USING AND SELECTING AMONG COPULAE

Frequentist and Bayesian perspectives

DANIEL BERG

University of Oslo & Norwegian Computing Center

Trial lecture

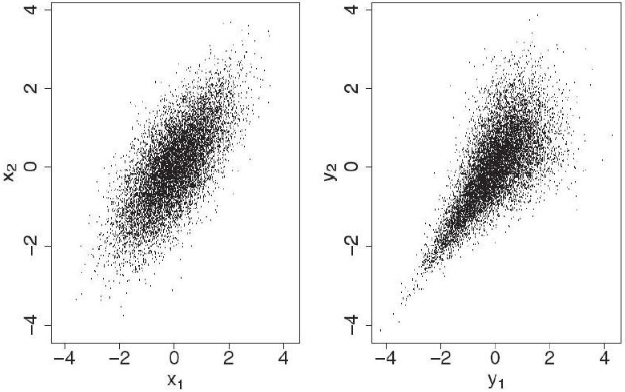
Oslo – 14.03.2008

Outline

- Introduction
- Parameter estimation
- Model selection
- Model evaluation
- Summary

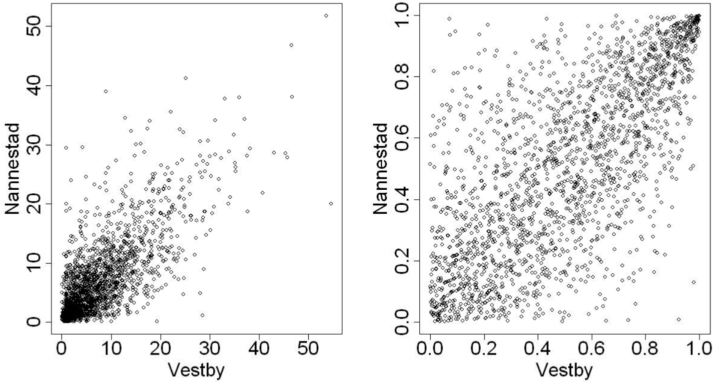
Introduction

Motivation



Introduction

Motivation



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.

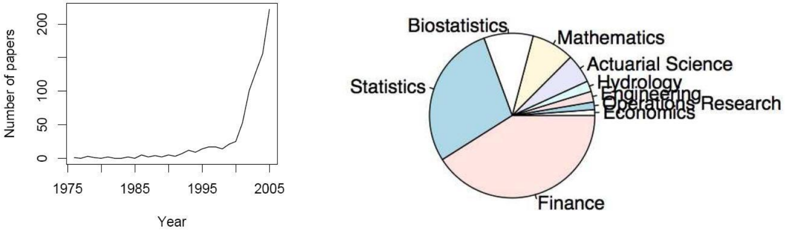


Figure: Based on a survey by Bourdeau-Brien (2007).

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 $C : [0,1]^d \rightarrow [0,1]$ 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:

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

  where R_{ji} is the rank of x_{ji} amongst $x_{1i}, \ldots, x_{ni}$.

- u_{ji} commonly referred to as *pseudo-observations* and models based on non-parametric margins and parametric copulas are referred to as *semi-parametric* copulas

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

Parameter estimation

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

Frequentist

- Method-of-moments
- Maximum likelihood
- Other methods: minimum distance, kernel smoothing, ...

Bayesian

- Posterior density

Parameter estimation

Method-of-moments (F)

- Moment m related to θ by one-to-one function g_m: $m = g_m(\theta; C)$
- If $\widehat{m}$ is a consistent estimator for m then $\widehat{\theta} = g_m^{-1}(\widehat{m}; C)$ 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

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

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

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

Parameter estimation

Maximum likelihood (F)

- In classical statistics, ML estimation is usually more efficient than method-of-moments
- Adaptation needed since inference is based on ranks ⇒ 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
- Particularly useful for multidimensional θ
- Genest et al. (1995) show consistency and that as $n \rightarrow \infty$:

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

- 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

Other methods (F)

- Minimum distance method:
  Introduced by Wolfowitz (1953) - minimize distance between empirical distribution and family of distributions for model under consideration.
- Example:
  $\widehat{\theta}$ is the parameter that minimizes some distance between $\widehat{C}$ and $C_{\widehat{\theta}}$

$$\widehat{\theta} = \arg\min_{\theta} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{W}_i - W_{\widehat{\theta},i} \right)^2 \right]$$

$$\widehat{W}_i = \widehat{C}(u_{i1}, \ldots, u_{id}), \quad W_{\widehat{\theta},i} = C_{\widehat{\theta}}(u_{i1}, \ldots, u_{id}), \quad u_{ji} = \widehat{F}_i(x_{ji})$$

- Kernel smoothing method:
  Derive smooth estimate of copula or its density without assuming a parametric form. Computationally expensive. See e.g. Gijbels and Mielniczuk (1990); Fermanian and Scaillet (2003).

Parameter estimation

Posterior density (B)

- 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

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

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

Parameter estimation

Posterior density (B)

- Applied to copulas:

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

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

- Pseudo-observations $\Rightarrow$ posterior pseudo-density?

$$P\{\theta|\boldsymbol{u}\} \propto L\{\boldsymbol{u}|\theta\} \cdot \pi(\theta)$$

$$L\{\boldsymbol{u}|\theta\} = \prod_{j=1}^{n} c_\theta \left\{ \widehat{F}_1(x_{j1}), \ldots, \widehat{F}_d(x_{jd})|\theta \right\}$$

Model selection

Models: $C_{k,\theta_k}(u_1, \ldots, u_d), \quad k = 1, \ldots, K, \quad \theta_k \in \Theta_k, \quad \dim(\theta_k) \geq 1$
Data: $\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jd}), \quad j = 1, \ldots, n$

Frequentist

- Akaike information criterion
- Pseudo-likelihood ratio tests

Bayesian

- Bayes factor
- Other methods: Deviance information criterion, Bayesian model averaging, ...

Model selection

Akaike information criterion (F)

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

- 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 (F)

- 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

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

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

- 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 (B)

- Idea: Compute posterior probability of copula model

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

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

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

Model selection

Bayes factor (B)

- Does not require preliminary estimation of θ_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 selection

Other methods (B)

- Several alternative versions of Bayes factors (*BF*) have been proposed; posterior *BF*, intrinsic *BF*, fractional *BF*, ...
- Bayes information criterion (*BIC*): Large sample approximation of Bayes factor assuming $\theta = \widehat{\theta}_{ML}$. Similar to *AIC*, does not depend on prior, easy to compute, etc.
- Deviance information criterion (*DIC*): Not an approximation of *BF*. Deviance based complexity criterion used as penalty in standard information criterion. Hierarchical modelling generalization of *AIC* and *BIC*. Particularly useful when posterior density have been obtained by MCMC. Approximately equal to *AIC* for models with little prior information.
- Bayesian model averaging - alternative to "mixed" copulas $C_m = w_1 C_1 + w_2 C_2$ estimated by EM algorithm (Hu, 2006)?
- Out-of-sample prediction accuracy

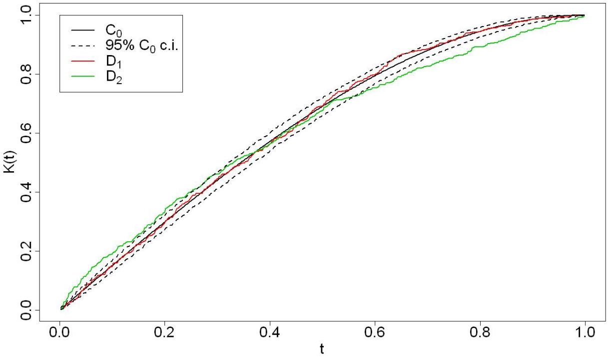
Model evaluation

- Given our "best" model - how good is it?
- Informal graphical diagnostics
- Goodness-of-fit tests
- Other methods: cross-validation, sensitivity analysis, out-of-sample validation, ...

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 $K(t) = P(C(u_1, \ldots, u_d) \leq t)$
- 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:

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

- 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

Model evaluation

Other methods

- Several other goodness-of-fit tests exist
- Cross-validation: systematically run through all possible splits of data and check residuals in each case
- Sensitivity analysis: change parameter estimator, change prior, ...
- Out-of-sample validation: prediction error

Summary and conclusion

- Bayesian methods particularly valuable when there is no or little data OR when we have strong prior information
- Most natural Bayesian model selection procedure is Bayes factor. Can select model without preliminary estimation of parameter.
- Model evaluation very important to evaluate model risk and identify need for better models
- Model selection and model evaluation are two related but different tasks
- What methods to use and in which order depends on the situation/problem

Further reading

- Introduction to copulas: Genest and Favre (2007)
- Extensive copula theory in general: Joe (1997); Nelsen (1999)
- Parameter estimation: Chen et al. (2006)
- Model selection (pseudo-likelihood ratio test): Chen and Fan (2005)
- Copula goodness-of-fit: Genest et al. (2008)
- Bayesian + copula: Huard et al. (2006); Pitt et al. (2006); Silva and Lopes (2008)

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

Chen, X., Y. Fan, and V. Tsyrennikov (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of The American Statistical Association 101*(475), 1228–1240.

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

Fermanian, J.-D. and O. Scaillet (2003). Nonparametric estimation of copulas for time series. *Journal of Risk 5*(4), 25–54.

Genest, C. and A.-C. Favre (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering 12*, 347–368.

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., B. Rémillard, and D. Beaudoin (2008). Omnibus goodness–of–fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics 42*, in press.

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.

Gijbels, I. and J. Mielniczuk (1990). Estimating the density of a copula function. *Commun. Stat: Theory Meth. 19*(2), 445–464.

Hu, L. (2006). Dependence patterns across financial markets: a mixed copula approach. *Applied Financial Economics, Taylor and Francis Journals 16*(10), 717–729.

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.

Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer Verlag.

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

Pitt, M., D. Chan, and R. Kohn (2006). Efficient bayesian inference for gaussian copula regression models. *Biometrika 93*(3), 537–554.

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

Silva, R. S. and H. Lopes (2008). Copula, marginal distributions and model selection: a bayesian note. *Statistics and Computing*. Forthcoming.

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

Wolfowitz, J. (1953). Estimation by the minimum distance method. *Ann. Inst. Stat. Math. 5*(1), 9–23.