



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Vojtěch Kika

**Statistická inference v modelech
mnohorozměrných rozdělání
založených na kopulích**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Marek Omelka, PhD.

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická statistika
a ekonometrie

Praha 2017

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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

Název práce: Statistická inference v modelech mnohorozměrných rozdělení založených na kopulích

Autor: Vojtěch Kika

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Marek Omelka, PhD., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato diplomová práce se zaměřuje na statistickou inferenci v modelech založených na kopulích. Popsány jsou základní pojmy teorie kopulí, následně jsou prezentovány metody pro statistickou inferenci. Ty jsou rozlišeny do tří hlavních skupin. První z nich jsou parametrické metody odhadu parametru kopule, které předpokládají plně parametrickou strukturu modelu, tedy jak pro sdružené, tak pro marginální rozdělení. Druhou skupinou jsou semiparametrické metody odhadu parametru kopule, které oproti parametrickým metodám nekladou parametrické předpoklady na marginální rozdělení. Poslední skupinou jsou testy dobré shody určené k testování hypotéz, zda zkoumaná kopule náleží do některé dané rodiny kopulí. Práce je v závěru doplněna o simulační studii, která zkoumá závislost pozorovaného pokrytí asymptotického intervalu spolehlivosti pro parametr kopule na rozsahu výběru. Pro tuto studii byla vybrána metoda založená na pseudověrohodnosti, tedy jedna z nejpoužívanějších semiparametrických metod odhadu. Ukazuje se, že pro většinu zkoumaných rodin kopulí je pozorované pokrytí velmi blízké teoretickému pro výběry rozsahu 50 a více. Pro Frankovu a Gumbel-Hougaardovu rodinu se ukázaly dostatečné již výběry o rozsahu 30 pozorování.

Klíčová slova: kopule, inference, odhad, testy dobré shody, simulační studie

Title: Statistical inference in multivariate distributions based on copula models

Author: Vojtěch Kika

Department: Department of Probability and Mathematical Statistics

Supervisor: Ing. Marek Omelka, PhD., Department of Probability and Mathematical Statistics

Abstract: This diploma thesis aims for statistical inference in copula based models. Basics of copula theory are described, followed by methods for statistical inference. These are divided into three main groups. First of them are parametric methods for copula parameter estimation which assume fully parametric structure, thus for both joint and marginal distributions. The second group consists of semiparametric methods for copula parameter estimation which, unlike parametric methods, do not require parametric structure for marginal distributions. The last group describes goodness-of-fit tests used for testing the hypothesis that considered copula belongs to some specific copula family. The thesis is accompanied by a simulation study that investigates the dependence of the observed coverage of the asymptotic confidence intervals for copula parameter on the sample size. Pseudolikelihood method was chosen for the simulation study since it is one of the most popular semiparametric methods. It is shown that sample size of 50 seems to be sufficient for the observed coverage to be close to the theoretical one. For Frank and Gumbel-Hougaard copula families even sample size of 30 gives us satisfying observed coverage.

Keywords: copula, inference, estimate, goodness of fit testing, simulation study

Rád bych zde poděkoval vedoucímu této práce, panu Ing. Marku Omelkovi, PhD., za jeho cenné rady a podnětné připomínky. Stejně tak bych rád poděkoval svým blízkým za podporu a trpělivost během tvorby této práce i během celého předešlého studia.

Obsah

Úvod	2
1 Kopule	3
1.1 Definice kopule	3
1.2 Základní vlastnosti kopulí	4
1.3 Kopule, distribuční funkce a náhodné veličiny	6
1.4 Grafické znázornění kopulí	11
1.5 Příklady kopulí	13
1.6 Zobecnění kopulí do více dimenzí	16
2 Statistická inference	18
2.1 Parametrické metody odhadu	18
2.1.1 Asymptotické rozdělení odhadu	19
2.2 Semiparametrické metody odhadu	21
2.2.1 Semiparametrické Z-odhady	21
2.2.2 Metoda nejmenší vzdálenosti	29
2.2.3 Momentová metoda	31
2.2.4 Porovnání semiparametrických metod	34
2.3 Testy dobré shody	35
2.3.1 Testy založené na empirické kopuli	35
2.3.2 Testy založené na Kendallově transformaci	36
2.3.3 Testy založené na Rosenblattově transformaci	37
3 Simulační studie	39
Závěr	43
Seznam použité literatury	45

Úvod

Problematika kopulí je aktuálně dynamicky se rozvíjícím oborem studia v oblasti mnohorozměrné statistiky. Kopule nabízí velice praktický způsob, jakým modelovat závislostní strukturu mezi dvěma a více veličinami. Zjednodušeně řečeno jsou kopule funkce, které spojují jednorozměrné marginální distribuční funkce náhodných veličin do jejich sdružené distribuční funkce. Na kopule se však dá dívat také přímo jako na sdruženou distribuční funkci takovou, že marginální rozdělení jsou rovnoměrná na intervalu $(0,1)$.

Podle známých informací použil poprvé pojem kopule americký matematik Abe Sklar v roce 1959 (viz Sklar, 1959). K dílčím výsledkům, které se pojí k teorii kopulí, však došli někteří autoři, například Hoeffding (1940), již dříve při studiu dvourozměrných rozdělení s pevně danými marginálními rozděleními. K rozšíření zájmu o kopule došlo v 90. letech 20. století vzhledem k možnostem jejich praktického uplatnění. Velké oblibě se kopule těší ve finančním sektoru, neboť umožňují odděleně modelovat jednorozměrná rozdělení (například chování tržních subjektů) a závislostní strukturu (to, jakým způsobem se subjekty na trhu ovlivňují). Dalším možným využitím kopulí v této oblasti může být modelování rozdělení ztrát ve více portfoliích půjček, dluhopisů nebo akcií. Také mimo finanční sektor nabízí autoři mnoho aplikací. Wang, Chang a Yeh (2009) používají kopule pro studium sdruženého rozdělení průtoku vody na soutoku dvou řek, čímž nabízí argumenty pro vhodné použití protipovodňových opatření. Onken a kol. (2009) používají kopule pro studium vzájemné závislosti mezi neurony, jakmile jsou vystaveny určitému druhu stimulace. Tyto i mnohé další práce ukazují užitečnost kopulí a motivují k jejich hlubšímu studiu.

Tato práce si klade za cíl prezentovat čtenářům základy teorie kopulí a následně přehledně a jednotně popsat existující metodologii vztahující se ke statistické inferenci v modelech založených na kopulích. Zabýváme se zde metodami odhadu parametrů kopulí a také testy dobré shody. Práce je doplněna simulačními studii zkoumajícími praktické vlastnosti vybrané metody odhadu. Celý text je členěn do kapitol a sekcí na základě používaných metod a tématických celků tak, aby práce mohla sloužit jak čtenáři se zájmem o komplexní pohled na statistickou inferenci v modelech založených na kopulích, tak čtenáři se zájmem o konkrétní statistickou metodu.

Vzhledem k absenci česky psané literatury na téma kopulí je tato diplomová práce psána v českém jazyce a obsahuje také základní pojmy teorie kopulí tak, aby mohla v případě potřeby sloužit zájemcům, kteří se doposud s kopulemi neseťkali. Například by tak mohla sloužit jako doplňkový studijní text pro studenty předmětu Mnohorozměrná analýza na Matematicko-fyzikální fakultě Univerzity Karlovy.

1. Kopule

1.1 Definice kopule

V úvodu tohoto textu budeme definovat pojem objektu kopule. Nejprve budeme teorii kopulí budovat ve dvourozměrném prostoru, později ukážeme, jakým způsobem lze celou tuto problematiku převést ze dvou do více dimenzí. Tato sekce je založena na knize Nelsen (2006). Nejprve potřebujeme definovat dvourozměrnou analogii neklesající funkce.

Definice 1. *Nechť $\overline{\mathbb{R}}$ značí rozšířenou reálnou osu, tedy interval $[-\infty, \infty]$. Nechť H je funkce, jejíž definiční obor tvoří podmnožinu $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$ a obor hodnot podmnožinu $\overline{\mathbb{R}}$. Potom H -objemem obdélníku $B = [x_1, x_2] \times [y_1, y_2]$ rozumíme hodnotu*

$$V_H(B) = H(x_1, y_1) + H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1),$$

kde vrcholy obdélníka, tedy body (x_1, y_1) , (x_2, y_2) , (x_1, y_2) a (x_2, y_1) leží v definičním oboru funkce H .

Definice 2. *Nechť H je funkce jako v definici 1. Řekneme, že H je 2-rostoucí, pokud pro každý obdélník B s vrcholy v definičním oboru H platí, že*

$$V_H(B) \geq 0.$$

Pro zjednodušení zápisu budeme dále také definovat speciální typ obdélníku.

Definice 3. *Jednotkovým čtvercem budeme nazývat obdélník $[0, 1] \times [0, 1]$ a budeme jej značit \mathbf{I}^2 . Analogicky budeme n -rozměrný interval $[0, 1]^n$ značit \mathbf{I}^n .*

Podotkněme, že obecně neplatí, že funkce, která je rostoucí v obou svých argumentech, je také 2-rostoucí a obdobně 2-rostoucí funkce nemusí být rostoucí v obou svých argumentech. Dokladem jsou následující příklady.

Příklad 1. Definujme funkci $H : \mathbf{I}^2 \rightarrow \mathbb{R}$ předpisem $H(x, y) = \log_2(1 + x + y)$. Tato funkce je rostoucí v obou svých argumentech, neboť logaritmus je rostoucí funkce. Nicméně $V_H(\mathbf{I}^2) = \log_2(1) + \log_2(3) - \log_2(2) - \log_2(2) = \log_2(3) - 2 < 0$, a tedy funkce H není 2-rostoucí.

Příklad 2. Definujme funkci $H : \mathbf{I}^2 \rightarrow \mathbb{R}$ předpisem $H(x, y) = (2x - 1)(2y - 1)$. Tato funkce je pro $y \in [0, 1/2)$ klesající v x . Nicméně pro libovolný obdélník v \mathbf{I}^2 ve tvaru $B = [x_1, x_2] \times [y_1, y_2]$ je jeho H -objem dán jako

$$\begin{aligned} V_H(B) &= (2x_1 - 1)(2y_1 - 1) + (2x_2 - 1)(2y_2 - 1) - (2x_1 - 1)(2y_2 - 1) \\ &\quad - (2x_2 - 1)(2y_1 - 1) = 4(x_2 - x_1)(y_2 - y_1) \geq 0, \end{aligned}$$

a tedy funkce H je 2-rostoucí.

Jedním ze způsobů, jakými můžeme popisovat chování jednotlivých složek dvourozměrných funkcí, jsou takzvané marginály.

Definice 4. Necht S_1 a S_2 jsou neprázdné podmnožiny $\overline{\mathbb{R}}$. Dále necht b_1 je největším prvkem množiny S_1 a b_2 je největším prvkem množiny S_2 . Poté řekneme, že funkce H definovaná na $S_1 \times S_2$ má marginály, které jsou dány jako funkce F a G tak, že S_1 a S_2 jsou po řadě jejich definičními obory a

$$\forall x \in S_1 : F(x) = H(x, b_2), \quad \forall y \in S_2 : G(y) = H(b_1, y).$$

Nyní můžeme přikročit k samotné definici dvourozměrné kopule.

Definice 5. (Dvourozměrnou) kopulí nazýváme funkci C splňující následující podmínky:

1. Definičním oborem funkce C je jednotkový čtverec \mathbf{I}^2 , oborem hodnot pak interval \mathbf{I} .
2. Funkce C je 2-rostoucí.
3. Pro každé $u, v \in \mathbf{I}$ platí

$$\begin{aligned} C(u, 0) = 0 = C(0, v), \\ C(u, 1) = u \quad \text{a} \quad C(1, v) = v. \end{aligned}$$

Můžeme si všimnout, že funkční hodnotu kopule C lze za přispění definice 1 snadno vyjádřit pomocí C -objemu jako $C(u, v) = V_C([0, u] \times [0, v])$.

Poznámka. Do zavedení m -rozměrné kopule v sekci 1.6 budeme slovem kopule rozumět dvourozměrnou kopuli.

1.2 Základní vlastnosti kopulí

V této sekci budou uvedeny některé ze základních vlastností kopulí. Následující lemma hovoří o monotonii kopulí ve svých argumentech.

Lemma 1. Necht C je kopule. Potom C je neklesající v obou svých argumentech.

Důkaz. Plyne z lemmatu 2.1.4 v knize Nelsen (2006). □

Povšimněme si, že zatímco u obecných 2-rostoucích funkcí tato monotonie v argumentech zaručena není, u kopulí je této vlastnosti již díky podmínce 3 z definice kopule dosaženo.

Následující věta pojednává o omezeních, které jsou dány pro funkční hodnoty kopule.

Věta 2 (Fréchet-Hoeffdingovy meze). Necht C je kopule. Potom pro každou dvojici (u, v) z \mathbf{I}^2 platí

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v). \quad (1.1)$$

Dále platí, že funkce $W(u, v) = \max(u + v - 1, 0)$ a $M(u, v) = \min(u, v)$ definované na \mathbf{I}^2 jsou kopule.

Důkaz. Důkaz nerovnosti v (1.1) uvádí Nelsen (2006, Theorem 2.2.3). Ukažme nyní, že spodní i horní mez nerovnosti jsou kopule. Podmínky 1 a 3 z definice kopule jsou zřejmě u obou mezí splněny, neboť definiční obor i obor hodnot jsou specifikovány a pro $(u, v) \in \mathbf{I}^2$ platí

$$\begin{aligned} W(u, 0) &= \max(u - 1, 0) = 0 = \max(v - 1, 0) = W(0, v), \\ M(u, 0) &= \min(u, 0) = 0 = \min(v, 0) = M(0, v) \end{aligned}$$

a také

$$\begin{aligned} W(u, 1) &= \max(u, 0) = u, & W(1, v) &= \max(0, v) = v, \\ M(u, 1) &= \min(u, 1) = u, & M(1, v) &= \min(0, v) = v. \end{aligned}$$

Zbývá ověřit, že jak W , tak M jsou 2-rostoucí funkce. Musí tedy platit, že pro každé pro $[u_1, u_2] \times [v_1, v_2] \subseteq \mathbf{I}^2$

$$\begin{aligned} V_W &= W(u_1, v_1) + W(u_2, v_2) - W(u_1, v_2) - W(u_2, v_1) \geq 0, \\ V_M &= M(u_1, v_1) + M(u_2, v_2) - M(u_1, v_2) - M(u_2, v_1) \geq 0. \end{aligned}$$

Zde je nutno přistoupit k rozboru možností, důkaz bude proto proveden pouze pro M . Pro W by následující rozbor probíhal analogicky.

1. Nechť $u_1 \leq u_2 \leq v_1 \leq v_2$. Potom $V_M = u_1 + u_2 - u_1 - u_2 = 0$.
2. Nechť $u_1 \leq v_1 \leq u_2 \leq v_2$. Potom $V_M = u_1 + u_2 - u_1 - v_1 = u_2 - v_1 \geq 0$.
3. Nechť $u_1 \leq v_1 \leq v_2 \leq u_2$. Potom $V_M = u_1 + v_2 - u_1 - v_1 = v_2 - v_1 \geq 0$.

Zbylé případy již plynou ze symetrie funkce M .

□

Věta 2 spolu se svým důkazem uvádí možnosti, jak v některých případech rozhodovat, zda je vybraná funkce kopule. Nesplňuje-li zadaná funkce nerovnost (1.1), pak nemůže být kopulí. Důkaz, že je zadaná funkce kopule, pak může být veden obdobně jako výše ověřením všech tří bodů definice.

Další základní vlastností kopulí je jejich spojitost.

Věta 3 (Spojitosť kopulí). *Nechť C je kopule. Potom pro každou dvojici bodů $(u_1, u_2), (v_1, v_2) \in \mathbf{I}^2$ platí*

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

C je tedy lipschitzovská, a tudíž také stejnoměrně spojitá na \mathbf{I}^2 .

Důkaz. Důkaz uvádí Nelsen (2006, Theorem 2.2.2) a plyne z trojúhelníkové nerovnosti a faktu, že C je 2-rostoucí.

□

1.3 Kopule, distribuční funkce a náhodné veličiny

V dosavadním průběhu tohoto textu jsme o kopulích mluvili jako o obecných funkcích. Dále bychom však chtěli využít kopule jakožto nástroj pro modelování vztahu vícerozměrného rozdělení a příslušných marginálních rozdělení. Zavedme tedy pojmy distribuční funkce a sdružené distribuční funkce.

Definice 6. Funkci F s definičním oborem $\overline{\mathbb{R}}$ nazveme *distribuční funkcí*, pokud

1. F je neklesající,
2. $F(-\infty) = 0$ a $F(\infty) = 1$.

V definici 6 nejsou kladeny žádné další podmínky. V tomto kontextu tedy nepožadujeme ani jednostrannou spojitost takto definované funkce. Obdobným způsobem můžeme definovat také sdruženou distribuční funkci.

Definice 7. Funkci H s definičním oborem $\overline{\mathbb{R}}^2$ nazveme *dvourozměrnou sdruženou distribuční funkcí*, pokud

1. H je 2-rostoucí,
2. $H(x, -\infty) = H(-\infty, y) = 0$ a $F(\infty, \infty) = 1$.

Jelikož největším prvkem $\overline{\mathbb{R}}$ je ∞ , vidíme, že funkce H má marginály dané jako $F(x) = H(x, \infty)$ a $G(y) = H(\infty, y)$. Následující lemma dokládá, že marginály dvourozměrné sdružené distribuční funkce jsou samy také distribuční funkce.

Lemma 4. Necht H je dvourozměrná sdružená distribuční funkce. Potom její marginály jsou distribuční funkce.

Důkaz. Necht $F(x) = H(x, \infty)$ je marginál funkce H . Ověříme pro F obě podmínky z definice distribuční funkce.

1. F je neklesající, pokud $\forall x_1, x_2 \in \overline{\mathbb{R}} : x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$. Vezměme tedy $x_1 \leq x_2$. Potom

$$\begin{aligned} F(x_2) - F(x_1) &= H(x_2, \infty) - H(x_1, \infty) \\ &= H(x_2, \infty) - H(x_1, \infty) + H(x_1, -\infty) - H(x_2, -\infty) \\ &= V_H([x_1, x_2] \times [-\infty, \infty]) \geq 0, \end{aligned}$$

neboť $H(x_1, -\infty) = H(x_2, -\infty) = 0$ a H je 2-rostoucí.

2. Z druhého bodu definice sdružené distribuční funkce snadno získáváme $F(-\infty) = H(-\infty, \infty) = 0$ a také $F(\infty) = H(\infty, \infty) = 1$.

Ověřením těchto bodů jsme dokázali, že F je distribuční funkce. □

Americký matematik Abe Sklar již v roce 1959 vyslovil větu, která popisuje, jakým způsobem kopule spojují dvourozměrnou sdruženou distribuční funkci a její marginály.

Věta 5 (Sklarova). *Nechť H je dvourozměrná sdružená distribuční funkce s marginály F a G . Potom existuje kopule C taková, že pro každé $x, y \in \overline{\mathbb{R}}$ platí*

$$H(x, y) = C(F(x), G(y)). \quad (1.2)$$

Jsou-li F a G spojité, pak je kopule C určena jednoznačně. V opačném případě je C určena jednoznačně na množině $\mathcal{H}(F) \times \mathcal{H}(G)$, kde $\mathcal{H}(F)$ značí obor hodnot funkce F a $\mathcal{H}(G)$ značí obor hodnot funkce G .

Dále, pokud C je kopule, F a G jsou distribuční funkce a platí (1.2), potom je funkce H dvourozměrná sdružená distribuční funkce s marginály F a G .

Důkaz. Důkaz uvádí Nelsen (2006, Theorem 2.3.3). □

Právě Sklarovi bývá často připisováno první použití označení „kopule“. Tento název má symbolizovat pouto, pár nebo spojení, které kopule vytváří mezi marginály.

Dosavadní část této sekce sice popisuje způsob, jakým kopule propojuje marginály distribučních funkcí, avšak distribuční funkce je libovolná funkce splňující zadané vlastnosti. Pro naše aplikace kopulí však bude užitečné chápat distribuční funkce jakožto funkce, které popisují náhodné veličiny. Pro reálnou náhodnou veličinu X na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathbb{P})$ definujeme její distribuční funkci jako $F(x) = P(X \leq x)$, kde $x \in \overline{\mathbb{R}}$. Analogicky definujeme dvourozměrnou sdruženou distribuční funkci reálných náhodných veličin X a Y definovaných na témže pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathbb{P})$ jako funkci $H(x, y) = P(X \leq x, Y \leq y)$, kde $x, y \in \overline{\mathbb{R}}$.

Poznámka. Kdykoliv budeme v následujícím textu hovořit o náhodných veličinách, budeme vždy předpokládat, že jsou to reálné náhodné veličiny definované na stejném pravděpodobnostním prostoru.

Chápeme-li nyní distribuční funkci jako distribuční funkci náhodné veličiny, můžeme Sklarovu větu přepsat v následujícím tvaru.

Věta 6 (Sklarova pro náhodné veličiny). *Nechť X a Y jsou náhodné veličiny s distribučními funkcemi F a G a dvourozměrnou sdruženou distribuční funkcí H . Potom existuje kopule C taková, že pro každé $x, y \in \overline{\mathbb{R}}$ je*

$$H(x, y) = C(F(x), G(y)).$$

Jsou-li F a G spojité, pak je kopule C určena jednoznačně. V opačném případě je C určena jednoznačně na množině $\mathcal{H}(F) \times \mathcal{H}(G)$.

Kopuli C danou náhodnými veličinami X a Y a větou 6 budeme označovat $C_{X,Y}$, abychom dali najevo její vztah k těmto náhodným veličinám.

Přirozenou otázkou, která se váže k aplikaci teorie kopulí na náhodné veličiny a jejich distribuční funkce, je ta, jakým způsobem ovlivňují transformace náhodných veličin tvar příslušné kopule. O problematice ryze monotónních transformací pojednává následující věta.

Věta 7 (O kopulích ryze monotónních transformací náhodných veličin). *Nechť X a Y jsou náhodné veličiny, jejichž distribuční funkce jsou spojité, a funkce $\alpha : \mathcal{H}(X) \rightarrow \overline{\mathbb{R}}$ a $\beta : \mathcal{H}(Y) \rightarrow \overline{\mathbb{R}}$ jsou spojité. Potom pro $u, v \in \mathbf{I}$ platí následující čtveřice tvrzení.*

1. *Jsou-li α a β rostoucí, pak*

$$C_{\alpha(X),\beta(Y)}(u, v) = C_{X,Y}(u, v).$$

2. *Je-li α rostoucí a β klesající, pak*

$$C_{\alpha(X),\beta(Y)}(u, v) = u - C_{X,Y}(u, 1 - v).$$

3. *Je-li α klesající a β rostoucí, pak*

$$C_{\alpha(X),\beta(Y)}(u, v) = v - C_{X,Y}(1 - u, v).$$

4. *Jsou-li α a β klesající, pak*

$$C_{\alpha(X),\beta(Y)}(u, v) = u + v - 1 + C_{X,Y}(1 - u, 1 - v).$$

Důkaz. Důkaz uvedeme pro bod 2. Pro zbývající případy je důkaz analogický. Nechť F_1, F_2, G_1, G_2 jsou po řadě distribuční funkce náhodných veličin $X, \alpha(X), Y, \beta(Y)$. Jelikož je α rostoucí funkce, platí pro $x, y \in \overline{\mathbb{R}}$

$$F_2(x) = \mathbf{P}(\alpha(X) \leq x) = \mathbf{P}(X \leq \alpha^{-1}(x)) = F_1(\alpha^{-1}(x))$$

a podobně, neboť β je klesající, platí

$$G_2(y) = \mathbf{P}(\beta(Y) \leq y) = \mathbf{P}(Y \geq \beta^{-1}(y)) = 1 - G_1(\beta^{-1}(y)),$$

kde poslední rovnost vyžaduje spojitost funkce G_1 . Dohromady proto

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

kde poslední rovnost vyžaduje spojitost funkce G_2 . Ta plyne ze spojitosti G_1 a faktu, že náhodná veličina, která vznikla spojitou, ryze monotónní transformací náhodné veličiny se spojitou distribuční funkcí, má spojitou distribuční funkci. Ze spojitosti F_2 a G_2 také plyne, že obor hodnot obou těchto funkcí je celý interval \mathbf{I} . Proto pro každé $u, v \in \mathbf{I}$ platí $C_{\alpha(X),\beta(Y)}(u, v) = u - C_{X,Y}(u, 1 - v)$. □

Připomeňme nyní Fréchet-Hoeffdingovy meze z věty 2. I tyto meze můžeme nyní využít v případě kopulí pro náhodné veličiny. Pro každou kopuli C a $u, v \in \mathbf{I}$ platí

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

Je-li H dvourozměrná sdružená distribuční funkce s marginály F a G , potom pro $x, y \in \overline{\mathbb{R}}$ platí

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y)). \quad (1.3)$$

Je otázkou, v jakém vztahu jsou náhodné veličiny X a Y s distribučními funkcemi F a G a se sdruženou distribuční funkcí H , pokud H nabývá jedné z Fréchet-Hoeffdingových mezí ve smyslu nerovnosti (1.3). Odpověď na tuto otázku shrnuje následující věta.

Věta 8. *Nechť X a Y jsou náhodné veličiny s distribučními funkcemi F a G a sdruženou distribuční funkcí H . Potom*

1. *Pro každé $x, y \in \overline{\mathbb{R}}$ platí*

$$H(x, y) = \max(F(x) + G(y) - 1, 0)$$

právě tehdy, když pro každé (u_1, v_1) a (u_2, v_2) z nosiče funkce H platí

$$u_1 < u_2 \Rightarrow v_1 \geq v_2.$$

2. *Pro každé $x, y \in \overline{\mathbb{R}}$ platí*

$$H(x, y) = \min(F(x), G(y))$$

právě tehdy, když pro každé (u_1, v_1) a (u_2, v_2) z nosiče funkce H platí

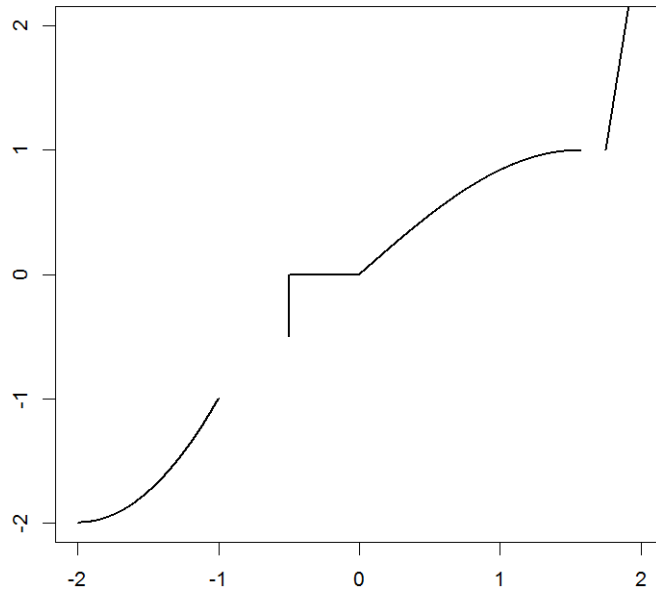
$$u_1 < u_2 \Rightarrow v_1 \leq v_2.$$

Důkaz. Viz Nelsen (2006, Theorems 2.5.4, 2.5.5.).

□

Obrázek 1.1 uvádí příklad nosiče splňujícího ekvivalenci z bodu 1 věty 8.

Formální matematický zápis zmíněné věty lze velmi snadno interpretovat. Nejnázne pak půjde tento vztah popsat na dvojici náhodných veličin X a Y s rovnoměrným rozdělením na intervalu $(0, 1)$. Pokud je sdružená distribuční funkce těchto náhodných veličin dána dolní Fréchet-Hoeffdingovou mezí ve smyslu (1.3), tedy funkcí $\max(F(x) + G(y) - 1, 0)$, potom platí, že $\mathbb{P}(X + Y = 1) = 1$. Takové náhodné veličiny nazýváme kontramonotónní. Tato situace obecně nastává právě tehdy, když je náhodná veličina Y klesající transformací náhodné veličiny X skoro jistě. V opačném případě, pokud je sdružená distribuční funkce těchto náhodných veličin dána horní Fréchet-Hoeffdingovou mezí ve smyslu (1.3), tedy funkcí $\min(F(x), G(y))$, platí, že $\mathbb{P}(X = Y) = 1$. Takové náhodné veličiny nazýváme komonotónní. Tato situace obecně nastává právě tehdy, když je náhodná veličina Y rostoucí transformací náhodné veličiny X skoro jistě.



Obrázek 1.1: Příklad nosiče sdružené distribuční funkce H , která je rovna dolní Fréchet-Hoeffdingově mezi.

Kopule přežití

V mnoha oblastech aplikací statistiky, ať už se jedná například o lékařské studie nebo finanční sektor, je předmětem zájmu spíše funkce přežití nežli distribuční funkce. Funkci přežití definujeme jako doplněk distribuční funkce, tedy jako $S(x) = 1 - F(x)$, kde F je distribuční funkce. Vidíme, že je-li S_X funkcí přežití náhodné veličiny X , pak platí $S_X(x) = \mathbf{P}(X > x)$. Funkce přežití proto nejčastěji popisuje, jaká je pravděpodobnost, že jedinec (pacient, klient) přežije zadanou dobu. Je-li také Y náhodná veličina s distribuční funkcí G a funkcí přežití S_Y , můžeme analogicky jako pro distribuční funkce definovat sdruženou funkci přežití jako $H_S(x, y) = \mathbf{P}(X > x, Y > y)$. Poté můžeme psát

$$\begin{aligned} H_S(x, y) &= 1 - F(x) - G(y) + H(x, y) = S_X(x) + S_Y(y) - 1 + C(F(x), G(y)) \\ &= S_X(x) + S_Y(y) - 1 + C(1 - S_X(x), 1 - S_Y(y)). \end{aligned}$$

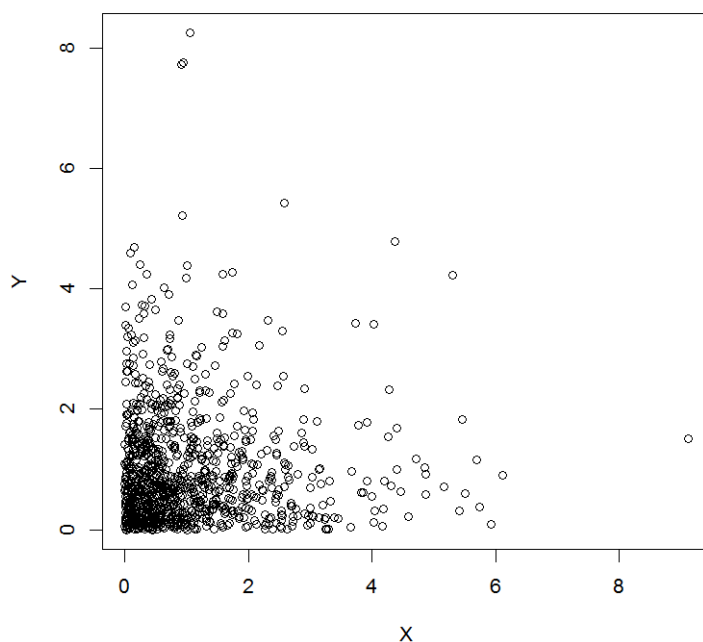
Pokud tedy definujeme funkci C_S předpisem

$$C_S(u, v) = u + v - 1 + C(1 - u, 1 - v),$$

kde $u, v \in \mathbf{I}$, obdržíme vztah mezi sdruženou funkcí přežití a funkcí C_S ve tvaru

$$H_S(x, y) = C_S(S_X(x), S_Y(y)).$$

Ze čtvrtého bodu věty 7 plyne, že pro spojitě náhodné veličiny X a Y je C_S také kopule. V obecném případě by bylo možné ukázat, že je C_S kopule, také ověřením podmínek z definice kopule. Taková úloha je snadným cvičením obdobným důkazu věty 2, proto jej zde vynecháváme.



Obrázek 1.2: Bodový graf pro nezávislé páry pozorování, jehož obě složky jsou vzájemně nezávislé náhodné veličiny z exponenciálního rozdělení $Exp(1)$.

1.4 Grafické znázornění kopulí

Pro praktickou práci s kopulemi je nutné představit nástroj, pomocí kterého budeme snadno schopni kopule vizualizovat a také graficky hodnotit nejen vhodnost použití konkrétní kopule, případně rodiny kopulí, ale také míru závislosti mezi sledovanými veličinami. Tuto problematiku uvedeme příkladem.

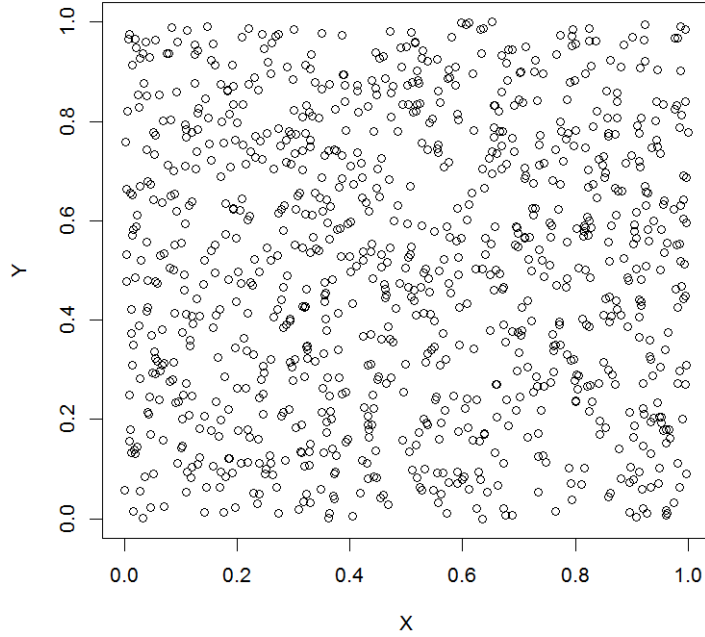
Příklad 3. Uvažujme dva vzájemně nezávislé náhodné výběry z exponenciálního rozdělení $X_1, \dots, X_{1000} \sim Exp(1)$ a $Y_1, \dots, Y_{1000} \sim Exp(1)$. Připomeňme, že distribuční funkcí tohoto rozdělení je funkce $F(x) = 1 - e^{-x}$ pro $x > 0$ a $F(x) = 0$ jinak. Vykresleme bodový graf pro náhodně vygenerované páry pozorování $(X_1, Y_1), \dots, (X_{1000}, Y_{1000})$.

Na obrázku 1.2 vidíme množinu bodů seskupených převážně poblíž počátku. Předmětem zájmu v tento moment je popsání závislosti mezi oběma veličinami. K tomuto účelu je však zřejmě takovéto vyobrazení nevhodné. Aplikujme tedy na každé pozorování funkci F jakožto distribuční funkci obou náhodných výběrů a vykresleme bodový graf pro transformované páry

$$(F(X_1), F(Y_1)), \dots, (F(X_{1000}), F(Y_{1000})).$$

Mezi oběma náhodnými výběry z obrázku 1.3 není patrná žádná závislost.

Příklad 3 dává návod na to, jakým způsobem postupovat při grafickém znázornění kopulí. Samotné znázornění párů získaných hodnot (realizací náhodných veličin X_1, \dots, X_n a Y_1, \dots, Y_n) vypovídá pouze o jejich sdruženém rozdělení, nikoliv o vzájemné závislosti. Známe-li tedy příslušné distribuční funkce F a G obou



Obrázek 1.3: Bodový graf pro nezávislé páry pozorování, jehož obě složky jsou transformace (pomocí distribuční funkce rozdělení $Exp(1)$) vzájemně nezávislých náhodných veličin z exponenciálního rozdělení $Exp(1)$.

náhodných výběrů, namísto dvojic (X_i, Y_i) pro $i \in \{1, \dots, n\}$ vykreslujeme dvojice $(F(X_i), G(Y_i))$. V praktických aplikacích však obvykle neznáme distribuční funkce F a G . Intuitivním řešením je v takové situaci nahrazení distribučních funkcí F a G jejich výběrovými protějšky

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(X_i \leq x), \quad G_n(y) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(Y_i \leq y), \quad (1.4)$$

kde $\mathbb{1}$ značí indikátorovou (charakteristickou) funkci. Povšimněme si, že oproti obvyklé definici empirické distribuční funkce je zde použito její přeskálování výrazem $n/(n+1)$. Cílem této úpravy je zamezení používání hodnot, které jsou na hranici definičního oboru kopule a mohly by způsobovat problémy při používání teorie maximální věrohodnosti (dále v sekci 2.2). Grafickým znázorněním kopulí je pak vykreslení párů $(F_n(X_i), G_n(Y_i))$. Složky těchto párů můžeme také přepsat jako

$$F_n(X_i) = \frac{R_i}{n+1}, \quad G_n(Y_i) = \frac{S_i}{n+1}, \quad i = 1, \dots, n,$$

kde R_i značí pořadí X_i ve výběru X_1, \dots, X_n a obdobně S_i značí pořadí Y_i ve výběru Y_1, \dots, Y_n . Odtud získáváme také jedno z možných vyjádření pro empirickou kopuli.

Definice 8. *Nechť X_1, \dots, X_n je náhodný výběr z rozdělení s distribuční funkcí F a Y_1, \dots, Y_n je náhodný výběr z rozdělení s distribuční funkcí G . Dále necht*

R_i značí pořadí X_i ve výběru X_1, \dots, X_n a obdobně S_i značí pořadí Y_i ve výběru Y_1, \dots, Y_n . Potom empirickou kopulí na základě těchto náhodných výběrů rozumíme funkci C_n definovanou na \mathbf{I}^2 předpisem

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right).$$

1.5 Příklady kopulí

V této sekci si uvedeme příklady kopulí, případně jejich rodin. Pro všechny argumenty (ne však parametry) kopulí v této sekci budeme automaticky předpokládat, že leží v intervalu \mathbf{I} .

Součinnová kopule (kopule nezávislosti)

Velmi jednoduchým, avšak důležitým příkladem kopule je součinnová kopule Π , která je na \mathbf{I}^2 definována předpisem

$$\Pi(u, v) = uv.$$

Tato kopule je často označována také jako kopule nezávislosti, neboť charakterizuje nezávislost dvojice spojitých náhodných veličin. Tento fakt můžeme shrnout pomocí následující věty.

Věta 9. *Nechť X a Y jsou absolutně spojitě náhodné veličiny. Potom jsou X a Y nezávislé právě tehdy, když $C_{X,Y} = \Pi$.*

Důkaz. Plyne snadno z věty 6 a faktu, že náhodné veličiny jsou nezávislé právě tehdy, když je jejich sdružená distribuční funkce rovna součinu marginálních distribučních funkcí.

□

Normální kopule

Nechť $\theta \in [-1, 1]$. Dále necht Φ značí distribuční funkci normovaného normálního rozdělení a $\Phi_2(\cdot, \cdot; \theta)$ značí distribuční funkci dvourozměrného normálního rozdělení s nulovou střední hodnotou a korelací mezi dvěma složkami θ . Potom je funkce definovaná předpisem

$$C(u, v; \theta) = \Phi_2(\phi^{-1}(u), \phi^{-1}(v); \theta)$$

kopule.

Fréchetova rodina kopulí

Nechť $\alpha, \beta \in \mathbf{I}$ jsou takové, že $\alpha + \beta \leq 1$. Potom je funkce definovaná předpisem

$$C(u, v; \alpha, \beta) = \alpha M(u, v) + (1 - \alpha - \beta) \Pi(u, v) + \beta W(u, v)$$

kopule. Připomeňme, že funkce M a W značí Fréchet-Hoeffdingovy meze a Π součinnovou kopuli. Parametrická rodina takovýchto kopulí se nazývá Fréchetova.

Mardiova rodina kopulí

Nechť $\theta \in [-1, 1]$ je parametr. Potom je funkce definovaná předpisem

$$C(u, v; \theta) = \frac{\theta^2(1 + \theta)}{2}M(u, v) + (1 - \theta^2)\Pi(u, v) + \frac{\theta^2(1 - \theta)}{2}W(u, v)$$

kopule. Funkce M a W opět značí Fréchet-Hoeffdingovy meze a Π součinnovou kopuli. Parametrická rodina takovýchto kopulí se nazývá Mardiova.

Cuadras-Augého rodina kopulí

Nechť $\theta \in \mathbf{I}$ je parametr. Potom je funkce definovaná předpisem

$$C(u, v; \theta) = (\min(u, v))^{\theta}(uv)^{1-\theta} = \begin{cases} uv^{1-\theta}, & u \leq v, \\ u^{1-\theta}v, & u \geq v \end{cases}$$

kopule. Parametrická rodina takovýchto kopulí se nazývá Cuadras-Augého. Můžeme si povšimnout, že volbou $\theta = 0$ získáme součinnovou kopuli Π a volbou $\theta = 1$ horní Fréchet-Hoeffdingovu mezi M .

Farlie-Gumbel-Morgensternova rodina kopulí

Nechť $\theta \in [-1, 1]$ je parametr. Potom je funkce definovaná předpisem

$$C(u, v; \theta) = uv + \theta uv(1 - u)(1 - v)$$

kopule. Parametrická rodina takovýchto kopulí se nazývá Farlie-Gumbel-Morgensternova. Nelsen (2006, str. 78) uvádí, že snadný funkční tvar zástupců této rodiny kopulí stojí za jejich velkou oblibou v praktických aplikacích.

Archimédovské kopule

Velmi širokou třídu kopulí tvoří takzvané Archimédovské kopule. Ty svou funkční formou pokrývají velké spektrum rodin kopulí. Pro jejich korektní zavedení budeme potřebovat následující definici.

Definice 9. *Nechť ϕ je spojitá, klesající funkce z \mathbf{I} do intervalu $[0, \infty]$ taková, že $\phi(1) = 0$. Potom pseudoinverzní funkcí (pseudoinverzí) k ϕ rozumíme funkci $\phi^{[-1]}$ z intervalu $[0, \infty]$ do \mathbf{I} definovanou předpisem*

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

Můžeme si všimnout, že $\phi^{[-1]}$ je na intervalu $[0, \infty]$ spojitá a nerostoucí, na intervalu $[0, \phi(0)]$ pak dokonce klesající. Dále také platí $\phi^{[-1]}(\phi(u)) = u$ na \mathbf{I} a

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

Pokud je tedy $\phi(0) = \infty$, pak $\phi^{[-1]} = \phi^{-1}$. Třída archimédovských kopulí je definována následovně.

Definice 10. *Nechť ϕ je funkce jako v definici 9 a $\phi^{[-1]}$ je její pseudoinverze. Dále předpokládejme, že ϕ je konvexní. Potom každou funkci C z \mathbf{I}^2 do \mathbf{I} danou předpisem*

$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \quad (1.5)$$

nazýváme Archimédovskou kopulí s generátorem ϕ .

Lze ukázat, že Archimédovská kopule definovaná v (1.5) je opravdu kopule. Odvození uvádí například (Nelsen, 2006, Theorem 4.1.4). Zde pouze podotkněme, že pro funkce ϕ , které nejsou konvexní, pak takto definovaná funkce C kopulí není.

Mezi příklady rodin kopulí, které náležejí do třídy Archimédovských kopulí, patří například

1. Claytonova rodina, která je pro $\theta \in [-1, \infty) \setminus \{0\}$ definována předpisem

$$C(u,v;\theta) = \left(\max(u^{-\theta} + v^{-\theta} - 1, 0) \right)^{-1/\theta}$$

a jejíž generátor je ve tvaru $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$. Claytonova rodina s parametrem $\theta = 0$ bývá často dodefinována jako součinná kopule Π , neboť

$$\lim_{\theta \rightarrow 0} \left(u^{-\theta} + v^{-\theta} - 1 \right)^{-1/\theta} = uv.$$

2. Ali-Mikhail-Haqova rodina, která je pro $\theta \in [-1, 1)$ definována předpisem

$$C(u,v;\theta) = \max \left(1 - \left((1-u)^\theta + (1-v)^\theta \right)^{1/\theta}, 0 \right)$$

a jejíž generátor je ve tvaru $\phi_\theta(t) = (1-t)^\theta$.

3. Gumbel-Hougaardova rodina, která je pro $\theta \in [1, \infty)$ definována předpisem

$$C(u,v;\theta) = \exp \left(\left((-\ln(u))^\theta + (-\ln(v))^\theta \right)^{1/\theta} \right)$$

a jejíž generátor je ve tvaru $\phi_\theta(t) = (-\ln(t))^\theta$.

4. Frankova rodina, která je pro $\theta \in (-\infty, \infty) \setminus \{0\}$ definována předpisem

$$C(u,v;\theta) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$$

a jejíž generátor je ve tvaru $\phi_\theta(t) = -\ln((e^{-\theta t} - 1)/(e^{-\theta} - 1))$. Také Frankova rodina s parametrem $\theta = 0$ může být dodefinována jako součinná kopule Π , neboť

$$\lim_{\theta \rightarrow 0} -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) = uv.$$

1.6 Zobecnění kopulí do více dimenzí

V této sekci ukážeme, jakým způsobem mohou být kopule rozšířeny do více-rozměrného prostoru. K tomu bude nejprve nutné zobecnit definice 1 a 2. Necht $\overline{\mathbb{R}}^m$ značí rozšířený m -rozměrný prostor reálných čísel, tedy m -rozměrný interval $[-\infty, \infty] \times \dots \times [-\infty, \infty]$.

Definice 11. *Necht H je funkce, jejíž definiční obor tvoří podmnožinu $\overline{\mathbb{R}}^m$ a obor hodnot podmnožinu \mathbb{R} . Potom H -objemem m -rozměrného intervalu*

$$B = [a_1, b_1] \times \dots \times [a_m, b_m]$$

rozumíme hodnotu

$$V_H(B) = \sum \text{sgn}(\mathbf{c})H(\mathbf{c}),$$

kde sčítáme přes všechny vrcholy \mathbf{c} m -rozměrného intervalu B . Tedy přes takové vektory $\mathbf{c} = (c_1, \dots, c_m)$, kde pro každé $i = 1, \dots, m$ je $c_i = a_i$ nebo $c_i = b_i$. Znaménko vrcholu \mathbf{c} je pak dáno jako

$$\text{sgn}(\mathbf{c}) = \begin{cases} -1, & \text{je-li } \sum_{k=1}^m \mathbb{1}(c_k = a_k) \text{ liché,} \\ 1, & \text{je-li } \sum_{k=1}^m \mathbb{1}(c_k = a_k) \text{ sudé.} \end{cases}$$

Zobecnění 2-rostoucích funkcí můžeme nyní definovat následovně.

Definice 12. *Necht H je funkce jako v definici 11. Řekneme, že H je m -rostoucí, pokud pro každý obdélník B s vrcholy v definičním oboru H platí, že*

$$V_H(B) \geq 0.$$

Také marginály můžeme definovat pro funkce více než dvou proměnných.

Definice 13. *Necht S_1, \dots, S_m jsou neprázdné podmnožiny $\overline{\mathbb{R}}$. Dále necht b_i je největším prvkem množiny S_i pro $i = 1, \dots, m$. Poté řekneme, že funkce H definovaná na $S_1 \times \dots \times S_m$ má jednorozměrné marginály, které jsou dány jako funkce $H_i, i = 1, \dots, m$ tak, že S_i je definičním oborem H_i a*

$$\forall x \in S_i : H_i(x) = H(b_1, \dots, b_{i-1}, x, b_{i+1}, \dots, b_m).$$

Analogicky jako v definici 13 bychom mohli definovat také vícerozměrné marginály. Například dvourozměrný marginál by pro $i = 1, \dots, m$ a $j = 1, \dots, m$ takové, že $i < j$, byl dán jako funkce dvou proměnných definovaná na $S_i \times S_j$ předpisem

$$\forall x \in S_i, \forall y \in S_j : H_{ij}(x, y) = H(b_1, \dots, b_{i-1}, x, b_{i+1}, \dots, b_{j-1}, y, b_{j+1}, \dots, b_m).$$

Definice 14. *Funkci C nazýváme m -rozměrnou kopulí, pokud splňuje následující podmínky:*

1. *Definičním oborem funkce C je \mathbf{I}^m , oborem hodnot pak interval \mathbf{I} .*
2. *Funkce C je m -rostoucí.*

3. Pro každé $\mathbf{u} \in \mathbf{I}^m$ platí $C(\mathbf{u}) = 0$, pokud alespoň pro jedno $i \in \{1, \dots, m\}$ je $u_i = 0$.

4. Funkce C má jednorozměrné marginály C_i pro $i = 1, \dots, m$, které splňují

$$C_i(u) = u \text{ pro každé } u \in \mathbf{I}.$$

Existují-li smíšené parciální derivace

$$c(u_1, \dots, u_m) = \frac{\partial^m C(u_1, \dots, u_m)}{\partial u_1 \dots \partial u_m}, \quad (1.6)$$

pak funkci c nazveme hustotou kopule.

Stejně jako v dvourozměrném případě jsou m -rozměrné kopule stejnoměrně spojitě a neklesající ve všech svých argumentech. Také Fréchet-Hoeffdingovy meze mají svou vícerozměrnou variantu. Pro kopuli C a každé $\mathbf{u} \in \mathbf{I}^m$ platí

$$W^m(\mathbf{u}) \leq C(\mathbf{u}) \leq M^m(\mathbf{u}),$$

kde

$$\begin{aligned} W^m(\mathbf{u}) &= \max(u_1 + \dots + u_m - m + 1, 0), \\ M^m(\mathbf{u}) &= \min(u_1, \dots, u_m). \end{aligned}$$

Zatímco horní mez M^m je opět kopule, pro $m \geq 3$ již W^m kopule není, jak uvádí Nelsen (2006, str. 47).

Sklarova věta pro m náhodných veličin má následující tvar.

Věta 10 (Sklarova pro m náhodných veličin). *Necht Y_1, \dots, Y_m jsou náhodné veličiny s distribučními funkcemi F_1, \dots, F_m a sdruženou distribuční funkcí H . Potom existuje m -rozměrná kopule C taková, že pro každé $\mathbf{y} \in \mathbb{R}^m$ platí*

$$H(y_1, \dots, y_m) = C(F_1(y_1), \dots, F_m(y_m)).$$

Jsou-li F_1, \dots, F_m spojitě, pak je C určena jednoznačně.

Důkaz. Viz Rüschendorf (2009, kapitola 2). □

Také charakterizace nezávislosti pomocí součinné kopule se přenáší do obecné dimenze. Necht Y_1, \dots, Y_m jsou absolutně spojitě náhodné veličiny. Potom jsou Y_1, \dots, Y_m nezávislé právě tehdy, když je jejich kopule daná jako m -rozměrná součinná kopule $\Pi^m(\mathbf{u}) = u_1 \cdot \dots \cdot u_m$.

Závěrem zde definujeme m -rozměrnou empirickou kopuli.

Definice 15. *Necht $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ jsou nezávislé, stejně rozdělené náhodné vektory, kde pro $i = 1, \dots, n$ je $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^\top$. Potom empirickou kopulí na základě těchto náhodných vektorů rozumíme funkci C_n definovanou na \mathbf{I}^m předpisem*

$$C_n(u_1, \dots, u_m) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(F_{1n}(Y_{i1}) \leq u_1, \dots, F_{mn}(Y_{im}) \leq u_m),$$

kde F_{jn} značí empirickou distribuční funkci j -té složky definovanou analogicky jako v (1.4).

2. Statistická inference

V celé této kapitole se budeme věnovat statistické inferenci v modelech založených na kopulích. Kapitola je rozdělena do třech hlavních sekcí lišících se přístupem a používanými metodami. Tyto sekce budou věnovány nejprve parametrickým metodám pro odhad parametrů kopule, dále semiparametrickým metodám kombinujícím metody založené na maximální věrohodnosti a empirické odhady, závěrem se pak zaměříme na testy dobré shody.

2.1 Parametrické metody odhadu

V této sekci budou představeny parametrické přístupy k modelování kopulí. Budeme se věnovat dvěma z nich, a to konkrétně metodě založené na maximální věrohodnosti a dále také metodě inferenčních funkcí pro marginály navrhanou v knize Joe (1997, kapitola 10).

Předpokládejme v této sekci následující strukturu dat. Nechť $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ je posloupnost nezávislých, stejně rozdělených náhodných vektorů s m složkami, tedy $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^\top$ pro každé $i \in \{1, \dots, n\}$. Nechť pro každé $j \in \{1, \dots, m\}$ označuje $F_j(\cdot; \boldsymbol{\alpha}_j)$ jednorozměrnou distribuční funkci náhodné veličiny Y_{1j} známou až na parametr $\boldsymbol{\alpha}_j$ dimenze k_j z parametrického prostoru $\Xi_j \subseteq \mathbb{R}^{k_j}$. Dále nechť C je kopule náležící do známé rodiny $\{C(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, která sdružuje tato jednorozměrná rozdělení, tedy dle Sklarovy věty platí

$$F(\mathbf{y}_1; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m, \boldsymbol{\theta}) = C(F_1(y_{11}; \boldsymbol{\alpha}_1), \dots, F_m(y_{1m}; \boldsymbol{\alpha}_m); \boldsymbol{\theta}).$$

Vektorem všech parametrů je pak $\boldsymbol{\eta} = (\boldsymbol{\alpha}_1^\top, \dots, \boldsymbol{\alpha}_m^\top, \boldsymbol{\theta}^\top)^\top$ jakožto vektor délky $k = \sum_{j=1}^m k_j + p$ z parametrického prostoru $\Upsilon \subseteq \mathbb{R}^k$. Pokud $c(\cdot; \boldsymbol{\theta})$ označuje hustotu kopule $C(\cdot; \boldsymbol{\theta})$ podle její definice v (1.6), nabývá hustota náhodného vektoru \mathbf{Y}_1 tvaru

$$f(\mathbf{y}_1; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m, \boldsymbol{\theta}) = c(F_1(y_{11}; \boldsymbol{\alpha}_1), \dots, F_m(y_{1m}; \boldsymbol{\alpha}_m); \boldsymbol{\theta}) \prod_{j=1}^m f_j(y_{1j}; \boldsymbol{\alpha}_j),$$

kde f_j značí (jednorozměrnou) hustotu Y_{1j} . Na základě napozorovaných dat pak můžeme vypočítat jak logaritickou věrohodnost pro jednotlivé složky (pouze na základě marginálů)

$$\ell_j(\boldsymbol{\alpha}_j) = \sum_{i=1}^n \log(f_j(Y_{ij}; \boldsymbol{\alpha}_j)), \quad (2.1)$$

tak zároveň logaritickou věrohodnost pro sdružené rozdělení

$$\ell(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m, \boldsymbol{\theta}) = \sum_{i=1}^n \log(f(\mathbf{Y}_i; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m, \boldsymbol{\theta})).$$

V tomto okamžiku se metoda maximální věrohodnosti a metoda inferenčních funkcí pro marginály rozcházejí. Při použití klasického postupu metodou maximální věrohodnosti získáme odhady parametrů $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m, \boldsymbol{\theta}$ jako

$$\left(\widehat{\boldsymbol{\alpha}}_{1,n}^\top, \dots, \widehat{\boldsymbol{\alpha}}_{m,n}^\top, \widehat{\boldsymbol{\theta}}_n^\top \right)^\top = \arg \max_{(\boldsymbol{\alpha}_1^\top, \dots, \boldsymbol{\alpha}_m^\top, \boldsymbol{\theta}^\top)^\top \in \Upsilon} \ell(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m, \boldsymbol{\theta}).$$

Za určitých podmínek je takto nalezený odhad řešením soustavy rovnic

$$\left(\frac{\partial \ell}{\partial \boldsymbol{\alpha}_1^\top}, \dots, \frac{\partial \ell}{\partial \boldsymbol{\alpha}_m^\top}, \frac{\partial \ell}{\partial \boldsymbol{\theta}^\top} \right)^\top = \mathbf{0}_k,$$

kde $\mathbf{0}_k$ značí vektor nul délky k . Těmito podmínkami jsou – diferencovatelnost logaritnické věrohodnosti ℓ podle všech složek $\boldsymbol{\eta}$, otevřenost parametrického prostoru Υ a omezenost shora funkce ℓ .

V případě metody inferenčních funkcí pro marginály je postup dvoukrokový. V prvním kroku nalezneme odhady parametrů $\tilde{\boldsymbol{\alpha}}_{1,n}, \dots, \tilde{\boldsymbol{\alpha}}_{m,n}$ pomocí maximalizace logaritnické věrohodnosti pro jednotlivé složky z (2.1), tedy

$$\tilde{\boldsymbol{\alpha}}_{j,n} = \arg \max_{\boldsymbol{\alpha}_j \in \Xi_j} \ell_j(\boldsymbol{\alpha}_j).$$

Takto získané odhady poté v druhém kroku dosadíme do celkové věrohodnosti ℓ , kterou již maximalizujeme pouze přes parametr $\boldsymbol{\theta}$. Formálně zapsáno

$$\tilde{\boldsymbol{\theta}}_n = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\tilde{\boldsymbol{\alpha}}_{1,n}, \dots, \tilde{\boldsymbol{\alpha}}_{m,n}, \boldsymbol{\theta}).$$

Takto nalezený odhad je za určitých podmínek řešením soustavy rovnic

$$\left(\frac{\partial \ell_1}{\partial \boldsymbol{\alpha}_1^\top}, \dots, \frac{\partial \ell_m}{\partial \boldsymbol{\alpha}_m^\top}, \frac{\partial \ell}{\partial \boldsymbol{\theta}^\top} \right)^\top = \mathbf{0}_k.$$

Těmito podmínkami jsou – existence uvedených parciálních derivací, otevřenost Υ a omezenost shora funkcí $\ell_1, \dots, \ell_m, \ell$.

V obecném případě se odhady $(\hat{\boldsymbol{\alpha}}_{1,n}^\top, \dots, \hat{\boldsymbol{\alpha}}_{m,n}^\top, \hat{\boldsymbol{\theta}}_n^\top)^\top$ a $(\tilde{\boldsymbol{\alpha}}_{1,n}^\top, \dots, \tilde{\boldsymbol{\alpha}}_{m,n}^\top, \tilde{\boldsymbol{\theta}}_n^\top)^\top$ neshodují. Joe (1997, str. 300) uvádí, že pro mnohorozměrné normální rozdělení dané normální kopulí (viz sekce 1.5) s korelační maticí parametrizovanou vektorem $\boldsymbol{\theta}$ a jednorozměrnými marginály s rozdělením $\mathcal{N}(\mu_j, \sigma_j^2)$ pro $j = 1, \dots, m$, kde $\boldsymbol{\alpha}_j = (\mu_j, \sigma_j^2)^\top$, se však oba typy odhadů shodují.

2.1.1 Asymptotické rozdělení odhadu

Abychom byli schopni pro zkoumané parametry počítat asymptotické konfidenční množiny a testovat hodnoty parametrů, potřebujeme znát asymptotické rozdělení odhadnutých parametrů. Předpokládejme, že $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ jsou nezávislé, stejně rozdělené náhodné vektory z rozdělení se sdruženou distribuční funkcí $F(\cdot; \boldsymbol{\eta})$ a hustotou $f(\cdot; \boldsymbol{\eta})$. Buď $\mathbf{g} : S_Y \times \Upsilon \rightarrow \mathbb{R}^k$ funkce, kde S_Y je obor hodnot náhodného vektoru \mathbf{Y}_1 a Υ je parametrický prostor tvořený možnými hodnotami parametru $\boldsymbol{\eta}$. Také předpokládejme, že soustava odhadovacích rovnic pro odhad $\tilde{\boldsymbol{\eta}}_n$ parametru $\boldsymbol{\eta}$ má tvar

$$\sum_{i=1}^n \mathbf{g}(\mathbf{Y}_i; \tilde{\boldsymbol{\eta}}_n) = \mathbf{0}_k. \quad (2.2)$$

Tento odhad identifikuje parametr $\boldsymbol{\eta}_Y$ takový, že $\mathbf{E} \mathbf{g}(\mathbf{Y}_1; \boldsymbol{\eta}_Y) = \mathbf{0}_k$. Tedy pokud definujeme \mathbf{g} například předpisem

$$g_j(\mathbf{y}; \boldsymbol{\eta}) = \frac{\partial \log f(\mathbf{y}; \boldsymbol{\eta})}{\partial \eta_j}, \quad j = 1, \dots, k,$$

dostáváme odhadovací rovnice z (2.2) přesně ve tvaru pro maximálně věrohodný odhad.

Pokud \mathbf{g} definujeme jako

$$\mathbf{g}(\mathbf{y}; \boldsymbol{\eta}) = \left(\frac{\partial \log f_1(\mathbf{y}; \boldsymbol{\eta})}{\partial \boldsymbol{\alpha}_1^\top}, \dots, \frac{\partial \log f_m(\mathbf{y}; \boldsymbol{\eta})}{\partial \boldsymbol{\alpha}_m^\top}, \frac{\partial \log f(\mathbf{y}; \boldsymbol{\eta})}{\partial \boldsymbol{\theta}^\top} \right)^\top,$$

dostáváme odhadovací rovnice z (2.2) ve tvaru pro metodu inferenčních funkcí pro marginály.

Nechť $\partial \mathbf{g} / \partial \boldsymbol{\eta}^\top$ je matice parciálních derivací funkce \mathbf{g} , tedy matice, která má na pozici (i, j) prvek $\partial g_i(\mathbf{y}; \boldsymbol{\eta}) / \partial \eta_j$. Dále předpokládejme, že funkce \mathbf{g} , rozdělení náhodného vektoru \mathbf{Y}_1 a parametrický prostor Υ splňují podmínky regularity pro Z-odhady, které zmiňují Jurečková, Sen a Picek (2013, věta 5.1.). Asymptotické výsledky teorie Z-odhadů nám poté dávají, že pro $n \rightarrow \infty$ platí

$$\sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_Y) \xrightarrow{d} \mathcal{N}_k \left(\mathbf{0}, (\mathbb{D}_g)^{-1} \mathbb{M}_g ((\mathbb{D}_g)^{-1})^\top \right),$$

kde \xrightarrow{d} značí konvergenci v distribuci,

$$\mathbb{D}_g = \mathbb{E} \left(\partial \mathbf{g}(\mathbf{Y}; \boldsymbol{\eta}_Y) / \partial \boldsymbol{\eta}^\top \right), \quad \mathbb{M}_g = \mathbb{E} \left(\mathbf{g}(\mathbf{Y}; \boldsymbol{\eta}_Y) \mathbf{g}^\top(\mathbf{Y}; \boldsymbol{\eta}_Y) \right)$$

a k je dimenze $\boldsymbol{\eta}$. Podotkněme, že v případě, že hledáme maximálně věrohodný odhad (tedy speciální případ Z-odhadu), lze podmínky regularity konkrétněji uvádět ve tvaru, v jakém je uvádí prof. Anděl (2007) v předpokladech věty 7.93.

Posledním krokem, který nám zbývá učinit, abychom byli schopni provádět inferenci o $\boldsymbol{\eta}$, je odhad uvedené asymptotické rozptylové matice. Možným přístupem je proto využít techniky jackknife. Více o této technice viz Efron (1987). Pokud pro $i = 1, \dots, n$ označíme $\tilde{\boldsymbol{\eta}}_{n-1}^{(i)}$ jako odhad $\boldsymbol{\eta}$ na základě dat $\mathbf{Y}_1, \dots, \mathbf{Y}_{i-1}, \mathbf{Y}_{i+1}, \dots, \mathbf{Y}_n$, nabízí Joe (1997, str. 302) odhad pro matici

$$n^{-1} (\mathbb{D}_g)^{-1} \mathbb{M}_g ((\mathbb{D}_g)^{-1})^\top \tag{2.3}$$

ve tvaru

$$\sum_{i=1}^n \left(\tilde{\boldsymbol{\eta}}_{n-1}^{(i)} - \tilde{\boldsymbol{\eta}} \right) \left(\tilde{\boldsymbol{\eta}}_{n-1}^{(i)} - \tilde{\boldsymbol{\eta}} \right)^\top.$$

Použitím tohoto odhadu jsme pak již zcela standardním způsobem schopni počítat konfidenční množiny pro $\boldsymbol{\eta}$ a testovat hypotézy o hodnotách $\boldsymbol{\eta}$, případně jejich částech.

Dalším přístupem pro získání asymptotických intervalů spolehlivosti je využití bootstrapu. Zvolme B dostatečně velké a pro každé $b = 1, \dots, B$ vygenerujme náhodný výběr $\mathbf{Y}_{1,b}^*, \dots, \mathbf{Y}_{n,b}^*$ jako prostý náhodný výběr s opakováním z výběru $\mathbf{Y}_1, \dots, \mathbf{Y}_n$. Pro každé $b = 1, \dots, B$ dále nalezneme odhad $\tilde{\boldsymbol{\eta}}_{n,b}^*$. Odhad matice (2.3) je pak dán jako

$$\frac{1}{B-1} \sum_{b=1}^B \left(\tilde{\boldsymbol{\eta}}_{n,b}^* - \bar{\boldsymbol{\eta}}_{n,B}^* \right) \left(\tilde{\boldsymbol{\eta}}_{n,b}^* - \bar{\boldsymbol{\eta}}_{n,B}^* \right)^\top, \quad \text{kde } \bar{\boldsymbol{\eta}}_{n,B}^* = \frac{1}{B} \sum_{b=1}^B \tilde{\boldsymbol{\eta}}_{n,b}^*.$$

Stejně tak by bootstrap mohl být využit přímo pro počítání intervalů spolehlivosti pro složky $\boldsymbol{\eta}$ (více viz van der Vaart, 2000, kapitola 23).

Porovnání odhadu metodou inferenčních funkcí pro marginály oproti odhadu metodou maximální věrohodnosti zkoumal ve své práci Xu (1996). Jedním typem porovnání bylo porovnání asymptotických rozptylových matic, druhým porovnání charakteristik z Monte Carlo simulací odhadů. Pro všechny zkoumané modely se ukázalo, že metoda inferenčních funkcí pro marginály je ve svých charakteristikách (například odhad střední čtvercové chyby odhadu) kvalitativně velmi blízká metodě maximální věrohodnosti, a proto je vzhledem ke své mnohem nižší výpočetní náročnosti a větší numerické stabilitě preferovaným způsobem parametrického odhadu parametru $\boldsymbol{\eta}$ v modelech založených na kopulích.

2.2 Semiparametrické metody odhadu

V případě, že nejsme schopni definovat parametrickou strukturu marginálů, případně máme dostatečné množství pozorování pro kvalitní neparametrický odhad marginálních rozdělání, plně parametrické metody definované v sekci 2.1 nejsou vhodným přístupem k odhadu. V této situaci je pak možno využít kombinace parametrických a neparametrických metod způsobem, že parametrickou strukturu předepisujeme pouze kopuli, nikoliv však jednotlivým marginálním rozděláním.

2.2.1 Semiparametrické Z-odhady

V úvodu této části podotkněme, že název semiparametrické Z-odhady je zde použit pro odhady, které jsou v anglicky psané literatuře nazývány jako „rank approximate Z-estimators“.

Maximálně pseudověrohodný odhad

Předpokládejme nyní, že $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ je posloupnost nezávislých, stejně rozdělených náhodných vektorů s m složkami, tedy $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^\top$ pro každé $i \in \{1, \dots, n\}$. Nechť pro každé $j \in \{1, \dots, m\}$ je F_j jednorozměrná distribuční funkce náhodné veličiny Y_{1j} . Dále nechť C je absolutně spojitá kopule náležící do známé rodiny $\{C(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, která sdružuje tato jednorozměrná rozdělání, tedy platí

$$F(\mathbf{y}_1; \boldsymbol{\theta}) = C(F_1(y_{11}), \dots, F_m(y_{1m}); \boldsymbol{\theta}).$$

Hustota náhodného vektoru \mathbf{Y}_1 je poté ve tvaru

$$f(\mathbf{y}_1; \boldsymbol{\theta}) = c(F_1(y_{11}), \dots, F_m(y_{1m}); \boldsymbol{\theta}) \prod_{j=1}^m f_j(y_{1j}),$$

kde c značí hustotu kopule C . Pokud definujeme empirickou distribuční funkci pro j -tý marginál na základě $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ jako

$$F_{jn}(y) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(Y_{ij} \leq y),$$

můžeme psát logaritmickou pseudověrohodnost ve tvaru

$$\ell_{PS}(\boldsymbol{\theta}) = \sum_{i=1}^n \log c(F_{1n}(Y_{i1}), \dots, F_{mn}(Y_{im}); \boldsymbol{\theta}) + \sum_{i=1}^n \sum_{j=1}^m \log(f_j(Y_{ij})), \quad (2.4)$$

kde f_j značí hustotu Y_{1j} , která nezávisí na parametru $\boldsymbol{\theta}$. Tato funkce se nazývá pseudověrohodnost z důvodu, že ačkoliv má obecnou strukturu obdobnou klasické věrohodnosti, je pouze její aproximací. Představme si, že by argumenty hustoty c v (2.4) byly dány jako $c(F_1(Y_{i1}), \dots, F_m(Y_{im}); \boldsymbol{\theta})$, tedy bychom použili skutečné marginální distribuční funkce namísto těch empirických. Označme tyto argumenty

$$U_{ij} = F_j(Y_{ij}) \quad \text{pro } i = 1, \dots, n \text{ a } j = 1, \dots, m.$$

Rozdělení nezávislých, stejně rozdělených náhodných vektorů $\mathbf{U}_1, \dots, \mathbf{U}_n$, kde $\mathbf{U}_i = (U_{i1}, \dots, U_{im})^\top$, by poté bylo rozdělení s distribuční funkcí C , se složkami, které jsou rovnoměrně rozdělené na \mathbf{I} . Takto upravená funkce ℓ_{ps} by pak dávala skutečnou věrohodnost pro parametr $\boldsymbol{\theta}$.

V našem případě však, jelikož neznáme skutečné marginály, používáme při výpočtu pseudověrohodnosti takzvaná pseudopozorování, která jsou dána jako

$$\widehat{U}_{ij} = F_{jn}(Y_{ij}) \quad \text{pro } i = 1, \dots, n \text{ a } j = 1, \dots, m. \quad (2.5)$$

Náhodné vektory $\widehat{\mathbf{U}}_1, \dots, \widehat{\mathbf{U}}_n$ již nejsou nezávislé a také jejich složky již nejsou rovnoměrně rozdělené na \mathbf{I} .

Odhad založený na maximalizaci pseudověrohodnosti v (2.4), má tedy tvar

$$\widehat{\boldsymbol{\theta}}_n^{PS} = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell_{PS}(\boldsymbol{\theta}).$$

Pokud budeme derivaci hustoty c podle parametru $\boldsymbol{\theta}$ značit \mathbf{c}' , tedy

$$\mathbf{c}' = (\partial c / \partial \theta_1, \dots, \partial c / \partial \theta_p)^\top,$$

splňoval by odhad na základě této pseudověrohodnosti soustavu rovnic

$$\sum_{i=1}^n \frac{\mathbf{c}'(\widehat{U}_{i1}, \dots, \widehat{U}_{im}; \widehat{\boldsymbol{\theta}}_n^{PS})}{c(\widehat{U}_{i1}, \dots, \widehat{U}_{im}; \widehat{\boldsymbol{\theta}}_n^{PS})} = \mathbf{0}_p. \quad (2.6)$$

Obecný semiparametrický Z-odhad

Analogicky k teorii Z-odhadů můžeme vektorovou funkci \mathbf{c}'/c na levé straně rovnice v (2.6) nahradit obecnější funkcí $\boldsymbol{\phi}(\cdot; \boldsymbol{\theta}) : (0,1)^m \rightarrow \mathbb{R}^p$ splňující podmínku

$$\int_{(0,1)^m} \boldsymbol{\phi}(\mathbf{u}; \boldsymbol{\theta}) dC(\mathbf{u}; \boldsymbol{\theta}) = \mathbf{0}_p. \quad (2.7)$$

Poznámka. Zde, v rovnici (2.7), i v dalším průběhu textu budeme integrálem vektorové funkce rozumět vektor integrálů jednotlivých složek. Podobně integrálem matice \mathbb{A} budeme rozumět matici, jejíž prvky jsou dány jako integrály prvků matice \mathbb{A} .

Hledaný odhad $\widehat{\boldsymbol{\theta}}_n^{RAZ}$ je pak řešením soustavy rovnic

$$\sum_{i=1}^n \boldsymbol{\phi}(\widehat{U}_{i1}, \dots, \widehat{U}_{im}; \widehat{\boldsymbol{\theta}}_n^{RAZ}) = \mathbf{0}_p. \quad (2.8)$$

Takovýto odhad budeme nazývat semiparametrický Z-odhad (v anglické literatuře označován jako „rank approximate Z-estimator“, odkud plyne jeho symbolické

značení). Jak uvádí Tsukahara (2005, sekce 3.1.), v obecném případě nemusí řešení soustavy (2.8) existovat. V tomto případě autor doporučuje odhad $\hat{\theta}_n^{RAZ}$ volit jako

$$\hat{\theta}_n^{RAZ} = \arg \min_{\theta \in \Theta} \left\| \sum_{i=1}^n \phi(\hat{U}_{i1}, \dots, \hat{U}_{im}); \theta \right\|,$$

kde $\|\cdot\|$ značí euklidovskou normu.

Zaměříme se nyní na vlastnosti odhadu $\hat{\theta}_n^{RAZ}$. K tomuto účelu potřebujeme nejprve sérii definic a podmínek regularity.

Definice 16. *Definujme následující:*

1. Necht \mathcal{Q} označuje systém spojitých funkcí q na intervalu $[0, 1]$, které jsou kladné na intervalu $(0, 1)$, symetrické okolo $1/2$, rostoucí na $[0, 1/2]$ a splňují $\int_0^1 (q(t))^{-2} < \infty$.
2. Funkci $r : (0, 1) \rightarrow (0, \infty)$ nazveme u -křivkou, pokud je symetrická okolo $1/2$ a klesající na intervalu $(0, 1/2]$.
3. Pro $\beta \in (0, 1)$ a u -křivku r definujme funkci

$$r_\beta(t) = \begin{cases} r(\beta t), & 0 < t \leq 1/2, \\ r(1 - \beta(1 - t)), & 1/2 < t \leq 1. \end{cases} \quad (2.9)$$

Pokud pro každé $\beta > 0$ na nějakém okolí 0 existuje konstanta M_β taková, že $r_\beta \leq M_\beta r$ na intervalu $(0, 1)$, nazývá se r reprodukující u -křivka. Systém reprodukujících u -křivek označíme \mathcal{R} .

Uveďme si příklady funkcí, které jsou obsaženy v systémech \mathcal{Q} a \mathcal{R} .

Příklad 4. Definujme pro $\alpha \in (0, 1/2)$ funkci $q(\cdot; \alpha) : [0, 1] \rightarrow \mathbb{R}$ předpisem

$$q(t; \alpha) = (t(1 - t))^\alpha.$$

Pro libovolné $\alpha \in (0, 1/2)$ je potom $q(\cdot; \alpha)$ zřejmě symetrická okolo $1/2$, kladná na $(0, 1)$ i spojitá na $[0, 1]$. Derivace $q(t; \alpha)$ podle t je rovna $\alpha (t(1 - t))^{\alpha-1} (1 - 2t)$, a tato derivace je tedy kladná na $(0, 1/2)$. Funkce $q(\cdot; \alpha)$ je proto vzhledem ke své spojitosti zprava v 0 rostoucí na $[0, 1/2]$. Dále platí také

$$\int_0^1 (q(t; \alpha))^{-2} = \int_0^1 (t(1 - t))^{-2\alpha} = B(1 - 2\alpha, 1 - 2\alpha) < \infty,$$

kde B značí beta funkci. Funkce $q(\cdot; \alpha)$ tedy pro každé $\alpha \in (0, 1/2)$ náleží do systému \mathcal{Q} .

Příklad 5. Definujme pro $\tau > 0, D > 0$ funkci $r(\cdot; D, \tau) : (0, 1) \rightarrow (0, \infty)$ předpisem

$$r(t; D, \tau) = D (t(1 - t))^{-\tau}.$$

Tato funkce je pro každé $\tau > 0, D > 0$ zřejmě symetrická okolo $1/2$ a obdobně jako v příkladu 4 lze ukázat, že je klesající na intervalu $[0, 1/2]$. Funkce $r(\cdot; D, \tau)$

je proto u-křivka. Definujme nyní funkci $r_\beta(\cdot; D, \tau)$ předpisem v (2.9). Potom pro libovolné pevné $\beta \in (0,1)$ a $t \in (0, 1/2]$ platí

$$\frac{r_\beta(t; D, \tau)}{r(t; D, \tau)} = \frac{r(\beta t; D, \tau)}{r(t; D, \tau)} = \frac{D(\beta t(1-\beta t))^{-\tau}}{D(t(1-t))^{-\tau}} = \left(\frac{1-t}{\beta(1-\beta t)} \right)^\tau.$$

Tato funkce je spojitá na intervalu $(0, 1/2]$, neboť pro každé $\beta \in (0,1)$ a $t \in (0, 1/2]$ je $\beta t < 1$, a její derivace podle t je dána jako

$$\begin{aligned} \frac{\partial}{\partial t} \frac{r_\beta(t; D, \tau)}{r(t; D, \tau)} &= \tau \left(\frac{1-t}{\beta(1-\beta t)} \right)^{\tau-1} \frac{-\beta(1-\beta t) + \beta^2(1-t)}{(\beta(1-\beta t))^2} \\ &= \tau \left(\frac{1-t}{\beta(1-\beta t)} \right)^{\tau-1} \frac{\beta(1-\beta)}{(\beta(1-\beta t))^2} < 0. \end{aligned}$$

Funkce $r_\beta(t; D, \tau)/r(t; D, \tau)$ je proto klesající v t na intervalu $(0, 1/2]$. Proto

$$\sup_{t \in (0, 1/2]} \frac{r_\beta(t; D, \tau)}{r(t; D, \tau)} = \lim_{t \rightarrow 0} \frac{r_\beta(t; D, \tau)}{r(t; D, \tau)} = \frac{1}{\beta^\tau}.$$

Dohromady tedy pro pevné $\beta \in (0,1)$ platí, že

$$r_\beta(t; D, \tau) \leq \frac{1}{\beta^\tau} r(t; D, \tau), \quad t \in (0, 1/2].$$

Zcela analogicky bychom mohli ukázat, že pro pevné $\beta \in (0,1)$ platí, že také

$$r_\beta(t; D, \tau) \leq \frac{1}{\beta^\tau} r(t; D, \tau), \quad t \in (1/2, 1).$$

Nalezli jsme tedy konstantu požadovanou v bodu 3 definice 16, a proto pro každé $\tau > 0, D > 0$ náleží funkce $r(\cdot; D, \tau)$ do systému \mathcal{R} .

Uvedme nyní předpoklady nutné k platnosti asymptotických výsledků uvedených níže. Podmínkami regularity pro semiparametrický Z-odhad budeme označovat následujících pět požadavků na rodinu kopulí $\{C(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, parametrický prostor Θ a funkci ϕ .

(R-1) Necht' jsou pro každé $\boldsymbol{\theta} \in \Theta$ složky funkce $\phi(\cdot; \boldsymbol{\theta}) : (0,1)^m \rightarrow \mathbb{R}^p$ spojitě diferencovatelné s parciálními derivacemi $\phi_j^{(i)} = \partial \phi_j / \partial u_i$ pro $i = 1, \dots, m$ a $j = 1, \dots, p$. Dále necht' existují funkce $r_{ij} \in \mathcal{R}, \tilde{r}_{ij} \in \mathcal{R}$ a $q_{ij} \in \mathcal{Q}$ splňující pro každé $\mathbf{u} \in (0,1)^m$

$$\begin{aligned} |\phi_j(\mathbf{u}; \boldsymbol{\theta})| &\leq \prod_{l=1}^m r_{lj}(u_l), \\ |\phi_j^{(i)}(\mathbf{u}; \boldsymbol{\theta})| &\leq \tilde{r}_{ij}(u_i) \prod_{k \neq i} r_{kj}(u_k), \\ \int_{(0,1)^m} \left(\prod_{l=1}^m r_{lj}(u_l) \right)^2 dC(\mathbf{u}; \boldsymbol{\theta}) &< \infty, \\ \int_{(0,1)^m} q_{ij}(u_i) \tilde{r}_{ij}(u_i) \prod_{k \neq i} r_{kj}(u_k) dC(\mathbf{u}; \boldsymbol{\theta}) &< \infty \end{aligned}$$

pro $i = 1, \dots, m$ a $j = 1, \dots, p$.

(R-2) Je-li θ_0 skutečná hodnota parametru θ , předpokládejme, že platí

$$\int_{(0,1)^m} \phi(\mathbf{u}; \theta) dC(\mathbf{u}; \theta_0) = \mathbf{0}_p \Leftrightarrow \theta = \theta_0.$$

(R-3) Pro každé $\theta \in \Theta$ platí

$$\int_{(0,1)^m} (\phi(\mathbf{u}; \theta))^2 dC(\mathbf{u}; \theta) < \infty.$$

(R-4) $\phi(\cdot; \theta)$ je diferencovatelná vzhledem k θ s Jakobiho maticí

$$\mathbb{D}_\phi(\mathbf{u}; \theta) = \left(\frac{\partial \phi_i(\mathbf{u}; \theta)}{\partial \theta_j} \right)_{i,j=1}^p.$$

$\mathbb{D}_\phi(\mathbf{u}; \theta)$ je dále spojitá v obou svých argumentech a Frobeniova norma $\|\mathbb{D}_\phi(\mathbf{u}; \theta)\|$ je shora omezená $dC(\cdot; \theta)$ integrovatelnou majorantou $h(\mathbf{u})$.

(R-5) Čtvercová matice \mathbb{A}_0 řádu p ve tvaru

$$\mathbb{A}_0 = \int_{(0,1)^m} \mathbb{D}_\phi(\mathbf{u}; \theta_0) dC(\mathbf{u}; \theta_0)$$

je regulární.

Příklad 6. Ověřme podmínku regularity (R-1) pro odhad založený na pseudověrohodnosti a dvourozměrnou Farlie-Gumbel-Morgensternovu rodinu kopulí (viz sekce 1.5) v případě, kdy parametr θ leží v intervalu $\Theta_R = [-1 + \epsilon, 1 - \epsilon]$ pro nějaké $\epsilon > 0$. Vzhledem k tomu, že ϕ má pouze jednu složku, nebudeme pro zjednodušení uvádět index (v podmínce regularity označován jako j), který se k ní váže.

Hustota kopule z Farlie-Gumbel-Morgensternovy rodiny je pro $u, v \in \mathbf{I}$ dána jako

$$c(u, v; \theta) = \frac{\partial^2 C(u, v; \theta)}{\partial u \partial v} = 1 + \theta(1 - 2u)(1 - 2v).$$

Její derivace podle θ je ve tvaru

$$c'(u, v; \theta) = \frac{\partial c(u, v; \theta)}{\partial \theta} = (1 - 2u)(1 - 2v).$$

Dohromady pak máme

$$\phi(u, v; \theta) = \frac{c'(u, v; \theta)}{c(u, v; \theta)} = \frac{(1 - 2u)(1 - 2v)}{1 + \theta(1 - 2u)(1 - 2v)}.$$

Parciální derivaci funkce ϕ podle jejích argumentů stačí vyjádřit pouze pro jeden z nich, neboť je ϕ ve svých argumentech symetrická. Platí

$$\begin{aligned} \phi^{(1)}(u, v; \theta) &= \frac{(4v - 2)(1 + \theta(1 - 2u)(1 - 2v)) - (1 - 2u)(1 - 2v)\theta(4v - 2)}{(1 + \theta(1 - 2u)(1 - 2v))^2} \\ &= \frac{4v - 2}{(1 + \theta(1 - 2u)(1 - 2v))^2}. \end{aligned}$$

Odhadujeme-li shora funkci $|\phi|$, dostáváme

$$|\phi(u, v; \theta)| = \left| \frac{(1-2u)(1-2v)}{1+\theta(1-2u)(1-2v)} \right| \leq \frac{1}{1-|\theta|} = \frac{1}{\epsilon}.$$

Podobně můžeme postupovat také pro funkci $|\phi^{(1)}|$, kde platí

$$|\phi^{(1)}(u, v; \theta)| = \left| \frac{4v-2}{(1+\theta(1-2u)(1-2v))^2} \right| \leq \frac{2}{(1-|\theta|)^2} = \frac{2}{\epsilon^2}.$$

Vidíme, že $|\phi|$ i $|\phi^{(1)}|$ jsou na $(0,1)^2$ omezené. Definujme funkce $r_1, r_2, \tilde{r}_1, \tilde{r}_2, q_1$ a q_2 předpis

$$\begin{aligned} r_1(u) = r_2(u) = \tilde{r}_1(u) = \tilde{r}_2(u) &= \left| u - \frac{1}{2} \right| + \frac{\sqrt{2}}{\epsilon}, & u \in (0,1), \\ q_1(v) = q_2(v) &= (v(1-v))^{1/3}, & v \in [0,1]. \end{aligned}$$

Všechny 4 požadavky v podmínce regularity (R-1) jsou nyní splněny, neboť pro každé $u, v \in (0,1)$ platí

$$\begin{aligned} |\phi(u, v; \theta)| &\leq \frac{1}{\epsilon} \leq \left(\left| u - \frac{1}{2} \right| + \frac{\sqrt{2}}{\epsilon} \right) \left(\left| v - \frac{1}{2} \right| + \frac{\sqrt{2}}{\epsilon} \right) = r_1(u)r_2(v), \\ |\phi^{(1)}(u, v; \theta)| &\leq \frac{2}{\epsilon^2} \leq \left(\left| u - \frac{1}{2} \right| + \frac{\sqrt{2}}{\epsilon} \right) \left(\left| v - \frac{1}{2} \right| + \frac{\sqrt{2}}{\epsilon} \right) = \tilde{r}_1(u)r_2(v) \end{aligned}$$

a podmínky na integrovatelnost v (R-1) jsou splněny, jelikož funkce $r_1, r_2, \tilde{r}_1, \tilde{r}_2, q_1$ a q_2 jsou omezené, stejně jako kopule $C(u, v; \theta)$ pro každé $u, v \in (0,1)$.

Závěrem tohoto příkladu podotkněme, že vzhledem k omezenosti funkcí ϕ a $\phi^{(1)}$ jsme mohli funkce r_1, r_2, \tilde{r}_1 a \tilde{r}_2 volit také jako omezené, díky čemuž byla jejich integrovatelnost okamžitě zajištěna. Jako funkce q_1 a q_2 pak stačilo brát libovolné funkce z \mathcal{Q} , které jsou omezené na $(0,1)$.

Následující věta shrnuje asymptotické výsledky pro odhad $\hat{\boldsymbol{\theta}}_n^{RAZ}$.

Věta 11. *Jsou-li splněny podmínky regularity (R-1) až (R-5), potom s pravděpodobností blížící se 1, jakmile $n \rightarrow \infty$, existuje řešení $\hat{\boldsymbol{\theta}}_n^{RAZ}$ soustavy (2.8), které konverguje ke skutečné hodnotě parametru $\boldsymbol{\theta}_0$. Dále, jakmile $n \rightarrow \infty$, potom*

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_n^{RAZ} - \boldsymbol{\theta}_0 \right) \xrightarrow{d} \mathcal{N}_p \left(\mathbf{0}, (\mathbb{A}_0)^{-1} \mathbb{Q}_0 \left((\mathbb{A}_0)^{-1} \right)^\top \right),$$

kde \mathbb{A}_0 je definováno v podmínce regularity (R-5) a

$$\mathbb{Q}_0 = \text{var} \left(\boldsymbol{\phi}(\mathbf{U}; \boldsymbol{\theta}_0) + \sum_{j=1}^m \int_{(0,1)^m} \boldsymbol{\phi}^{(j)}(\mathbf{u}; \boldsymbol{\theta}_0) (\mathbb{1}(U_j \leq u_j) - u_j) dC(\mathbf{u}; \boldsymbol{\theta}_0) \right),$$

kde $\boldsymbol{\phi}^{(j)} = (\phi_1^{(j)}, \dots, \phi_p^{(j)})^\top$ jsou definovány v podmínce regularity (R-1) a \mathbf{U} je náhodný vektor z rozdělení s distribuční funkcí $C(\cdot; \boldsymbol{\theta}_0)$.

Důkaz. Viz Tsukahara (2005, Theorem 1). □

Abychom byli schopni tvrzení věty 11 aplikovat, je nutné uvést také konzistentní odhad rozptylové matice $(\mathbb{A}_0)^{-1} \mathbb{Q}_0 ((\mathbb{A}_0)^{-1})^\top$. Tsukahara (2005, sekce 3.1.) uvádí, že takovýmto odhadem je matice řádu p ve tvaru $(\hat{\mathbb{A}})^{-1} \hat{\mathbb{Q}} ((\hat{\mathbb{A}})^{-1})^\top$. Odhad matice $\hat{\mathbb{A}}$ je dán jako

$$\hat{\mathbb{A}} = \int_{(0,1)^m} (\mathbb{D}_\phi(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{RAZ})) dC(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{RAZ}). \quad (2.10)$$

Odhad $\hat{\mathbb{Q}}$ je dán jako výběrový rozptyl

$$\hat{\mathbb{Q}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{Q}_i - \bar{\mathbf{Q}}_n) (\mathbf{Q}_i - \bar{\mathbf{Q}}_n)^\top, \quad (2.11)$$

kde pro $i = 1, \dots, n$ je

$$\mathbf{Q}_i = \phi(\hat{\mathbf{U}}_i; \hat{\boldsymbol{\theta}}_n^{RAZ}) + \sum_{j=1}^m \int_{(0,1)^m} \phi^{(j)}(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{RAZ}) (\mathbb{1}(\hat{U}_{ij} \leq u_j) - u_j) dC(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{RAZ})$$

a $\hat{\mathbf{U}}_i$ jsou pseudopozorování definovaná v (2.5).

Příklad 7. Uvažujme, že chceme odhadovat jednorozměrný parametr θ v dvou-
rozměrné kopuli $C(\cdot; \theta)$ náležící do zvolené rodiny $\{C(\cdot; \theta); \theta \in \Theta \subseteq \mathbb{R}\}$. Před-
pokládejme, že budeme parametr θ odhadovat pomocí teorie maximálně pseudo-
věrohodných odhadů. Samotný odhad $\hat{\theta}_n^{PS}$ nalezneme jako řešení soustavy rovnic
(2.6). Ukažme si, jakým způsobem nalézt odhad asymptotického rozptylu odhadu
 $\hat{\theta}_n^{PS}$. Připomeňme, že pseudověrohodný odhad je speciálním případem semiparam-
etrického Z-odhadu, kde je pro jednorozměrný parametr funkce ϕ dána jako
 $\phi(\mathbf{u}; \theta) = c'(\mathbf{u}; \theta)/c(\mathbf{u}; \theta)$.

Nejprve se zaměříme na odhad matice \mathbb{A}_0 ve větě 11, kterou v tomto případě
pro jednorozměrný parametr označíme jako A_0 . Po dosažení konkrétního tvaru
funkce ϕ získáváme, že

$$A_0 = \int_{(0,1)^2} \frac{\partial^2 \log c(\mathbf{u}; \theta_0)}{\partial \theta^2} dC(\mathbf{u}; \theta_0). \quad (2.12)$$

Za podmínek analogických těm, které ve větě 7.27 uvádí prof. Anděl (2007), pak
můžeme vyjádřit

$$\begin{aligned} A_0 &= - \int_{(0,1)^2} \left(\frac{c'(\mathbf{u}; \theta_0)}{c(\mathbf{u}; \theta_0)} \right)^2 dC(\mathbf{u}; \theta_0) = - \mathbb{E}_{C(\cdot; \theta_0)} \left(\frac{c'(\mathbf{U}; \theta_0)}{c(\mathbf{U}; \theta_0)} \right)^2 \\ &= - \text{var}_{C(\cdot; \theta_0)} \left(\frac{c'(\mathbf{U}; \theta_0)}{c(\mathbf{U}; \theta_0)} \right), \end{aligned} \quad (2.13)$$

kde poslední rovnost plyne z podmínky regularity (R-2). Ekvivalentem A_0 v teorii
maximální věrohodnosti je Fisherova informace. Vidíme, že v tomto případě je

však prvek A_0 záporný. Tento fakt vychází z obecného tvaru A_0 tak, jak jej uvádíme pro Z-odhady. Pokud bychom požadovali, aby A_0 bylo kladné, volili bychom funkci ϕ ve tvaru $\phi(\mathbf{u}; \theta) = -c'(\mathbf{u}; \theta)/c(\mathbf{u}; \theta)$. Nahradíme-li v (2.13), podobně jako to činí Genest, Ghoudi a Rivest (1995, kapitola 3), skutečný parametr θ_0 odhadem $\hat{\theta}_n^{PS}$ a \mathbf{U} získanými pseudopozorováními $\widehat{\mathbf{U}}_i$ pro $i = 1, \dots, n$, dostáváme použitím výběrového rozptylu odhad

$$\widehat{A} = -\frac{1}{n-1} \sum_{i=1}^n \left(\frac{c'(\widehat{\mathbf{U}}_i; \hat{\theta}_n^{PS})}{c(\widehat{\mathbf{U}}_i; \hat{\theta}_n^{PS})} \right)^2.$$

Nyní přejdeme k odhadu matice \mathbb{Q}_0 ve větě 11, kterou v tomto případě pro jednorozměrný parametr označíme jako Q_0 . Opět dosazením konkrétní funkce ϕ dostáváme Q_0 ve tvaru

$$\text{var} \left(\frac{\partial \log c(\mathbf{U}; \theta_0)}{\partial \theta} + \underbrace{\sum_{j=1}^2 \int_{(0,1)^2} \frac{\partial^2 \log c(\mathbf{u}; \theta_0)}{\partial \theta \partial u_j} (\mathbb{1}(U_j \leq u_j) - u_j) dC(\mathbf{u}; \theta_0)}_{I_j} \right).$$

Počítejme integrál uvedený v předchozí rovnici pomocí per partes integrace vzhledem k u_j . Tento integrál je pak roven

$$\begin{aligned} I_j &= \int_0^1 \left(\int_0^1 \frac{\partial^2 \log c(\mathbf{u}; \theta_0)}{\partial \theta \partial u_j} (\mathbb{1}(U_j \leq u_j) - u_j) c(\mathbf{u}; \theta_0) du_j \right) du_{3-j} \\ &= \int_0^1 \left(\left[\frac{\partial \log c(\mathbf{u}; \theta_0)}{\partial \theta} (\mathbb{1}(U_j \leq u_j) - u_j) c(\mathbf{u}; \theta_0) \right]_{u_j=0}^1 \right) du_{3-j} \end{aligned} \quad (2.14)$$

$$- \int_0^1 \left(\int_0^1 \frac{\partial \log c(\mathbf{u}; \theta_0)}{\partial \theta} (-1) c(\mathbf{u}; \theta_0) du_j \right) du_{3-j} \quad (2.15)$$

$$- \int_0^1 \left(\int_0^1 \frac{\partial \log c(\mathbf{u}; \theta_0)}{\partial \theta} (\mathbb{1}(U_j \leq u_j) - u_j) \frac{\partial c(\mathbf{u}; \theta_0)}{\partial u_j} du_j \right) du_{3-j}. \quad (2.16)$$

Integrál v (2.14) je roven 0, neboť jak $\mathbb{1}(U_j \leq 1) - 1 = 1 - 1 = 0$, tak rovněž $\mathbb{1}(U_j \leq 0) - 0 = 0 - 0 = 0$. Integrál v (2.15) je roven 0, neboť se jedná přesně o podmínku regularity (R-2). Dohromady je tedy integrál I_j dán pouze jako výraz v (2.16). Hodnotu Q_0 můžeme nyní přepsat jako

$$\text{var} \left(\frac{c'(\mathbf{U}; \theta_0)}{c(\mathbf{U}; \theta_0)} - \sum_{j=1}^2 \int_{(0,1)^2} \frac{c'(\mathbf{u}; \theta_0)}{c(\mathbf{u}; \theta_0)} (\mathbb{1}(U_j \leq u_j) - u_j) \frac{c^{(j)}(\mathbf{u}; \theta_0)}{c(\mathbf{u}; \theta_0)} dC(\mathbf{u}; \theta_0) \right).$$

Nahradme opět skutečný parametr θ_0 odhadem $\hat{\theta}_n^{PS}$ a vektory \mathbf{u} a \mathbf{U} získanými pseudopozorováními $\widehat{\mathbf{U}}_i$ pro $i = 1, \dots, n$. Pokud navíc integrál uvedený výše počítáme vzhledem k empirické kopuli C_n namísto k $C(\cdot; \theta_0)$, dostáváme použitím výběrového rozptylu odhad

$$\widehat{Q} = \frac{1}{n-1} \sum_{i=1}^n (Q_i - \overline{Q}_n) (Q_i - \overline{Q}_n)^\top,$$

kde pro $i = 1, \dots, n$ je

$$Q_i = \frac{c'(\widehat{\mathbf{U}}_i; \widehat{\theta}_n^{PS})}{c(\widehat{\mathbf{U}}_i; \widehat{\theta}_n^{PS})} - \sum_{j=1}^2 \sum_{k=1}^n \frac{c'(\widehat{\mathbf{U}}_i; \widehat{\theta}_n^{PS})}{c(\widehat{\mathbf{U}}_i; \widehat{\theta}_n^{PS})} (\mathbb{1}(\widehat{U}_{ij} \leq \widehat{U}_{kj}) - \widehat{U}_{kj}) \frac{c^{(j)}(\widehat{\mathbf{U}}_i; \widehat{\theta}_n^{PS})}{n \cdot c(\widehat{\mathbf{U}}_i; \widehat{\theta}_n^{PS})}.$$

Konečně, odhadem asymptotického rozptylu odhadu $\widehat{\theta}_n^{PS}$ můžeme označit

$$\widehat{\sigma}^2 = \frac{\widehat{Q}}{\widehat{A}^2}. \quad (2.17)$$

Další příklady odhadovacích funkcí

V úvodu této podsekcce jsme poznamenali, že přístup pomocí semiparametrických Z-odhadů je přímočarým zobecněním pseudověrohodného odhadu z (2.6). Tsukahara (2005, sekce 3.1.) uvádí, že pro většinu používaných rodin kopulí je v případě volby $\phi(\mathbf{u}, \theta) = \mathbf{c}'(\mathbf{u}, \theta)/c(\mathbf{u}, \theta)$ splněna pětice podmínek regularity (R-1) až (R-5), a tedy je výše uvedená asymptotická teorie platná. Jaké jsou však další možné volby funkce ϕ ? Předpokládejme, že $m = 2$ a $p = 1$, tedy, že máme dvourozměrnou kopuli s jednorozměrným parametrem z rodiny $\{C(\cdot, \theta); \theta \in \Theta\}$. Definujme v tomto případě

$$\begin{aligned} \tau(C; \theta) &= 4 \int_{(0,1)^2} C(\mathbf{u}; \theta) dC(\mathbf{u}; \theta) - 1, \\ \rho(C; \theta) &= 12 \int_{(0,1)^2} u_1 u_2 dC(\mathbf{u}; \theta) - 3. \end{aligned} \quad (2.18)$$

Funkce τ je nazývána Kendallovo τ , funkce ρ pak Spearmanovo ρ . Obě tyto funkce jsou měřítky síly závislosti mezi dvěma veličinami. Více informací k těmto funkcím, případně k jejich porovnání s Pearsonovým korelačním koeficientem lze nalézt v práci Embrechts, McNeil a Straumann (1999). Zdefinujme dále funkci $\phi_\tau(\mathbf{u}; \theta) = 4C(\mathbf{u}; \theta) - 1 - \tau(C; \theta)$ a funkci $\phi_\rho(\mathbf{u}; \theta) = 12u_1 u_2 - 3 - \rho(C; \theta)$. Obě tyto funkce pak zřejmě splňují podmínku (2.7) a příslušné odhady patří mezi semiparametrické Z-odhady. Označme takto vzniklé odhady parametru θ po řadě jako τ semiparametrický Z-odhad a ρ semiparametrický Z-odhad.

2.2.2 Metoda nejmenší vzdálenosti

V případě, že nejsme přesvědčeni o správnosti zvolené rodiny kopulí, do které patří ta námi modelovaná, je alternativou k použití semiparametrických Z-odhadů použití metod založených na minimalizaci vzdálenosti. V anglicky psané literatuře jsou takové odhady nazývány „minimum distance estimators“. Výhodou takovýchto odhadů je, že v případě, že naše představa o parametrické rodině kopulí $\{C(\cdot, \theta); \theta \in \Theta \subset \mathbb{R}^p\}$ není správná, odhaduje metoda nejmenší vzdálenosti takové θ , které v určitém smyslu (jak uvidíme níže) minimalizuje vzdálenost mezi skutečnou kopulí a $\{C(\cdot, \theta); \theta \in \Theta\}$.

Předpokládejme opět, že $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ je posloupnost nezávislých, stejně rozdělených náhodných vektorů s m složkami, tedy $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^\top$ pro každé $i \in \{1, \dots, n\}$. Dále necht C značí skutečnou kopuli příslušející náhodnému vektoru \mathbf{Y}_1 a $\{C(\cdot, \theta); \theta \in \Theta \subset \mathbb{R}^p\}$ je parametrická rodina kopulí, kterou chceme

používat pro modelování dat. Zdůrazněme, že nepředpokládáme, že C náleží do této parametrické rodiny. Definujme dále funkcionál T na prostoru všech kopulí předpisem

$$T(D) = \arg \min_{\boldsymbol{\theta} \in \Theta} d(D, C(\cdot; \boldsymbol{\theta})),$$

kde d značí nějakou vzdálenost mezi rozděleními na $(0, 1)^m$. Ve zbylé části této práce budeme používat Cramér-von Misesovu vzdálenost

$$d(D, C) := d_{CvM}(D, C) = \int_{(0,1)^m} (D(\mathbf{u}) - C(\mathbf{u}))^2 d\mathbf{u}.$$

Další možnou volbou vzdálenosti může být například Kolmogorov-Smirnovova vzdálenost, která je definována jako

$$d_{KS}(D, C) = \sup_{\mathbf{u} \in (0,1)^m} |D(\mathbf{u}) - C(\mathbf{u})|.$$

Pokud C_n značí empirickou kopuli na základě $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ (viz definice 15), pak definujeme odhad parametru $\boldsymbol{\theta}$ metodou nejmenší vzdálenosti jako

$$\hat{\boldsymbol{\theta}}_n^{MD} = T(C_n).$$

I pro tento odhad jsme za určitých podmínek schopni nalézt asymptotické rozdělení. Označme vnitřek parametrického prostoru Θ jako $\text{int}(\Theta)$.

Podmínkami regularity pro odhad metodou nejmenší vzdálenosti budeme označovat následujících pět požadavků na rodinu kopulí $\{C(\cdot, \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$, parametrický prostor Θ a vzdálenost d .

(D-1) Pro každou kopuli D je zobrazení $\mathbf{t} \mapsto d(D, C(\cdot; \mathbf{t}))$ spojitě.

(D-2) $d(C(\cdot; \boldsymbol{\theta}), C(\cdot; \mathbf{t})) = 0$ právě tehdy, když $\boldsymbol{\theta} = \mathbf{t}$.

(D-3) Pro každé $\boldsymbol{\theta} \in \text{int}(\Theta)$ existuje funkce $\boldsymbol{\delta}_\theta : \mathbb{R}^m \rightarrow \mathbb{R}^p$ taková, že každá její složka náleží do prostoru $L^2((0, 1)^m)$ a

$$\int_{(0,1)^m} |C(\mathbf{u}; \mathbf{t}) - C(\mathbf{u}; \boldsymbol{\theta}) - (\mathbf{t} - \boldsymbol{\theta})^\top \boldsymbol{\delta}_\theta(\mathbf{u})|^2 d\mathbf{u} = o(\|\mathbf{t} - \boldsymbol{\theta}\|^2).$$

(D-4) Zobrazení $\boldsymbol{\theta} \mapsto \boldsymbol{\delta}_\theta$ je pro každé $\boldsymbol{\theta} \in \text{int}(\Theta)$ spojitě podle kvadratického středu, tedy

$$\lim_{\mathbf{t} \rightarrow \boldsymbol{\theta}} \int_{(0,1)^m} \|\boldsymbol{\delta}_t(\mathbf{u}) - \boldsymbol{\delta}_\theta(\mathbf{u})\|^2 d\mathbf{u} = 0.$$

(D-5) Matice $\int_{(0,1)^m} \boldsymbol{\delta}_\theta(\mathbf{u}) \boldsymbol{\delta}_\theta^\top(\mathbf{u}) d\mathbf{u}$ je pro každé $\boldsymbol{\theta} \in \text{int}(\Theta)$ regulární.

Nejčastěji může být funkce $\boldsymbol{\delta}_\theta$ volena tak, že její i -tá složka je dána jako $(\boldsymbol{\delta}_\theta(\mathbf{u}))_i = \partial C(\mathbf{u}; \boldsymbol{\theta}) / \partial \theta_i$. Definujme dále funkci $\boldsymbol{\gamma}_\theta : \mathbb{R}^m \rightarrow \mathbb{R}^p$ předpisem

$$\boldsymbol{\gamma}_\theta(\mathbf{u}) = \left(\int_{(0,1)^m} \boldsymbol{\delta}_\theta(\mathbf{v}) \boldsymbol{\delta}_\theta^\top(\mathbf{v}) d\mathbf{v} \right)^{-1} \boldsymbol{\delta}_\theta(\mathbf{u}).$$

Nyní jsme připraveni zformulovat větu o asymptotické normalitě odhadu metodou nejmenší vzdálenosti, jejíž obecnější variantu předkládá Tsukahara (2005, Theorem 3).

Věta 12. *Nechť jsou splněny podmínky regularity (D-1) až (D-5). Potom pro $n \rightarrow \infty$ platí*

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_n^{MD} - T(C) \right) \xrightarrow{d} \mathcal{N}_p(\mathbf{0}, \mathbb{W}_0),$$

kde

$$\mathbb{W}_0 = \text{var} \left(\int_{(0,1)^m} \boldsymbol{\gamma}_{\boldsymbol{\theta}}(\mathbf{u}) \left(\mathbb{1}(\mathbf{U} \leq \mathbf{u}) - \sum_{j=1}^m C^{(j)}(\mathbf{u}; \boldsymbol{\theta}) \mathbb{1}(\mathbf{U}_j \leq \mathbf{u}_j) \right) d\mathbf{u} \right),$$

$C^{(j)}(\mathbf{u}; \boldsymbol{\theta}) = \partial C(\mathbf{u}; \boldsymbol{\theta}) / \partial u_j$ a \mathbf{U} je náhodný vektor z rozdělení s distribuční funkcí C .

Důkaz. Viz Tsukahara (2005, sekce 3.2.4.). □

Zde pouze podotkněme, že člen $T(C)$ vyjadřuje přesně objekt našeho zájmu, tedy hodnotu parametru $\boldsymbol{\theta}$, která minimalizuje $d(C, C_{\boldsymbol{\theta}})$.

I v tomto případě Tsukahara (2005, sekce 3.2.4.) uvádí konzistentní odhad $\widehat{\mathbb{W}}$ asymptotické rozptylové matice. Takovýto odhad je dán jako výběrový rozptyl

$$\widehat{\mathbb{W}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{W}_i - \overline{\mathbf{W}}_n) (\mathbf{W}_i - \overline{\mathbf{W}}_n)^{\top},$$

kde pro $i = 1, \dots, n$ je

$$\mathbf{W}_i = \int_{(0,1)^m} \boldsymbol{\gamma}_{\hat{\boldsymbol{\theta}}_n^{MD}}(\mathbf{u}) \left(\mathbb{1}(\widehat{\mathbf{U}}_i \leq \mathbf{u}) - \sum_{j=1}^m C^{(j)}(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{MD}) \mathbb{1}(\widehat{\mathbf{U}}_{ij} \leq \mathbf{u}_j) \right) d\mathbf{u}$$

a $\widehat{\mathbf{U}}_i$ jsou pseudopozorování definovaná v (2.5).

2.2.3 Momentová metoda

Další ze skupiny semiparametrických přístupů je momentová metoda, která aplikuje standardní postupy momentového odhadu na teorii kopulí. V anglicky psané literatuře jsou takové odhady nazývány „copula moment estimators“. Tuto metodu zkoumali Brahimí a Necir (2012), a tato sekce je shrnutím nejdůležitějších poznatků uvedených v této práci.

Předpokládejme znovu stejnou strukturu dat jako v podsekcí 2.2.2 s tím, že navíc předpokládáme, že rodina kopulí $\{C(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$ splňuje takzvanou podmínku uspořádání, tedy

$$\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \Rightarrow (\forall \mathbf{u} \in (0,1)^m : C(\mathbf{u}; \boldsymbol{\theta}_1) < (\text{nebo } >) C(\mathbf{u}; \boldsymbol{\theta}_2)). \quad (2.19)$$

Definujme dále k -tý moment kopule $C(\cdot; \boldsymbol{\theta})$ předpisem

$$M_k(\boldsymbol{\theta}) = \int_{(0,1)^m} (C(\mathbf{u}; \boldsymbol{\theta}))^k dC(\mathbf{u}; \boldsymbol{\theta})$$

pro $k = 1, 2, \dots$. Povšimněme si, že vzhledem k tomu, že pro každé $\mathbf{u} \in (0, 1)^m$ je $0 \leq C(\mathbf{u}; \boldsymbol{\theta}) \leq 1$, existuje konečný k -tý moment kopule pro libovolné k . Na základě k -tých momentů můžeme snadno definovat funkci $\mathbf{M} : \Theta \rightarrow \mathbb{R}^p$ předpisem $\mathbf{M}(\boldsymbol{\theta}) = (M_1(\boldsymbol{\theta}), \dots, M_p(\boldsymbol{\theta}))^\top$. Vzhledem k podmínce uspořádání je funkce \mathbf{M} prostá. Tedy existuje inverzní zobrazení k \mathbf{M} a momentový odhad $\hat{\boldsymbol{\theta}}_n^{CM}$ parametru $\boldsymbol{\theta}$ můžeme definovat jako

$$\hat{\boldsymbol{\theta}}_n^{CM} = \mathbf{M}^{-1} \left(\left(\widehat{M}_1(\boldsymbol{\theta}), \dots, \widehat{M}_p(\boldsymbol{\theta}) \right)^\top \right),$$

kde $\widehat{M}_k(\boldsymbol{\theta})$ značí výběrový odhad k -tého momentu. Nalezení momentového odhadu můžeme popsat v následujících krocích.

1. Spočteme hodnoty pseudopozorování $\widehat{\mathbf{U}}_i$, která jsou definovaná v (2.5).
2. Nalezneme pro $k = 1, \dots, p$ výběrové odhady momentů kopule

$$\widehat{M}_k = \frac{1}{n} \sum_{i=1}^n \left(C_n(\widehat{\mathbf{U}}_i) \right)^k,$$

kde C_n značí empirickou kopuli definovanou v definici 15

3. Momentový odhad $\hat{\boldsymbol{\theta}}_n^{CM}$ nalezneme jako $\boldsymbol{\theta}$ řešící soustavu rovnic

$$\begin{aligned} M_1(\boldsymbol{\theta}) &= \widehat{M}_1, \\ &\vdots \\ M_p(\boldsymbol{\theta}) &= \widehat{M}_p. \end{aligned}$$

Obdobně jako u předchozích metod představíme podmínky regularity nutné k platnosti asymptotických vlastností tohoto odhadu. Nejprve však označme $\mathbf{L}(\mathbf{u}; \boldsymbol{\theta}) = (L_1(\mathbf{u}; \boldsymbol{\theta}), \dots, L_p(\mathbf{u}; \boldsymbol{\theta}))^\top$, kde

$$L_k(\mathbf{u}; \boldsymbol{\theta}) = (C(\mathbf{u}; \boldsymbol{\theta}))^k - M_k(\boldsymbol{\theta}).$$

Podmínkami regularity pro odhad momentovou metodou budeme označovat následující tři požadavky na rodinu kopulí $\{C(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$ a parametrický prostor Θ .

- (M-1) Necht $\boldsymbol{\theta}_0$ je skutečnou hodnotou parametru $\boldsymbol{\theta}$. Předpokládejme pak, že platí

$$\int_{(0,1)^m} \mathbf{L}(\mathbf{u}; \boldsymbol{\theta}) dC(\mathbf{u}; \boldsymbol{\theta}_0) = \mathbf{0}_p \Leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

- (M-2) Funkce $\mathbf{L}(\cdot; \boldsymbol{\theta})$ je diferencovatelná vzhledem k $\boldsymbol{\theta}$ s Jakobiho maticí

$$\mathbb{D}_{\mathbf{L}}(\mathbf{u}; \boldsymbol{\theta}) = \left(\frac{\partial L_i(\mathbf{u}; \boldsymbol{\theta})}{\partial \theta_j} \right)_{i,j=1}^p.$$

$\mathbb{D}_{\mathbf{L}}(\mathbf{u}; \boldsymbol{\theta})$ je dále spojitá v obou svých argumentech a Frobeniova norma $\|\mathbb{D}_{\mathbf{L}}(\mathbf{u}; \boldsymbol{\theta})\|$ je shora omezená $dC(\cdot; \boldsymbol{\theta})$ integrovatelnou majorantou $h(\mathbf{u})$.

(M-3) Čtvercová matice \mathbb{B}_0 řádu p ve tvaru

$$\mathbb{B}_0 = \int_{(0,1)^m} \mathbb{D}_{\mathbf{L}}(\mathbf{u}; \boldsymbol{\theta}_0) dC(\mathbf{u}; \boldsymbol{\theta}_0)$$

je regulární.

Nyní jsme již připraveni formulovat větu o asymptotické normalitě odhadu momentovou metodou.

Věta 13. *Nechť je splněna podmínka uspořádání (2.19) a podmínky regularity (M-1) až (M-3). Potom momentové odhady $\hat{\boldsymbol{\theta}}_n^{CM}$ existují s pravděpodobností blížící se 1, jakmile $n \rightarrow \infty$. Navíc pro $n \rightarrow \infty$ platí*

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_n^{CM} - \boldsymbol{\theta}_0 \right) \xrightarrow{d} \mathcal{N}_p \left(\mathbf{0}, (\mathbb{B}_0)^{-1} \mathbb{Z}_0 \left((\mathbb{B}_0)^{-1} \right)^\top \right),$$

kde matice \mathbb{B}_0 je definována v podmínce regularity (M-3),

$$\mathbb{Z}_0 = \text{var} \left(\mathbf{L}(\mathbf{U}; \boldsymbol{\theta}_0) + \sum_{j=1}^m \int_{(0,1)^m} \mathbf{V}_j(\boldsymbol{\theta}_0) (\mathbb{1}(U_j \leq u_j) - u_j) dC(\mathbf{u}; \boldsymbol{\theta}_0) \right),$$

$$\mathbf{V}_j(\boldsymbol{\theta}) = (\partial(C(\mathbf{u}; \boldsymbol{\theta}))^1 / \partial u_j, \dots, \partial(C(\mathbf{u}; \boldsymbol{\theta}))^p / \partial u_j)^\top$$

a \mathbf{U} je náhodný vektor z rozdělení s distribuční funkcí $C(\cdot; \boldsymbol{\theta}_0)$.

Důkaz. Viz Brahimí a Necir (2012, Theorem 1). □

Abychom byli schopni tvrzení věty 13 aplikovat v praxi, je nutné uvést konzistentní odhad rozptylové matice $(\mathbb{B}_0)^{-1} \mathbb{Z}_0 \left((\mathbb{B}_0)^{-1} \right)^\top$. Brahimí a Necir (2012, Remark 1) navrhnou tento odhad jako čtvercovou matice $(\hat{\mathbb{B}})^{-1} \hat{\mathbb{Z}} \left((\hat{\mathbb{B}})^{-1} \right)^\top$ řádu p . Odhad $\hat{\mathbb{B}}$ je dán jako

$$\hat{\mathbb{B}} = \int_{(0,1)^m} \left(D_{\mathbf{L}}(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{CM}) \right) dC(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{CM}).$$

Odhad $\hat{\mathbb{Z}}$ je dán jako výběrový rozptyl

$$\hat{\mathbb{Z}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}}_n) (\mathbf{Z}_i - \bar{\mathbf{Z}}_n)^\top,$$

kde pro $i = 1, \dots, n$ je

$$\mathbf{Z}_i = \mathbf{L}(\hat{\mathbf{U}}_i; \hat{\boldsymbol{\theta}}_n^{CM}) + \sum_{j=1}^m \int_{(0,1)^m} \mathbf{V}_j(\hat{\boldsymbol{\theta}}_n^{CM}) (\mathbb{1}(\hat{U}_{ij} \leq u_j) - u_j) dC(\mathbf{u}; \hat{\boldsymbol{\theta}}_n^{CM})$$

a $\hat{\mathbf{U}}_i$ jsou pseudopozorování definovaná v (2.5).

2.2.4 Porovnání semiparametrických metod

Mnozí autoři se ve svých textech zabývali porovnáním mezi jednotlivými semiparametrickými metodami pomocí simulačních studií. Další častou metodou, kterou jsme se v předešlém textu nezabývali, je metoda založená na Kendallově τ definovaném v (2.18). Pokud definujeme navíc výběrové Kendallovo τ předpisem

$$\tau_{2,n} = -1 + \frac{4}{n(n-1)} \sum_{i \neq j} \mathbb{1}(Y_{i1} \leq Y_{j1}, Y_{i2} \leq Y_{j2}),$$

můžeme nalézt odhad parametru θ jako takové θ , které řeší rovnici $\tau(C, \theta) = \tau_{2,n}$. Takový odhad nazýváme odhad metodou inverzního τ a značíme $\hat{\theta}_n^{INV}$.

Tsukahara (2005, kapitola 4) se ve své simulační studii zaměřil na porovnání vychýlení a střední čtvercové chyby pro různé rodiny kopulí, různé hodnoty jejich parametrů a různé metody odhadu těchto parametrů. Byly uvažovány pouze dvourozměrné kopule s jednorozměrným parametrem. Odhad vychýlení na základě N náhodných výběrů je dán jako

$$\widehat{\text{Bias}} = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_i - \theta_0$$

a odhad střední čtvercové chyby jako

$$\widehat{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta_0)^2.$$

Porovnány byly čtyři rodiny kopulí, a to Claytonova, Gumbel-Hougaardova, Frankova a Plackettova. Zkoumané metody byly – metoda založená na pseudo-věrohodnosti, τ semiparametrický Z-odhad, metoda inverzního τ , metoda nejmenší vzdálenosti (Cramér-von Misesova vzdálenost) a metoda nejmenší vzdálenosti (Kolmogorov-Smirnovova vzdálenost). Autorovy výsledky ukazují, že ve většině případů má metoda založená na pseudověrohodnosti nejmenší čtvercovou chybu, zatímco nejmenší vychýlení je patrné u τ semiparametrického Z-odhadu. Naopak nejhorší vlastnosti vykazuje metoda nejmenší vzdálenosti založená na Kolmogorov-Smirnovově vzdálenosti. Obecně si lze také povšimnout, že jak vychýlení, tak střední čtvercová chyba odhadu rostou, jakmile skutečná hodnota parametru θ udává silnější míru pozitivní závislosti mezi veličinami.

Brahimi a Necir (2012, kapitola 4) ve své práci uvádí simulační studii, ve které je zkoumáno vychýlení a odmocnina střední čtvercové chyby u kopulí náležících do Gumbelovy rodiny. Zkoumané metody byly – metoda založená na pseudo-věrohodnosti, momentová metoda a takzvaná τ - ρ metoda, která je zobecněním metody inverzního τ pro dvourozměrný parametr θ . Více o poslední jmenované metodě uvádí ve své práci Brahimi a Necir (2012, kapitola 1). Výsledky této simulační studie ukazují, že metoda založená na pseudověrohodnosti a momentová metoda dosahují lepších výsledků (v obou hodnotících parametrech) než τ - ρ metoda. Výhodou momentové metody je v tomto případě výpočetní nenáročnost, neboť odhady parametrů v Gumbelově rodině pomocí momentové metody mohou být vypočteny explicitně. Poznamenejme však, že jiné rodiny kopulí zkoumány

nebyly, nelze tedy s dostatečnou jistotou doporučit momentovou metodu oproti jiným.

Závěrem této podsekce podotkněme, že kromě vzájemného porovnání semiparametrických metod se někteří z autorů zabývali také porovnání parametrického a semiparametrického přístupu. Simulační studie za tímto účelem provedli například Kim, Silvapulle a Silvapulle (2007). V první části svého článku porovnávají pro zvolené rodiny kopulí (Ali-Mikhail-Haqova, Claytonova, Frankova, Gumbelova, Joeva, Plackettova) odhady středních čtvercových chyb odhadů parametrů za podmínky, že u parametrických metod je parametrická podoba marginálů definována správně. Sledujeme-li, jaký je poměr takových odhadů středních čtvercových chyb mezi metodou inferenčních funkcí pro marginály a metodou založenou na pseudověrohodnosti, vidíme, že se pohybuje obvykle v rozmezí mezi 0,85 a 0,95. V druhé části této simulační studie autoři posuzovali mimo jiné odhad vychýlení, pokud je parametrická podoba marginálů definována nesprávně. Výsledky ukazují, že vychýlení pro parametrické metody je pro nesprávně specifikované marginály výrazně vyšší než pro metodu založenou na pseudověrohodnosti. Celkově proto autoři doporučují preferovat metodu založenou na pseudověrohodnosti oproti parametrickým metodám uvedeným v sekci 2.1, které jsou citlivé na správnou specifikaci marginálních rozdělání.

2.3 Testy dobré shody

V dosavadním průběhu této kapitoly jsme se zaměřovali pouze na odhadování parametru θ za předpokladu, že zkoumaná kopule C patří do rodiny $\{C(\cdot; \theta); \theta \in \Theta\}$. Druhou otázkou, kterou bychom se při modelování pomocí kopulí měli zabývat, je zda kopule C opravdu patří do rodiny $\{C(\cdot; \theta); \theta \in \Theta\}$. Uvedeme zde několik variant testů používaných pro testování nulové hypotézy $H_0 : C \in \{C(\cdot; \theta); \theta \in \Theta\}$.

Předpokládejme opět, že $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ je posloupnost nezávislých, stejně rozdělených náhodných vektorů s m složkami, tedy $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^\top$ pro každé $i \in \{1, \dots, n\}$. Dále necht C značí skutečnou kopuli příslušející náhodnému vektoru \mathbf{Y}_1 a $\{C(\cdot; \theta); \theta \in \Theta \subset \mathbb{R}^p\}$ je parametrická rodina kopulí, kterou chceme používat pro modelování dat, a necht \widehat{U}_i jsou pseudopozorování definovaná v (2.5).

2.3.1 Testy založené na empirické kopuli

Intuitivní postup pro testování dobré shody je založen na porovnání empirické kopule C_n s kopulí $C(\cdot; \widehat{\theta}_n)$, kde $\widehat{\theta}_n$ je odhad parametru θ za platnosti H_0 , tedy například odhad založený na pseudověrohodnosti. Základem pro takovéto testování je proces

$$\mathbb{C}_n = \sqrt{n} \left(C_n - C(\cdot; \widehat{\theta}_n) \right).$$

Jeho použití však například Fermanian (2005, kapitola 1) označuje vzhledem ke složitosti asymptotického rozdělení za nepraktické. Genest a Rémillard (2008) přesto zkoumali využití tohoto procesu při testování a navrhují jeho použití

v Cramér-von Misesově statistice a Kolomogorov-Smirnovově statistice. Tvar těchto statistik je postupně

$$S_n = \int_{(0,1)^m} (\mathbb{C}_n(\mathbf{u}))^2 dC_n(\mathbf{u})$$

a

$$T_n = \sup_{\mathbf{u} \in (0,1)^m} |\mathbb{C}_n(\mathbf{u})|.$$

Asymptotická rozdělení statistik S_n a T_n jsou závislá na limitním chování procesu \mathbb{C}_n . Genest a Rémillard (2008, Theorem 1) ukazují konvergenci tohoto procesu (za určitých podmínek regularity) a také konzistenci testů založených na statistikách S_n a T_n ve smyslu, že pokud nulová hypotéza neplatí, je s pravděpodobností blížící se 1, jakmile $n \rightarrow \infty$, zamítnuta. Nevýhodou těchto metod je, že asymptotické rozdělení statistik S_n a T_n za platnosti H_0 závisí na rodině $\{C(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$ a také na hodnotě neznámého parametru $\boldsymbol{\theta}$. Přibližné (dané asymptotickým rozdělením) p-hodnoty je pak možné získat za použití parametrického bootstrapu. Tento postup blíže popisují Genest, Rémillard a Beaudoin (2009, Appendix A).

2.3.2 Testy založené na Kendallově transformaci

Další z možných přístupů k testování dobré shody je založen na takzvané Kendallově transformaci. Genest, Quessy a Rémillard (2006, kapitola 1) navrhuji uvažovat zobrazení z \mathbb{R}^m do $[0,1]$ definované předpisem

$$\mathbf{Y} = (Y_1, \dots, Y_m)^\top \mapsto V = H(\mathbf{Y}) = C(F_1(Y_1), \dots, F_m(Y_m)),$$

kde H je sdružená distribuční funkce náhodného vektoru \mathbf{Y} a F_i je marginální distribuční funkce složky Y_i . Toto zobrazení bývá nazýváno Kendallova transformace. Genest a Rivest (1993, kapitola 3) ukazují, že pokud K značí (jednorozměrnou) distribuční funkci náhodné veličiny V , lze ukázat, že K může být neparametricky odhadnuta pomocí empirické distribuční funkce na základě transformovaných pseudopozorování ve tvaru $\hat{V}_1 = C_n(\hat{\mathbf{U}}_1), \dots, \hat{V}_n = C_n(\hat{\mathbf{U}}_n)$. Označme tuto empirickou distribuční funkci jako

$$K_n(v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\hat{V}_i \leq v) \quad \text{pro } v \in [0,1].$$

Genest, Rémillard a Beaudoin (2009, sekce 3.2.) uvádějí, že K_n je konzistentním odhadem distribuční funkce K . Za platnosti nulové hypotézy je sdružené rozdělení vektoru $\mathbf{U} = F_1(Y_1), \dots, F_m(Y_m)^\top$ dáno jako $C(\cdot, \boldsymbol{\theta})$ pro $\boldsymbol{\theta} \in \Theta$. Kendallova transformace $C(\mathbf{U}; \boldsymbol{\theta})$ má tedy distribuční funkci, kterou označíme jako $K_\boldsymbol{\theta}$.

Pomocí měření odlišnosti K_n a parametrického odhadu $K_{\hat{\boldsymbol{\theta}}_n}$ můžeme testovat nulovou hypotézu ve tvaru $H'_0 : K \in \{K_\boldsymbol{\theta}, \boldsymbol{\theta} \in \Theta\}$. Obecně může nastat situace, kdy ačkoliv kopule C nepatří do zkoumané parametrické rodiny $\{C(\cdot; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$, distribuční funkce K Kendallovy transformace již může patřit do $\{K_\boldsymbol{\theta}, \boldsymbol{\theta} \in \Theta\}$. Platí proto, že $H_0 \subset H'_0$. Pokud tedy zamítáme H'_0 , zamítáme také H_0 . Nezamítnutí H'_0 však neznamená, že by H_0 neměla být zamítnuta. Z tohoto důvodu nejsou testy založené na procesu

$$\mathbb{K}_n = \sqrt{n} (K_n - K_{\hat{\boldsymbol{\theta}}_n})$$

v obecném případě konzistentní. Genest, Quessy a Rémillard (2006, sekce 4.2.) nabízejí postupy pro výpočet p-hodnot i přes nekonzistenci takových testů. Podobně jako v případě testů založených na empirické kopuli jsou testové statistiky založeny na Cramér-von Misesově statistice a Kolmogorov-Smirnovově statistice. Zde jsou testové statistiky ve tvaru

$$S_n^{(K)} = \int_{(0,1)} (\mathbb{K}_n(u))^2 dK_{\hat{\theta}_n}(u)$$

a

$$T_n^{(K)} = \sup_{u \in (0,1)} |\mathbb{K}_n(u)|.$$

Asymptotická rozdělení těchto testových statistik jsou opět závislá na kopuli $C(\cdot; \theta)$ a samotném parametru θ . Přibližné p-hodnoty opět nejsou tabelovány a mohou být nalezeny pouze pomocí simulačních metod. Genest, Rémillard a Beaudoin (2009, Appendix B) uvádějí postup pro použití parametrického bootstrapu v tomto případě.

2.3.3 Testy založené na Rosenblattově transformaci

Dalším typem transformace, na které mohou být založeny testy dobré shody, je Rosenblattova transformace. Rosenblattovou transformací danou kopulí C rozumíme zobrazení $\mathcal{R} : (0,1)^m \rightarrow (0,1)^m$, které vektoru $\mathbf{u} \in (0,1)^m$ přiřadí vektor $\mathcal{R}(\mathbf{u}) = (e_1, \dots, e_m)^\top$, kde $e_1 = u_1$ a pro každé $i \in \{2, \dots, m\}$ je

$$e_i = \left(\frac{\partial^{i-1} C(u_1, \dots, u_i, 1, \dots, 1)}{\partial u_1 \dots \partial u_{i-1}} \right) / \left(\frac{\partial^{i-1} C(u_1, \dots, u_{i-1}, 1, \dots, 1)}{\partial u_1 \dots \partial u_{i-1}} \right).$$

Vlastností takové transformace je to, že \mathcal{U} má rozdělení s distribuční funkcí C právě tehdy, když rozdělení $\mathcal{R}(\mathcal{U})$ je m -rozměrná kopule nezávislosti

$$\Pi((e_1, \dots, e_m)^\top) = e_1 \cdot \dots \cdot e_m.$$

Na základě tohoto pozorování je nulová hypotéza $H_0 : \mathcal{U} \sim C \in \{C(\cdot; \theta), \theta \in \Theta\}$ ekvivalentní nulové hypotéze $H_0^* : \mathcal{R}_\theta(\mathcal{U}) \sim \Pi$ pro nějaké $\theta \in \Theta$, kde $\mathcal{R}_\theta(\cdot)$ značí Rosenblattovu transformaci danou kopulí $C(\cdot, \theta)$. K testu H_0^* můžeme využít, že transformovaná pseudopozorování $\widehat{\mathbf{R}}_1 = \mathcal{R}_{\hat{\theta}_n}(\widehat{\mathbf{U}}_1), \dots, \widehat{\mathbf{R}}_n = \mathcal{R}_{\hat{\theta}_n}(\widehat{\mathbf{U}}_n)$ mohou být za H_0^* interpretována jako výběr z kopule Π . Tato interpretace je však pouze přibližná, neboť tato transformovaná pseudopozorování nejsou vzájemně nezávislá. Empirická distribuční funkce

$$D_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\widehat{\mathbf{R}}_i \leq \mathbf{u})$$

by pak měla být v určitém smyslu blízka kopuli Π . Testování dobré shody na základě Rosenblattovy transformace mohou být proto založena na zkoumání této blízkosti. Genest, Rémillard a Beaudoin (2009, kapitola 4) nabízí dva typy testových statistik, obě založené na Cramér-von Misesově přístupu. Prvním typem testové statistiky je

$$S_n^{(C)} = n \int_{(0,1)^m} (D_n(\mathbf{u}) - \Pi(\mathbf{u}))^2 dD_n(\mathbf{u}) = \sum_{i=1}^n (D_n(\widehat{\mathbf{R}}_i) - \Pi(\widehat{\mathbf{R}}_i))^2,$$

kde druhá rovnost plyne z toho, že funkce D_n je po částech konstantní se skoky velikosti $1/n$ přesně v bodech $\widehat{\mathbf{R}}_1, \dots, \widehat{\mathbf{R}}_n$. Druhým typem testové statistiky pak je

$$S_n^{(B)} = n \int_{(0,1)^m} (D_n(\mathbf{u}) - \Pi(\mathbf{u}))^2 d\mathbf{u}.$$

Stejně jako u výše uvedených testů je asymptotické rozdělení těchto statistik závislé na kopuli $C(\cdot; \boldsymbol{\theta})$ a samotném parametru $\boldsymbol{\theta}$. I zde uvádějí Genest, Rémillard a Beaudoin (2009, Appendix D) postup pro použití parametrického bootstrapu.

3. Simulační studie

Cílem v této kapitole uvedené simulační studie je ověřit praktickou platnost teoretických výsledků uvedených v předešlém průběhu této práce. Konkrétně se zde zaměříme na vliv rozsahu výběru na platnost asymptotické teorie odhadu parametrů kopulí.

Pro tuto simulační studii je vybrána metoda založená na pseudo-věrohodnosti, neboť se jedná o nejuniverzálnější metodu. Výhody této metody oproti jiným jsou shrnuty v podsekcí 2.2.4. Simulační studie, které byly uskutečněny v rámci zde citovaných i dalších textů, se však primárně zaměřují na porovnání této metody s dalšími metodami pro odhadování parametrů kopulí. Rozsah výběru pak bývá volen jako dostatečně velký a hlavním předmětem zájmu jsou odhady vychýlení a střední čtvercové chyby tak, jak jsou uvedeny v podsekcí 2.2.4. My se zde zaměříme na to, zda je v závislosti na rozsahu výběru dodržováno pokrytí intervalu spolehlivosti pro jednorozměrný parametr θ na základě vygenerovaných hodnot z předem zvolené kopule.

Přesný postup při tvorbě této simulační studie je následující.

1. Vygenerujeme $N = 10\,000$ náhodných výběrů z rozdělení se známou sdruženou distribuční funkcí $C(\cdot, \theta)$. Uvažovány jsou dvourozměrné kopule s jednorozměrným parametrem, konkrétně
 - normální kopule s parametrem korelace $\theta \in \{0,2; 0,5; 0,8\}$,
 - Farlie-Gumbel-Morgensternova (dále FGM) rodina kopulí s parametrem $\theta \in \{-0,45; 0,45; 0,9\}$,
 - Claytonova rodina kopulí s parametrem $\theta \in \{0,5; 2; 8\}$,
 - Ali-Mikhail-Haqova rodina kopulí s parametrem $\theta \in \{-0,5; 0,4; 0,71\}$ (dále jako AMH),
 - Gumbel-Hougaardova rodina kopulí s parametrem $\theta \in \{1,25; 2,5; 5\}$,
 - Frankova rodina kopulí s parametrem $\theta \in \{1,87; 5,75; 18,2\}$.

Všechny použité kopule jsou uvedeny v sekci 1.5. Způsoby, jakými lze generovat z rozdělení založených na kopulích, nabízí Nelsen (2006, kapitola 3). Parametry kopulí jsou voleny tak, aby Kendallovo τ příslušné kopule odpovídalo přibližně hodnotám 0,2; 0,5 a 0,8. Pouze v případě AMH rodiny kopulí a FGM rodiny kopulí odpovídají parametry Kendallovu τ o hodnotách $-0,1; 0,1$ a $0,2$. Důvodem je struktura těchto rodin, která neumožňuje modelování silných závislostí. Pro všechny tyto kopule jsou generovány výběry s různými rozsahy. Ty jsou zvoleny jako $n \in \{5, 10, 20, 30, 50, 100, 300, 500\}$.

2. Pro každý výběr pomocí metody založené na pseudověrohodnosti nalezneme odhad $\hat{\theta}$ parametru θ a příslušný interval spolehlivosti založený na větě 11 a následném odhadu asymptotického rozptylu detailně uvedeném v příkladu 7. Ten je označen jako $\hat{\sigma}^2$ a uveden v (2.17).

Poté je interval spolehlivosti s asymptotickým pokrytím 95 % dán jako

$$\left(\hat{\theta} - u_{0.975} \frac{\hat{\sigma}}{\sqrt{n}}; \hat{\theta} + u_{0.975} \frac{\hat{\sigma}}{\sqrt{n}} \right),$$

kde $u_{0,975}$ značí 97,5% kvantil normovaného normálního rozdělení.

3. Spočteme pozorované pokrytí jako podíl počtu intervalů spolehlivosti, které pokryly skutečnou hodnotu parametru θ , a celkového počtu N vygenerovaných náhodných výběrů. V případě, že z numerických důvodů nejsme schopni získat všechny odhady, případně intervaly spolehlivosti, definujeme pozorované pokrytí jako podíl počtu intervalů spolehlivosti, které pokryly skutečnou hodnotu parametru θ , a celkového počtu skutečně nalezených intervalů spolehlivosti.

Získané pozorované pokrytí je uvedeno v tabulce 3.1. K situaci uvedené v bodu 3, kdy jsme vinou výpočetních problémů nebyli schopni nalézt interval spolehlivosti pro θ , obecně nastávalo pouze u Frankovy kopule v případě malých výběrů. Z 10 000 vygenerovaných výběrů rozsahu 5 z Frankovy kopule s parametrem $\theta = 18,2$ jsme získali pouze 6 474 intervalů spolehlivosti. V dalších čtyřech případech jsme získali mezi 9 a 10 tisíci intervalů spolehlivosti. Jednalo se o $\theta = 1,87$ a $\theta = 5,75$ u výběrů s rozsahem 5 a dále také o $\theta = 5,75$ a $\theta = 18,2$ u výběrů s rozsahem 10. Nicméně, byl-li rozsah výběru alespoň 20, u všech rodin jsme získali plný počet intervalů spolehlivosti, a výsledky tudíž nemohly být zkresleny.

V tabulce 3.1 můžeme vidět, že ve všech rodinách se s rostoucím rozsahem výběru blíží pozorované pokrytí tomu teoretickému, tedy 95 %. U Gumbel-Hougaardovy rodiny a Frankovy rodiny tomu tak je již pro relativně malé rozsahy výběru, přibližně od $n = 30$. Relativně rychlou konvergenci k teoretickému pokrytí můžeme pozorovat také u Claytonovy rodiny. U normálních kopulí a AMH rodiny vidíme postupný nárůst pozorovaného pokrytí s rostoucím rozsahem výběru. U obou rodin je při rozsahu výběru $n = 50$ pozorované pokrytí rovno přibližně 90 %, až při rozsahu výběru $n = 500$ se pohybuje okolo 94 %. Odlišným případem je pak FGM rodina, kde je naopak pro malé rozsahy výběru pozorované pokrytí 100 %, případně téměř 100 %. Až u výběrů rozsahu $n = 100$ a více pozorujeme pozorované pokrytí blízké tomu teoretickému.

Celkově můžeme říci, že pro výběry s rozsahem $n = 100$ a více je u všech zkoumaných rodin kopulí velmi dobře dodržována teoretická hladina.

Ve druhé části této simulační studie rozšíříme naši oblast zájmu také o kopule ve více dimenzích. Postup simulací a výpočtů zůstává stejný jako v první části, zaměříme se však také na pětirozměrné kopule, abychom viděli, zda také rozměr kopule má vliv na rychlost konvergence k asymptotickému rozdělení. V této části již využijeme pouze Archimédovské kopule, které umožňují modelování slabých i silných závislostí, tedy Claytonovu, Gumbel-Hougaardovu a Frankovu rodinu. Funkční předpisy m -rozměrných Archimédovských kopulí kopírují dvourozměrný případ. Pro generátor ϕ a $\mathbf{u} = (u_1, \dots, u_m)^\top$ je tato kopule dána jako

$$C(\mathbf{u}) = \phi^{[-1]}(\phi(u_1) + \dots + \phi(u_m)).$$

Zvolené parametry i rozsahy výběrů zůstávají stejné jako v první části simulační studie, ačkoliv praktické použití například výběru s rozsahem 5 na modelování závislosti 5 veličin nemůže být doporučeno.

Tabulka 3.2 uvádí pozorované pokrytí pro pětirozměrné kopule. Gumbel-Hougaardova rodina ukazuje stejně jako ve dvourozměrném případě pozorované pokrytí velmi blízké 95 %, a to již pro malé rozsahy výběru. U Claytonovy rodiny

Tabulka 3.1: Pozorované pokrytí 95% intervalů spolehlivosti založených na pseudověrohodnosti pro parametr θ vybraných dvourozměrných kopulí. Vyjádřeno v procentech.

Rozsah výběru	5	10	20	30	50	100	300	500
Normální kopule								
$\theta = 0,2$	59,8	71,8	84,5	87,8	91,2	93,2	94,5	94,6
$\theta = 0,5$	68,8	77,5	84,4	86,8	89,3	91,7	93,3	93,9
$\theta = 0,8$	84,4	87,9	89,8	90,6	91,5	92,6	93,5	94,2
FGM rodina								
$\theta = -0,45$	100,0	99,9	98,8	97,2	97,2	95,5	94,4	94,7
$\theta = 0,45$	100,0	99,7	98,4	97,2	96,6	95,3	94,2	94,8
$\theta = 0,9$	100,0	99,5	97,8	97,8	98,3	98,3	98,3	98,3
Claytonova rodina								
$\theta = 0,5$	69,5	86,1	91,8	93,5	94,5	94,8	94,4	94,7
$\theta = 2$	81,0	95,7	95,2	94,8	94,4	94,3	94,7	94,4
$\theta = 8$	56,0	90,4	93,0	93,5	94,3	94,9	95,9	95,1
AMH rodina								
$\theta = -0,5$	71,9	85,4	90,9	92,7	94,6	95,7	94,1	94,2
$\theta = 0,4$	93,5	76,3	83,2	87,5	90,1	92,8	94,1	94,4
$\theta = 0,71$	100,0	99,4	77,6	80,5	87,3	91,3	93,4	94,0
Gumbel-Hougaardova rodina								
$\theta = 1,25$	97,5	98,0	97,2	97,4	94,8	95,1	95,2	95,5
$\theta = 2,5$	89,7	95,9	95,4	95,2	94,9	95,1	94,9	95,1
$\theta = 5$	56,8	92,9	94,4	94,3	95,0	95,5	95,0	95,1
Frankova rodina								
$\theta = 1,87$	95,9	96,3	95,3	95,1	95,2	95,0	95,1	94,9
$\theta = 5,75$	97,6	97,1	96,2	95,9	95,5	95,2	95,0	95,0
$\theta = 18,2$	96,2	94,5	95,2	95,5	96,0	96,1	95,8	95,4

naopak vidíme, že ve většině měřených případů je pozorované pokrytí menší než požadovaných 95 %. Obzvláště tomu tak je u parametru $\theta = 0,5$ odpovídajícímu nízké míře závislosti, kde bylo pro všechny výběry s rozsahy do 30 pozorované pokrytí menší než 90 %. Výjimku v tomto ohledu tvoří Claytonova rodina s parametrem $\theta = 8$, kde pro velké rozsahy výběru (100 a více) vidíme pozorované pokrytí přibližně 97 %. Spolehlivost pokrytí pro Frankovu rodinu se v pětirozměrném případě ukazuje také jako velmi dobrá. Pro všechny hodnoty parametru je již od výběrů rozsahu 20 pozorovaná hladina velmi blízká 95 %. Obdobně jako u Claytonovy rodiny pozorujeme u velkých rozsahů pro největší hodnotu parametru ($\theta = 18,2$) pozorované pokrytí typicky mezi 97 a 98 %.

Také v pětirozměrném případě jsme nebyli z numerických důvodů schopni vždy určit interval spolehlivosti. Řádově stovky neurčitelných intervalů se vyskytly u všech tří rodin pro malé rozsahy výběru (do 20 včetně). Pouze u Frankovy rodiny s hodnotou parametru $\theta = 18,2$ jsme obdrželi pouze 9939 intervalů pro rozsah výběru 30 a 9989 intervalů pro rozsah výběru 50.

Tabulka 3.2: Pozorované pokrytí 95% intervalů spolehlivosti založených na pseudověrohodnosti pro parametr θ vybraných pětirozměrných kopulí. Vyjádřeno v procentech.

Rozsah výběru	5	10	20	30	50	100	300	500
Claytonova rodina								
$\theta = 0,5$	81,1	85,0	89,3	89,9	90,6	91,8	93,2	94,2
$\theta = 2$	92,1	92,1	91,8	91,7	92,3	91,9	93,0	93,1
$\theta = 8$	91,7	90,2	91,2	93,3	94,9	96,5	97,3	97,2
Gumbel-Hougaardova rodina								
$\theta = 1,25$	97,4	96,1	93,5	94,8	95,5	95,7	96,1	95,6
$\theta = 2,5$	95,7	95,9	95,6	95,5	95,2	95,3	95,3	95,1
$\theta = 5$	95,5	95,7	95,7	96,1	96,4	97,1	96,5	96,6
Frankova rodina								
$\theta = 1,87$	79,5	90,9	94,5	94,5	94,5	95,0	94,9	94,8
$\theta = 5,75$	95,0	96,5	96,6	96,4	96,1	96,2	95,5	95,2
$\theta = 18,2$	96,2	95,3	97,1	97,5	98,1	98,4	97,7	97,3

Poznámka. Tato simulační studie byla uskutečněna s využitím balíčku copula ve statistickém programu R (R Core Team, 2016). Kód, který byl využit pro provedení simulací, je uveden jako elektronická příloha této diplomové práce. Dalším zájemcům autor tento kód rád poskytne na vyžádání.

Závěr

V první části této diplomové práce jsme zavedli pojem kopule a uvedli některé základní a důležité vlastnosti kopulí. Součástí této sekce je také přehled nejpoužívanějších rodin kopulí a návod, podle kterého je možné graficky znázorňovat dvourozměrné kopule. Pro větší názornost byla většina obecných poznatků o kopulích formulována primárně právě pro dvourozměrné kopule. Sekce 1.6 pak přináší zobecnění pro vícerozměrné kopule spolu se stručnými komentáři, zda se vlastnosti dvourozměrných kopulí přenáší také do obecného vícerozměrného prostoru.

Hlavní kapitola této práce pak popisuje metody pro statistickou inferenci v modelech založených na kopulích. Nejprve jsme představili parametrické metody odhadu parametru kopule, a to metodu maximální věrohodnosti a metodu inferenčních funkcí pro marginály. Tyto metody předpokládají parametrickou strukturu jak pro kopuli, tak pro jednotlivá marginální rozdělení. S využitím teorie Z-odhadů jsme pro obě metody současně formulovali tvar asymptotického rozdělení odhadu parametru kopule a nabídli možnosti, jakými je možné počítat asymptotické rozptylové matice takových odhadů.

Dále jsme se zabývali semiparametrickými metodami pro odhad parametru kopule, které vyžadují parametrickou strukturu pro kopuli, avšak nikoliv pro marginální rozdělení. Nabídnuty byly tři metody, a to semiparametrické Z-odhady („rank approximate Z-estimators“), odhady metodou nejmenší vzdálenosti („minimum distance estimators“) a momentové odhady („copula moment estimators“). U všech tří metod jsme opět formulovali patřičná asymptotická rozdělení odhadů včetně všech potřebných předpokladů. Na základě výsledků několika simulačních studií v literatuře jsme porovnali tyto metody. Ukázalo se, že pravděpodobně nejspolehlivější volbou je odhad založený na pseudověrohodnosti, tedy speciální případ semiparametrického Z-odhadu, který ukazuje dobré vlastnosti jak při porovnání se zbylými semiparametrickými metodami, tak při porovnání s parametrickými metodami.

V závěru teoretické části této práce jsme se zaměřili na testy dobré shody, které slouží k testování hypotézy, zda zkoumaná kopule náleží do některé dané rodiny kopulí. Vybrány byly tři různé přístupy k testům dobré shody, a to testy založené na empirické kopuli, testy založené na Kendallově transformaci a testy založené na Rosenblattově transformaci. Všechny metody byly popsány a doplněny o odkazy na literaturu popisující praktickou implementaci těchto testů pomocí parametrického bootstrapu.

Vybrané teoretické poznatky byly poté ověřeny pomocí simulační studie, která zkoumala závislost pozorovaného pokrytí asymptotického intervalu spolehlivosti pro parametr kopule na rozsahu výběru. Pro tuto studii byla vzhledem k důvodům uvedeným výše vybrána metoda založená na pseudověrohodnosti. Ukázali jsme, že pro většinu zkoumaných rodin kopulí je pozorované pokrytí velmi blízké teoretickému pro výběry rozsahu 50 a více. Pro Frankovu a Gumbel-Hougaardovu rodinu se v tomto směru ukázaly dostatečně již výběry o rozsahu 30 pozorování.

Cílem práce bylo přehledně popsat používané metody statistické inference v modelech založených na kopulích. Tyto metody jsou založeny na dostupné literatuře, která je vždy (na začátku i v průběhu sekcí) řádně odkazována. Pro

větší přehlednost a srozumitelnost celého textu bylo veškeré značení sjednoceno. Teoretické poznatky jsou dále rozšířeny o vysvětlující komentáře, detailní popisy postupů a příklady tak, abychom čtenáři co nejvíce usnadnili porozumění této problematice. V úvodní kapitole jsou také provedeny některé z důkazů, které byly v literatuře vynechány. Simulační studie v závěru práce navíc přináší doporučení o potřebném rozsahu výběru pro spolehlivou funkčnost intervalů spolehlivosti pro parametr kopule založených na pseudověrohodnosti.

Možným rozšířením této diplomové práce by mohlo být popsání dalších metod statistické inference, prohloubení doposud známých poznatků o kopulích, případně aplikace prezentovaných metod na reálná data.

Seznam použité literatury

- ANDĚL, J. (2007). *Základy matematické statistiky*. Druhé opravené vydání. Matfyzpress, Praha. ISBN 80-7378-001-1.
- BRAHIMI, B. a NECIR, A. (2012). A semiparametric estimation of copula models based on the method of moments. *Statistical Methodology*, **9**(4), 467–477. ISSN 1572-3127.
- EFRON, B. (1987). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial Mathematics. ISBN 0898711797.
- EMBRECHTS, P., MCNEIL, A. a STRAUMANN, D. (1999). Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value At Risk And Beyond*. Cambridge University Press. ISBN 978-0521781800. 176–223.
- FERMANIAN, J. D. (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis*, **95**(1), 119–152. ISSN 0047-259X.
- GENEST, C. a RÉMILLARD, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'IHP Probabilités et statistiques*, **44**(6), 1096–1127. ISSN 0246-0203.
- GENEST, C., RÉMILLARD, B. a BEAUDOIN, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**(2), 199–213. ISSN 0167-6687.
- GENEST, C. a RIVEST, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, **88** (423), 1034–1043. ISSN 0162-1459.
- GENEST, C., GHOUDI, K. a RIVEST, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, pages 543–552.
- GENEST, C., QUÉSSY, J.-F. a RÉMILLARD, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, **33**(2), 337–366. ISSN 0303-6898.
- HOEFFDING, W. (1940). *Masstabinvariante Korrelationstheorie*. Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin 5.
- JOE, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis. ISBN 9780412073311.
- JUREČKOVÁ, J., SEN, P. K. a PICEK, J. (2013). *Methodology in robust and nonparametric statistics*. CRC Press. ISBN 978-1-4398-4069-6.

- KIM, G., SILVAPULLE, M. J. a SILVAPULLE, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, **51**(6), 2836 – 2850. ISSN 0167-9473.
- NELSEN, R. B. (2006). *An Introduction to Copulas*. Springer Series in Statistics. Springer. ISBN 9780387286594.
- ONKEN, A., GRÜNEWÄLDER, S., MUNK, M. H. J. a OBERMAYER, K. (2009). Analyzing short-term noise dependencies of spike-counts in macaque prefrontal cortex using copulas and the flashlight transformation. *PLOS Computational Biology*, **5**(11), 1–13. ISSN 1553-734X.
- R CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RÜSCHENDORF, L. (2009). On the distributional transform, sklar’s theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, **139** (11), 3921–3927. ISSN 0378-3758.
- SKLAR, M. (1959). *Fonctions de Répartition À N Dimensions Et Leurs Marges*. Université Paris 8.
- TSUKAHARA, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics*, **33**(3), 357–375. ISSN 0319-5724.
- VAN DER VAART, A. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN 9780521784504.
- WANG, C., CHANG, N.-B. a YEH, G.-T. (2009). Copula-based flood frequency (coff) analysis at the confluences of river systems. *Hydrological Processes*, **23** (10), 1471–1486. ISSN 1099-1085.
- XU, J. (1996). *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*. PhD thesis, University of British Columbia.