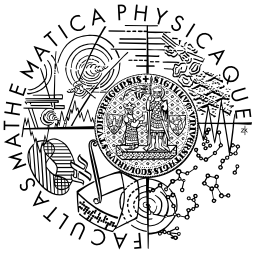


MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Filip Šťastný

Dvourozměrná rozdělení při daných marginálech

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Michal Pešta, Ph.D.
Studijní program: Finanční matematika
Studijní obor: MFMP

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ....

Podpis autora

Mé upřímné poděkování patří panu doc. RNDr. Michalu Peštovi, Ph. D. za poskytování rad a doporučení během psaní této práce, kterých si velmi vážím, a také za vstřícnost při komunikaci.

Název práce: Dvourozměrná rozdělení při daných marginálech

Autor: Filip Šťastný

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Michal Pešta, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Jedním z nástrojů pro studium závislosti mezi náhodnými veličinami jsou kopule. Při modelování vícerozměrných veličin je pomocí Sklarovy věty možné prostřednictvím kopulí modelovat zvlášť marginální rozdělení a vztah mezi nimi, tento přístup nám tak umožňuje rozdělit si konstrukci vícerozměrných rozdělení na tyto dva faktory. Při pevných marginálních rozděleních pak konstrukce spočívá pouze ve volbě vhodné kopule. Tato práce se zabývá kopulemi v případě dvourozměrných rozdělení s danými spojitými marginálními rozděleními a je zaměřena na parametrické kopule, především prostřednictvím Archimédovských kopulí. Jsou zde uvedeny základní vlastnosti kopulí a Sklarova věta, která umožňuje jejich studium ve stochastickém kontextu. Dále jsou zde ve spojitosti s kopulemi studovány míry závislostí Kendallovo tau, Spearmanovo rho a koeficienty závislosti chvostů. Na závěr se práce zabývá metodami pro odhad neznámých parametrů kopulí, které jsou ilustrovány na dvou příkladech.

Klíčová slova: dvourozměrná rozdělení marginální rozdělení kopule

Title: Two-dimensional distributions for given margins

Author: Filip Šťastný

Department: Department of probability and mathematical statistics

Supervisor: doc. RNDr. Michal Pešta, Ph.D., Department of probability and mathematical statistics

Abstract: One of the tools for study of dependence between random variables are copulas. While modelling multidimensional variables it is possible using Sklar's theorem to model through copulas marginal distributions and relationship between them separately, this approach thus enables us to split construction of multi-dimensional distributions into these two factors. With marginal distributions fixed, the construction is consisting of appropriate copula choice only. This thesis deals with copulas in the case of two-dimensional distributions with continuous fixed marginal distributions and is focused on parametrical copulas, mainly through Archimedean copulas. Basic properties of copulas with Sklar's theorem, which enables studying copulas in stochastic context, are presented here. Further, measures of dependence such as Kendall's tau, Spearman's rho and coeficients of tail dependence are in connection with copulas studied in this thesis. At the end, the thesis deals with methods of estimation unknown parameters, which are ilustrated on two examples.

Keywords: two-dimensional distributions marginal distribution copula

Obsah

Úvod **2**

1 Kopule **3**
1.1 Definice a základní vlastnosti kopulí 3
1.2 Sklarova věta 4
1.3 Archimédovské kopule 7

2 Závislost **9**
2.1 Míry závislosti 9
2.2 Rodiny kopulí 15

3 Aplikace na data **19**

Závěr **26**

Seznam použité literatury **27**

Úvod

Jedním z hlavních cílů ve statistice a pravděpodobnosti je modelování náhodných veličin, ve vícedimenzionálním případě modelování zahrnuje dva faktory, kterými jsou marginální rozdělení náhodných veličin a vztah mezi nimi. Modelování marginálních rozdělení a vztahu náhodných veličin najednou může být mnohdy složité, výhodnějším přístupem by mohlo být uvažovat oba faktory zvlášť. Jedním z hlavních nástrojů, který nám tento přístup umožňuje jsou kopule (z angl. „couple“). Jak anglický název napovídá, kopule v jistém smyslu „párují“ marginální veličiny do vícerozměrných veličin a obsahují tak informaci o jejich závislosti. Na počátku teorie kopulí stál v polovině minulého století americký matematik Abe Sklar (1925-2020), který celý koncept zavedl, od té doby našly kopule uplatnění v modelování v rámci mnoha oblastí, jako příklady uveďme finančnictví, hydrologii nebo geologii.

Předpokládáme-li pevně daná marginální rozdělení, pak se konstrukce vícerozměrných rozdělení redukuje na studium závislostí mezi uvažovanými náhodnými veličinami, tedy na studium kopulí, které jsou tak předmětem této práce. Kopule jsou často studovány za předpokladu spojitých marginálních rozdělení, což přináší některé výhody jako je například jednoznačnost, i zde se omezíme na tento předpoklad, dále se také budeme zabývat jen dvoudimezionálním případem.

Různé kopule vyjadřují různé typy závislostí, proto je teorie kopulí spojena právě se studiem závislostí mezi náhodnými veličinami. Tomu odpovídá i vztah, který je mezi kopulemi a některými mírami závislosti jako jsou Kendallovo tau nebo Spearmanovo rho.

Při studiu kopulí se častokrát uvažují kopule, které jsou závislé na parametru, tyto parametrické kopule pak generují tzv. rodiny. Zavedení těchto parametrů nám pak umožňuje prostřednictvím jejich odhadování aplikovat kopule na reálná data.

V této práci se tedy snažíme alespoň částečně pokrýt právě uvedené aspekty spojené s teorií kopulí. V úvodní kapitole jsou kopule formálně zadefinovány a uvedeny jejich základní vlastnost, dále je zde vyslovena Sklarova věta, na které je teorie kopulí založena, první kapitolu uzavírají Archimédovské kopule, konkrétní třída kopulí ve specifickém tvaru. Druhá kapitola je věnována závislosti a konkrétním příkladům, v první části jsou studovány některé míry a typy závislostí v kontextu kopulí, ve druhé části je pak uvedeno několik rodin kopulí společně s jejich vlastnostmi. Třetí kapitola je věnována odhadování neznámých parametrů kopulí, jsou zde zmíněny některé metody pro odhadování, vybrané metody jsou pak ilustrovány na dvou příkladech.

1. Kopule

V úvodní kapitole se budeme zabývat speciálním typem funkcí dvou proměnných, které se nazývají kopule. Nejprve zavedeme potřebné definice a uvedeme vybrané vlastnosti kopulí, následně vyslovíme stěžejní větu, která dá do souvislosti dvourozměrná rozdělení, jejich marginální rozdělení a kopule – Sklarovu větu. Na závěr se podrobněji podíváme na třídu kopulí s názvem Archimédovské kopule.

1.1 Definice a základní vlastnosti kopulí

Obdélníkem v $\mathbb{R}^2$ budeme označovat kartézský součin $[x_1,y_1] \times [x_2,y_2]$ dvou uzavřených intervalů z $\mathbb{R}$, vrcholy tohoto obdélníku jsou body (x_1,y_1), (x_1,y_2), (x_2,y_1), $(x_2,y_2) \in \mathbb{R}^2$. Dále symbolem $\mathbf{I}$ značíme jednotkový interval [0,1]. Předtím než uvedeme samotnou definici kopulí zavedeme si pojem 2-*rostoucí funkce*.

Definice 1. *Reálná funkce dvou proměnných H s definičním oborem $D \subseteq \mathbb{R}^2$ je 2-rostoucí, pokud pro všechny obdélníky B s vrcholy ležícími v D je*

$$V_H(B) = H(x_2,y_2) - H(x_1,y_2) - H(x_2,y_1) + H(x_1,y_1) \geq 0.$$

Nyní už definice kopulí.

Definice 2. *Kopule je funkce $C : \mathbf{I}^2 \to \mathbf{I}$, která splňuje následující vlastnosti:*

1. *C je 2-rostoucí,*
2. *Pro každé $u \in \mathbf{I}$ a každé $v \in \mathbf{I}$ je $C(u,0) = C(0,v) = 0$, $C(u,1) = u$ a $C(1,v) = v$.*

Druhá vlastnost z uvedené definice nám udává chování kopule na hranici jednotkového čtverce – grafem je zešikmený čtyřúhelník v jednotkové krychli $\mathbf{I}^3$ s vrcholy (0,0,0), (1,0,0), (1,1,1), (0,1,0).

Následující tvrzení ukazuje, že kopule lze shora i zdola omezit, z příslušných mezí získáme konkrétní příklady kopulí.

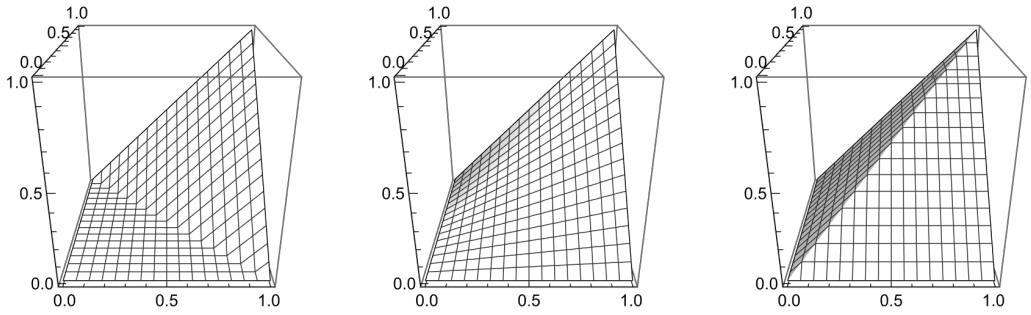
Tvrzení 1. *Nechť C je kopule, pak pro všechna $(u,v) \in \mathbf{I}^2$ platí:*

$$\max(u + v - 1,0) \leq C(u,v) \leq \min(u,v).$$

Důkaz. Nechť (u,v) je libovolný bod z $\mathbf{I}^2$. Je $C(u,v) \leq C(u,1) = u$ a $C(u,v) \leq C(1,v) = v$, dohromady dostáváme $C(u,v) \leq \min(u,v)$. Dále, vzhledem k tomu, že C je 2-*rostoucí*, je $V_C([u,1] \times [1,v]) \geq 0$, po rozepsání výrazu na levé straně této nerovnosti, získáme $C(u,v) \geq u + v - 1$ a společně s nezáporností kopule je $C(u,v) \geq \max(u + v - 1,0)$.

□

Funkce, které tvoří meze pro kopule jsou též kopulemi a dále v textu se s nimi setkáme, budeme je značit $M(u,v) = \min(u,v)$ resp. $W(u,v) = \max(u +$



Obrázek 1.1: Grafy kopulí zleva W, Π a M.

$v - 1{,}0)$, v literatuře nesou název Fréchetova-Hoeffdingova horní resp. dolní mez. Další významnou kopulí je tzv. součinová kopule tvaru $\Pi(u,v) = uv$. Všechny tři zmiňované kopule jsou zobrazeny na obr. 1.1.

Jednou ze základních vlastností kopulí je spojitost, kterou dostaneme jako důsledek lipschitzovské spojitosti, jež je obsahem následujícího tvrzení. Důkaz je založen na několika jednoduchých vlastnostech 2-rostoucích funkcí, které splňují podmínku analogickou podmínce $C(u,0) = C(0,v)$ pro kopule. Příslušná tvrzení s důkazy lze najít v knize Nelsen (2010).

Tvrzení 2. *Nechť C je kopule, pak pro všechna $(u_1,u_2),(v_1,v_2) \in \mathbf{I}^2$ platí:*

$$|C(u_2,v_2) - C(u_1,v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

Tedy C je absolutně spojitá na $\mathbf{I}^2$.

Na závěr této sekce zavedeme částečné uspořádání na množině všech kopulí. Uspořádání následně využijeme ve druhé kapitole.

Definice 3. *Nechť C_1 a C_2 jsou kopule. Řekneme, že C_1 je menší než C_2, píšeme $C_1 \preceq C_2$, pokud $C_1(u,v) \leq C_2(u,v)$ pro všechna $(u,v) \in \mathbf{I}^2$. Analogicky, řekneme, že C_1 je větší než C_2 při opačných nerovnostech.*

Z tvrzení 1 dostáváme, že M je největší a W je nejmenší ze všech kopulí ve smyslu předchozí definice.

1.2 Sklarova věta

Jak již bylo naznačeno v úvodu kapitoly, na Sklarově větě je založena veškerá aplikace kopulí uvažovaná v této práci. Využívá kopule k vyjádření sdružené distribuční funkce dvou náhodných veličin pomocí jejích marginálních distribučních funkcí.

Pro přehlednost a jednoznačnost značení uvedeme nejprve příslušné definice sdružené a marginální distribuční funkce pro reálné náhodné veličiny včetně definice kvantilové funkce, kterou budeme využívat při vyjadřování kopule. V dalším textu budeme automaticky předpokládat, že všechny náhodné veličiny jsou reálné.

Definice 4. *Pro náhodnou veličinu X definujeme její distribuční funkci F z $\mathbb{R}$ do $\mathbf{I}$ jako $F(x) = \mathsf{P}(X \leq x)$ pro všechna $x \in \mathbb{R}$.*

Funkci F^{-1} z $\mathbf{I}$ do $\mathbb{R}$ danou předpisem $F^{-1}(t) = \inf\{x; F(x) \geq t\}$ pro $t \in \mathbf{I}$ nazýváme kvantilová funkce náhodné veličiny X.

Definice 5. *Pro náhodné veličiny X a Y definujeme jejich sdruženou distribuční funkci H z $\mathbb{R}^2$ do $\mathbf{I}$ jako $H(x,y) = \mathsf{P}(X \leq x, Y \leq y)$ pro všechna $(x,y) \in \mathbb{R}^2$.*

Funkce $F(x) = \lim_{y\to\infty} H(x,y)$ a $G(y) = \lim_{x\to\infty} H(x,y)$ nazýváme marginální distribuční funkce.

Věta 3 (Sklarova)**.** *Nechť X a Y jsou náhodné veličiny s distribučními funkcemi F a G a sdruženou distribuční funkcí H. Pak existuje kopule C taková, že*

$$H(x,y) = C(F(x),G(y)),\ (x,y) \in \mathbb{R}^2. \tag{1.1}$$

Pokud F a G jsou spojité, pak C je určena jednoznačně, jinak je určena jednoznačně na $RanF \times RanG$.

Naopak, pokud C je kopule a F a G jsou distribuční funkce, pak funkce definovaná jako v 1.1 je sdružená distribuční funkce nějakého vektoru a její marginální distribuční funkce jsou F a G.

Důkaz. Důkaz věty v této (stochastické) verzi je uveden v knize McNeil, Frey a Embrechts (2015).

□

Máme-li dánu sdruženou distribuční funkci a tedy i její marginální distribuční funkce, pak je možné vyjádřit příslušnou kopuli ze vzorce 1.1 podle následujícího důsledku Sklarovy věty.

Důsledek. Nechť H je sdružená distribuční funkce se spojitými marginálními distribučními funkcemi F a G a kopulí C, pak pro všechna $(u,v) \in \mathbf{I}^2$ platí $C(u,v) = H(F^{-1}(u),G^{-1}(v))$.

V následujícím příkladu z Nelsen (2010) máme dánu sdruženou distribuční funkci a snažíme se vyjádřit příslušnou kopuli, použijeme k tomu předchozí důsledek.

Příklad. Uvažujme distribuční funkci

$$H(x,y) = \begin{cases} \frac{(x+1)(e^y-1)}{x+2e^y-1}, & (x,y) \in [-1,1] \times [0,\infty], \\ 1 - e^{-y}, & (x,y) \in (1,\infty] \times [0,\infty], \\ 0, & \text{jinak.} \end{cases}$$

Marginální distribuční funkce F resp. G získáme limitním přechodem pro $y \to \infty$ resp. $x \to \infty$:

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{(x+1)}{2}, & x \in [-1,1], \\ 1, & x > 1 \end{cases} \text{ a } G(y) = \begin{cases} 0, & y < 0, \\ 1 - e^{-y}, & y \geq 0. \end{cases}$$

Kvantilové funkce jsou $F^{-1}(u) = 2u - 1$ a $G^{-1}(v) = -ln(1 - v)$ pro $u, v \in \mathbf{I}$. F a G jsou spojité, tedy dle důsledku 1.2 kopuli C příslušnou distribuční funkci H dostaneme jako $C(u,v) = H(F^{-1}(u),G^{-1}(v)) = uv/(u + v - uv)$, $(u,v) \in \mathbf{I}^2$.

Uvažujme nyní H, F, G a C jako v předchozím důsledku, pro rozdělení vektoru $(F(X),G(Y))^T$ je $\mathsf{P}(F(X) \leq u,G(Y) \leq v) = \mathsf{P}(X \leq F^{-1}(u),Y \leq G^{-1}(v)) = H(F^{-1}(u),G^{-1}(v)) = C(u,v)$. Tedy při formálním rozšíření kopule C na $\mathbb{R}^2$ lze C chápat jako sdruženou distribuční funkci vektoru $(F(X),G(Y))^T$, přičemž $F(X)$ i $G(Y)$ mají rovnoměrné rozdělení na (0,1). Někdy, nedojde-li ke zmatení, budeme tedy kopulí $C(u,v)$ rozumět sdruženou distribuční funkci vektoru $(U,V)^T$, kde U a V mají rovnoměrné rozdělení na (0,1).

Další vlastností kopulí je jejich invariance vůči rostoucím transformacím příslušných náhodných veličin.

Tvrzení 4. *Nechť X a Y jsou spojité náhodné veličiny s kopulí $C_{X,Y}$ a nechť α resp. β jsou rostoucí funkce na $RanX$ resp. $RanY$. Pak $C_{\alpha(X),\beta(Y)} = C_{X,Y}$.*

Důkaz. Označme F_1 resp. G_1 distribuční funkce X resp. Y a F_2 resp. G_2 distribuční funkce $\alpha(X)$ resp. $\beta(Y)$. Pak pro všechna $(x,y) \in \mathbb{R}^2$

$$\begin{aligned} C_{\alpha(X),\beta(Y)}(F_2(x),G_2(y)) &= \mathsf{P}(\alpha(X) \leq x,\beta(Y) \leq y) = \mathsf{P}(X \leq \alpha^{-1}(x),Y \leq \beta^{-1}(y)) \\ &= C_{X,Y}(F_1(\alpha^{-1}(x)),G_1(\beta^{-1}(y))) = C_{X,Y}(F_2(x),G_2(y)), \end{aligned}$$

kde poslední rovnost plyne z výpočtu $F_2(x) = \mathsf{P}(\alpha(X) \leq x) = \mathsf{P}(X \leq \alpha^{-1}(x)) = F_1(\alpha^{-1}(x))$ pro všechna $x \in \mathbb{R}$, analogicky pro G_2. Vzhledem ke spojitosti F_2 a G_2 je $RanF_2 \times RanG_2 = \mathbf{I}^2$ a tedy rovnost platí pro všechny body z $\mathbf{I}^2$.

□

Pokud bychom ve tvrzení 4 uvažovali obě funkce klesající, případně jednu klesající a druhou rostoucí, dostali bychom deterministický vztah mezi kopulemi, např. pro obě funkce klesající je $C_{\alpha(X),\beta(Y)}(u,v) = u + v - 1 + C_{X,Y}(1-u,1-v)$.

V návaznosti na Sklarovu větu se krátce zaměříme na stochastickou interpretaci kopulí Π, M a W, která bude užitečná v kapitole 2 při zkoumání závislostí mezi dvěma náhodnými veličinami.

Tvrzení 5. *Nechť X a Y jsou spojité náhodné veličiny se sdruženou distribuční funkcí H, marginálními distribučními funkcemi F a G a kopulí C, pak:*

1. *X a Y jsou nezávislé právě tehdy, když $C = \Pi$,*
2. *Y je skoro jistě rostoucí funkcí X právě tehdy, když $C = M$,*
3. *Y je skoro jistě klesající funkcí X právě tehdy, když $C = W$.*

Důkaz. 1. X a Y jsou nezávislé právě tehdy, když $H(x,y) = F(x)\,G(y)$ pro všechna $(x,y) \in \mathbb{R}^2$, a to je ekvivalentní s $C(u,v) = uv = \Pi(u,v)$ pro všechna $(u,v) \in \mathbb{R}^2$.

Podrobnosti k důkazu 2. a 3. je možné najít v knize Nelsen (2010).

□

Pokud pro dvě náhodné veličiny bude jedna skoro jistě rostoucí funkcí druhé, budeme říkat, že jsou komonotónní, podobně v případě, že jedna bude skoro jistě klesající funkcí druhé, budeme říkat, že jsou kontramonotónní. Z předchozího tvrzení tedy máme, že $C = M$ právě tehdy, když X a Y jsou komonotónní a analogicky pro kontramonotonicitu.

1.3 Archimédovské kopule

Jedním z účelů této práce je zaměřit se na parametrický přístup při konstrukci dvourozměrných rozdělení pomocí kopulí. K tomu využijeme především speciální třídu kopulí zvanou Archimédovské kopule. Budeme vycházet ze situace, kdy je možné pomocí nějaké funkce φ kopuli ztransformovat tak, že platí $\varphi(C(u,v)) = \varphi(u) + \varphi(v)$. Vlastnosti funkce φ a příslušných kopulí jsou obsahem této sekce.

Dále budeme pracovat s funkcemi, jejichž definiční obor je jednotkový interval, pokud budeme uvažovat funkci, která obecně není definována v bodě 0, budeme ji automaticky chápat jako funkci, která je v bodě nula dodefinovaná příslušnou limitou zprava.

Definice 6. *Nechť φ je spojitá, klesající funkce z $\mathbf{I}$ do $[0,\infty]$ a $\varphi(1) = 0$. Pak pseudo-inverz funkce φ je funkce $\varphi^{[-1]}$ z $[0,\infty]$ na $\mathbf{I}$ daná předpisem*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & t \in [0,\varphi(0)), \\ 0, & t \in [\varphi(0),\infty]. \end{cases}$$

Následující tvrzení nám poskytuje nutnou a postačující podmínku pro funkci φ, aby funkce C ze vzorce $\varphi(C(u,v)) = \varphi(u) + \varphi(v)$ byla kopulí, důkaz je uveden v knize Nelsen (2010).

Tvrzení 6. *Nechť φ je spojitá, klesající funkce z $\mathbf{I}$ do $[0,\infty]$, $\varphi(1) = 0$ a nechť $\varphi^{[-1]}$ je pseudo-inverz funkce φ. Pak funkce $C : \mathbf{I}^2 \to \mathbf{I}$, $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ je kopule právě tehdy, když φ je konvexní.*

Nyní už definice Archimédovské kopule.

Definice 7. *Nechť φ je spojitá, ryze klesající funkce z $\mathbf{I}$ do $[0,\infty]$ a $\varphi(1) = 0$. Kopuli C tvaru $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ nazýváme Archimédovská kopule. Funkci φ nazýváme generátor kopule C, v případě, že $\varphi(0) = \infty$, hovoříme o ryzím generátoru kopule C.*

Kopule závislé na jednom či více parametrech tvoří tzv. rodiny kopulí. Pro Archimédovské kopule získáme příslušné rodiny tak, že za generátor kopule zvolíme parametrickou funkci. V následujícím příkladu ukážeme konstrukci konkrétní rodiny Archimédovských kopulí.

Příklad. Uvažujme funkci $\varphi_\theta(x) = (x^{-1/\theta} - 1)^\theta$, $x \in \mathbf{I}$, tato funkce je spojitá, ryze klesající a $\varphi_\theta(1) = 0$ pro $\theta \in [1,\infty)$, splňuje tedy předpoklady tvrzení 6. Dále $\varphi_\theta(0) = \infty$, tedy $\varphi_\theta^{[-1]}(t) = \varphi_\theta^{-1}(t) = (t^{1/\theta} + 1)^{-\theta}$, $t \in [0,\infty)$. Dosazením do vzorce z tvrzení 6 získáme rodinu Archimédovských kopulí

$$C_\theta(u,v) = [1 + \{(u^{-1/\theta} - 1)^\theta + (v^{-1/\theta} - 1)^\theta\}^{1/\theta}]^{-\theta}.$$

Pokud jako generátor zvolíme $\varphi(x) = 1 - x$, $x \in \mathbf{I}$, získáme kopuli W, podobně, volbou $\varphi(x) = -\log(x)$, $x \in \mathbf{I}$, získáme kopuli Π, tedy kopule W a Π jsou Archimédovské. Avšak, na rozdíl od kopulí W a Π, M Archimédovská není, nicméně M je častokrát limitním případem v rámci některých rodin, uvažujeme-li limitu pro parametr.

Na konci první kapitoly jsme zavedli částečné uspořádání na množině všech kopulí. Nyní se zaměříme na částečné uspořádání konkrétně pro Archimédovské

kopule. Uvažujme obecnou rodinu kopulí $\{C_\theta\}_{\theta\in\Theta}$, kde $\Theta \subset \mathbb{R}$. Pokud pro θ_1, $\theta_2 \in \Theta$, $\theta_1 \leq \theta_2$ platí $C_{\theta_1} \preceq C_{\theta_2}$ (resp. $C_{\theta_1} \succeq C_{\theta_2}$), říkáme, že $\{C_\theta\}_{\theta\in\Theta}$ je pozitivně (resp. negativně) uspořádaná.

V případě Archimédovských kopulí máme k dispozici nutnou a postačující podmínku pro určení zda daná rodina je pozitivně (nebo negativně) uspořádaná, důkaz tvrzení 7 lze najít v knize Nelsen (2010).

Tvrzení 7. *Nechť C_1 a C_2 jsou Archimédovské kopule s generátory φ_1 a φ_2. Pak $C_1 \preceq C_2$ právě tehdy, když $\varphi_1 \circ \varphi_2^{[-1]}(x+y) \leq \varphi_1 \circ \varphi_2^{[-1]}(x) + \varphi_1 \circ \varphi_2^{[-1]}(y)$ pro všechna x, $y \in [0,\infty)$.*

Vlastnosti, kterou splňuje funkce $\varphi_1 \circ \varphi_2^{[-1]}$ z předchozího tvrzení, se říká subaditivita (funkce f je subaditivní, pokud $f(x+y) \leq f(x) + f(y)$ pro všechna x, $y \in [0,\infty)$). Existuje několik zjednodušujících tvrzení pro určení, zda funkce $\varphi_1 \circ \varphi_2^{[-1]}$ je subaditivní, my je zde pro jednoduchost neuvádíme a v následující kapitole při vypisování vlastností jednotlivých Archimédovských rodin uvádíme jejich uspořádanost bez důkazu, příslušný výpočet by byl založen na předchozím tvrzení.

2. Závislost

Kopule v jistém smyslu vyjadřují strukturu závislosti mezi dvěma náhodnými veličinami. V této kapitole nejprve uvedeme ve spojitosti s kopulemi některé míry závislosti, které využijeme ve třetí kapitole, a zmíníme jejich možné interpretace. Ve druhé části se zaměříme na konkrétní rodiny kopulí, uvedeme hodnoty zmiňovaných měr a pro ilustraci budou zobrazeny výsledky generování z těchto kopulí pro různé parametry.

2.1 Míry závislosti

Všechny míry závislosti, které zmíníme v této kapitole se vztahují pouze ke kopulím, tj. nezávisí na marginálních rozděleních. Ve spojitosti s prvními dvěma mírami – Kendallovo tau a Spearmanovo rho – využijeme pojmu souladnosti. Řekneme, že dva body (x_1,y_1), $(x_2,y_2) \in \mathbb{R}^2$ jsou souladné, pokud $(x_1-x_2)(y_1-y_2) > 0$ a nesouladné, pokud $(x_1 - x_2)(y_1 - y_2) < 0$. V kontextu náhodných veličin, budeme-li uvažovat, že (x_1,y_1), (x_2,y_2) jsou dvě nezávislé realizace náhodného vektoru (X,Y), jsou, neformálně řečeno, náhodné veličiny X a Y souladné, pokud X má tendenci růst (a klesat) společně s Y a analogicky pro nesouladné náhodné veličiny. To motivuje příslušné definice zmiňovaných měr. Pro náhodný vektor (X,Y) jeho nezávislou stochastickou kopií rozumíme vektor (X',Y'), který je nezávislý s (X,Y) a má stejné rozdělení.

Definice 8. *Nechť (X,Y) je náhodný vektor a (X',Y') je jeho nezávislá stochastická kopie. Pak Kendallovo tau definujeme jako*

$$\tau_{X,Y} = P((X - X')(Y - Y') > 0) - P((X - X')(Y - Y') < 0).$$

V následujícím tvrzení získáme vzorec pro výpočet Kendallova tau. Jak bylo zmíněno v úvodu sekce, Kendallovo tau závisí pouze na příslušné kopuli, tedy i příslušný vzorec bude záviset jen na kopuli.

Tvrzení 8. *Nechť (X,Y) je náhodný vektor se spojitými marginálními distribučními funkcemi a kopulí C. Pak*

$$\tau_{X,Y} = 4 \iint_{\mathbf{I}^2} C(u,v)\,\mathrm{d}C(u,v) - 1.$$

Důkaz. Nejprve označme F a G marginální distribuční funkce vektoru (X,Y). Z definice Kendallova tau je

$$\begin{aligned}
\tau_{X,Y} &= \mathsf{P}((X - X')(Y - Y') > 0) - \mathsf{P}((X - X')(Y - Y') < 0) \\
&= 2\,\mathsf{P}((X - X')(Y - Y') > 0) - 1 \\
&= 2(\mathsf{P}(X > X',Y > Y') + \mathsf{P}(X < X',Y < Y')) - 1 \\
&= 4\,\mathsf{P}(X > X',Y > Y') - 1,
\end{aligned}$$

kde druhá rovnost plyne ze spojitosti (X,Y) a poslední rovnost z toho, že (X',Y') je nezávislá kopie (X,Y). Připomeňme, že distribuční funkce vektoru (X,Y) je ve

tvaru $C(F(x),G(y))$, máme tedy

$$\begin{aligned}\mathsf{P}(X > X',Y > Y') &= \mathsf{P}(X < X',Y < Y') \\ &= \iint_{\mathbb{R}^2} \mathsf{P}(X' < x,Y' < y)\,\mathrm{d}C(F(x),G(y)) \\ &= \iint_{\mathbb{R}^2} C(F(x),G(y))\,\mathrm{d}C(F(x),G(y)) = \iint_{\mathbf{I}^2} C(u,v)\,\mathrm{d}C(u,v),\end{aligned}$$

kde jsme v poslední rovnosti použili substituci $u = F(x)$, $v = G(y)$. Celkem tedy $\tau_{X,Y} = 4\iint_{\mathbf{I}^2} C(u,v)dC(u,v) - 1$.

□

Dále v textu budeme někdy místo značení $\tau_{X,Y}$ používat τ_C, kde C je příslušná kopule náhodných veličin X a Y.

V případě Archimédovských kopulí je možné Kendallovo tau vyjádřit pomocí alternativního vzorce, který závisí pouze na příslušném generátoru kopule. Důkaz následujícího tvrzení je uveden v knize Nelsen (2010).

Tvrzení 9. *Nechť X a Y jsou náhodné veličiny s Archimédovskou kopulí s generátorem φ. Pak*

$$\tau_{X,Y} = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,\mathrm{d}t. \tag{2.1}$$

Příklad. Vypočítáme Kendallovo tau pro rodinu Archimédovských kopulí danou předpisem $C_\theta(u,v) = \{\max(u^{-\theta}+v^{-\theta}-1,0)\}^{-1/\theta}$ s generátorem $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta}-1)$, kde $\theta \in [-1,\infty) \setminus \{0\}$, tato rodina se nazývá Claytonova rodina. Nejprve

$$\frac{\varphi(t)}{\varphi'(t)} = \frac{\frac{1}{\theta}(t^{-\theta}-1)}{-t^{-(\theta+1)}} = \frac{t^{(\theta+1)}-t}{\theta},$$

následně dosadíme do vzorce 2.1:

$$\tau_{C_\theta} = 1 + 4\int_0^1 \frac{t^{(\theta+1)}-t}{\theta}dt = 1 + \frac{4}{\theta}\left(\frac{1}{\theta+2} - \frac{1}{2}\right) = \frac{\theta}{\theta+2}.$$

Limitním případem pro $\theta \to 0$ je kopule $C_0 = \Pi$, pro kterou $\tau_\Pi = 0$. Máme tedy $\tau_{C_\theta} = \theta/(\theta+2)$ pro všechna $\theta \in [-1,\infty)$.

Nyní přejdeme k definici Spearmanova rho. Definice je založena na stejném principu jako pro Kendallovo tau s tím rozdílem, že náhodné veličiny X' a Y' nyní nemají stejné sdružené rozdělení jako X a Y, ale jsou nezávislé.

Definice 9. *Nechť (X,Y) a $(\tilde{X},\tilde{Y})$ jsou nezávislé náhodné vektory takové, že $X \overset{d}{=} \tilde{X}$, $Y \overset{d}{=} \tilde{Y}$ a $\tilde{X}$ a $\tilde{Y}$ jsou nezávislé. Pak Spearmanovo rho definujeme jako*

$$\rho_{X,Y} = 3(\mathsf{P}((X-\tilde{X})(Y-\tilde{Y}) > 0) - \mathsf{P}((X-\tilde{X})(Y-\tilde{Y}) < 0)).$$

Podobně jako pro Kendallovo tau uvedeme vzorec pro výpočet Spearmanova rho, který závisí pouze na příslušné kopuli.

Tvrzení 10. *Nechť (X,Y) je náhodný vektor se spojitými marginálními distribučními funkcemi a kopulí C. Pak*

$$\rho_{X,Y} = 12\iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v) - 3 = 12\iint_{\mathbf{I}^2} C(u,v)\,\mathrm{d}u\,\mathrm{d}v - 3. \tag{2.2}$$

Důkaz. Na začátku důkazu postupujeme analogicky jako v důkazu tvrzení 8, dostáváme tedy $\rho_{X,Y} = 6(\mathsf{P}(X > \tilde{X},Y > \tilde{Y}) + \mathsf{P}(X < \tilde{X},Y < \tilde{Y})) - 3$. Dále upravíme zvlášť oba sčítance

$$\begin{aligned}\mathsf{P}(X > \tilde{X},Y > \tilde{Y}) = \mathsf{P}(\tilde{X} < X,\tilde{Y} < Y) &= \iint_{\mathbb{R}^2} \mathsf{P}(\tilde{X} < x,\tilde{Y} < y)\,\mathrm{d}C(F(x),G(y)) \\ &= \iint_{\mathbb{R}^2} \mathsf{P}(\tilde{X} < x,\tilde{Y} < y)\,\mathrm{d}C(F(x),G(y)) \\ &= \iint_{\mathbb{R}^2} \Pi(F(x),G(y))\,\mathrm{d}C(F(x),G(y)) = \iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v),\end{aligned}$$

kde jsme v poslední rovnosti použili substituci $u = F(x)$, $v = G(y)$. Pro druhý sčítanec je

$$\begin{aligned}\mathsf{P}(X < \tilde{X},Y < \tilde{Y}) &= \iint_{\mathbb{R}^2} \mathsf{P}(\tilde{X} > x,\tilde{Y} > y)\,\mathrm{d}C(F(x),G(y)) \\ &= \iint_{\mathbb{R}^2} (1 - F(x) - G(y) + \Pi(F(x),G(y)))\,\mathrm{d}C(F(x),G(y)) \\ &= \iint_{\mathbf{I}^2} (1 - u - v + uv)\,\mathrm{d}C(u,v) \\ &= 1 - \frac{1}{2} - \frac{1}{2} + \iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v) = \iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v),\end{aligned}$$

kde jsme v předposlední rovnosti využili toho, že $C(u,v)$ je sdružená distribuční funkce vektoru (U,V), kde U a V mají rovnoměrné rozdělení na (0,1). Celkem máme

$$\rho_{X,Y} = 6\left(\iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v) + \iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v)\right) - 3 = 12\iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v) - 3.$$

Druhou rovnost ve tvrzení získáme prohozením rolí C a Π.

□

Na Spearmanovo rho lze pohlížet ještě jiným způsobem, pro náhodné veličiny X a Y s distribučními funkcemi F a G a kopulí C se jedná o Pearsonův korelační koeficient ρ_P veličin $F(X)$ a $G(Y)$, tj. veličin jejichž distribuční funkce je kopule C.

Tvrzení 11. *Nechť náhodné veličiny X a Y mají spojité distribuční funkce F a G. Pak $\rho_{X,Y} = \rho_P(F(X),G(Y))$.*

Důkaz. Náhodné veličiny $F(X)$ a $G(Y)$ mají rovnoměrné rozdělení na (0,1), tedy $\mathsf{E}\,[F(X)] = \mathsf{E}\,[G(Y)] = 1/2$ a $\mathrm{var}(F(X)) = \mathrm{var}(G(Y)) = 1/12$. Z tvrzení 10 je

$$\begin{aligned}\rho_{X,Y} &= 12\iint_{\mathbf{I}^2} uv\,\mathrm{d}C(u,v) - 3 = 12\,\mathsf{E}\,[UV] - 3 = \\ &= \frac{\mathsf{E}\,[UV] - 1/4}{1/12} = \frac{\mathsf{E}\,[UV] - \mathsf{E}\,[U]\,\mathsf{E}\,[V]}{\sqrt{\mathrm{var}(U)}\sqrt{\mathrm{var}(V)}} = \rho_P(F(X),G(Y)).\end{aligned}$$

□

Uveďme nyní bez důkazu výčet několika vlastností Kendallova tau a Spearmanova rho, které jsou shrnuty v následujícím tvrzení. Pro obě míry použijeme symbol $\sigma_{X,Y}$, tj. $\sigma_{X,Y} \in \{\tau_{X,Y}, \rho_{X,Y}\}$.

Tvrzení 12. *Nechť X a Y náhodné veličiny, pak $\sigma_{X,Y}$ splňuje:*

1. $-1 \leq \sigma_{X,Y} \leq 1$,
2. $\sigma_{X,Y} = \sigma_{Y,X}$,
3. *pro kopule C_1 a C_2 takové, že $C_1 \preceq C_2$ je $\sigma_{C_1} \leq \sigma_{C_2}$,*
4. *pokud Y je s.j. rostoucí funkcí X, pak $\sigma_{X,Y} = \sigma_M = 1$,*
5. *pokud Y je s.j. klesající funkcí X, pak $\sigma_{X,Y} = \sigma_W = -1$,*
6. *pokud X a Y jsou nezávislé, pak $\sigma_{X,Y} = \sigma_\Pi = 0$,*
7. *pokud α a β jsou obě s.j. rostoucí, nebo klesající, pak $\sigma_{\alpha(X),\beta(Y)} = \sigma_{X,Y}$, pokud jedna z funkcí je rostoucí a druhá klesající, pak $\sigma_{\alpha(X),\beta(Y)} = -\sigma_{X,Y}$.*

Pro lepší interpretaci měr Kendallovo tau a Spearmanovo rho zadefinujeme dva druhy závislosti mezi náhodnými veličinami a uvedeme jejich vztah s danými mírami.

Uvažujme sdruženou pravděpodobnost, že $X \leq x$ a zároveň $Y \leq y$, tuto pravděpodobnost porovnáme s případem, kdy X a Y jsou nezávislé, to motivuje pojem kvadrantová závislost.

Definice 10. *Nechť X a Y jsou náhodné veličiny. X a Y jsou pozitivně kvadrantově závislé (PKZ), pokud platí:*

$$P(X \leq x, Y \leq y) \geq P(X \leq x)\, P(Y \leq y) \text{ pro všechna } (x,y) \in \mathbb{R}^2.$$

Ekvivalentně:

$$P(X > x, Y > y) \geq P(X > x)\, P(Y > y) \text{ pro všechna } (x,y) \in \mathbb{R}^2.$$

Analogicky definujeme negativní kvadrantovou závislost při opačných nerovnostech mezi pravděpodobnostmi.

Pro náhodné veličiny X a Y z definice 10 označme F a G jejich distribuční funkce, H sdruženou distribuční funkci a C příslušnou kopuli. Pokud F, G a H jsou spojité, pak je možné vlastnost PKZ zapsat jako $H(x,y) \geq F(x)G(y)$ pro všechna $(x,y) \in \mathbb{R}^2$, to je ekvivalentní s $C(u,v) \geq uv$ pro všechna $(u,v) \in \mathbf{I}^2$. Pro vlastnost PKZ musí být příslušné nerovnosti splněny pro všechna $(x,y) \in \mathbb{R}^2$, při alternativním vyjádření pomocí kopulí pro všechna $(u,v) \in \mathbf{I}^2$, nicméně pojem PKZ můžeme podobně použít i lokálně pro jednotlivé body $(u,v) \in \mathbf{I}^2$. Z tvrzení 10 máme $\rho_{X,Y} = 12 \iint_{\mathbf{I}^2} C(u,v)\,\mathrm{d}u\,\mathrm{d}v - 3$, tento výraz lze vyjádřit také jako $12 \iint_{\mathbf{I}^2} C(u,v) - uv\,\mathrm{d}u\,\mathrm{d}v$. Tedy při lokálním pohledu na PKZ může být Spearmanovo rho interpretováno jako průměrná kvadrantová závislost.

Z vyjádření $\rho_{X,Y} = 12 \iint_{\mathbf{I}^2} C(u,v) - uv\,\mathrm{d}u\,\mathrm{d}v = 12 \iint_{\mathbf{I}^2} C(u,v) - \Pi(u,v)\,\mathrm{d}u\,\mathrm{d}v$ je také vidět geometrický význam, Spearmanovo rho udává „objem“ (angl. signed volume) mezi kopulí C a kopulí pro nezávislost Π. V kontextu částečného uspořádání na množině kopulí tedy máme, že pokud $C_1 \preceq C_2$, pak se zvětší objem mezi C_2 a Π oproti objemu mezi C_1 a Π a tedy i $\rho_{C_2} \geq \rho_{C_1}$, čímž dostáváme vlastnost 3 ve tvrzení 12.

Pro Kendallovo tau zadefinujeme pojem závislost věrohodnostního poměru.

Definice 11. *Nechť X a Y jsou spojité náhodné veličiny se sdruženou hustotou h. Pak X a Y jsou pozitivně závislé ve smyslu věrohodnostního poměru (PZVP), pokud pro všechna x_1, y_1, x_2, $y_2 \in \mathbb{R} \cup \{\infty\}$ taková, že $x_1 \leq x_2$ a $y_1 \leq y_2$ platí:*

$$h(x_1,y_1)h(x_2,y_2) \geq h(x_1,y_2)h(x_2,y_1).$$

Analogicky pro X a Y definujeme negativní závislost věrohodnostního poměru (NZVP) při opačné nerovnosti.

Stejně jako při kvadrantové závislosti i zde je nerovnost v definici 11 vyžadována pro všechna $(x,y) \in \mathbb{R}^2$, opět ale pojem PZVP můžeme podobně použít pro jednotlivé body $(x,y) \in \mathbb{R}^2$. Pokud náhodné veličiny X a Y mají sdruženou hustotu h, pak se dá ukázat (důkaz viz. Nelsen (2010)), že pro Kendallovo tau je

$$\tau_{X,Y} = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{y_1}\int_{-\infty}^{x_1} h(x_1,y_1)h(x_2,y_2) - h(x_1,y_2)h(x_2,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}x_2\,\mathrm{d}y_2.$$

Tedy při lokálním pohledu na PZVP může být Kendallovo tau interpretováno jako průměrná závislost věrohodnostního poměru.

Pro porovnání obou zadefinovaných pojmů uveďme jen, že závislost věrohodnostního poměru je „silnější“ vlastností než kvadrantová závislost.

Přejděme k poslední uvažované míře závislosti mezi dvěma náhodnými veličinami, která se nazývá závislost chvostů. Na rozdíl od Kendallova tau a Spearmanova rho, které v jistém smyslu zahrnovaly všechny hodnoty X a Y, závislost chvostů je založena pouze na extrémních hodnotách uvažovaných veličin, neformálně, udává, jestli extrémně velké (resp. malé) hodnoty jedné veličiny jsou spojeny s extrémně velkými (resp. malými) hodnotami druhé veličiny.

Definice 12. *Nechť X a Y jsou spojité náhodné veličiny s distribučními funkcemi F a G. Pak koeficient závislosti horního chvostu definujeme jako limitu (pokud existuje):*

$$\lambda_U = \lim_{t\to 1^-} P(Y > G^{-1}(t)|X > F^{-1}(t)).$$

Analogicky koeficient závislosti dolního chvostu definujeme jako limitu (pokud existuje)

$$\lambda_L = \lim_{t\to 0^+} P(Y \leq G^{-1}(t)|X \leq F^{-1}(t)).$$

Pokud $\lambda_U > 0$ (resp. $\lambda_L > 0$), pak říkáme, že X a Y jsou závislé v horním (resp. dolním) chvostu. V opačném případě říkáme, že X a Y jsou nezávislé v horním (resp. dolním) chvostu.

V následujícím tvrzení získáme podobně jako pro Kendallovo tau a Spearmanovo rho vzorec pro výpočet závislosti chvostů.

Tvrzení 13. *Nechť X a Y jsou spojité náhodné veličiny s distribučními funkcemi F a G a kopulí C a nechť limity v definici 12 existují. Pak*

$$\lambda_U = 2 - \lim_{t\to 1^-} \frac{1-C(t,t)}{1-t}$$

a

$$\lambda_L = \lim_{t\to 0^+} \frac{C(t,t)}{t}.$$

Důkaz. Z definice λ_U a podmíněné pravděpodobnosti je

$$\begin{aligned}\lambda_U &= \lim_{t\to 1^-} \frac{\mathsf{P}(Y > G^{-1}(t), X > F^{-1}(t))}{\mathsf{P}(X > F^{-1}(t))} = \lim_{t\to 1^-} \frac{\mathsf{P}(G(Y) > t, F(X) > t)}{\mathsf{P}(F(X) > t)} \\ &= \lim_{t\to 1^-} \frac{1 - 2t + C(t,t)}{1-t} = 2 - \lim_{t\to 1^-} \frac{1 - C(t,t)}{1-t},\end{aligned}$$

kde ve třetí nerovnosti využíváme toho, že $G(Y)$ a $F(X)$ má rovnoměrné rozdělení na (0,1) a $(F(X),G(Y))$ má rozdělení s distribuční funkcí C.

Podobně pro λ_L máme

$$\begin{aligned}\lambda_L &= \lim_{t\to 0^+} \frac{\mathsf{P}(Y \le G^{-1}(t), X \le F^{-1}(t))}{\mathsf{P}(X \le F^{-1}(t))} = \lim_{t\to 0^+} \frac{\mathsf{P}(G(Y) \le t, F(X) \le t)}{\mathsf{P}(F(X) \le t)} \\ &= \lim_{t\to 0^+} \frac{C(t,t)}{t}.\end{aligned}$$

□

V případě, že kopule C z tvrzení 13 je Archimédovská, je ve tvaru $C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ pro nějaký generátor φ. Dosazením tohoto vyjádření do 2.3 a 2.4 dostaneme pro Archimédovské kopule

$$\lambda_U = 2 - \lim_{t\to 1^-} \frac{1 - \varphi^{[-1]}(2\varphi(t))}{1-t} = 2 - \lim_{t\to 0^+} \frac{1 - \varphi^{[-1]}(2t)}{1 - \varphi^{[-1]}(t)} \tag{2.3}$$

a

$$\lambda_L = \lim_{t\to 0^+} \frac{\varphi^{[-1]}(2\varphi(t))}{t} = \lim_{t\to\infty} \frac{\varphi^{[-1]}(2t)}{\varphi^{[-1]}(t)}. \tag{2.4}$$

Příklad. Uvažujme Claytonovu rodinu kopulí s generátorem $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta} - 1)$, kde $\theta \in [-1,\infty) \setminus \{0\}$. Vypočítáme koeficient závislosti horního a dolního chvostu. Využijeme k tomu vzorce 2.3 a 2.4 uvedené v předchozím odstavci. Nejprve $\varphi^{[-1]}(x) = \max((\theta x + 1)^{1/\theta}, 0)$. Pro libovolné $\delta > 0$ je pro všechna $x \in [0,\delta)$ $(\theta x + 1)^{-1/\theta} > 0$, při výpočtu λ_U tedy nemusíme uvažovat maximum:

$$\lambda_U = 2 - \lim_{x\to 0^+} \frac{1 - (\theta 2x + 1)^{-1/\theta}}{1 - (\theta x + 1)^{-1/\theta}} = 2 - \lim_{x\to 0^+} \frac{\frac{1}{\theta}(2\theta x + 1)^{-1/\theta - 1} 2\theta}{\frac{1}{\theta}(\theta x + 1)^{-1/\theta - 1}\theta} = 2 - 2 = 0.$$

Výpočet λ_L rozdělíme na dvě části pro $\theta > 0$ a $\theta \in [-1,0)$. Pokud $\theta > 0$, pak $(\theta x - 1)^{-\frac{1}{\theta}} > 0$ pro všechna $x \in [0,\infty)$, tedy i zde nemusíme uvažovat maximum:

$$\lambda_L = \lim_{x\to\infty} \frac{(\theta 2x + 1)^{-1/\theta}}{(\theta x + 1)^{-1/\theta}} = \lim_{x\to\infty} \left(\frac{\theta 2x + 1}{\theta x + 1}\right)^{-1/\theta} = 2^{-1/\theta}.$$

Pro $\theta \in [-1,0)$ využijeme vzorec z tvrzení 13. Pro $t \le 2^{1\theta}$ je $[\max(2t^{-\theta} - 1, 0)] = 0$ a využijeme L'Hôpitalovo pravidlo:

$$\lambda_L = \lim_{t\to 0^+} \frac{(\max(2t^{-\theta} - 1, 0))^{-1/\theta}}{t} = \lim_{t\to 0^+} \frac{0}{t} = 0.$$

Z definice závislosti chvostů plyne, že pro kopuli $C_0 = \Pi$ je $\lambda_U = \lambda_L = 0$, zahrneme tedy do celkového výsledku i hodnotu $\theta = 0$, celkem máme:

$$\lambda_U = 0 \text{ a } \lambda_L = \begin{cases} 2^{-\frac{1}{\theta}}, & \theta > 0, \\ 0, & \theta \in [-1,0]. \end{cases}$$

2.2 Rodiny kopulí

Nyní se zaměříme na některé vybrané rodiny kopulí. Pro každou rodinu uvedeme hodnoty Kendallova tau, Spearmanova rho, pokud bude vztah jednoduchý, a koeficientů závislosti chvostů, případně zmíníme některé další vlastnosti. Na základě toho se pokusíme určit pro jaké typy dat by jednotlivé rodiny mohly být vhodné, což využijeme ve třetí kapitole, kde budeme pracovat s reálnými daty. Pro ilustraci budou u každé rodiny zobrazeny výsledky generování pro různé parametry. Rodiny kopulí byly vybrány na základě nejčastějšího používání v praxi, dále na základě toho, aby všechny možné hodnoty Kendallova tau a Spearmanova rho pokrývaly co největší část intervalu $[-1,1]$, a aby byly zastoupeny různé (ne)závislosti chvostů.

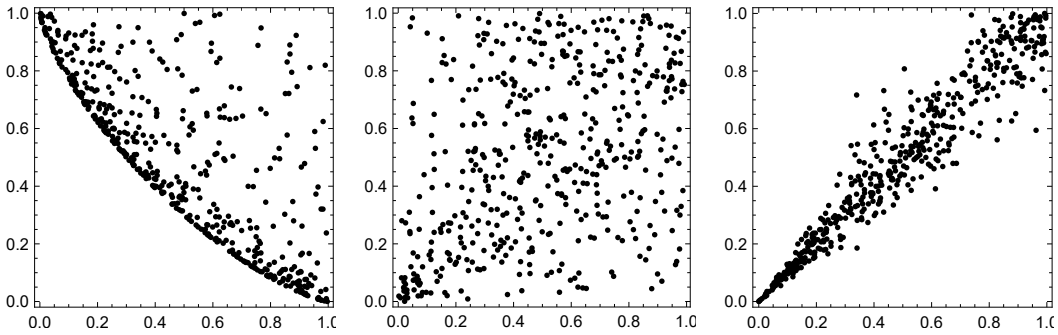
Claytonova rodina

- $C_\theta(u,v) = (\max(u^{-\theta}+v^{-\theta}-1,0))^{-1\theta}$, $\varphi_\theta(t) = \frac{1}{\theta}(t^{-\theta}-1)$, kde $\theta \in [-1,\infty)\backslash\{0\}$
- limitní případy: $C_{-1} = W$, $C_0 = \Pi$, $C_\infty = M$
- Claytonova rodina je pozitivně uspořádaná, to jest pokud $\theta_1 \leq \theta_2$, pak $C_{\theta_1} \preceq C_{\theta_2}$. Společně s předchozím bodem tedy dostáváme, že Claytonova rodina „interpoluje“ případy od kontramonotonicity přes nezávislost po komonotonicitu. Všimněme si ještě, že pro $\theta \in [-1,0)$ je nosičem rozdělení množina $\{(u,v) \in (0,1) : u^{-\theta} + v^{-\theta} - 1 > 0\} \subset (0,1)$, tedy rozšíření Claytonovy rodiny pro záporné hodnoty parametru θ není vhodné pro modelování dat.
- $\tau_\theta = \frac{\theta}{\theta+2}$
- $\lambda_U = 0$, $\lambda_L = \begin{cases} 2^{-1/\theta}, & \theta > 0, \\ 0, & \theta \in [-1,0] \end{cases}$

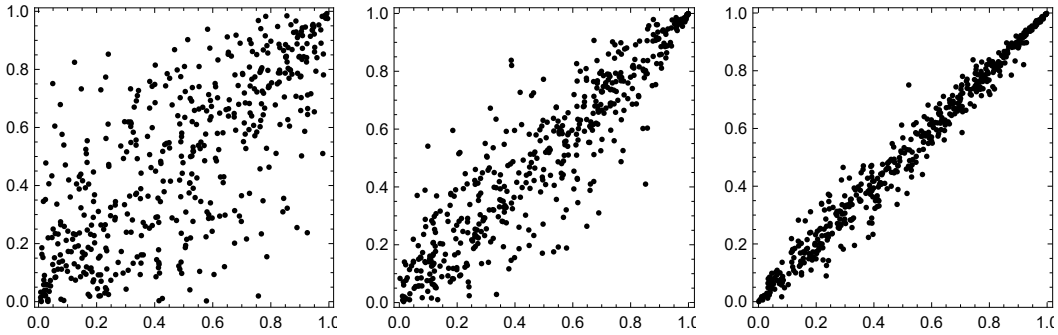
Claytonova rodina kopulí by mohla být vhodná k modelování situací, kdy velmi malé hodnoty jedné veličiny jsou spojeny s velmi malými hodnotami veličiny druhé (závislost v dolním chvostu), zatímco pro velké hodnoty jsou veličiny nezávislé. Pro zvětšující se hodnoty parametru je na obrázku 2.1 zřetelnější závislost v dolním chvostu.

Gumbelova rodina

- $C_\theta(u,v) = \exp(-\{(-\ln u)^\theta + (-\ln v)^\theta\}^{1/\theta})$, $\varphi_\theta(t) = (-\ln t)^\theta$, kde $\theta \in [1,\infty)$
- limitní případy: $C_1 = \Pi$, $C_\infty = M$
- Gumbelova rodina je stejně jako Claytonova rodina pozitivně uspořádaná. Na rozdíl od Claytonovy rodiny ale tato rodina „interpoluje“ pouze případy od nezávislosti po komonotonicitu.
- $\tau_\theta = 1 - \frac{1}{\theta}$
- $\lambda_U = 2 - 2^{1/\theta}$, $\lambda_L = 0$



Obrázek 2.1: Náhodný výběr z Claytonovy kopule o rozsahu 1000 pro hodnoty parametru θ pořadě $-3/4$, 1 a 10.



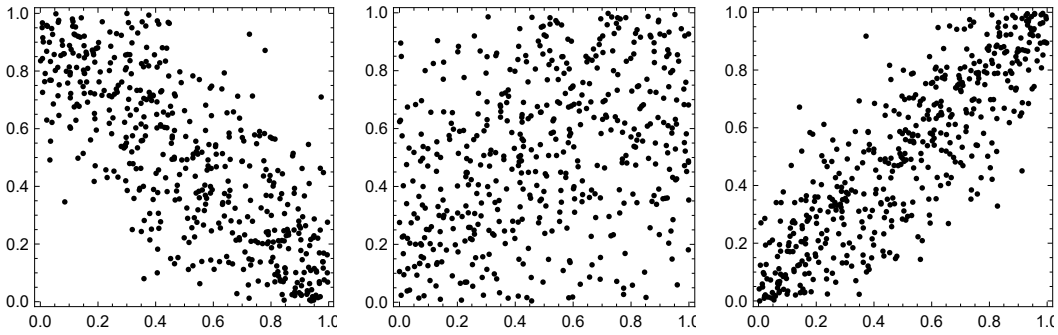
Obrázek 2.2: Náhodný výběr z Gumbelovy kopule pro hodnoty parametru θ pořadě 2, 4 a 10.

Vzhledem ke třetímu bodu není tato rodina vhodná pro modelování vztahu mezi dvěma veličinami s negativní závislostí. Ohledně závislosti chvostů máme opačnou situaci než v předchozím případě, Gumbelova rodina kopulí je závislá v horním chvostu, což je možné pozorovat na obrázku 2.2.

Frankova rodina

- $C_\theta(u,v) = -\frac{1}{\theta}\ln(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$, $\varphi_\theta(t) = -\ln(\frac{e^{\theta t}-1}{e^{\theta}-1})$, $\theta \in (-\infty,\infty)\setminus\{0\}$
- limitní případy: $C_{-\infty} = W$, $C_0 = \Pi$, $C_\infty = M$
- Frankova rodina je pozitivně uspořádaná a stejně jako Claytonova rodina „interpoluje“ případy od kontramonotonicity přes nezávislost po komonotonicitu
- $\tau_\theta = 1 - \frac{4}{\theta}(1 - D_1(\theta))$, $\rho_\theta = 1 - \frac{12}{\theta}(D_1(\theta) - D_2(\theta))$ kde $D_k(x)$ je Debyeho funkce $D_k(x) = kx^{-k}\int_0^x \frac{t^k}{e^t-1}\,\mathrm{d}t$, $k = 1{,}2$
- $\lambda_U = 0$, $\lambda_L = 0$

Frankovu rodinu lze použít pro modelování veličin, které jsou nezávislé v obou chvostech. V levých dolních a pravých horních rozích na obrázku 2.3 nepozorujeme takové koncentrace bodů jako v případě Claytonovy respektive Gumbelovy rodiny.



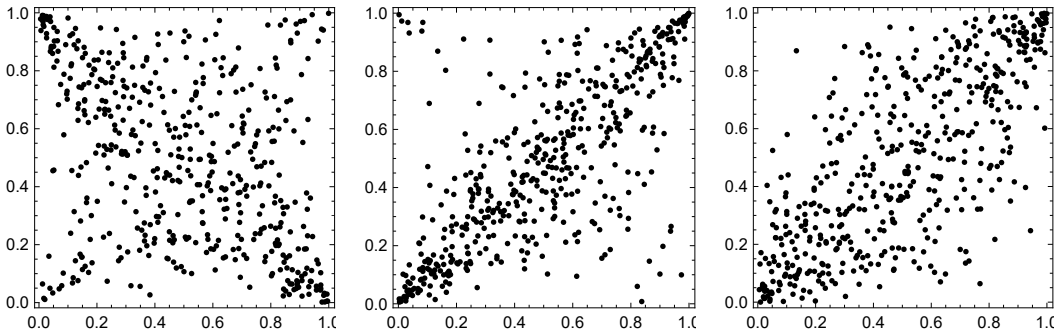
Obrázek 2.3: Náhodný výběr z Frankovy kopule o rozshau 1000 pro hodnoty parametru θ pořadě -7, 1 a 10.

Rodina t kopulí

- Rodina t kopulí je odvozená z dvourozměrného Studentova t rozdělení s parametry ν stupňů volnosti a pozitivně definitní maticí $P = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}$. Tato rodina kopulí se od předešlých rodin liší především tím, že příslušné kopule jsou vyjádřeny implicitně, k získání (implicitního) tvaru t kopule využijeme důsledek 1.2 Sklarovy věty, který aplikujeme na dvourozměrné t rozdělení $t_2(\nu,P)$, v obecném tvaru tedy dostaneme $C_{\nu,P}(u,v) = T_{\nu,P}(T_\nu^{-1}(u),T_\nu^{-1}(v))$, kde T_ν je distribuční funkce (jednorozměrného) t rozdělení s ν stupni volnosti a $T_{\nu,P}$ je sdružená distribuční funkce dvourozměrného t rozdělení s ν stupni volnosti a maticí P. Dále budeme t kopuli místo $C_{\nu,P}$ značit prostřednictvím σ, tj. $C_{\nu,\sigma}$.
- $C_{\nu,\sigma}(u,v) = \int_{-\infty}^{T_\nu^{-1}(u)} \int_{-\infty}^{T_\nu^{-1}(v)} \frac{\Gamma((\nu+2)/2)}{\Gamma(\nu/2)\pi\nu\sqrt{1-\sigma^2}} (1 + \frac{x^2-2\sigma xy+y^2}{\nu(1-\sigma^2)})^{-(\nu+2)/2} \,\mathrm{d}x\,\mathrm{d}y$ pro $\sigma \in (-1,1)$ a $\nu > 0$
- limitní případy: $C_{\nu,0} = W$, $C_{\nu,1} = M$
- limitními případy jsou opět kopule W a M, ale, na rozdíl od předchozích rodin, kopule Π není speciálním případem této rodiny
- $\tau_{\nu,\sigma} = \frac{2}{\pi} \arcsin(\sigma)$
- $\lambda := \lambda_U = \lambda_L = 2\, T_{\nu+1}\left(-\sqrt{(\nu+1)(1-\sigma)/(1+\sigma)}\right)$
- při pevném σ je λ klesající funkcí ν a při pevném ν je λ rostoucí funkcí σ

Rodina t kopulí doplňuje předchozí případy tím, že je závislá v obou chvostech, jak může být patrné na obrázku 2.4.

Další rodiny společně s jejich vlastnostmi je možné najít např. v knize Joe (2015) nebo v článku Cuevas, Yela a Achcar (2019).



Obrázek 2.4: Náhodný výběr z t kopule o rozsahu 1000 pro hodnoty parametru (ν, σ) pořadě $(1, -1/4)$, $(1, 3/4)$ a $(50, 3/4)$.

3. Aplikace na data

V rámci této kapitoly aplikujeme teorii kopulí na data. Při modelování rozdělení vícerozměrných dat může někdy být vhodné rozdělit si případ na dvě části, tou první je modelování marginálních rozdělení daných veličin, druhá část je pak modelování závislostí mezi jednotlivými veličinami. Zaměříme se na druhou část, v rámci které využijeme teorii kopulí, konkrétně se zaměříme na odhadování neznámých parametrů kopulí. Teoretické postupy budou ilustrovány na reálných datech.

V celé kapitole budeme uvažovat náhodný výběr $(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$ dvourozměrných vektorů, kde $\boldsymbol{X}_i = (X_{i,1},X_{i,2})^\top$ pro $i = 1,\ldots,n$. Případně budeme jako $\boldsymbol{X} = (X_1,X_2)^\top$ značit generický vektor a jeho složky.

Začneme momentovou metodou pro odhad parametrů kopulí, výhodou této metody je, že není potřeba znát marginální rozdělení zkoumaných veličin.

Momentová metoda Při této metodě využijeme teoretického vztahu mezi Kendallovým tau nebo Spearmanovým rho a parametrem jednotlivých rodin kopulí. Nejprve na základě dat odhadneme uvažovanou míru závislosti a poté z příslušného vztahu vyjádříme odhadnutou hodnotu parametru.

Pro Kendallovo tau používáme následující odhad:

$$\hat{\tau}_{X_1,X_2} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \operatorname{sign}((X_{i,1}-X_{j,1})(X_{i,2}-X_{j,2})).$$

Dosazením do příslušných vztahů mezi Kendallovým tau a parametrem rodin kopulí uvažovaných v předchozí sekci dostaneme odhady, které jsou pro Archimédovské kopule uvedeny v tabulce 3.1. Pro t kopuli dostaneme odhad $\hat{\sigma} = \sin(\frac{\pi}{2}\hat{\tau})$.

Pro Spearmanovo rho používáme odhad:

$$\hat{\rho}_{X_1,X_2} = \frac{12}{n(n^2-1)} \sum_{i=1}^{n} (\operatorname{rank}(X_{i,1}) - \frac{1}{2}(n+1))(\operatorname{rank}(X_{i,2}) - \frac{1}{2}(n+1)).$$

Stejně jako pro Kendallovo tau dostaneme odhad neznámého parametru kopule dosazením do příslušného vztahu mezi Spearmanovým rho a parametrem.

Další způsob jak lze odhadnout neznámé parametry je využít metodu maximální věrohodnosti.

Metoda maximální věrohodnosti Tato metoda je založena na maximalizaci součinu hustot v napozorovaných hodnotách vůči neznámým parametrům $\boldsymbol{\eta} \in \mathbb{R}^d$,

odhad	Clayton	Gumbel	Frank
$\hat{\theta}$	$\frac{2\hat{\tau}}{1-\hat{\tau}}$	$(1-\hat{\tau})^{-1}$	$g(\hat{\tau})^*$

*Pozn:** $\hat{\theta}$ získáme vyjádřením ze vztahu $\hat{\tau} = 1 - 4/\hat{\theta}(1 - D_1(\hat{\theta}))$.

Tabulka 3.1: Odhad parametru θ jednotlivých rodin kopulí pomocí Kendallova tau.

$d \in \mathbb{N}$. Funkci, kterou maximalizujeme nazýváme *věrohodnostní funkce*, značíme $L_n(\boldsymbol{\eta})$. Někdy navíc pro zjednodušení věrohodnostní funkci zlogaritmujeme, tuto funkci pak nazýváme *logaritmická věrohodnost*, značíme $\ell_n(\boldsymbol{\eta})$. Příslušný odhad je pak hodnota $\hat{\boldsymbol{\eta}}_n$, ve které $L_n(\boldsymbol{\eta})$ (resp. $\ell_n(\boldsymbol{\eta})$) nabývá svého maxima.

Na rozdíl od metody momentů, zde jsou při odhadování parametrů zahrnuty i marginální rozdělení. V závislosti na předpokladech o marginálních rozděleních je možné rozlišovat dva modely – parametrický a semi-parametrický. Při parametrickém modelu předpokládáme, že známe tvar marginálních distribučních funkcí $F(\cdot;\boldsymbol{\mu})$ a $G(\cdot;\boldsymbol{\nu})$, které závisí na neznámých parametrech $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\nu} \in \mathbb{R}^q$, p, $q \in \mathbb{N}$. Pak dle Sklarovy věty je $H(x,y;\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\nu}) = C_{\boldsymbol{\theta}}(F(x;\boldsymbol{\mu}),G(y;\boldsymbol{\nu}))$, kde $H(\cdot,\cdot;\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\nu})$ je distribuční funkce rozdělení, ze kterého pochází náhodný výběr, a pro hustotu platí $h(x,y;\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\nu}) = c_{\boldsymbol{\theta}}(F(x;\boldsymbol{\mu}),G(y;\boldsymbol{\nu}))f(x;\boldsymbol{\mu})g(y;\boldsymbol{\nu})$, kde $f(\cdot;\boldsymbol{\mu})$ a $g(\cdot;\boldsymbol{\nu})$ jsou příslušné hustoty marginálních rozdělení a $c_{\boldsymbol{\theta}}$ je hustota rozdělení s distribuční funkcí $C_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \mathbb{R}^r$, $r \in \mathbb{N}$, hustotu $c_{\boldsymbol{\theta}}(u,v)$ lze získat jako $c_{\boldsymbol{\theta}}(u,v) = \partial C_{\boldsymbol{\theta}}(u,v)/\partial u \partial v$, pokud existuje. Věrohodnostní funkce, případně logaritmická věrohodnost, je pak tvaru

$$L_n(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\nu}) = \prod_{i=1}^{n} c_{\boldsymbol{\theta}}(F(x_{i,1};\boldsymbol{\mu}),G(x_{i,2};\boldsymbol{\nu}))f(x_{i,1};\boldsymbol{\mu})g(x_{i,1};\boldsymbol{\nu}),$$

případně

$$\ell_n(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\nu}) = \sum_{i=1}^{n} \log(c_{\boldsymbol{\theta}}(F(x_{i,1};\boldsymbol{\mu}),G(x_{i,2};\boldsymbol{\nu})))+ \\ + \sum_{i=1}^{n} log(f(x_{i,1};\boldsymbol{\mu})) + \sum_{i=1}^{n} log(g(x_{i,1};\boldsymbol{\nu}))$$

a odhady neznámých parametrů (speciálně odhad parametrů $\boldsymbol{\theta}$) jsou

$$(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\nu}}) = \underset{(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\nu})\in P}{\arg\max}\, L_n(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\nu}),$$

kde P je parametrický prostor.

Odhadování všech parametrů najednou může být v praxi výpočetně náročné, v tom případě je možné využít metodu IFM (inference function for margins), při které se nejdříve odhadnou neznámé parametry marginálních rozdělení a poté při těchto pevných odhadnutých parametrech se odhadne zbývající parametr kopule, více k metodě IFM lze najít v knize Joe (2015). Tuto metodu také použijeme v rámci prvního příkladu.

V semi-parametrickém modelu o marginálních rozděleních nic nepředpokládáme. Aplikujeme metodu maximální (pseudo) věrohodnosti přímo na kopuli $C_{\boldsymbol{\theta}}$, jakožto sdruženou distribuční funkci. K tomu potřebujeme pozorované hodnoty z tohoto rozdělení, avšak hodnoty, které pozorujeme pochází z rozdělení $(X_1,X_2)^\top$ (jehož kopule je C), tedy místo náhodného výběru $((X_{1,1},X_{1,2})^\top,\ldots,(X_{n,1},X_{n,2})^\top)$ uvažujeme tzv. pseudo-výběr $((F(X_{1,1}),G(X_{1,2}))^\top,\ldots,((F(X_{n,1}),G(X_{n,2}))^\top)$, kde F a G jsou marginální distribuční funkce veličin X_1 a X_2 (při využití pseudo-výběru, pak tuto metodu nazýváme metoda maximální pseudo-věrohodnosti). Vzhledem k tomu, že F a G jsou neznámé, je nutné uvažovat jejich odhady, různé konstrukce pseudo-výběrů se pak liší metodou odhadů F a G. V našem případě pro obecnou distribuční funkci F budeme stejně jako v knize McNeil a kol. (2015)

uvažovat odhad $\hat{F}_n(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbb{1}[X_i \leq x]$, který se od klasického odhadu liší tím, že ve jmenovateli je $(n+1)$ místo n, tato úprava zajistí, že v pseudo-výběru nedostaneme vektor $(1,1)^\top$, ve kterém mohou některé hustoty kopulí nabývat hodnoty ∞.

Příklad. Uvažujme Frankovu rodinu $C_\theta(u,v) = -\frac{1}{\theta}\ln(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$, $\theta \geq 1$. Pak hustotu získáme jako

$$c_\theta(u,v) = \frac{\partial C_\theta(u,v)}{\partial u \partial v} = \frac{\theta(1-e^{-\theta})e^{-\theta(x+y)}}{(e^{-\theta(x+y)} - e^{-\theta x} - e^{-\theta y} + e^{-\theta})^2} \text{ pro } (u,v) \in \mathbf{I}^2, \text{ 0 jinak.}$$

Věrohodnostní funkce je tvaru

$$L_n(\theta) = \prod_{i=1}^{n} \frac{\theta(1-e^{-\theta})e^{-\theta(u_i+v_i)}}{(e^{-\theta(u_i+v_i)} - e^{-\theta u_i} - e^{-\theta v_i} + e^{-\theta})^2},$$

kde (u_i,v_i), $i = 1,\ldots,n$, jsou hodnoty z pseudo-výběru. Zlogaritmování předchozí rovnosti získáme logaritmickou věrohodnost

$$\ell_n(\theta) = n\log(\theta) + n\log(1-e^{-\theta}) - \theta\sum_{i=1}^{n}(u_i+v_i) - \\ - 2\sum_{i=1}^{n}\log(e^{-\theta(u_i+v_i)} - e^{-\theta u_i} - e^{-\theta v_i} + e^{-\theta}).$$

Odhad neznámého parametru θ je pak $\hat{\theta}_n = \arg\max_{\theta \geq 1} l_n(\theta)$.

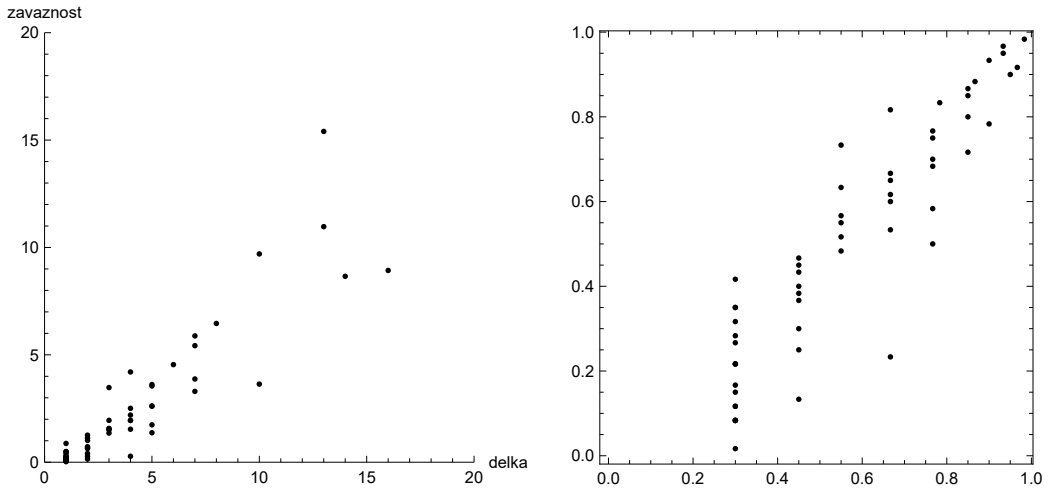
Pokud má kopule více parametrů je možné zkombinovat momentovou metodu s metodou maximální věrohodnosti, např. pro t kopuli bychom parametr σ odhadli pomocí momentové metody a zbylý parametr ν pomocí metody maximální věrohodnosti při pevném odhadnutém $\hat{\sigma}$.

Pomocí uvedených metod lze pro uvažované rodiny kopulí odhadnout neznámý parametr, zbývá určit, která z odhadnutých kopulí by pro pozorovaná data mohla být tou „nejvhodnější“. Pro každou rodinu kopulí je možné použít test dobré shody, který testuje hypotézu, že data pocházejí z rozdělení s danou rodinou kopulí proti alternativě, že data z daného rozdělení nepocházejí. Na základě tohoto testu lze pro napozorovaná data vyloučit některé rodiny, pro které test zamítá hypotézu, podrobnější informace k testu dobré shody a jeho porovnání s další metodou pro výběr kopulí lze najít např. v článku Fang a kol. (2014). V případě momentové metody se dále spoléháme především na apriorní předpoklady o datech (například závislosti chvostů), na základě kterých zvolíme vhodnou rodinu a poté odhadneme neznámý parametr. Při odhadování pomocí metody maximální věrohodnosti je možné využít nástroje, které se v obecném případě používají pro výběr z více modelů, jako příklady uveďme Akaikovo informační kritérium (AIC) a Bayesovo informační kritérium (BIC), které budeme využívat v rámci našich dat. AIC a BIC jsou definovány jako $\text{AIC} = -2l_n(\hat{\boldsymbol{\theta}}_n) + 2k$, $\text{BIC} = -2l_n(\hat{\boldsymbol{\theta}}_n) + k\ln(n)$ kde $\hat{\boldsymbol{\theta}}_n$ jsou odhadnuté parametry a k je počet odhadovaných parametrů. Při volbě „nejvhodnější“ kopule na základě AIC pak volíme tu s nejnižší hodnotou AIC, stejně tak pro výběr na základě BIC.

Uvedené postupy pro odhady parametrů nyní ve dvou příkladech aplikujeme na reálná data. Budeme uvažovat rodiny kopulí ze sekce 2.2, v obou příkladech uvedeme odhady pomocí momentové metody, v prvním příkladu využijeme navíc metodu maximální věrohodnosti se semi-parametrickým přístupem, ve druhém příkladu pak využijeme parametrický přístup. Odhady pomocí Kendallova tau, Spearmanova rho, metody IFM a maximální pseudo-věrohodnosti budeme značit pořadě $\hat{\theta}_\tau$, $\hat{\theta}_\rho$, $\hat{\theta}_{\mathrm{IFM}}$ a $\hat{\theta}_{\mathrm{MPL}}$.

Příklad 1 Prvním příkladem je aplikace na meteorologické sucho. Jednou z možností jak měřit meteorologické sucho je pomocí indexu SPI (Standardized Precipitation Index), který je konstruován na základě měsíčních úhrnů srážek v dané lokalitě, podrobnosti konstrukce indexu SPI jsou uvedeny v článku (Hao a kol., 2017). Jedno období sucha začíná okamžikem, kdy index SPI klesne pod předem danou a mez a končí okamžikem, kdy SPI nad danou mez opět vzroste. Pro období sucha pak pozorujeme jeho délku, tj. dobu, kterou SPI stráví pod danou mezí, a závažnost, která je definovaná jako záporný součet hodnot SPI za dané období, další informace k modelování délky a závažnosti sucha včetně konkrétní aplikace je možné najít v článku Lee a kol. (2012).

V našem případě uvažujeme 3-měsíční index SPI založen na měsíčních úhrnech srážek v kanadském městě Repentigny, provincie Quebec, za období 1973-2022 a jako mez byla zvolena hodnota 0, data pak sestávají z 59 pozorování délky a závažnosti sucha v dané oblasti, úhrny srážek byly získány prostřednictvím systému Mathematica (viz. Wolfram Research, 2021), index SPI pak sestrojen pomocí systému R (viz. R Core Team, 2021). Pro modelování délky období sucha se v praxi nejčastěji používá exponenciální rozdělení (Exp(λ)), pro závažnost pak gamma rozdělení ($\Gamma(a,p)$), tento příklad slouží především k ilustraci odhadu parametru kopulí, proto budeme tato rozdělení předpokládat i v našem případě. Data jsou zobrazena na obrázku 3.1 včetně zobrazení příslušného pseudo-výběru, pro větší přehlednost není v levém grafu zobrazeno odlehlé pozorování o souřadnicích $(68; 96{,}5376)$. Obrázek naznačuje pozitivní závislost mezi délkou a závažností.



Obrázek 3.1: Vlevo graf délky a závažnosti období sucha bez odlehlého pozorování $(68; 96{,}537)$, vpravo příslušný pseudo-výběr.

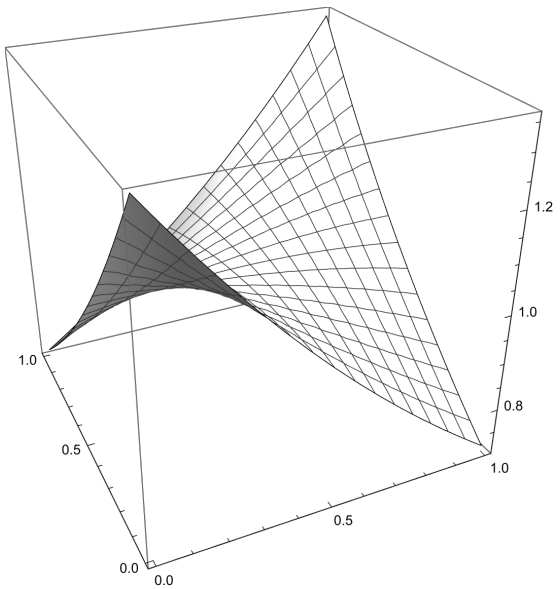
Odhady Kendallova tau a Spearmanova rho jsou $\hat{\tau} = 0{,}804$ a $\hat{\rho} = 0{,}920$. V rámci marginálních rozdělení je pro délku odhad parametru exponenciálního rozdělení $\hat{\lambda} = 0{,}197$, pro závažnost je odhad parametrů gamma rozdělení $\hat{a} = 0{,}121$ a $\hat{p} = 0{,}483$. Pro t kopuli pak dostaneme pomocí metody IFM odhady $\hat{\nu} = 166{,}948$ a $\hat{\rho} = 0{,}168$, hodnoty informačních kritérií jsou AIC = 0,841 a BIC = 4,997. Pro ostatní rodiny jsou odhady pomocí momentové metody a metody IFM společně s hodnotami AIC a BIC uvedeny v tabulce 3.2.

rodina	$\hat{\theta}_\tau$	$\hat{\theta}_\rho$	$\hat{\theta}_{\text{IFM}}$	AIC	BIC
Clayton	0,921	0,912	6,414	312,490	314.567
Gumbel	1,461	1,462	1,030	1,330	3,408
Frank	3,094	3,026	0,639	0,785	2,863

Tabulka 3.2: Odhady neznámých parametrů

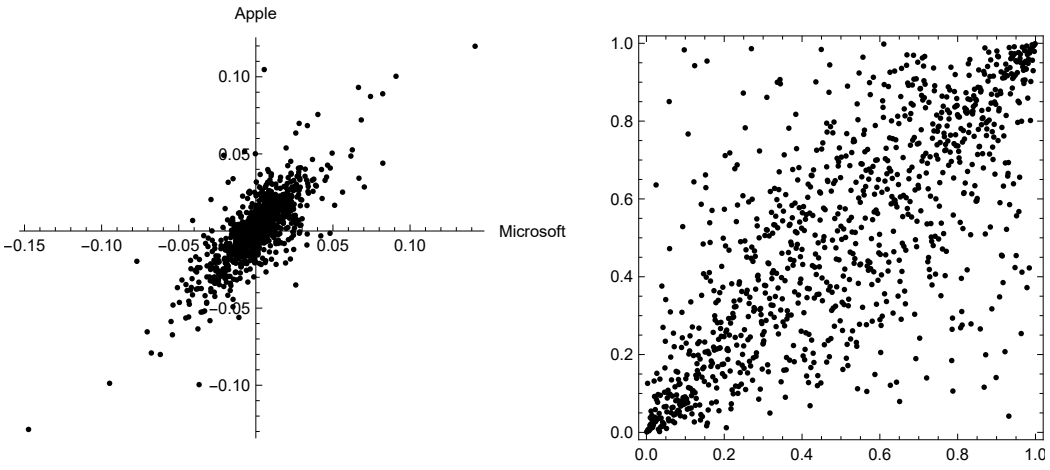
Poznámka. Při druhém kroku v rámci odhadu parametrů kopulí pomocí metody IFM se ve ztransformovaných datech (původní data ztransformovaná pomocí odhadnutých marginálních distribučních funkcí) vyskytovaly hodnoty blízké bodu (1,1), proto byla ztransformovaná data přenásobena konstantou $n/(n+1) = 59/60$ (podobně jako při semi-parametrickém přístupu)).

Na základě hodnoty AIC i BIC bychom pak jako nejvhodnější z uvažovaných kopulí zvolili Frankovu kopuli s parametrem $\hat{\theta} = 0{,}639$, která vyjadřuje pozitivní závislost a je nezávislá v obou chvostech, hustota Frankovy kopule s odhadnutým parametrem je znázorněna na obrázku 3.2. Avšak poznamenejme, že příslušná kopule nemusí být vhodná pro popis skutečnosti vzhledem k tomu, že jsme omezili výběr kopulí na čtyři rodiny, výsledky také mohou být ovlivněny předpokladem o marginálních rozděleních.



Obrázek 3.2: Hustota Frankovy kopule s parametrem $\hat{\theta} = 0{,}6399$

Příklad 2 Ve druhém příkladu jako data použijeme denní výnosy akcií společností Microsoft a Apple za roky 2019-2022, celkový počet pozorování je 1007, data byla získána prostřednictvím systému Mathematica (viz. Wolfram Research, 2021). Výnosy jsou zobrazeny na obrázku 3.3 včetně zobrazení příslušného pseudo-výběru. Grafické znázornění naznačuje pozitivní závislost mezi výnosy a také závislost v dolním i horním chvostu, na základě tohoto pozorování by pro modelování závislosti mohla být z uvažovaných rodin vhodná rodina t kopulí.



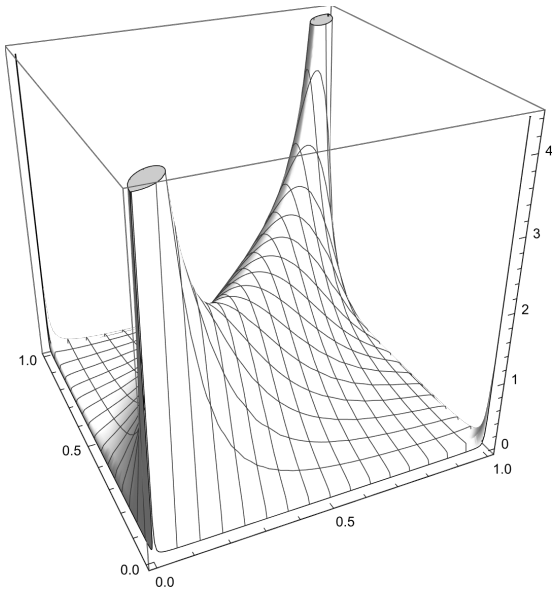
Obrázek 3.3: Vlevo výnosy společností Microsoft a Apple v letech 2019-2022, vpravo příslušný pseudo-výběr.

Odhady Kendallova tau a Spearmanova rho jsou $\hat{\tau} = 0{,}565$ a $\hat{\rho} = 0{,}739$. Pro t kopuli je odhad ν a σ pomocí maximální pseudo-věrohodnosti $\hat{\nu} = 2{,}667$ a $\hat{\sigma} = 0{,}775$, hodnoty informačních kritérií jsou $AIC = -1009{,}958$ a $BIC = -1000{,}128$. Odhady parametrů ostatních rodin pomocí momentové metody a maximální pseudo-věrohodnosti jsou společně s hodnotami AIC a BIC uvedeny v tabulce 3.3.

rodina	$\hat{\theta}_\tau$	$\hat{\theta}_\rho$	$\hat{\theta}_{MLE}$	AIC	BIC
Clayton	2,599	2,476	2,599	-821,264	-816,349
Gumbel	2,300	2,237	2,214	-870,252	-865,338
Frank	7,066	6,526	7,183	-838,437	-833,522

Tabulka 3.3: Odhady neznámých paramterů

Při volbě nejvhodnější z uvažovaných kopulí na základě AIC i na základě BIC bychom pak zvolili t kopuli s parametry $\hat{\nu} = 2{,}667$ a $\hat{\sigma} = 0{,}775$, která je závislá v dolním i horním chvostu, to souhlasí s naší úvahou založenou na obrázku 3.3. Hustota t kopule s odhadnutými parametry je znázorněna na obrázku 3.4, všimněme si chování hustoty v obou chvostech. Podobně jako v příkladu 1 poznamenejme, že vzhledem k omezenému počtu uvažovaných rodin, nemusí námi vybraná kopule vhodně reflektovat skutečnost.



Obrázek 3.4: Hustota t kopule s parametry $\hat{\nu} = 2{,}6673$ a $\hat{\sigma} = 0{,}7755$.

Závěr

V této práci byly uvedeny základy teorie kopulí a s nimi spojené náležitosti. V první kapitole byly kopule zadefinovány a uvedeny některé základní vlastnosti, za zmínku stojí například jejich spojitost, díky které je práce s kopulemi jednodušší. Kopule byly shora i zdola omezeny, tyto meze pak společně s kopulí pro nezávislost tvořili fundamentální kopule v celé práci. Zavedený pojem kopule pak byl použit ve Sklarově větě, která využívá kopule v kontextu marginálních a sdružených distribučních funkcí. Na závěr kapitoly byly představeny Archimédovské kopule, se kterými se, jak bylo možné usoudit, prostřednictvím jejich generátorů jednoduše pracuje, pomocí generátorů jsme také získali přímočarý postup pro konstrukci různých rodin.

Ve druhé kapitole byla nejprve zkoumána spojitost kopulí s Kendallovým tau a Spearmanovým rho, ukázalo se, že kopule a obě míry spolu úzce souvisí, byly zmíněny vzorce pro výpočet obou měr závislé pouze na příslušné kopuli. Následně byla s pomocí dvou typů závislosti uvedena možná interpretace uvažovaných měr a geometrická interpretace kopulí. Část o závislosti pak byla uzavřena zaměřením se na extrémní hodnoty prostřednictvím závislostí chvostů. V závěru druhé kapitoly byly i v rámci přípravy na třetí kapitolu uvedeny čtyři příklady rodin kopulí, na kterých jsme mohli vidět různé vlastnosti různých rodin.

Ve třetí kapitole byla uvedena metoda momentů pro odhad parametru kopule, jejíž výhodou, jak jsme viděli, byla nezávislost na marginálních rozděleních, druhou metodou pro odhady pak byla metoda maximální věrohodnosti, u které byl rozlišen parametrický a semi-parametrický přístup, v rámci této metody byla uvedena metoda IFM, která je méně výpočetně náročná. Dále byly zmíněny některé možnosti při výběru vhodné kopule. Obě metody pak byly aplikovány na příklad z hydrologie, kde byly modelovány délka a závažnost sucha, a na příklad z finančnictví, kde byly kopule použity na výnosy společností Microsoft a Apple.

Seznam použité literatury

Cuevas, J. R. T., Yela, J. P. a Achcar, J. A. (2019). A Method to Select Bivariate Copula Functions. *Revista Colombiana de Estadistica*, **42**, 61–80.

Fang, Y., Madsen, L. a Liu, L. (2014). Comparison of Two Methods to Check Copula Fitting. *IAENG International Journal of Applied Mathematics*, **44**, 53–61.

Hao, Z., Hao, F., Sing, V. P., Ouyang, W. a Cheng, H. (2017). An integrated package for drought monitoring, prediction and analysis to aid drought modeling and assessment. *Enviromental Modelling and Software*, **91**, 199–209.

Joe, H. (2015). *Dependence Modeling with Copulas*. 1st Edition. Chapman and Hall/CRC, New York. ISBN 9780429103186.

Lee, T., Modares, R. a Ouarda, T. B. M. J. (2012). Data-based analysis of bivariate copula tail dependence for drought duration and severity. *Hydrological Processes*, **27**, 1454–1463.

McNeil, A. J., Frey, R. a Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton University Press, New Yersey. ISBN 978-0-691-16627-8.

Nelsen, R. B. (2010). *An Introduction to Copulas*. Second Edition. Springer New York, New York. ISBN 978-1-4419-2109-3.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Wolfram Research, I. (2021). *Mathematica*. Version 13.0. Wolfram Research, Inc., Champaign, Illinois.