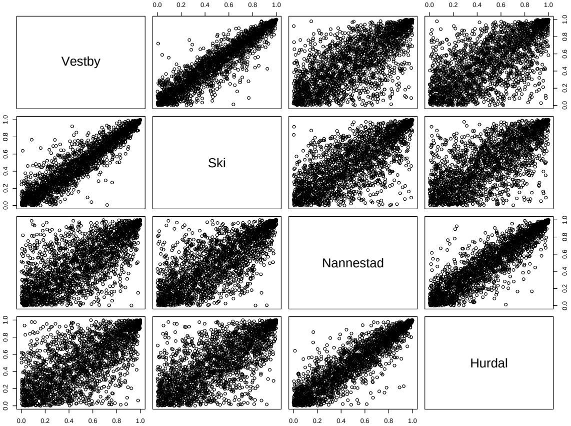
Examples

Precipitation data



Examples

Precipitation data



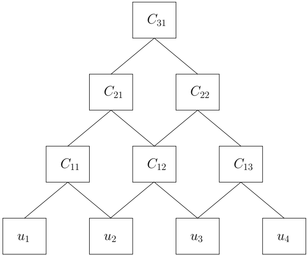
Extensions

Pair-copula constructions

- Originally proposed by Joe (1997) and later discussed in detail by Bedford and Cooke (2002, 2001); Kurowicka and Cooke (2006) (simulation) and Aas et al. (2007) (inference).
- Allows for the specification of $d(d-1)/2$ bivariate copulae of which the first $d-1$ are unconditional and the rest conditional.
- The bivariate copulae involved do not have to belong to the same class.

Extensions

Pair-copula constructions (pcc)



- C_{21} is the copula of $F(u_1|u_2)$ and $F(u_3|u_2)$.
- C_{22} is the copula of $F(u_2|u_3)$ and $F(u_4|u_3)$.
- C_{31} is the copula of $F(u_1|u_2, u_3)$ and $F(u_4|u_2, u_3)$.
- Bedford and Cooke (2002) introduced *vines* as tree structures to help organize the many different constructions.

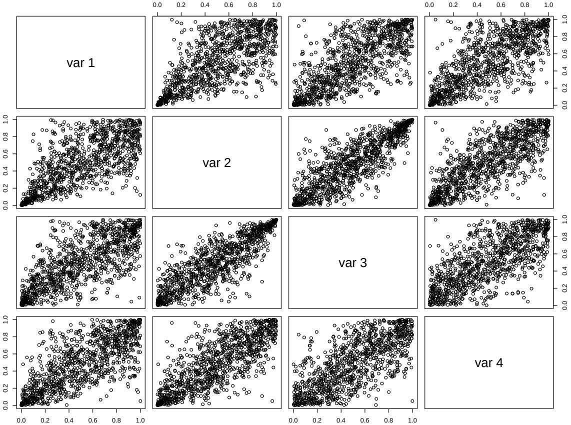
Extensions

Pair-copula constructions (pcc)

```
> x = SimulateCopulae(n=1000,d=4,construction=list(type="dpcc",copula=c("clayton","gumbel
> pairs(x)
```

Extensions

Pair-copula constructions (pcc)



Summary

- Correlation coefficient only a measure of linear dependence
- Empirical evidence calls for alternative models for mv dependence besides the multinormal
- Copulas is a very flexible and promising tool - but still alot of research needed
- Criticism: copulas are static
- Use with cause and be critical! Do goodness-of-fit exercises.
- Paul Embrechts: "copulas form a most useful concept for a lot of applied modeling, they do not yield, however, a panacea for the construction of useful and wellâunderstood multivariate dfs, and much less for multivariate stochastic processes."

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2007). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics 42*, In press.

Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence 32*, 245–268.

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

Chen, X. and Y. Fan (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Canadian Journal of Statistics 33*, 389–414.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendence. *Acad. Royal Bel., Bull. Class. Sci.*, 5e *série 65*, 274–292.

Embrechts, P., A. McNeil, and D. Straumann (1999, May). Correlation and Dependence in Risk Management: Properties and Pitfalls. *RISK*, 69–71.

Genest, C., K. Ghoudi, and L. P. Rivest (1995). A semi–parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82*, 543–552.

Genest, C. and B. Werker (2002). Conditions for the asymptotic semiparametric effciency of an omnibus estimator of dependence parameters in copula models. In C. M. Cuadras, J. Fortiana, and J. A. Rodríguez-Lallela (Eds.), *Distributions with Given Marginals and Statistical Modelling*, Kluwer, Dordrecht, The Netherlands, pp. 103–112.

Huard, D., G. Évin, and A.-C. Favre (2006). Bayesian copula selection. *Computational Statistics & Data Analysis 51*, 809–822.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Klaassen, C. A. J. and J. A. Wellner (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli 3*, 55–77.

Kurowicka, D. and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. New York: Wiley.

Oakes, D. (1994). Multivariate survival distributions. *Journal of Nonparametric Statistics 3*, 343–354.

Shih, J. H. and T. A. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics 51*, 1384–1399.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris 8*, 299–231.

Daniel Berg
daniel@danielberg.no
www.danielberg.no