Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Petr Zahradník

Kopule a korelace

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Zdeněk Hlávka, Ph.D.
Studijní program: Obecná matematika

2007

Děkuji vedoucímu mé práce, panu Zdeňku Hlávkovi, za zajímavé téma, cenné rady a v neposlední řadě za trpělivost.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 21.5.2007 Petr Zahradník

Obsah

Název práce: Kopule a korelace
Autor: Petr Zahradník
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Mgr. Zdeněk Hlávka, Ph.D.
e-mail vedoucího práce: hlavka@karlin.mff.cuni.cz

Abstrakt: Hlavním cílem této práce je vyšetřit vztah různých korelačních koeficientů k parametrům kopulí pro různé rodiny kopulí. V první části zopakujeme definici a základní vlastnosti kopulí, v druhé části si shrneme dnes užívané „korelační koeficienty“. Nakonec se dostáváme k jednotlivým rodinám kopulí a speciálním vyšetřováním nalezených vztahů jejich parametrů k námi navrženým koeficientům.
Klíčová slova: kopule, korelace, odhady parametrů kopulí.

Title: Copulas and Correlation
Author: Petr Zahradník
Department: Department of Probability and Mathematical Statistics
Supervisor: Mgr. Zdeněk Hlávka, Ph.D.
Supervisor‘ s e-mail address: hlavka@karlin.mff.cuni.cz

Abstract: The main goal of the thesis is to investigate the relationship between the correlation coefficient (or measure of association) and copula parameters for several parametric copulas. In the first part of this work we revise the definition and basic properties of copulas. Further we summarize commonly used coefficients and finally reach the goal of the thesis.

Keywords: copula, correlation, measure of association, estimator.

Kapitola 1

Kopule

1.1 Úvod

V této práci se seznámíme s kopulemi a jejich užití při studiu závislosti. Práce předpokládá čtenáře sice minimálně, ale přesto erudovaného v matematické statistice. Pojmy, které jsou v bakalářském ročníku notoricky známé, už se nedefinují, základní tvrzení, používaná implicitně, se neopakují. Vše ostatní je vysvětleno podrobně.

1.2 Definice a seznámení

Kopule jsou funkce, které pojí jednorozměrné marginální distribuční funkce k mnohorozměrným distribučním funkcím. Byly poprvé představeny v roce 1959 A. Sklarem jako odpověď M. Fréchetovi týkající se vztahu mezi vícerozměrnými pravděpodobnostními distribučními funkcemi a jejich méněrozměrnými marginály. Blíže k historii viz [11]. Užití kopulí v článku [3], který pro ně našel dobré využití ve financích, znamenalo jejich bouřlivý rozvoj, který trvá dodnes a skýtá mnoho otevřených (a vzhledem k uplatnění pro mnohé matematiky lákavě lukrativních) témat.

Definice 1. Řekneme, že funkce $C : I^2 \to I$, kde I značí interval [0,1], je (dvojrozměrná) kopule, má-li následující vlastnosti:
(i) $C(0,t) = C(t,0) = 0$ a $C(1,t) = C(t,1) = t \quad \forall\, t \in I$
(ii) $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0 \quad \forall u_1,u_2,v_1,v_2 \in I$ taková, že $u_1 \leq u_2$ a $v_1 \leq v_2$.

Z definice okamžitě plyne, že C je neklesající v každé proměnné (položme u či v rovné nule) a spojitá (protože je dokonce stejnoměrně spojitá). Tento závěr nám plyne právě z monotonie a z trojúhelníkové nerovnosti. Mějme bez újmy na obecnosti $u_1 \leq u_2$ a $v_1 \leq v_2$:

$$
\begin{aligned}
|C(u_2, v_2) - C(u_1, v_1)| &\leq |C(u_2, v_2) - C(u_1, v_2)| + |C(u_1, v_2) - C(u_1, v_1)| \\
&\leq |C(u_2, 1) - C(u_1, 1)| + |C(1, v_2) - C(1, v_1)| \\
&= |u_2 - u_1| + |v_2 - v_1|.
\end{aligned}
$$

Všimněme si dalších základních vlastností, okamžitě plynoucích z definice:

Lemma 1.2.1. $\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v)\ \forall u, v \in I$.

Důkaz. Důkaz je elementární, první nerovnost: $C(1, 1) - C(1, v) - C(u, 1) + C(u, v) \geq 0$, odtud $C(u, v) \geq u + v - 1$ a kopule musí zároveň být nezáporná, druhá nerovnost: $C(u, v) \leq C(u, 1) = u$, pro v analogicky. □

Poznámka. Tato nerovnost je kopulovou verzí Fréchetovy–Hoeffdingovy nerovnosti vyjádřené pomocí distribučních funkcí. Mezím tedy budeme říkat Fréchetovy–Hoeffdingovy a značit je: $W(u, v) = \max(u + v - 1, 0)$, $M(u, v) = \min(u, v)$. Zmíníme zde ještě jednu charakteristickou kopuli, tzv. součinovou kopuli: $\Pi(u, v) = uv$.

Důležitost kopulí pro matematickou statistiku tkví v následujícím základním tvrzení:

Věta 1.2.2 (Sklar)**.** *Nechť H je dvojrozměrná distribuční funkce (d.f.) s marginálními d.f. F a G. Pak existuje kopule C taková, že $H(x, y) = C(F(x), G(y))$ pro všechna x a $y \in R$. Navíc, jsou-li F a G spojité, je C určena jednoznačně. Opačně, pro jakékoliv jednorozměrné d.f. F a G a jakoukoliv kopuli C, funkce H definovaná výše je dvojrozměrná d.f. s marginály F a G.*

Důkaz. Viz [9]. □

Od nyní dále mohou být všechna tvrzení nazývána důsledky. Poznamenejme, že kopule sama je dvojrozměrnou distribuční funkcí. Její marginály mají rovnoměrné rozdělení na I.

1.3 Charakterizace

Nyní buďte X a Y spojité náhodné veličiny se sdruženou d.f. H, marginálními d.f. po řadě F a G a kopulí C. Vztah veličin X a Y dává do souvislosti s význačnými kopulemi následující věta 1.3.1.

Věta 1.3.1. *(i) X a Y jsou nezávislé* $\Leftrightarrow C(u,v) = uv$;
(ii) Y je skoro jistě rostoucí funkcí X $\Leftrightarrow C(u,v) = \min(u,v)$;
(iii) Y je skoro jistě klesající funkcí X $\Leftrightarrow C(u,v) = \max(u+v-1,0)$.

Důkaz. (i) je zřejmé, zbytek viz například [8]. □

Věta 1.3.2. *Nechť X a Y jsou spojité náhodné veličiny s kopulí C_{XY}. Nechť α a β jsou ryze monotónní transformace na $RanX$ a $RanY$, kde $RanF$ rozumíme obraz zobrazení F. Potom platí:*

(i) Jsou-li obě transformace rostoucí, pak

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

(ii) Je-li α rostoucí a β klesající, pak

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

(iii) Je-li α klesající a β rostoucí, pak

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

(iv) Jsou li obě transformace klesající, pak

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

Důkaz. Snadný důkaz jen naznačíme pro nám později užitečný případ (i). Nechť F_1, G_1, F_2, G_2 značí distribuční funkce po řadě $X, Y, \alpha(X), \beta(Y)$. Máme rovnosti:

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

Díky faktu, že $RanF_2 = RanG_2 = I$ a spojitosti X a Y máme výsledek. Zbytek je podobnou rutinou, viz [9]. □

Máme-li kopuli, potom, jako důsledek Sklarovy věty, automaticky máme dvojrozměrnou distribuční funkci s jakýmikoliv marginálními d.f. si přejeme. To je jistě užitečné například v modelování či simulacích. Navíc, v duchu předchozí věty 1.3.2, neparametrická podstata závislosti mezi dvěma náhodnými veličinami je vyjádřena kopulí (ta je totiž invariantní na striktně rostoucí transformace). Studium neparametrické závislosti je tedy studium vlastností kopulí. Dejme si proto pár význačných typů (rodin) kopulí do zásoby.

1.4 Některé rodiny kopulí

Nejprve zmiňme jednu obecnou skupinu kopulí, kam patří mnoho rodin, Archimédovy kopule. K tomuto potřebujeme pojem pseudoinverzní funkce (pseudoinverze): Nechť je φ spojitá, ryze klesající funkce z I do $[0,\infty]$ taková, že $\varphi(1) = 0$. Pseudoinverzí funkce φ rozumíme funkci $\varphi^{[-1]}$ s definičním oborem $[0,\infty]$ danou předpisem

$$\varphi^{[-1]}(t) = \varphi^{-1}(t) \text{ pro } \forall t \in [0, \varphi(0)] \text{ a}$$
$$\varphi^{[-1]}(t) = 0 \text{ jinde.}$$

Definice 2 (Archimedes)**.** Nechť je φ spojitá, ryze klesající funkce z I do $[0,\infty]$ taková, že $\varphi(1) = 0$ a $\varphi^{[-1]}$ její pseudoinverze. Kopuli, kterou lze zapsat ve tvaru:

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

nazveme Archimedovou a funkci φ jejím generátorem.

Poznámka. V případě, že $\varphi(0) = \infty$, je samozřejmě $\varphi^{[-1]} = \varphi^{-1}$. Archimédovy kopule jsou symetrické a asociativní. Toto a další vlastnosti viz [9].

Představme si některé Archimedovy rodiny.

Definice 3 (GH)**.** Gumbelovou–Hougaardovou rodinou zveme kopule dané vztahem:

$$C_\theta(u,v) = e^{-[(-\log u)^\theta + (-\log v)^\theta]^{\frac{1}{\theta}}}; \quad \theta \in [1,\infty].$$

Definice 4 (AMH)**.** Do rodiny Ali–Mikhail–Haq patří kopule dané vztahem:

$$C_\theta(u,v) = \frac{uv}{1 - \theta(1-u)(1-v)}; \quad \theta \in [-1,1).$$

Definice 5 (Clayton)**.** Claytonovou rodinou zveme kopule zapsané následujícím vzorcem:

$$C_\theta(u,v) = \max[(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0]; \quad \theta \in [-1, \infty) \setminus \{0\}.$$

Definice 6 (Frank)**.** Kopule Frankovy rodiny jsou definované vztahem:

$$C_\theta(u,v) = -\frac{1}{\delta} \log \left[1 + \frac{(e^{-\delta u} - 1)(e^{-\delta v} - 1)}{e^{-\delta} - 1}\right]; \quad \theta \neq 0.$$

Dále se nám bude hodit rodina patřící mezi obecnou skupinu kopulí s kvadratickými částmi, tzn. kopule obecně dané vzorcem $C(u,v) = a(v)u^2 + b(v)u + c(v)$ pro kvadratické části v proměnné u.

Definice 7 (FGM)**.** Vzorec pro rodinu zvanou Farlie–Gumbel–Morgenstern je:

$$C_\theta(u,v) = uv + \theta uv(1-u)(1-v); \quad \theta \in [-1, 1].$$

Zmiňme si také některé rodiny s dvěma parametry:

Definice 8 (Fréchet)**.** Fréchetovou rodinu zveme kopule splňující:

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

$\alpha, \beta \in [0,1], \alpha + \beta \leq 1.$

Definice 9 (MO)**.** Marshallova–Olkinova rodina je jméno pro kopule se vzorcem:

$$C_{\alpha,\beta}(u,v) = \min(u^{(1-\alpha)}v, uv^{(1-\beta)}); \quad \alpha, \beta \in [0,1].$$

Kapitola 2

Korelace

Úkolem korelační analýzy je odhalit vztahy mezi náhodnými veličinami. V této kapitole si uvedeme několik přístupů, jak tyto vztahy a závislost měřit. Nejobvykleji se zkoumají vztahy lineární.

2.1 Pearsonův korelační koeficient

Jedná se o nejčastěji užívaný korelační koeficient a je definován, jsou-li X a Y náhodné veličiny s konečnými druhými momenty a kladnými rozptyly, takto:

Definice 10.

$$\varrho = cov(X,Y)/\sqrt{(varX)(varY)}.$$

Lze snadno ukázat, že $\varrho(X,Y) = \varrho(Y,X)$ a že při lineárních transformacích X a Y korelační koeficient nanejvýš změní znaménko. Platí, viz [1], že $-1 \leq \varrho \leq 1$ a také, že $\varrho = 1$ (resp. $\varrho = -1$), právě když $Y = a + bX$ s pravděpodobností 1, kde $b > 0$ (resp. $b < 0$).

Definice 11. Mějme náhodný výběr $(X_1, Y_1), \ldots, (X_n, Y_n)$ z nějakého dvojrozměrného rozdělení. Označme $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$; $S_X^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$; $S_{XY} = \frac{1}{n-1}\sum_{k=1}^{n}((X_k - \overline{X})(Y_k - \overline{Y}))$; výběrovou verzí Pearsonova koeficientu nazveme

$$r = S_{XY}/\sqrt{S_X^2 S_Y^2}.$$

U následujících koeficientů se budeme snažit slovnímu spojení „korelační koeficient“ úzkostlivě vyhýbat, v moderní zahraniční literatuře se toto spojení totiž používá pouze u vztahu lineárního, zatímco pro jiné monotónní vztahy se užívá termín „measure of association“.

2.2 Kendallovo τ

Hodnocení výběrového korelačního koeficientu je vázáno na splnění předpokladu, že výběr pochází z normálního rozdělení (viz. [1]). Tento předpoklad však dost často bývá silně porušen, leckdy nejde hodnoty uvedených náhodných veličin stanovit přesně a někdy dokonce jejich hodnoty neznáme vůbec a máme k dispozici jen jejich pořadí.

Definice 12. Mějme dvě pozorování $(x_i, y_i), (x_j, y_j)$ spojitého náhodného vektoru (X, Y). Řekneme, že tato pozorování jsou konkordantní, platí-li $x_i < x_j$ a zároveň $y_i < y_j$ nebo $x_i > x_j$ a zároveň $y_i > y_j$. V jiném případě řekneme, že jsou diskordantní. Připomeňme, že ve spojitém rozdělení je pravděpodobnost shody $(x_i = x_j)$ rovna nule.

Kendallovo τ je potom definováno jako pravděpodobnost konkordance mínus pravděpodobnost diskordance, tedy:

Definice 13. Buďte $(X_1, Y_1), (X_2, Y_2)$ nezávislé a stejně rozdělené náhodné vektory, každý se sdruženou d.f. H. Potom je Kendallovo τ definováno následovně:

$$\tau = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0).$$

Definice 14. Mějme $(x_1, y_1), ..., (x_n, y_n)$ náhodný výběr pozorování vektoru (X, Y) o rozsahu n. Máme $\binom{n}{2}$ různých párů (x_i, y_i) a (x_j, y_j) pozorování a každý pár je buď konkordantní, nebo diskordantní. Nechť c značí počet konkordantních a d počet diskordantních párů. Výběrová verze Kendallova tau je potom

$$t = (c - d)/(c + d) = (c - d)/\tbinom{n}{2}.$$

Poznámka. Známe-li jen pořadí veličin, výběrovou verzi Kendallova koeficientu můžeme zapsat i bez pojmu konkordance, zápis však vypadá trochu ďábelsky:

$$t = \frac{1}{n(n-1)} \sum \sum_{i \neq j} (\operatorname{sgn}(r_i - r_j) \operatorname{sgn}(q_i - q_j));$$

kde r_i a q_i jsou pořadí veličin X_i a Y_i.

2.3 Spearmanovo ρ

Spearmanovo ρ je, stejně jako Kendallovo τ, založeno na konkordanci a diskordanci. Abychom získali jeho hodnotu, mějme $(X_1, Y_1), (X_2, Y_2)$ a (X_3, Y_3) tři nezávislé náhodné vektory, každý se sdruženou distribuční funkcí H, jejíž marginálními d.f. jsou F a G, a kopulí C. Spearmanovo ρ je úměrné pravděpodobnosti konkordance minus pravděpodobnost diskordance dvou vektorů (X_1, Y_1) (X_2, Y_3), to znamená dvojice vektorů se stejnými marginálními d.f., ale jeden se sdruženou distribuční funkcí H, zatímco složky druhého vektoru jsou nezávislé:

Definice 15. S předpoklady výše zmíněnými definujme:

$$\rho = 3P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0].$$

Je jasné, že v definici by nám posloužil vektor (X_3, Y_2) namísto vektoru (X_2, Y_3) stejně dobře. Zarazila-li čtenáře v definici konstanta 3 (proč ne třeba 3,99?), je to dobře, ještě si jí všimneme (viz větu 3.0.7).

Výběrová verze Spearmanova ρ:

Definice 16. Výběrová verze se definuje jako Pearsonův výběrový koeficient aplikovaný na pořadí složek dvojrozměrného vektoru, přesněji, mějme $(X_1, Y_1), \ldots, (X_n, Y_n); n \in N$ výběr ze spojitého dvojrozměrného rozdělení. Seřaďme výběr podle první složky. Nechť $R_1, \ldots, R_n$ jsou pořadí veličin $Y_1, \ldots, Y_n$. Potom lze, jak se můžeme dočíst v [1], výběrová verze Spearmanova ρ upravit na následující přehledný tvar:

$$r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (i - R_i)^2.$$

2.4 Blomqvistovo β

Uveďme si zde na závěr třetí neparametrický koeficient, opět založený na konkordanci.

Definice 17. Buď (X, Y) náhodný vektor.

$$\beta = P[(X - \tilde{x})(Y - \tilde{y}) > 0] - P[(X - \tilde{x})(Y - \tilde{y}) < 0].$$

kde $\tilde{x}$ a $\tilde{y}$ znamenají po řadě mediány náhodných veličin X a Y.

Výběrovou verzi β zavedeme intuitivně:

Definice 18. Mějme $(X_1, Y_1), \ldots, (X_n, Y_n); n \in N$ výběr z dvojrozměrného rozdělení se sdruženou distribuční funkcí F a spojitými marginálními d.f. Mějme dvojrozměrnou kontingenční tabulku, kde dělícími body jsou výběrové mediány jednotlivých marginálních rozdělení. Buď n_1 počet napozorovaných dvojic, které patří do prvního či třetího kvadrantu, n_2 buď počet dvojic z kvadrantu druhého a čtvrtého, $n_1 + n_2 = n$. Potom výběrová verze Blomqvistova β je:

$$b = \frac{n_1 - n_2}{n}.$$

Výhody tohoto koeficientu oproti předchozím koeficientům jsou ryze praktické. Zaprvé, jak uvidíme v lemmatu 2.4.1, hodnotu β lze explicitně spočíst, kdykoliv máme explicitně vyjádřenou kopuli, což rozhodně není pravidlem u předchozích koeficientů τ a ρ (viz [6]). Zadruhé je tu nízká výpočtová náročnost jeho odhadu. Definice t či r_S napovídají, že implementace na spočtení budou mít časovou složitost $O(n^2)$, zatímco časová náročnost spočtení odhadu b Blomqvistova β, u kterého je třeba spočíst medián jednorozměrných marginálních d.f., poroste s n pouze lineárně.

Lemma 2.4.1. *Buďte X a Y spojité náhodné veličiny se sdruženou d.f. H a s marginálními d.f. po řadě F a G a kopulí C. Potom máme $F(\tilde{x}) = G(\tilde{y}) = 1/2$ a $\beta = 4C(\frac{1}{2}, \frac{1}{2}) - 1$.*

Důkaz.

$$\begin{aligned}
\beta &= 2P((X - \tilde{x})(Y - \tilde{y}) > 0) - 1 \\
&= 2(P[X > \tilde{x}, Y > \tilde{y}] + P[X < \tilde{x}, Y < \tilde{y}]) - 1 \\
&= 2(H(\tilde{x}, \tilde{y}) + [1 - F(\tilde{x}) - G(\tilde{y}) + H(\tilde{x}, \tilde{y}]) - 1 \\
&= 4H(\tilde{x}, \tilde{y}) - 1 = 4C(F(\tilde{x}), G(\tilde{y})) - 1 \\
&= 4C(\frac{1}{2}, \frac{1}{2}) - 1.
\end{aligned}$$

□

Kapitola 3

Závislost

Ukažme si nejprve, v duchu předchozí úvahy o vhodnosti kopulí při zkoumání neparametrické závislosti, že u dvou náhodných veličin nám znalost jejich kopule ještě nestačí k určení Pearsonova korelačního koeficientu ϱ.

Příklad 3.0.2. Mějme náhodné veličiny X a Y, X měj rovnoměrné rozdělení na intervalu $[0,1]$ a nechť $Y = X$. Jak okamžitě plyne z definice, $\varrho_{XY} = 1$. Transformujme náhodné veličiny: $\alpha(X) = X$, $\beta(Y) = Y^2$. Transformace jsou rostoucí, tedy z věty 1.3.2 nám plyne:

$$C_{X,Y^2} = C_{X,Y} = C_{X,X} = M.$$

Ovšem $\varrho_{\alpha(X)\beta(Y)}$ je rozhodně různé od jedné:

$$cov(X,Y^2) = EXY^2 - EXEY^2 = EX^3 - EXEX^2 = 1/4 - 1/6 = 1/12,$$
$$varX = EX^2 - (EX)^2 = 1/12 \text{ a také: } varY^2 = EX^4 - (EX^2)^2 = 4/45.$$

Odtud dostáváme výsledek:

$$\varrho_{\alpha(X)\beta(Y)} = \frac{\sqrt{15}}{4}.$$

Poznámka. Je nám tedy jasné, že při snaze určit Pearsonův korelační koeficient bez marginálních rozdělení neuspějeme. Lze ukázat i rostoucí transformace, které korelační koeficient sníží o dost více, tento příklad je vskutku základní. Ukažme si ale, podrobněji v [10], že kopule i tak při zkoumání Pearsonova korelačního koeficientu mohou mít podstatnou úlohu:

Buď (X, Y) dvojrozměrný náhodný vektor s kopulí C a marginálními d.f. F a G. Potom kovariance X a Y lze vyjádřit takto:

$$cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(C[F(x),G(x)] - \Pi[F(x),G(x)])dxdy$$

a vzhledem k Fréchetovým–Hoeffdingovým mezím máme dále tvrzení: Mějme fixní marginální d.f. F a G (opravdu *fixní*, srovnejme s příkladem 3.0.2). ϱ_C vždy leží mezi koeficienty odpovídajícími rozdělením s mezními kopulemi W a M:

$$\varrho_W \leq \varrho_C \leq \varrho_M$$

a zavedeme-li na množině všech kopulí částečné uspořádání (množina kopulí je tzv. ČUM), pak je ϱ (jakožto funkce kopule) neklesající. ČUM se zavede intuitivně: $C_1 \prec C_2$, jestliže $C_1(u,v) \leq C_2(u,v)$ $\forall u, v \in I$.

Je na čase zdůraznit, že Kendallův či Spearmanův koeficient problémy při ryze monotónních transformacích nedělají, pravděpodobnost, že $X_1 \geq X_2$ je totiž při rostoucí transformaci α rovna pravděpodobnosti, že $\alpha(X_1) \geq \alpha(X_2)$. To nám napovídá, že takové koeficienty budou hrát ve vztazích ke kopulím veledůležitou roli. Povaha koeficientů jako je Kendallův, to jest, že, nepřesně řečeno, nezávisí na rozdělení, jim dává přízvisko „neparametrické“.

Abychom přesně ukázali roli, kterou kopule hrají při studiu neparametrických koeficientů, zaveďme si tzv. funkci konkordance (shody) Q, která je rozdílem pravděpodobností konkordance a diskordance mezi dvěma spojitými náhodnými vektory (X_1, Y_1) a (X_2, Y_2) s obecně rozdílnými sdruženými d.f. H_1 a H_2, ale se stejnými marginály F a G. Ukažme, že takováto funkce závisí na rozdělení vektorů (X_1, Y_1) a (X_2, Y_2) jen skrz jejich kopule:

Věta 3.0.3. *Nechť jsou (X_1, Y_1) a (X_2, Y_2) nezávislé spojité náhodné vektory se sdruženými d.f. po řadě H_1 a H_2, s marginály F (od X_1 a X_2) a G (od Y_1 a Y_2). Nechť C_1 a C_2 značí kopule po řadě (X_1, Y_1) a (X_2, Y_2), tzn. $H_1(x, y) = C_1(F(x), G(y))$ a $H_2(x, y) = C_2(F(x), G(y))$. Nechť Q značí rozdíl pravděpodobností konkordance a diskordance (X_1, Y_1) a (X_2, Y_2), to jest, nechť*

$$Q = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Potom

$$Q = Q(C_1, C_2) = 4\iint_{I^2} C_2(u,v)dC_1(u,v) - 1.$$

Důkaz. Podrobně si zde ukážeme důkaz, který je k nalezení v hrubší formě v [9]. Jelikož jsou náhodné vektory spojité, je pravděpodobnost shodného pozorování nula, $1-P[(X_1-X_2)(Y_1-Y_2)>0]=P[(X_1-X_2)(Y_1-Y_2)<0]$ a tedy

$$Q = 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1.$$

Ovšem $P[(X_1-X_2)(Y_1-Y_2)>0]=P[X_1>X_2,Y_1>Y_2]+P[X_1<X_2,Y_1<Y_2]$ a na spočítání těchto pravděpodobností můžeme zkusit podmiňovat a využít integrování přes sdruženou d.f. jednoho z vektorů, například (X_1, Y_1). Máme:

$$\begin{aligned}
P[X_1 > X_2, Y_1 > Y_2] &= P[X_2 < X_1, Y_2 < Y_1] \\
&= \iint_{R^2} P[X_2 < X_1, Y_2 < Y_1 | X_1 = x, Y_1 = y] dH_1(x, y) \\
&= \iint_{R^2} P[X_2 < x, Y_2 < y] dH_1(x, y) \\
&= \iint_{R^2} P[X_2 < x, Y_2 < y] dC_1(F(x), G(y)) \\
&= \iint_{R^2} C_2(F(x), G(y)) dC_1(F(x), G(y))
\end{aligned}$$

a substitucí $u = F(x)$, $v = G(y)$ dostaneme:

$$P[X_1 > X_2, Y_1 > Y_2] = \int\int_{I^2} C_2(u, v) dC_1(u, v).$$

Stejným způsobem,

$$\begin{aligned}
&P[X_1 < X_2, Y_1 < Y_2] = \\
&= P[X_2 < x, Y_2 < y | X_1 = x, Y_1 = y] \\
&= \iint_{R^2} P[X_2 > x, Y_2 > y] dC_1(F(x), G(y)) \\
&= \iint_{R^2} [1 - F(x) - G(y) + C_2(F(x), G(y))] dC_1(F(x), G(y)) \\
&= \iint_{I^2} [1 - u - v + C_2(u, v)] dC_1(u, v).
\end{aligned}$$

C_1 je sdružená distribuční funkce (U, V) rovnoměrně rozdělených náhodných veličin na $(0, 1)$, zřejmě tedy $E(U) = E(V) = 1/2$ a proto:

$$P[X_1 < X_2, Y_1 < Y_2] = 1 - \frac{1}{2} - \frac{1}{2} + \iint_{I^2} C_2(u,v)dC_1(u,v)$$

$$\text{a tedy } P[(X_1 - X_2)(Y_1 - Y_2) > 0] = 2\iint_{I^2} C_2(u,v)dC_1(u,v)$$

a tvrzení plyne dosazením. □

Poznámka. Funkce Q hraje důležitou roli, shrňme si její vlastnosti: Je symetrická ve svých proměnných a, vzpomeneme-li na ČUM, je neklesající v každém svém argumentu.

Jako okamžitý důsledek spojení definice Kendallova τ (viz definici 13) a výsledků předchozí věty 3.0.3 máme následující tvrzení.

Věta 3.0.4. *Buďte X a Y spojité náhodné veličiny, jejichž kopule je C. Potom dostáváme následující návod k počítání Kendallova τ pro X a Y:*

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

Tento integrál lze interpretovat jako střední hodnotu funkce $C(U,V)$ rovnoměrně rozdělených náhodných veličin U a V, jejichž sdružená distribuční funkce je C, tj. $\tau_{X,Y} = 4E(C(U,V)) - 1$.

Důkaz. Zřejmý. □

Příklad 3.0.5. Spočtěme si, co dává funkce Q pro dvojice základních kopulí M, W a Π. Uvědomme si nejprve, $d\Pi(u,v) = dudv$ a že nosičem M je pouze hlavní diagonála $u = v$ v I^2. Toto zjištění je zřejmé, když si uvědomíme, že míra indukovaná kopulí M každého otevřeného obdélníku, který leží celý nad nebo pod touto hlavní diagonálou, je nula. Stejně tak nosič W je druhá diagonála, $v = 1 - u$. Takže pokud máme funkci g definovanou a integrovatelnou na I^2, platí:

$$\iint_{I^2} g(u,v)dM(u,v) = \int_0^1 g(u,u)du; \iint_{I^2} g(u,v)dW(u,v) = \int_0^1 g(u,1-u)du.$$

Proto:

$$Q(M,M) = 4\iint_{I^2} \min(u,v)dM(u,v) - 1 = 4\int_0^1 udu - 1 = 1;$$

$$Q(M,\Pi) = 4\iint_{I^2} uvdM(u,v) - 1 = 4\int_0^1 u^2du - 1 = 1/3;$$

$$Q(M,W) = 4\iint_{I^2} \max(u+v-1,0)dM(u,v) - 1 = 4\int_{1/2}^{1}(2u-1)du - 1 = 0;$$
$$Q(W,\Pi) = 4\iint_{I^2} uvdW(u,v) - 1 = 4\int_0^1 u(1-u)du - 1 = -1/3;$$
$$Q(W,W) = 4\iint_{I^2} \max(u+v-1,0)dW(u,v) - 1 = 4\int_0^1 0du - 1 = -1;$$
$$Q(\Pi,\Pi) = 4\iint_{I^2} uvd\Pi(u,v) - 1 = 4\int_0^1\int_0^1 uvdudv - 1 = 0.$$

Buď nyní C libovolná kopule. Protože funkce konkordance je rozdílem dvou pravděpodobností, je jasné, že $Q(C,C) \in [-1,1]$ a jako důsledek ČUM a hodnot funkce Q pro význačné kopule M, W a Π v tomto příkladě plyne, že:

$$Q(C,M) \in [0,1], Q(C,W) \in [-1,0] \text{ a } Q(C,\Pi) \in [-1/3,1/3].$$

Pro počítání Kendallova τ tedy obvykle musíme spočíst leckdy netriviální dvojný integrál. Pro Archimedovské kopule je situace poněkud jednodušší:

Lemma 3.0.6. *Buďte X a Y náhodné veličiny s Archimédovskou kopulí C generovanou funkcí ϕ z Φ. Kendallovo τ pro X a Y je potom dáno vzorcem:*

$$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}$$

Důkaz. Lze nalézt v [7]. □

Poznámka. Jak R.B.Nelsen v [9] píše, pro velkou třídu kopulí , jež splňují podmínku integrovatelnosti $(\partial C/\partial u).(\partial C/\partial v)$ na I^2, se leckdy také hodí k počítání vyjádření Kendallova τ takto:

$$\tau_C = 1 - 4\iint_{I^2} \frac{\partial C(u,v)}{\partial u}\frac{\partial C(u,v)}{\partial v}dudv$$

Přistupme k dalšímu koeficientu– Spearmanovu ρ (viz definici 15). Připomeňme si, že sdružená d.f. (X_1,Y_1) je $H(x,y)$, zatímco u (X_2,Y_3) je to $F(x)G(y)$. Proto kopule (X_2,Y_3) je Π a užijeme-li větu 3.0.3 a symetrii Q, okamžitě máme tvrzení:

Věta 3.0.7. *Buďte X a Y spojité náhodné veličiny, jejichž kopule je C. Potom:*

$$\begin{aligned}\rho_{XY} &= 3Q(C,\Pi) = 12\iint_{I^2} uv dC(u,v) - 3\\ &= 3Q(\Pi,C) = 12\iint_{I^2} C(u,v) dudv - 3.\end{aligned}$$

Důkaz. Prosté dosazení. □

Tady si uvědomíme, že číslo tři v definici 15 Spearmanova ρ je tedy jen normalizační konstanta, jelikož již víme, že $Q(C,\Pi) \in [-1/3, 1/3]$ a zdá se nám nemravné, aby se koeficient, vyjadřující závislost dvou veličin, pohyboval pouze v tomto rozmezí. Radši máme interval $[-1,1]$. Všimněme si také, že definice Spearmanova ρ a jeho výběrové verze jsou korektní. Kdybychom si ušetřili práci s osvětou a i Spearmanovo ρ pro X a Y definovali tak, že to je Pearsonův korelační koeficient aplikovaný na $U = F(X)$ a $V = G(Y)$, máme $E(U) = E(V) = 1/2, VarU = VarV = 1/12$, sdružená d.f. U a V je C a tedy:

$$\rho = \frac{E(UV) - 1/4}{1/12} = 12E(UV) - 3 = 12\iint_{I^2} uvdC - 3.$$

Poznámka. V [4] nalezneme důkaz vztahu pro počítání Spearmanova ρ: Nechť X a Y jsou spojité náhodné veličiny s kopulí C. Potom:

$$\rho_{XY} = 3 - 6\iint_{I^2}\left[u\frac{\partial C}{\partial u}(u,v) + v\frac{\partial C}{\partial u}(u,v)\right] dudv.$$

3.1 Jednoparametrické rodiny kopulí a korelace

Jako úvod si zde uvedeme jeden nový výsledek z roku 2007, který je vyvrcholením snahy (viz [9] kapitola 5.1.3) o nalezení vztahu mezi ρ a τ a dá se nalézt v [4]. Hledáme podíl Spearmanova ρ a Kendallova τ v rozděleních s jednoparametrickou kopulí.

Věta 3.1.1. *Nechť $C(\theta, u, v)$ je rodina kopulí s jedním reálným parametrem, který patří do intervalu, jehož je nula vnitřním bodem a nechť $C(0, u, v) =$*

uv. Nechť ρ_θ a τ_θ značí Spearmanovo ρ a Kendallovo τ pro kopuli $C(\theta, u, v)$. Jestliže (a) $\partial C/\partial\theta, \partial/\partial\theta[u(\partial C/\partial u) + v(\partial C/\partial v)]$ a $\partial/\partial\theta[(\partial C/\partial u)(\partial C/\partial v)]$ jsou spojité na $J \times I$ pro nějaký interval J centrovaný v nule a dále platí (b) $\iint_{I^2} \partial C/\partial\theta(0, u, v)dudv \neq 0$, potom:

$$\lim_{\theta \to 0} \frac{\rho_\theta}{\tau_\theta} = \frac{3}{2}$$

Vyzdvihněme, že pro $\theta \to 0$ se kopule C_θ blíží k součinové, tedy „k nezávislosti“. Počítejme nyní naše zavedené neparametrické koeficienty přímo z parametrů jednotlivých rodin kopulí.

Lemma 3.1.2 (FGM). *Nechť je C_θ členem rodiny FGM. C_θ je absolutně spojitá distribuční funkce, tedy*

$$dC_\theta(u, v) = \frac{\partial^2 C_\theta(u, v)}{\partial u \partial v} = [1 + \theta(1 - 2u)(1 - 2v)]dudv,$$

odkud máme, po chvíli jednoduchého, ale pracného počítání (nebo přizvání PC ke spolupráci), že: $\int\int_{I^2} C_\theta(u, v)dC_\theta(u, v) = \iint_{I^2}(uv + \theta uv(1 - u)(1 - v))(1 + \theta(1 - 2u)(1 - 2v))dudv = \frac{1}{4} + \frac{\theta}{18}$, $\iint_{I^2} uv + \theta uv(1 - u)(1 - v)dudv = 1/4 + \theta/36$; výpočet β je triviální dosazení. Platí tedy:

$$\tau_\theta = \frac{2\theta}{9}; \quad \rho_\theta = \frac{\theta}{3}; \quad \beta_\theta = \frac{\theta}{4}.$$

Je zřejmé, že jednotlivé podíly koeficientů v tomto jednoduchém případě nezávisí na θ a podíl ρ a τ je konstantně $3/2$. Vzhledem k tomu, že např. $\tau_\theta \in [-2/9, 2/9]$, je jasné, jak zdůrazňuje Joe v [6], že možnosti modelování s touto kopulí jsou omezené.

Lemma 3.1.3 (AMH). *Spočítejme nyní naše neparametrické koeficienty pro rodinu AMH.*

$$\int_0^1 \log\left[\frac{(1 - \theta(1 - t))}{t}\right] / \frac{\frac{\theta - 1}{t}}{1 - \theta(1 - t)} dt = -\frac{\theta + (1 - \theta)^2 \log(1 - \theta)}{6\theta^2};$$

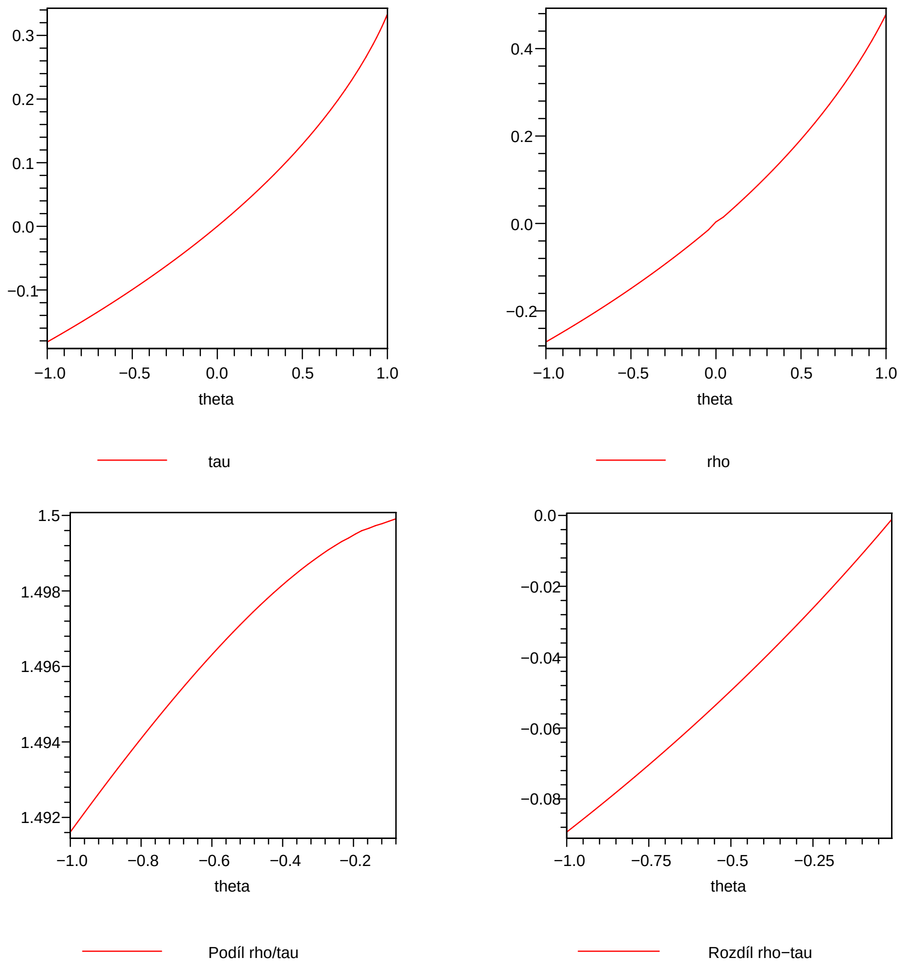
tudíž dostaneme výsledek pro τ:

$$\tau_\theta = \frac{3\theta - 2}{3\theta} - \frac{2(1 - \theta^2}{3\theta^2} \log(1 - \theta),$$

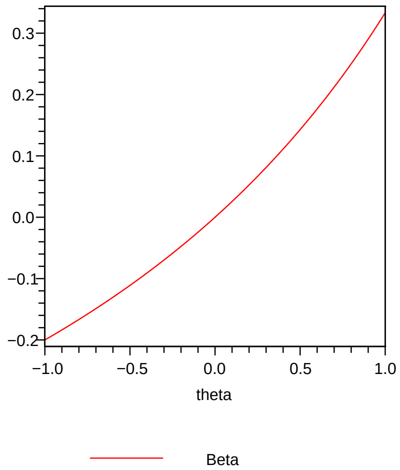
ρ získáme ze základního vztahu v tvrzení 3.0.7,

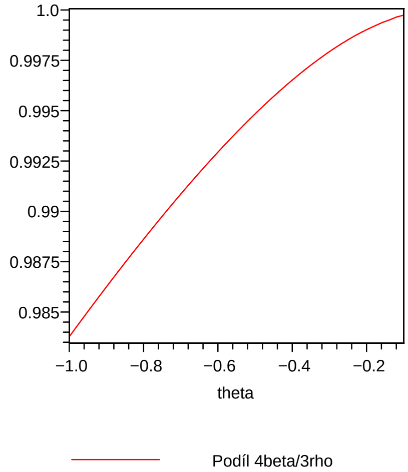
$$\rho_\theta = \frac{12(1+\theta)}{\theta^2}\int_1^{1-\theta}\frac{\log t}{1-t} - \frac{24(1-\theta)}{\theta^2}\log(1-\theta) - \frac{3(\theta+12)}{\theta},$$

a β_θ nám opět dělá pouze minimum práce: $\beta_\theta = \frac{\theta}{4-\theta}$. Grafické znázornění:



S nadšením tedy potvrzujeme, že tvrzení věty 3.1.1 je na našem grafu „Podíl rho/tau“ na obrázku výrazné. Grafy podílu a rozdílu koeficientů τ a ρ jsou vykresleny jen na pro záporné hodnoty parametru θ z důvodu, že v nule výraz není definovaný, náhodné veličiny se zde blíží nezávislosti. Samozřejmě, že graf pro kladné hodnoty θ by měl vypovídající schopnost identickou. Zapojme do studia i parametr β:

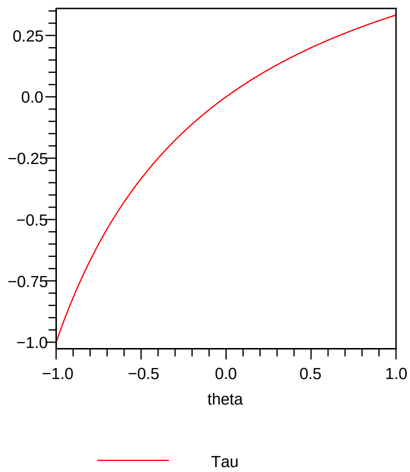


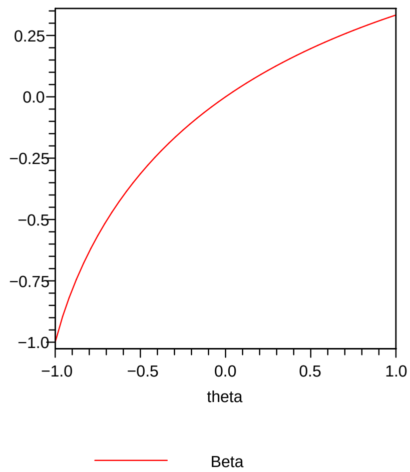


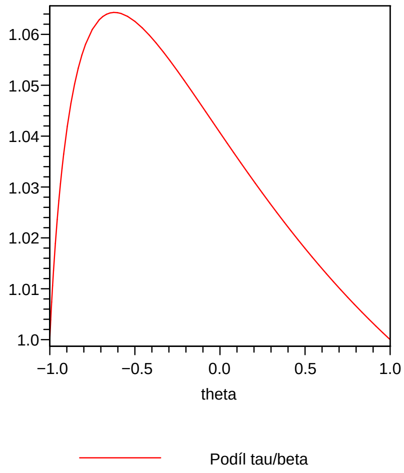
Zde vidíme, že pro θ blízká nule lze směle používat na místo τ či ρ koeficient β (vynásobený příslušnou konstantou). To není náhoda, stejně jako u Frankovy rodiny (viz lemma 3.1.6) je teoretický podklad této varianty vysvětlen pomocí Taylorových rozvojů v [9].

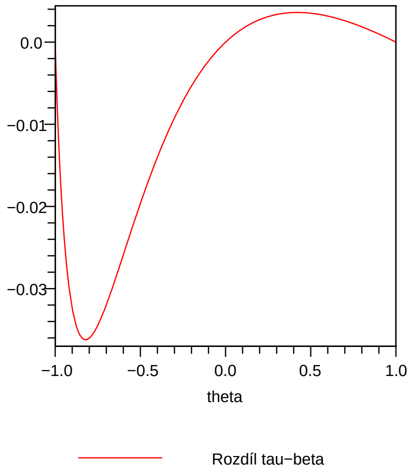
Lemma 3.1.4 (Clayton)**.** *Claytonova rodina je, jak víme, Archimédova, čili zjistíme-li její generátor, už se nám počítá lehce. Generátorem jest funkce* $\varphi(t) = \theta^{-1}(t^{-\theta} - 1)$, *tedy* $\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{t^{\theta+1}-t}{\theta}$ *pro* $\theta \neq 0$ *a* $\frac{\varphi_0(t)}{\varphi'_0(t)} = t \log t$, *spolu s dosazením do vzorce pro* β *tedy máme:*

$$\tau_\theta = \frac{\theta}{\theta+2}, \quad \beta_\theta = 4(2.2^\theta - 1)^{-1/\theta} - 1.$$





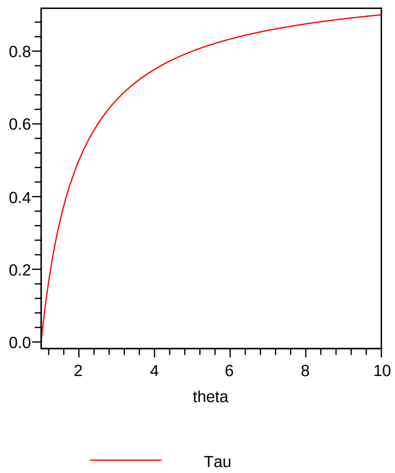


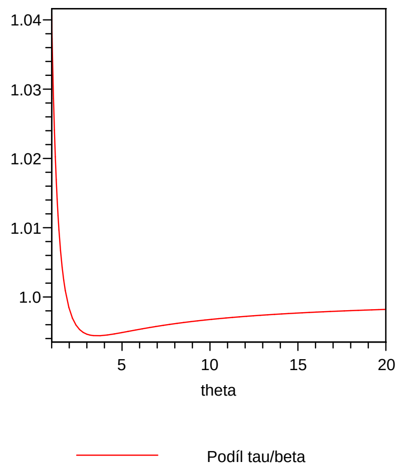


Lemma 3.1.5 (GH)**.** *Generátorem Gumbelovy–Hougaardovy rodiny je funkce* $(-\log t)^\theta$, *máme tedy, pro* $\theta \geq 1$, $\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{t \log t}{\theta}$, $C(\frac{1}{2}, \frac{1}{2}) = \exp(-2^{1/\theta} \log 2)$ *a proto:*

$$\tau_\theta = \frac{\theta - 1}{\theta}, \qquad \beta_\theta = 4 \exp(-2^{1/\theta} \log 2) - 1$$

a dostáváme následující grafy, z kterých je vidět, že τ *lze velmi dobře nahradit koeficientem* β*:*





Všimněme si, že koeficient τ nabývá pouze kladných hodnot, což tuto rodinu činí specifickou.

Lemma 3.1.6 (Frank)**.** *Zkoumejme Frankovu rodinu. Jak se lze dočíst v [5], počítání dopadne takto:*

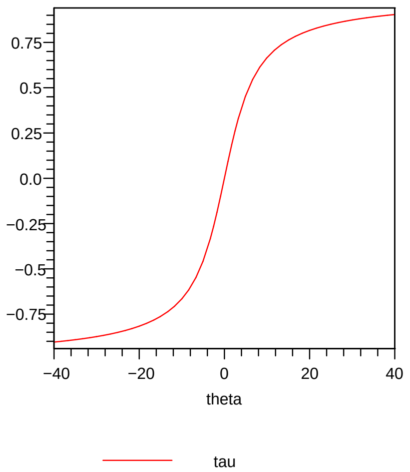
$$\tau_\theta = 1 - \frac{4}{\theta}[1 - D_1(\theta)],$$

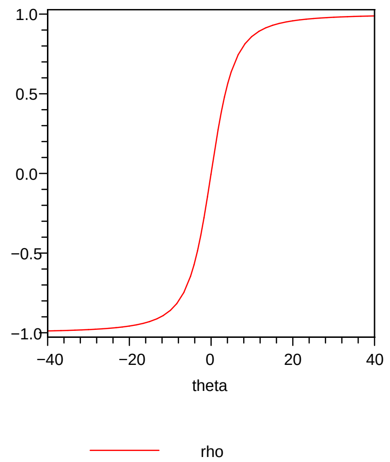
$$\rho_\theta = 1 - \frac{12}{\theta}[D_1(\theta) - D_2(\theta)],$$

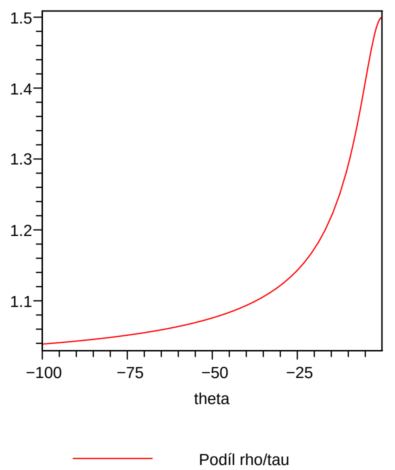
kde $\forall$ *n přirozená:*

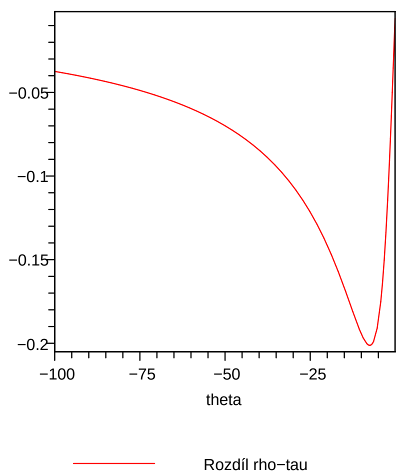
$$D_n(x) = \frac{n}{x^n}\int_0^x \frac{t^n}{e^t - 1}dt$$

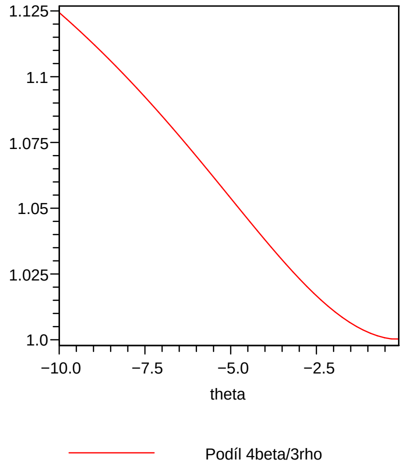
je tzv. „Debyeho funkce“.

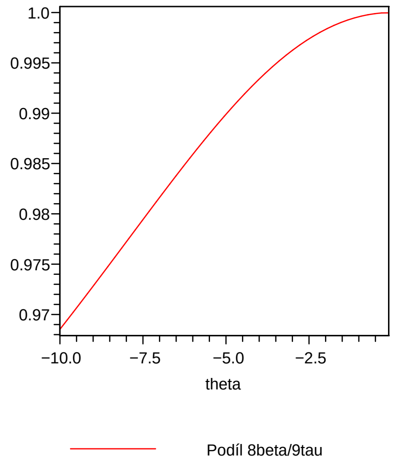












Naše pozorování jsou znovu ve shodě s teoretickými očekáváními. Grafy podílu a rozdílu jsou zde opět vykresleny pro θ jdoucí k nule zleva, kde se naše kopule blíží nezávislosti. Kdybychom v nule naše křivky uměle vyhladili, graf podílu (resp. rozdílu) by vypadal jako graf „pěkné“ sudé (resp. liché) funkce. V [5] lze najít i zajímavou simulační studii týkající se této rodiny.

Abychom nezůstali jen u jednoparametrických rodin, které mají široké využití, zmíníme tu některé speciální výsledky i u dvouparametrických rodin, které jsou přece jen na okraji zájmu.

3.2 Dvouparametrické rodiny a korelace

R.B. Nelsen v [9] předkládá následující dokázaná tvrzení:

Lemma 3.2.1 (Fréchet). *Pro náhodný vektor (X, Y) s kopulí patřící do Fréchetovy rodiny platí:*

$$\tau_{\alpha,\beta} = \frac{(\alpha - \beta)(\alpha + \beta + 2)}{3},$$

$$\rho_{\alpha,\beta} = \alpha - \beta.$$

Lemma 3.2.2 (MO). *Pro náhodný vektor (X, Y) s kopulí patřící do Marshallovy–Olkinovy rodiny platí:*

$$\tau_{\alpha,\beta} = \frac{\alpha\beta}{\alpha - \alpha\beta + \beta},$$

$$\rho_{\alpha,\beta} = \frac{3\alpha\beta}{2\alpha - \alpha\beta + 2\beta}.$$

3.3 Závěrečné zamyšlení

Z našich počítání a grafů se okamžitě skýtá otázka jistého obrácení problému. Přesné metody již nejsou náplní této práce, ale přesto zhruba nakoukneme do jistě zajímavých sfér a pro autora i čtenáře je to návod k budoucímu zamyšlení. Nabízím následující postup: Mějme sadu n pozorování náhodného vektoru (X_1, X_2) s neznámou kopulí C, je libovolné, zda je to plodnost odrůdy jahod hnojené ledkem respektive jahodkou, či vývoj cen akcií na pražské a londýnské burze. Chceme zjistit, do které parametrické rodiny kopulí patří C, popř. testovat hypotézu:

$$H_0 : C \in C_\theta : \theta \in \Theta \text{ oproti } H_1 : C \notin C_\theta : \theta \in \Theta.$$

Při rozsáhlých datech jde již s velkou přesností zkoumat empirická marginální rozdělení a neparametrické vztahy mezi jednotlivými výběry. Empiricky odhadněme například Kendallovo τ (pomocí t, viz definici 14, samozřejmě, že se dají užít i další neparametrické koeficienty, jejich výběr bude záležet na okolnostech). Jak vidíme, u mnoha rodin lze jejich parametr (empirický parametr označme $\hat{\theta}$) z Kendallova τ explicitně spočíst (v této práci viz GH, Clayton), u jiných, jako je například rodina AMH, bude určitě třeba nějaké numerické postupy, které nám inverzní počítání parametru kopule z Kendallova τ umožní. Vezměme jednu dvojici dat a užijme naše empirické distribuční funkce $\hat{F}_i$. Přichází chvíle, kdy se nám bude opět hodit Sklarova věta. Mělo by platit:

$$\hat{F}(x, y) = C_{\hat{\theta}}(\hat{F}_1(x), \hat{F}_2(y)),$$

kde $C_{\hat{\theta}}$ je jedna z námi navrhovaných Archimédovských kopulí. Nyní již zbývá pomocí numerických či grafických postupů rozhodnout, do které rodiny tedy naše hledaná kopule opravdu patří. Toto důležité rozhodování je zatím nevyřešená záležitost, kterou se zabývají v dnešních dnech mnohé články po světě využívající různé simulace a postupy, které by vystačily na samostatnou práci. Narazíme tu na metody zvané Akaike's Information Criterion (AIC), Goodness–Of–Fit (GOF) atp. Právě GOF se dostal do centra zájmu a bylo k němu navrženo několik přístupů. Tato problematika je již těžší. Čtenáře proto odkáži na [2], práci, která v úvodu přístupy shrnuje, uvádí patřičnou literaturu a sama dochází k zajímavým závěrům. Určitě by se pro výběr správné rodiny kopulí leccos dalo vytěžit i z chování vztahů jednotlivých koeficientů, z grafů to vypadá, že by tyto vztahy měly být charakteristické. Tento přístup však ještě čeká na detailní promyšlení.

Literatura

[1] Anděl,J.: *Základy matematické statistiky*, Matfyzpress, Praha, 2005.

[2] Berg D., Bakken H.: *A copula goodness-of-fit approach based on the conditional probability integral transformation, verze preprint na http://www.danielberg.no/* 2007.

[3] Embrechts, P., A. McNeil, and D. Straumann: *Correlation and dependence in risk management: Properties and pitfalls; Proceedings XXXth International ASTIN Colloquium* (1999) 227-250.

[4] Fredricks G. A., Nelsen R. B.: *On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables, Journal of Statistical Planning and Inference* **137** (2007) 2143–2150.

[5] Genest, C.: *Frank's family of bivariate distributions, Biometrika* **74** (1987) 759–764.

[6] Joe, H.: *Multivariate models and dependence concepts. London: Chapman and Hall/CRC* Monographs on Statistics and Applied Probability **73** 1997.

[7] Genest C., MacKay J.: *The Joy of Copulas: Bivariate distributions with uniform marginals, Amer. Statist.* **40 (4)** (1986) 280–283.

[8] Mikusinksi P., Sherwood H., Taylor M.D.: *The Fréchet bounds revisited, Real Analysis in Exchange* **17** (1991-92) 759–764.

[9] Nelsen, R.B.: *An Introduction to Copulas*, Springer Verlag, New York, 1999.

[10] Pfeifer D., Nešlehová J.: *Modeling and generating dependent risk processes for IRM and DFA*, ASTIN COLLOQUIUM BERLIN, 2003.

[11] Sklar, A.: *Random variables, distribution functions and copulas- a personal look backward and forward, Distributions with Fixed Marginals and Related Topics*, Institute od Mathematical Statistics, Hayward, 1996.