VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Rasa Balčiūnaitė

Jungčių taikymas transporto priemonių valdytojų civilinės atsakomybės privalomojo draudimo žalų modeliavimui

Magistro darbas

VILNIUS 2006

Matematinės analizės katedra

Darbo vadovas **prof. V. Paulauskas** ____________________
(parašas)

Darbas apgintas ______*2006 m. birželio mėn. 1 d.*______
Gynimo posėdžio protokolo Nr. ____________________

Registravimo Nr. __________________
2006-05-20 ______________________

Turinys

Reziumė

In this Master work the concept of copulas as a tool for modelling relationships among multivariate outcomes is introduced. A copula is a function that links univariate margins to their multivariate distribution. Copulas were introduced in 1959. The literature on the statistical properties and application of copulas has been developing rapidly in recent years.
In this Master work basic properties of copulas are described, then several families of copulas and relationships to measures of dependences. Later procedure for selecting the parametric family of Archimedean copulas is illustrated by using Lithuanian Motor Third Party Liability insurance data losses and expenses. For these data it is shown how to fit copulas according to nonparametric procedure which was proposed by Genest and Rivest [1].

Šio darbo tema yra jungčių (angl. *copula*) panaudojimas ryšiams tarp daugiamačių dydžių modeliuoti. Jungtis yra funkcija, kuri sujungia kelių atsitiktinių dydžių marginalinius skirstinius į bendrą daugiamatę funkciją. Jungties sąvoka pirmą kartą statistikoje įvesta 1959 m.
Šiame darbe aprašomos pagrindinės jungčių savybės, keletas jungčių šeimų, išskiriant atskirą šeimą - Archimedo jungtis, taip pat priklausomumo matai tarp atsitiktinių dydžių. Vėliau tinkamos jungties pritaikymo turimam duomenų rinkiniui procedūra iliustruojama nagrinėjant transporto priemonių valdytojų civilinės atsakomybės privalomojo draudimo žalų ir išlaidų žaloms administruoti duomenis.

1 Įvadas

Jungtis (angl. *copula*) yra funkcija, kuri sujungia marginalines pasiskirstymo funkcijas į daugiamatę pasiskirstymo funkciją. Žodis "copula" yra lotyniškas daiktavardis, kuris reiškia "sąsaja", "nuoroda", "ryšys". Jis naudojamas gramatikoje ir logikoje norint aprašyti tą teiginio dalį, kuri sieja veiksnį ir tarinį. Žodį "jungtis" (angl. *copula*) matematine ir statistine prasme pirmą kartą panaudojo Sklaras (Sklar, 1959) jo vardu pavadintoje teoremoje, apibūdinančioje funkcijas, kurios susieja vienmates pasiskirstymo funkcijas į jų bendras daugiamates pasiskirstymo funkcijas. Tačiau pati funkcija pradėta minėti jau anksčiau Wassily Hoeffding'o (Hoeffding), Frechet'o (Frechet), Dall'Aglio ir kituose darbuose. Jungčių pagalba galima konstruoti dvimačių pasiskirstymų šeimas ir analizuoti priklausomumo ryšius tarp atsitiktinių dydžių.

1997 m. jungtys pradėtos taikyti draudimo bei finansų srityse priklausomų dydžių modeliavime. Dažnai aktualus uždavinys yra kelių rūšių draudimo bendrų nuostolių įvertinimas, tinkamas žalų įvertinimas, kai žalos išskaidomos į atskiras rūšis, bendro žalų ir išlaidų žaloms administruoti skirstinio nusakymas; taip pat bendros rizikos įvertinimas gyvybės draudimo sutartyje, kai sutartimi apdraudžiami du asmenys. Visais čia išvardintais atvejais dydžiai yra priklausomi ir neteisingos prielaidos apie jų priklausomumą bei modelio parinkimas gali duoti klaidingus rezultatus vėliau skaičiuojant rizikos matus. Todėl šiems atvejams nagrinėti yra tinkamos jungtys, kurios leidžia modeliuoti priklausomus dydžius.

Šio darbo tikslas yra jungčių pagalba surasti bendrą dviejų priklausomų dydžių pasiskirstymo funkciją, naudojant transporto priemonių valdytojų civilinės atsakomybės privalomojo draudimo žalų duomenis.

Antrame skyriuje įvedama jungties sąvoka, aprašomos pagrindinės jų savybės. Trečiame ir ketvirtame skyriuose aptariamos jungčių šeimos, išskiriant atskirą - Archimedo jungčių klasę. Penktame skyriuje nusakomi priklausomumo tarp atsitiktinių dydžių matai.

Šeštame skyriuje nagrinėjamos jungties pritaikymo duomenims neparametrinė ir parametrinė procedūros. Vėliau tinkamos parametrinės Archimedo jungties parinkimas turimam duomenų rinkiniui iliustruojamas analizuojant transporto priemonių valdytojų civilinės atsakomybės privalomojo draudimo žalas bei išlaidas žaloms administruoti.

2 Jungtys (angl. *copulas*)

Neformaliai jungtys gali būti apibrėžtos taip: tegul X ir Y yra tolydūs atsitiktiniai dydžiai, turintys pasiskirstymo funkcijas $F(x) = P(X \leq x)$ ir $G(y) = P(Y \leq y)$, jų simultaninė pasiskirstymo funkcija $H(x,y) = P(X \leq x, Y \leq y)$. Kiekvienam taškui $(x,y) \in [-\infty, \infty]$ galime priskirti taškus, esančius $\mathbf{I^3}$ (čia ir toliau tekste simboliu $\mathbf{I}$ bus žymimas intervalas $[0,1]$) su koordinatėmis $(F(x), G(y), H(x,y))$. Funkcija, kuri plokštumos taškams $F(x)$ ir $G(y)$ priskiria tašką, esantį $\mathbf{I}$, yra jungtis. Jungtys taip pat žinomos kaip priklausomumo funkcijos ar tolygūs atvaizdavimai. Formalus apibrėžimas yra toks [6]:

Apibrėžimas 2.1. *Jungtis (dvimatė) yra funkcija $C : \mathbf{I^2} \to \mathbf{I}$ tokia, kad:*

1. $C(0,x) = C(x,0) = 0$ *ir* $C(1,x) = C(x,1) = x$ *kiekvienam* $x \in \mathbf{I}$*;*
2. C *yra 2-didėjanti: tegul* $a, b, c, d \in \mathbf{I}$*, jei* $a \leq b$ *ir* $c \leq d$*, tai*

$$V_C([a,b] \times [c,d]) = C(b,d) - C(a,d) - C(b,c) + C(a,c) \geq 0.$$

2-oje savybėje esanti funkcija V_C vadinama stačiakampio $[a,b] \times [c,d]$ *C-tūriu.* Ekvivalenčiai, jungtis yra vienetiniame kvadrate apibrėžta dvimatė pasiskirstymo funkcija, kurios marginaliniai skirstiniai yra tolygūs intervale $\mathbf{I}$. Pastebime, kad jungtis C duoda tikimybės matą ant $\mathbf{I^2}$ per funkciją $V_C([0,u] \times [0,v]) = C(u,v)$.
Pavyzdžiui, funkcija $\Pi(u,v) = uv$ tenkina 1-ąją ir 2-ąją 2.1 apibrėžimo savybes, taigi yra jungtis. Stačiakampio Π-tūris V_Π yra jo plotas. Jungtis Π vadinama sandaugos jungtimi ir turi svarbią statistinę interpretaciją.
Aukščiau aprašyti formalus ir neformalus apibrėžimai yra susieti Sklar'o teorema (1959), kuri iš dalies paaiškina ir jungčių svarbą statistiniame modeliavime. Ji parodo, kad daugiamatės pasiskirstymo funkcijos išreiškiamos per marginalines pasiskirstymo funkcijas.

Apibrėžimas 2.2. *Funkcija F, apibrėžta srityje $\bar{\mathbf{R}}$, vadinama pasiskirstymo funkcija, jei*

1. F *yra nemažėjanti,*
2. $F(-\infty) = 0$, $F(+\infty) = 1$.

Apibrėžimas 2.3. *Simultaninė pasiskirstymo funkcija (angl. joint distribution function) yra funkcija H, apibrėžta srityje $\bar{\mathbf{R}}^\mathbf{2}$, tokia, kad:*

1. H *yra 2-didėjanti (angl. 2-increasing),*

2. $H(x,-\infty) = H(-\infty,y) = 0$, $H(+\infty,+\infty) = 1$.

H turi marginalines funkcijas F ir G, lygias $F(x) = H(x,\infty)$ ir $G(y) = H(\infty,y)$.

Teorema 2.1 (Sklaro teorema, [6])**.** *Tegul H yra dvimatė pasiskirstymo funkcija, turinti marginalines pasiskirstymo funkcijas F ir G. Tuomet egzistuoja jungtis C tokia, kad $H(x,y) = C(F(x),G(y))$. Atvirkščiai, bet kokioms pasiskirstymo funkcijoms F ir G 2.3 apibrėžime apibrėžta funkcija H yra dvimatė pasiskirstymo funkcija, kurios marginalinės pasiskirstymo funkcijos yra F ir G. Be to, jei F ir G tolydžios, tai C yra vienintelė.*

Kaip išvadas iš 2-osios 2.1 apibrėžimo savybės bet kokiai jungčiai turime:

1. C yra nemažėjanti kiekvieno kintamojo atžvilgiu;

2. C tenkina Lipšico sąlygą: kiekvienam $a,b,c,d \in I$

$$|C(b,d) - C(a,c)| \leq |b-a| + |d-c|.$$

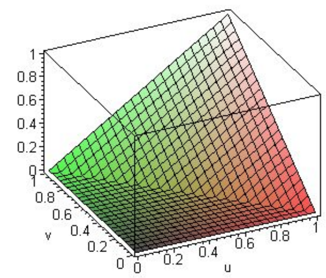
Taigi jungtys yra tolygiai tolydžios.
Duotai simultaninei pasiskirstymo funkcijai H, turinčiai marginalines pasiskirstymo funkcijas F ir G, apibrėžtas taip, kaip Sklaro teoremoje, galime sukonstruoti atitinkamą jungtį: $C(u,v) = H(F^{(-1)}(u), G^{(-1)}(v))$, kur $F^{(-1)}$ yra funkcijos F tolydi iš dešinės ir turinti ribą iš kairės (*cadlag*) atvirkštinė funkcija, apibrėžta $F^{(-1)}(u) = sup\{x|F(x) \leq u\}$ (analogiškai ir $G^{(-1)}$). Jei X ir Y tolydūs atsitiktiniai dydžiai, turintys aukščiau paminėtas pasiskirstymo funkcijas, tai C yra atsitiktinių dydžių $U = F(X)$ ir $V = G(Y)$ simultaninė pasiskirstymo funkcija ($F(x)$ ir $G(y)$ tolygiai pasiskirstę intervale $[0,1]$).
Tuomet, kai jungtį C nagrinėjame kaip dvimatę pasiskirstymo funkciją $\mathbf{I^2}$ ir ji turi simultaninį tankį $p(u,v) = \partial^2 C(u,v)/\partial u \partial v$, tai C yra absoliučiai tolydi. Kitu atveju C gali būti singuliari arba turėti ir absolučiai tolydžią ir singuliariąją komponentes.
Parodoma, kad jei H dvimatė pasiskirstymo funkcija, turinti marginalines pasiskirstymo funkcijas F ir G, tai

$$max\{F(x) + G(y) - 1, 0\} \leq H(x,y) \leq min\{F(x), G(y)\}$$

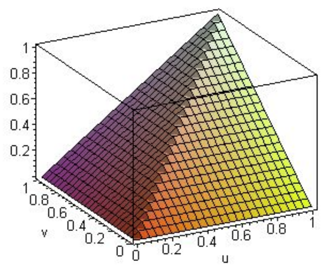
arba (kai $H(x,y) = C(F(x),G(y))$)

$$W(u,v) = max\{u+v-1, 0\} \leq C(u,v) \leq min\{u,v\} = M(u,v).$$

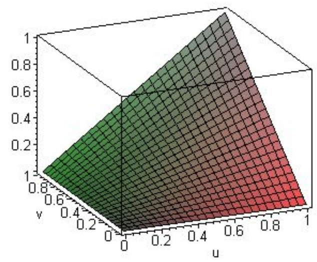
Ši nelygybė žinoma Fréchet-Hoeffding ribų nelygybės vardu, o funkcijos W ir M - apatinė ir viršutinė Fréchet-Hoeffding ribos. Be to šios funkcijos



1 pav.: Apatinė Fréchet-Hoeffding riba



2 pav.: Viršutinė Fréchet-Hoeffding riba



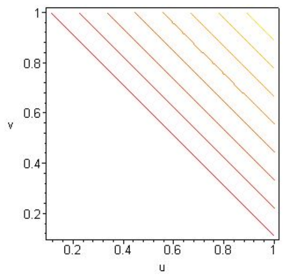
3 pav.: Sandaugos jungtis

pačios yra jungtys. Bet kokios jungties grafikas yra tolydus paviršius vienetinio kubo erdvėje, apribotas Fréchet-Hoeffding ribų grafikais, t.y., paviršiais $z = W(u, v)$ ir $z = M(u, v)$ (žr., 1 - 3 pav.).

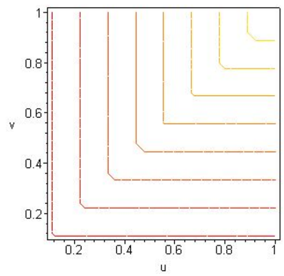
Jungtis galime vaizduoti ir lygio diagramomis (angl. *contour diagram*), t.y., lygio aibių (angl. *level sets*) - aibių iš $\mathbf{I^2}$, apibrėžtų lygybe $C(u, v) = const.$, pasirinktoms konstantoms iš intervalo $\mathbf{I}$, grafikais. 4 - 6 brėžiniuose pavaizduotos jungčių M, W, Π lygio diagramos. Taškai $(t, 1)$ ir $(1, t)$ yra lygio aibės, atitinkančios konstantą t, kraštiniai taškai. Dėl šios priežasties diagramoje nereikia žymėti lygio aibių. 2.1 apibrėžimo 2-ojoje savybėje apibrėžtos sąlygos $C(1, t) = t = C(t, 1)$ aiškiai parodo kiekvienos lygio aibės konstantas. Be to, kiekvienai duotai jungčiai C ir duotam t iš intervalo $\mathbf{I}$, lygio aibė $\{(u, v) \in \mathbf{I^2} | C(u, v) = t\}$ turi gulėti užtušuotame trikampyje (7 pav.), kurio ribos yra lygio aibės, apibrėžtos $M(u, v) = t$ ir $W(u, v) = t$.

Jungtys Π, W ir M turi svarbias statistines interpretacijas. Tegul X ir Y tolydūs atsitiktiniai dydžiai. Tada:

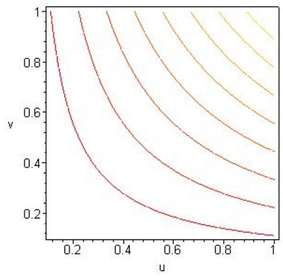
1. X ir Y jungtis yra $M(u, v)$ tada ir tik tada, kai X ir Y - beveik tikrai didėjanti vienas kito funkcija;



4 pav.: W lygio diagrama



5 pav.: M lygio diagrama



6 pav.: Π lygio diagrama

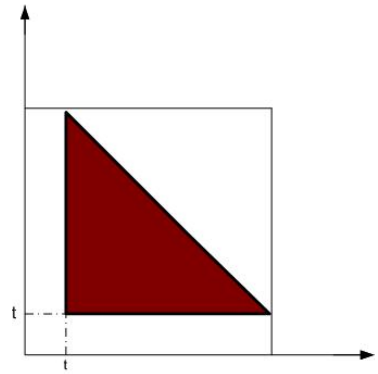
2. X ir Y jungtis yra $W(u, v)$ tada ir tik tada, kai X ir Y - beveik tikrai mažėjanti vienas kito funkcija;

3. X ir Y jungtis yra $\Pi(u, v)$ tada ir tik tada, kai X ir Y - nepriklausomi atsitiktiniai dydžiai.

Fréchet-Hoeffding ribų nelygybės $W(u, v) \leq C(u, v) \leq M(u, v)$ pagalba kiekvienai jungčiai C ir kiekvienam u, v iš $\mathbf{I}$ galima įvesti dalinę tvarką jungčių aibėje:

Apibrėžimas 2.4. *Jei C_1 ir C_2 yra jungtys, tai sakoma, kad C_1 yra* mažesnė *nei C_2 (arba C_2 yra* didesnė *nei C_1), žymima $C_1 \prec C_2$ ($C_2 \succ C_1$), jei $C_1(u, v) \leq C_2(u, v)$ kiekvienam $u, v \in \mathbf{I}$.*

Fréchet-Hoeffding apatinė riba - jungtis W yra mažesnė už bet kurią jungtį, o Fréchet-Hoeffding viršutinė riba - jungtis M, yra didesnė už bet kurią kitą jungtį. Šis dalinis jungčių aibės sutvarkymas vadinamas suderinamumo tvarkiniu (angl. *concordance ordering*) ir svarbus nagrinėjant ryšius tarp jungčių bei atsitiktinių dydžių priklausomumą.
Jei α ir β yra didėjančios nuo X ir Y funkcijos atitinkamai (t.y., jei $\alpha(x_1) < \alpha(x_2)$, kai $x_1 < x_2$, $\beta(y_1) < \beta(y_2)$, kai $y_1 < y_2$), tai $\alpha(X)$ ir $\beta(Y)$ jungtis yra tokia pati, kaip X ir Y jungtis - t.y., $C_{\alpha(X),\beta(Y)} = C_{X,Y}$ - tai jungtis, kuri



7 pav.: Lygio aibių $\{(u,v) \in \mathbf{I}^2 | C(u,v) = t\}$ sritis

užfiksuoja "neparametrinę", "nepriklausančią nuo pasiskirstymo" arba "nekintančią nepriklausomai nuo skalės" priklausomybės tarp X ir Y prigimtį. Kai nors viena iš α ar β griežtai mažėjanti, jungtis pasikeičia atitinkamu būdu:

1. Jei α griežtai didėjanti, β griežtai mažėjanti, tai

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

2. Jei α griežtai mažėjanti, β griežtai didėjanti, tai

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

3. Jei ir α, ir β griežtai mažėjančios, tai

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

3 Jungčių šeimos

Turėdami jungčių rinkinį, naudodami Sklaro teoremą, galime sukonstruoti dvimačius pasiskirstymus, turinčius pasirinktus marginalinius skirstinius. Literatūroje galima rasti daug jungčių klasių pavyzdžių, dauguma kurių yra šeimų narės su vienu ar daugiau realių parametrų (dažnai žymima C_α, $C_{\alpha,\beta}$ ir pan.). Pristatysime keletą parametrinių jungčių šeimų.

3.1 Farlie - Gumbel - Morgenstern šeima

$$C_\alpha(u,v) = uv + \alpha uv(1-u)(1-v), \alpha \in [-1, -1]$$

Tai vienintelės jungtys, kurių funkcinė forma yra u ir v kvadratinis polinomas. Bendras šeimos žymėjimas FGM.
FGM šeimos nariai yra simetriniai, t.y., $C_\alpha(u,v) = C_\alpha(v,u)$ kiekvienam $(u,v) \in \mathbf{I^2}$. Sakoma, jog atsitiktinių dydžių pora (X,Y) sukeičiama (angl. *exchangeable*), jei vektoriai (X,Y) ir (Y,X) yra vienodai pasiskirstę. Vienodai pasiskirsčiusiems tolydiems atsitiktininiams dydžiams sukeičiamumas yra ekvivalentu jungties simetriškumui.

8 pav.: FGM jungtis su parametru 0.1 ir jos lygio diagrama

3.2 Kubinės u ir v jungtys

$C(u,v) = uv + uv(1-u)(1-v)[\alpha uv + \beta u(1-v) + \gamma v(1-u) + \delta(1-u)(1-v)]$,
kur $\alpha, \beta, \gamma, \delta$ realiosios konstantos parinktos taip, kad taškai (α,β), (α,γ), (δ,β), (δ,γ) gulėtų aibėje $[-1,2]\times[-2,1]\cup\{(x,y)|x^2-xy+y^2-3x+3y \leq 0\}$. Kai $\alpha = \beta = \gamma = \delta = \theta$, C yra kvadaratinė ir priklauso FGM šeimai. Priešingai nei FGM jungtys, kubinės u ir v jungtys gali būti asimetriškos.

3.3 Normaliosios jungtys

Pažymime $N_\rho(x,y)$ dvimatę normaliojo pasiskirstymo funkciją su koreliacijos koeficientu ρ. Tuomet C_ρ yra funkciją N_ρ atitinkanti jungtis, užrašoma lygybe $C_\rho = N_\rho(\Phi^{(-1)}(u), \Phi^{(-1)}(v))$ (čia Φ standartinio normaliojo skirstinio pasiskirstymo funkcija). Kadangi nėra tikslios $\Phi^{(-1)}$ išraiškos, negalime surasti ir tikslios N_ρ išraiškos. Tačiau N_ρ galima suskaičiuoti apytiksliai tam, kad sukonstruotumėme dvimates pasiskirstymo funkcijas turinčias tokią pačią priklausomybės struktūrą kaip standartinis normalusis skirstinys, bet su ne normaliniais marginaliais skirstiniais.

Apibrėžimas 3.1. *Atsitiktinių dydžių porai X, Y su marginalinėmis pasiskirstymo funkcijomis F, G atitinkamai ir simultanine pasiskirstymo funkcija H, marginalinės išgyvenamumo funkcijos $\overline{F}$, $\overline{G}$ ir simultaninė išgyvenamumo funkcija $\overline{H}$ apibrėžiamos $\overline{F}(x) = P[X > x]$, $\overline{G}(y) = P[Y > y]$ ir*

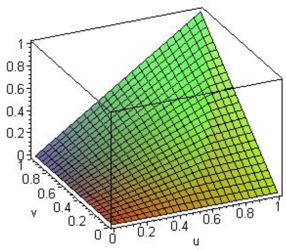
$\overline{H}(x,y) = P[X > x, Y > y]$ *atitinkamai. Funkcija* $\hat{C}$*, sujungianti marginalines išgyvenamumo funkcijas į simultaninę išgyvenamumo funkciją, vadinama išgyvenamumo jungtimi (angl. survival copula):* $\overline{H}(x,y) = \hat{C}(\overline{F}(x), \overline{G}(y))$*.*

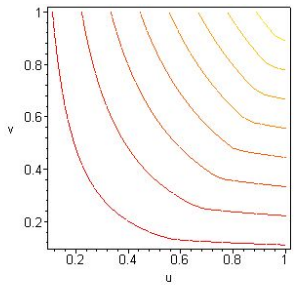
$\hat{C}$ yra jungtis ir ji su įprastine dydžių X ir Y jungtimi siejama lygybe $\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v)$.

3.4 Cuadras-Auge jungtys

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

Atvejis, kai $\alpha = \beta$ pirmą kartą pasirodė Cuadras ir Auge (1981) darbe. Šiuo atveju išgyvenamumo jungtys siejamos su Marshall ir Olkin (1967) dvimačiu eksponentiniu skirstiniu. $C_{\alpha,0} = C_{0,\beta} = \Pi$, $C_{1,1} = M$. Kai $\alpha, \beta \in (0,1)$, jungtis $C_{\alpha,\beta}$ turi ir absoliučiai tolydžiąją ir singuliariąją komponentes.





9 pav.: Cuadras-Auge jungtis su parametrais 0.7 ir 0.2 ir jos lygio diagrama

4 Archimedo jungtys

Archimedo jungtys plačiai naudojamos taikymuose (ypač finansų, draudimo srityse) dėl keletos priežasčių: paprasta jas sukonstruoti, daug jungčių šeimų priklauso šiai klasei, turi ne mažai gerų savybių (pvz., dauguma, bet ne visos, išsiplečia į aukštesnės dimensijos jungtis per asociatyvumo savybę). Archimedo jungtys pirmiausia pasirodė ne statistikoje, bet nagrinėjant tikimybines metrines erdves.

Tegul X ir Y tolydūs atsitiktiniai dydžiai su bendra pasiskirstymo funkcija H ir marginalinėmis pasiskirstymo funkcijomis F ir G. Kai X ir Y yra nepriklausomi, tai $H(x,y) = F(x)G(y)$ kiekvienam x, y iš $\overline{\mathbf{R}}$, ir tai vienintelis pavyzdys, kai bendra pasiskirstymo funkcija išsiskaido į funkcijų F ir G daugiklius.

Yra atvejų, kai H galime išreikšti kaip funkcijų, kurių argumentai yra marginalinės funkcijos F ir G, sumą, t.y., $\varphi(H(x,y)) = \varphi(F(x)) + \varphi(G(y))$. Jungtims ši išraiška atrodys taip:

$$\varphi(C(u,v) = \varphi(u) + \varphi(v).$$

Apibrėžimas 4.1 ([5]). *Tegul φ tolydi, griežtai mažėjanti funkcija iš* $\mathbf{I}$ *į* $[0,\infty]$ *tokia, kad* $\varphi(1) = 0$. *Funkcijos φ pseudo atvirkštinė (* angl. *pseudo-inverse) funkcija $\varphi^{[-1]}$ apibrėžiama:*

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

$\varphi^{[-1]}$ yra tolydi ir nemažėjanti intervale $[0,\infty]$ ir griežtai mažėjanti intervale $[0,\varphi(0)]$. Be to, $\varphi^{[-1]}(\varphi(u)) = u$ intervale $\mathbf{I}$ ir

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

Jeigu $\varphi(0) = \infty$, tai $\varphi^{[-1]} = \varphi^{-1}$.

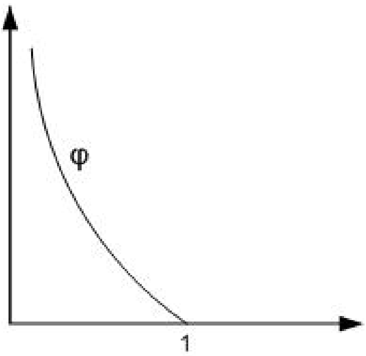
Lema 4.1 ([5]). *Tegul φ tolydi, griežtai mažėjanti funkcija iš* $\mathbf{I}$ *į* $[0,\infty]$ *tokia, kad* $\varphi(1) = 0$, *ir tegul $\varphi^{[-1]}$ būna funkcijos φ pseudo atvirkštinė kaip apibrėžta 4.1 apibrėžime. Tegul C yra funkcija iš* $\mathbf{I}^2$ *į* $\mathbf{I}$ *duota*

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

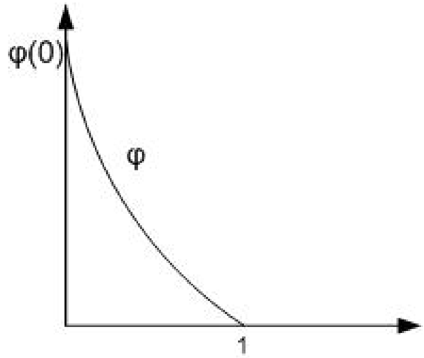
Tuomet $C(u,v) = \varphi^{[-1]}(\varphi(u)+\varphi(v))$ tenkina pirmąją 2.1 jungties apibrėžimo sąlygą. Jei papildomai, φ iškila, tuomet parodoma, jog C tenkina ir antrąją 2.1 apibrėžimo sąlygą. Taigi yra jungtis. Tokios jungtys vadinamos Archimedo (angl. Archimedean).

Funkcija φ vadinama jungties generatoriumi. Kai $\varphi(0) = \infty$, sakome, kad φ griežtas (angl. *strict*) generatorius. Šiuo atveju $\varphi^{[-1]} = \varphi^{-1}$ ir sakoma, jog jungtis $C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v))$ yra griežta. Jei $\varphi(0) \le \infty$ sakome, kad generatorius φ negriežtas (angl. *non-strict*) ir atitinkamai jungtis negriežta. 10 ir 11 brėžiniuose pateikiame griežtų ir negriežtų generatorių grafikų pavyzdžius.

Pavyzdys 4.1. *Turime generatorių $\varphi(t) = -ln(t)$, $t \in [0,1]$. $\varphi(0) = \infty$, todėl φ griežta. Iš to seka, kad $\varphi^{[-1]}(t) = \varphi^{-1}(t) = exp(-t)$, o jungties išraiška yra $C(u,v) = exp(-[(-lnu) + (-lnv)]) = uv = \Pi(u,v)$. Π yra griežta Archimedo jungtis.*



10 pav.: Griežtas generatorius



11 pav.: Negriežtas generatorius

Iš Archimedo jungčių savybių paminėtinos tokios, kaip jų simetriškumas, asociatyvumas, taip pat jei $c > 0$ bet kokia konstanta, tai $c\varphi$ taip pat yra jungties C generatorius.
1-oje lentelėje pateikiame pagrindinius žinomus Archimedo jungčių generatorius, jų generuojamas jungtis bei parametro galiojimo intervalus. Parametro α intervalas duotas toks, kad generatorius φ yra tolydus (kai kurioms parametro reikšmėms iš intervalo skaičiuojamos ribos).

1 lentelė: Archimedo jungtys

Eil. Nr.	$C_\alpha(u,v)$	$\varphi_\alpha(t)$	$\alpha \in$	Ar griežtas
1	$\max\left(\left[u^{-\alpha}+v^{-\alpha}-1\right]^{1/\alpha},0\right)$	$\frac{1}{\alpha}\left(t^{-\alpha}-1\right)$	$[-1,\infty)\setminus\{0\}$	$\alpha \geq 0$
2	$\max\left(1-[(1-u)^\alpha+(1-v)^\alpha]^{1/\alpha},0\right)$	$(1-t)^\alpha$	$[1,\infty)$	Ne
3	$\frac{uv}{1-\alpha(1-u)(1-v)}$	$\ln\frac{1-\alpha(1-t)}{t}$	$[-1,1)$	Taip
4	$\exp\left(-[(-\ln u)^\alpha+(-\ln v)^\alpha]^{1/\alpha}\right)$	$(-\ln t)^\alpha$	$[1,\infty)$	Taip
5	$-\frac{1}{\alpha}\ln\left(1+\frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1}\right)$	$-\ln\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}$	$(-\infty,\infty)\setminus\{0\}$	Taip
6	$1-[(1-u)^\alpha+(1-v)^\alpha-(1-u)^\alpha(1-v)^\alpha]^{1/\alpha}$	$-\ln[1-(1-t)^\alpha]$	$[1,\infty)$	Taip
7	$\max(\alpha uv+(1-\alpha)(u+v-1),0)$	$-\ln[\alpha t+(1-\alpha)]$	$(0,1]$	Ne
8	$\max\left[\frac{\alpha^2uv-(1-u)(1-v)}{\alpha^2-(\alpha-1)^2(1-u)(1-v)},0\right]$	$\frac{1-t}{1+(\alpha-1)t}$	$[1,\infty)$	Ne
9	$uv\exp(-\alpha\ln u\ln v)$	$\ln(1-\alpha\ln t)$	$(0,1]$	Taip
10	$uv/\left[1+(1-u^\alpha)(1-v^\alpha)\right]^{1/\alpha}$	$\ln(2t^{-\alpha}-1)$	$(0,1]$	Taip
11	$\max\left([u^\alpha v^\alpha-2(1-u^\alpha)(1-v^\alpha)]^{1/\alpha},0\right)$	$\ln(2-t^\alpha)$	$(0,1/2]$	Ne
12	$\left(1+[(u^{-1}-1)^\alpha+(v^{-1}-1)^\alpha]^{1/\alpha}\right)^{-1}$	$\left(\frac{1}{t}-1\right)^\alpha$	$[1,\infty)$	Taip
13	$\exp(1-[(1-\ln u)^\alpha+(1-\ln v)^\alpha-1]^{1/\alpha})$	$(1-\ln t)^\alpha-1$	$(0,\infty)$	Taip
14	$(1+[(u^{-1/\alpha}-1)^\alpha+(v^{-1/\alpha}-1)^\alpha]^{1/\alpha})^{-\alpha}$	$(t^{-1/\alpha}-1)^\alpha$	$[1,\infty)$	Taip
15	$\max(\{1-[(1-u^{1/\alpha})^\alpha+(1-v^{1/\alpha})^\alpha]^{1/\alpha}\}^\alpha,0)$	$(1-t^{1/\alpha})^\alpha$	$[1,\infty)$	Ne
16	$\frac{1}{2}(S+\sqrt{S^2+4\alpha})$, $S=u+v-1-\alpha(\frac{1}{u}+\frac{1}{v}-1)$	$\left(\frac{\alpha}{t}+1\right)(1-t)$	$[1,\infty)$	$\alpha>0$
17	$\left(1+\frac{\left[(1+u)^{-\alpha}-1\right]\left[(1+v)^{-\alpha}-1\right]}{2^{-\alpha}-1}\right)^{-1/\alpha}-1$	$-\ln\frac{(1+t)^{-\alpha}-1}{2^{-\alpha}-1}$	$(-\infty,\infty)\setminus\{0\}$	Taip
18	$\max\left(1+\alpha/\ln[e^{\alpha/(u-1)}+e^{\alpha/(v-1)}],0\right)$	$e^{\alpha/(t-1)}$	$[2,\infty)$	Ne
19	$\alpha/\ln(e^{\alpha/u}+e^{\alpha/v}-e^\alpha)$	$e^{\alpha/t}-e^\alpha$	$(0,\infty)$	Taip
20	$\left[\ln(\exp(u^{-\alpha})+\exp(v^{-\alpha})-e)\right]^{-1/\alpha}$	$\exp(t^{-\alpha})-e$	$(0,\infty)$	Taip
21	$1-(1-\{\max([1-(1-u)^\alpha]^{1/\alpha}+[1-(1-v)^\alpha]^{1/\alpha}-1,0)\}^\alpha)^{1/\alpha}$	$1-[1-(1-t)^\alpha]^{1/\alpha}$	$[1,\infty)$	Ne
22	$\max([1-(1-u^\alpha)\sqrt{1-(1-v^\alpha)}-(1-v^\alpha)\sqrt{1-(1-u^\alpha)}]^{1/\alpha},0)$	$\arcsin(1-t^\alpha)$	$(0,1]$	Taip

Kai kurios lentelėje išvardintos jungtys turi savo pavadinimus. 1-oji jungčių

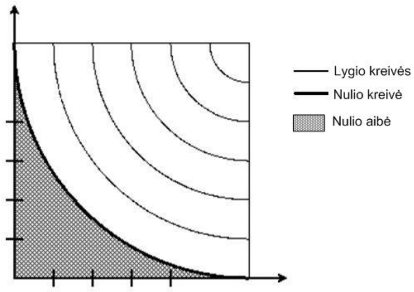
šeima vadinama Kleitono (Clayton) šeima, 3-oji - Ali-Mikhail-Haq, 4-oji - Gumbelo (Gumbel), 5-oji - Franko (Frank) jungčių šeima.
Archimedo jungtys plačiai naudojamos taikymuose (ypač finansų, draudimo ir pan. srityse) dėl savo paprastos formos ir gerų savybių (pvz., dauguma, bet ne visos, išsiplečia į aukštesnės dimensijos per asociatyvumo savybę).

4.1 Svarbiausios Archimedo jungčių savybės

Šiame skyrelyje patyrinėsime keletą pagrindinių Archimedo jungčių savybių. Patogumo dėlei pažymėkime Ω aibę tolydžių griežtai mažėjančių iškilų funkcijų φ iš $\mathbf{I}$ į $[0, \infty]$ tokių, kad $\varphi(1) = 0$.
Iškyla klausimas, kodėl ši jungčių klasė vadinama "Archimedo". Pagal Archimedo aksiomas teigiamiems realiems skaičiams, turime: jei a, b teigiami realūs skaičiai, tai egzistuoja sveikasis skaičius n toks, kad $na > b$. Archimedo jungtis elgiasi taip, kaip binarioji operacija intervale $\mathbf{I}$. Ja jungtis C bet kokiai porai taškų u, v iš $\mathbf{I}$ priskiria tašką $C(u, v)$, priklausantį intervalui $\mathbf{I}$. "Operacija" C yra komutatyvi, asociatyvi ir išlaiko tvarką, t.y., iš to, kad $u_1 \leq u_2$ ir $v_1 \leq v_2$ išplaukia, jog $C(u_1, v_1) \leq C(u_2, v_2)$. Kiekvienam u iš intervalo $\mathbf{I}$ galime rekursyviai apibrėžti $C - laipsnius$ (angl. C-powers) u_C^n tokiu būdu: $u_C^1 = u$, $u_C^{n+1} = C(u, u_C^n)$, čia $n \geq 1$. Pirmą kartą terminas "Archimedo" paminėtas 1965 m. Ling darbe.
Archimedo jungtys gali būti absoliučiai tolydžios arba turėti singuliarias komponentes. Bet kokios jungties C lygio aibė žymima $\{(u, v) \in \mathbf{I}^2 | C(u, v) = t\}$. Archimedo jungčiai ir $t > 0$ ši aibė susideda iš taškų, esančių ant *lygio kreivės* (angl. *level curve*) $\varphi(u) + \varphi(v) = \varphi(t)$ vienetiniame kvadrate $\mathbf{I}^2$, kuri jungia taškus $(1, t)$ ir $(t, 1)$. Dažnai lygio kreivė užrašoma pavidalu $v = L_t(u)$. Išreikšdami v kaip funkciją nuo u, gauname

$$v = L_t(u) = \varphi^{[-1]}(\varphi(t) - \varphi(u)) = \varphi^{-1}(\varphi(t) - \varphi(u)).$$

Paskutinis žingsnis, pakeičiant $\varphi^{[-1]}$ į φ^{-1} pagrįstas, nes $\varphi(t) - \varphi(u)$ yra intervale $[0, \varphi(0))$. Kai $t = 0$, aibė $\{(u, v) \in \mathbf{I}^2 | C(u, v) = 0)$ vadinama *nulio aibe* (angl. *zero set*) ir žymima $Z(C)$. Daugumai Archimedo jungčių $Z(C)$ yra tiesiog dvi tiesios atkarpos $\{0\} \times \mathbf{I}$ ir $\mathbf{I} \times \{0\}$. Kitoms Archimedo jungtims $Z(C)$ turi nelygų nuliui plotą ir tokios nulio aibės ribinė kreivė yra $\varphi(u) + \varphi(v) = \varphi(0)$, t.y., $v = L_0(u)$ vadinama C *nulio kreive* (angl. *zero curve*). Pateikiame pastarosios pavyzdį (12 pav.) - 2-os jungčių šeimos 1-oje lentelėje narys su parametru $\alpha = 2$. Šiuo atveju lygio kreivės ir nulio kreivė yra apskritimo ketvirčiai.
Visos Archimedo jungčių lygio kreivės yra iškilos.
Taip pat Archimedo jungtims galime surasti srities $\mathbf{I}^2$ C-matą, kuris gali būti ant lygio kreivės arba apimti plotą, esantį žemiau ar į kairę nuo lygio kreivės.



12 pav.: 2-os Archimedo jungties su parametru $\alpha = 2$ lygio kreivės ir nulio aibė

Teorema 4.1 ([5]). *Tegul C yra generatoriaus φ, priklausančio aibei Ω, sugeneruota jungtis. Tegul $K_C(t)$ žymi aibės $\{(u,v) \in \mathbf{I^2} | C(u,v) \leq t\}$, arba ekvivalenčiai aibės $\{(u,v) \in \mathbf{I^2} | \varphi(u) + \varphi(v) \geq \varphi(t)\}$, C-matą. Tada kiekvienam t iš $\mathbf{I}$,*

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}.$$

Suformuluosime teiginį, kuris vėliau bus naudinga nagrinėjant neparametrinio asociacijos mato - Kendalo τ versiją populiacijos atveju.

Teiginys 4.2 ([5]). *Tegul U ir V tolygūs $(0,1)$ atsitiktiniai dydžiai, kurių simultaninė pasiskirstymo funkcija yra generatoriaus φ iš Ω sugeneruota jungtis C. Tuomet funkcija K_C, apibrėžta aukščiau esančia lygybe, yra atsitiktinio dydžio $C(U,V)$ pasiskirstymo funkcija. Be to, funkcija*

$$K_C'(s,t) = \begin{cases} t, & s \leq t \\ t - \frac{\varphi(t)-\varphi(s)}{\varphi'(t^+)}, & s > t \end{cases}$$

yra simultaninė atsitiktinių dydžių U ir $C(U,V)$ pasiskirstymo funkcija.

Užbaigdami šį skyrių paminėsime, jog Archimedo jungčių parinkimui ir pritaikymui turimiems duomenims egzistuoja specialios procedūros, kurias vėliau aprašysime ir panaudosime empiriniame tyrime.

5 Priklausomumo ir asociacijos dydžių aprašymas

Šiame skyrelyje panagrinėsime būdus, kai jungtis galima panaudoti atsitiktinių dydžių priklausomumui, asociacijoms tirti.

Atsitikinių dydžių priklausomumas matuojamas įvairiais būdais. Daugelis asociacijos matų nekinta nepriklausomai nuo atsitiktinių dydžių transformacijų. Nekintantys nuo skalės priklausomumo dydžiai gali būti randami naudojant jungtis.
Be to, priklausomumo savybės ir dydžiai susiję tarpusavyje. Labiausiai žinomi skalės invariantai yra Kendalo τ (angl. *Kendall's tau*) ir Spirmeno ρ (angl. *Spearman's rho*) populiacijos atveju. Abu jie matuoja priklausomumo formą, žinomą kaip suderinamumas.

5.1 Suderinamumas (angl. *concordance*)

Tegul tolydžių atsitiktinių dydžių vektoriaus (X, Y) stebėjimai žymimi (x_i, y_i), (x_j, y_j). Sakome, kad (x_i, y_i) ir (x_j, y_j) suderinami, jei $x_i < x_j$ ir $y_i < y_j$ arba jei $x_i > x_j$ ir $y_i > y_j$. Taip pat sakome, kad (x_i, y_i) ir (x_j, y_j) yra nesuderinami (angl. *disconcordant*), jei $x_i < x_j$ ir $y_i > y_j$ arba, jei $x_i > x_j$ ir $y_i < y_j$. Suderinamumo apibrėžimą dar galime suformuluoti ir taip: (x_i, y_i) ir (x_j, y_j) yra suderinami, jei $(x_i - x_j)(y_i - y_j) > 0$ ir nesuderinami, jei $(x_i - x_j)(y_i - y_j) < 0$.

5.1.1 Kendalo τ (angl. *Kendall's tau*)

Priklausomumo matas Kendalo τ imties atveju apibrėžiamas taip: tegul $\{(x_1, y_1), (x_2, y_2) ..., (x_n, y_n)\}$ tolydžių atsitiktinių dydžių vektoriaus (X, Y) n stebėjimų imtis. Iš viso skirtingų stebėjimų porų (x_i, y_i) ir (x_j, y_j) imtyje yra C_n^2. Kiekviena tų porų yra suderinama arba nesuderinama. Pažymime c suderinamų porų skaičių, d - nesuderinamų porų skaičių. Tada Kendalo τ imties atvejui apibrėžiamas taip:

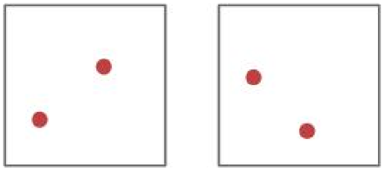
$$t = \frac{c - d}{c + d} = (c - d)/C_n^2.$$

Atitinkamai galime sakyti, kad t yra stebėjimų poros (x_i, y_i) ir (x_j, y_j), kuri atsitiktinai parinkta iš imties, suderinamumo tikimybė minus nesuderinamumo tikimybė. Kendalo τ populiacijos atveju apibrėžiamas panašiai.

Apibrėžimas 5.1 ([6]). *Turime tolydžių atsitiktinių dydžių vektorių (X, Y), šių dydžių pasiskirstymo funkcija H. Tegul (X_1, Y_1) ir (X_2, Y_2) nepriklausomi vienodai pasiskirstę atsitiktiniai vektoriai, kurių kiekvienas turi pasiskirstymo funkciją H. Tuomet Kendalo τ populiacijos atveju apibrėžiamas kaip suderinamumo tikimybė minus nesuderinamumo tikimybė:*

$$\tau = \tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Geometriškai du taškai (x_1, y_1) ir (x_2, y_2) plokštumoje yra suderinami, jei juos jungianti atkarpa turi teigiamą nuolydį ir nesuderinami, jei atkarpa turi neigiamą nuolydį.



13 pav.: Suderinamų (kairėje) ir nesuderinamų (dešinėje) taškų poros

Tam, kad parodytumėme, kokį vaidmenį atlieka jungtys suderinamumo ir asociacijos dydžių matavime, turime apibrėžti "suderinamumo funkciją" (angl. *concordance function*) Q. Ji yra skirtumas tarp tikimybių, jog du tolydžių atsitiktinių dydžių vektoriai (X_1, Y_1) ir (X_2, Y_2) su galimai skirtingomis simultaninėmis pasiskirstymo funkcijomis H_1 ir H_2, bet turintys vienodas marginalines pasiskirstymo funkcijas F ir G, yra suderinami ir nesuderinami. Tada parodysime, kad ši funkcija priklauso nuo (X_1, Y_1) ir (X_2, Y_2) pasiskirstymų tik per jų jungtis.

Teorema 5.1 ([5])**.** *Tegul (X_1, Y_1) ir (X_2, Y_2) nepriklausomi atsitiktiniai vektoriai, kurių komponentės X_1 ir X_2 turi marginalinę pasiskirstymo funkciją F, o komponenės Y_1 ir Y_2 turi marginalinę pasiskirstymo funkciją G. Tariame, kad šie atsitiktiniai vektoriai turi simultanines pasiskirstymo funkcijas H_1 ir H_2. Tegul C_1 ir C_2 yra atitinkamai dydžių (X_1, Y_1) ir (X_2, Y_2) jungtys tokios, kad $H_1(x, y) = C_1(F(x), G(y))$ ir $H_2(x, y) = C_2(F(x), G(y))$. Pažymime Q skirtumą tarp tikimybių, jog poros (X_1, Y_1) ir (X_2, Y_2) yra suderinamos ir nesuderinamos. t.y., tegul*

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

Tada

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

Keletas funkcijos Q savybių:

1. Q yra simetrinė savo argumentų atžvlgiu: $Q(C_1, C_2) = Q(C_2, C_1)$.

2. Q yra nedidėjanti kiekvieno argumento atžvilgiu: jei $C_1 \prec C_1'$ ir $C_2 \prec C_2'$ kiekvienam (u, v) iš $\mathbf{I}^2$, tai $Q(C_1, C_2) \leq Q(C_1', C_2')$.

3. Jungtys suderinamumo funkcijoje gali būti pakeistos išgyvenamumo jungtimis, t.y., $Q(C_1, C_2) = Q(\hat{C}_1, \hat{C}_2)$.

Tarkime, C pasirinkta jungtis. Kadangi Q yra skirtumas tarp dviejų tikimybių, tai $Q(C, C) \in [-1, 1]$.
Apibendrinant visas turimas suderinamumo išraiškas, gaunama Kendalo τ išraišką populiacijos atveju.

Teorema 5.2 ([5])**.** *Tegul X ir Y tolygūs atsitiktiniai dydžiai, kurių jungtis C. Tuomet dydžių X ir Y Kendalo τ populiacijos atveju (žymėsime $\tau_{X,Y}$ arba τ_C) apibrėžiamas:*

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

Teoremoje esantis integralas gali būti interpretuojamas kaip atsitiktinių, tolygių intervale $(0, 1)$ dydžių U ir V, turinčių simultaninę pasiskirstymo funkciją C, funkcijos $C(U, V)$ tikėtina reikšmė, t.y.,

$$\tau_C = 4E(C(U, V)) - 1. \tag{1}$$

Norint surasti Kendalo τ, reikia skaičiuoti dvigubą integralą, tačiau Archimedo jungčių atveju situacija paprastesnė. Integralas supaprastinamas, jo išraiškoje panaudojant Archimedo jungties generatorių, kuris yra vieno kintamojo funkcija. Tai parodo žemiau suformuluota teorema.

Teorema 5.3 ([5])**.** *Tegul X ir Y yra tolygūs atsitiktiniai dydžiai su Archimedo jungtimi C, kurią generuoja generatorius φ iš aibės Ω. X ir Y Kendalo τ populiacijos atveju bus lygus*

$$\tau_C = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt. \tag{2}$$

Įrodymas. *Tegul U ir V tolygūs intervale $(0, 1)$ atsititiniai dydžiai, turintys simultaninę pasiskirstymo funkciją C. Tegul K_C žymi dydžio $C(U, V)$ pasiskirstymo funkciją. Tuomet pagal* (1) *turime, kad*

$$\tau_C = 4E(C(U, V)) - 1 = 4 \int_0^1 t dK_C(t) - 1.$$

Integruodami dalimis, gauname

$$\tau_C = 3 - 4 \int_0^1 K_C(t) dt.$$

Tačiau žinome, kad $C(U, V)$ pasiskirstymo funkcija $K_C(t)$ yra lygi

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}.$$

Iš to seka, jog

$$\tau_C = 3 - 4\int_0^1 \left[t - \frac{\varphi(t)}{\varphi'(t^+)}\right] dt = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

Paskutinėje lygybėje $\varphi(t^+)$ galime pakeisti į $\varphi(t)$, nes iškilos funkcijos yra beveik visur diferencijuojamos. $\diamondsuit$

5.1.2 Spirmeno ρ (angl. *Spearman's rho*)

Kaip ir Kendalo τ, priklausomumo matas Spirmeno ρ populiacijos atveju yra paremtas suderinamumu ir nesuderinamumu.

Apibrėžimas 5.1. *Tegul (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) yra trys nepriklausomi atsitiktiniai vektoriai, turintys tą pačią simultaninę pasiskirstymo funkciją H (kurios marginalinės funkcijos yra F ir G) ir jungtį C. Tuomet Spirmeno ρ populiacijos atveju apibrėžiamas kaip tikimybės, jog vektoriai (X_1, Y_1) ir (X_2, Y_3) yra suderinami ir tikimybės,jog jie nesuderinami, skirtumas:*

$$\rho = \rho_{X,Y} = 3\left(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]\right) \tag{3}$$

Dydžių (X_1, Y_1) pasiskirstymo funkcija yra $H(x, y)$, o dydžių (X_2, Y_3) pasiskirstymo funkcija yra $F(x)G(y)$ (nes X_2 ir Y_3 nepriklausomi). Todėl X_2 ir Y_3 jungtis yra Π. Iš to išplaukia, kad:

$$\rho_{X,Y} = \rho_C = 3Q(C, \Pi) = 12\int\int_{I^2} uv dC(u, v) - 3 = 12\int\int_{I^2} C(u, v) du dv - 3. \tag{4}$$

(3) lygbėje bei (4) lygybės pirmoje dalyje konstanta 3 yra normuojanti konstanta. Antrojoje (4) lygties dalyje esantį integralą galime interpretuoti kaip paviršiaus $z = C(u, v)$ tūrį virš $\mathbf{I^2}$, todėl ρ_C yra proporcingas tūris. Trečioji (4) lygybės dalis interpretuojama kaip tūris tarp paviršių $z = C(u, v)$ ir $z = \Pi$. ρ_C galima laikyti "vidutiniu atstumu" tarp X ir Y pasiskirstymo (kurį rodo C) ir nepriklausomumo (nepriklausomumą žymi Π).

5.1.3 Kiti suderinamumo matai

1920 m. Corrado Gini pasiūlė asociacijos matą imties atvejui, paremtą absoliučiu skirtumu tarp rangų. Populiacijos atveju atsitiktiniams dydžiams X ir Y, turintiems jungtį C, šis dydis yra:

$$\gamma = 2\int\int_{I^2} \left(|u + v - 1| - |u - v|\right) dC(u, v).$$

Šį matą, kaip ir Kendalo τ ar Spirmeno ρ, galime išreikšti per suderinamumo funkciją Q:

$$\gamma_{X,Y} = \gamma_C = Q(C,M) + Q(C,W).$$

Spirmeno $\rho = 3Q(C,\Pi)$ matuoja suderinamumo ryšį arba "atstumą" tarp X ir Y pasiskirstymo ir nepriklausomumo. Tuo tarpu, Gini $\gamma = Q(C,M) + Q(C,W)$ žymi suderinamumo ryšį arba "atstumą" tarp jungties C ir monotoniško priklausomumo (tai žymi jungtys M ir W).

Dar vieną papildomą suderinamumo dydį pasiūlė Blomqvist'as (Blomqvist, 1959). Tarkime, suderinamumo funkcijos išraiškoje, suderinamumo tikimybės ir nesuderinamumo tikimybės skirtume, nagrinėjami atsitiktinis vektorius ir fiksuotas taškas. T.y., turime:

$$P\left[(X-x_0)(Y-y_0)>0\right] - P\left[(X-x_0)(Y-y_0)<0\right]$$

kažkokiam pasirinktam taškui $(x_0, y_0) \in \mathbf{R^2}$. Blomqvist'as tyrinėjo tą atvejį, kai vietoj x_0 ir y_0 parinktos populiacijos medianos reikšmės. Šis dydis dažnai vadinamas "vidurio koreliacijos koeficientu" (angl. *medial correlation coefficient*). Žymima β ir apibrėžiama:

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

kur $\tilde{x}$ ir $\tilde{y}$ yra dydžių X ir Y medianos. Kai X ir Y yra tolydūs atsitiktiniai dydžiai, turintys jungtį C, β užrašoma:

$$\beta = \beta_C = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1.$$

Nors Blomqvist'o β priklauso tik nuo jungties reikšmės $\mathbf{I^2}$ centro taške, tačiau šiuo dydžiu galima aproksimuoti Kendalo τ ar Spirmeno ρ.

5.2 Kiti priklausomumo matai

Be aukščiau aprašyto suderinamumo yra ir kitokį priklausomumą tarp atsitiktinių dydžių nusakantys dydžiai, kurie šiame darbe nebus nagrinėjami. Tai - kvadranto priklausomumas (angl. *quadrant dependence*), uodegos monotoniškumas (angl. *tail monotonicity*), stochastinis monotoniškumas (angl. *stochastic monotonicity*), tikėtinumo santykio priklausomumas (angl. *likelihood ratio dependence*).

6 Jungties parinkimas ir pritaikymas duomenims

Šiame skyrelyje aprašysime tinkamos jungties parinkimą turimiems empiriniams duomenims. Turime dvimačius duomenis ir norime pritaikyti jungtį. Laikysimės prielaidos, jog duomenys gali būti tinkamai sumodeliuoti Archimedo jungčių pagalba.
Tarkime, turime dvimačių stebėjimų (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) n dydžio imtį. Šie dydžiai pasiskirstę pagal nežinomą dvimatę pasiskirstymo funkciją $H(x, y)$, ir turi marginalinius skirstinius $F(x)$, $G(y)$ bei Archimedo jungtį $C(u, v)$. Mūsų tikslas parinkti generatorių φ - vieno kintamojo funkciją, priklausančią nuo parametro α.
Jeigu laikome, kad duomenims pritaikomas Archimedo jungčių šeimos narys (nagrinėsime tik vieno parametro šeimas), tai galime atskirti marginalinių skirstinių $F(x)$ ir $G(y)$ parinkimą nuo parametro α įvertinimo. Įvertinimą galima atlikti vienu arba dviem žingsniais ir parametriniu arba neparametriniu metodu.
Dviejų žingsnių būdas susideda iš marginalinių skirstinių $F(x)$ ir $G(y)$ įvertinimo ir po to jungties parametro α įvertinimo. Marginalinius skirstinius taip pat galime vertinti parametriniu arba neparametriniu metodu.
Vertindami jungties parametrą vienu žingsniu galime naudoti parametrinį arba neparametrinį netodą. Parametrinis metodas yra didžiausio tikėtinumo metodas, kai vienu ir tuo pačiu žingsniu suskaičiuojame marginalinių skirstinių ir jungties parametrus. Tikėtinumo funkcija šiuo atveju turės pavidalą $L(\theta, \alpha; X, Y)$, kur θ simbolizuoja marginalinių pasiskirstymo funkcijų parametrus, o α jungties parametrą.
Neparametrinį vieno žingsnio metodą pasiūlė Genest ir Rivest (1993)[1]. Šis metodas neatsižvelgia į marginalinius X ir Y skirstinius. Parametras α įvertinamas naudojant Kendalo koreliacijos matą τ.
Plačiau aptarsime vieno žingsnio neparametrinį ir dviejų žingsnių parametrinį metodus.

6.1 Genest ir Rivest procedūra

Ieškodami Archimedo jungties, atsižvelgiame į faktą, kad jungtis $C_\varphi(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$ yra vienareikšmiškai nusakoma funkcija $K(t) = t - \frac{\varphi(t)}{\varphi'(t)}$, apibrėžta intervale $[0, 1]$. Be to, žinoma, kad $K(t)$ yra pasiskirstymo funkcija intervale $(0, 1)$. Taip pat turime, jog Kendalo τ populiacijos atveju Archimedo jungtims suskaičiuojamas, naudojantis išraiška $\tau_C = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt$.
Tariame, kad turime stebėjimų imtį (X_1, Y_1), (X_2, Y_2),..., (X_n, Y_n). Stebėjimai

turi dvimatę pasiskirstymo funkciją $H(x, y)$ ir tolydžius marginalinius skirstinius $F(x)$, $G(y)$ bei priklausomybės funkciją $C_\varphi(u, v)$. Dydžiai $U = F(X)$ ir $V = G(Y)$ yra tolygiai pasiskirstę. Norint parinkti tinkamą Archimedo jungties generatorių, dirbama su tarpiniais atsitiktiniais dydžiais $Z = H(X, Y)$, turinčiais pasiskirstymo funkciją $K(z) = P(Z \leq z) = P(C(F(x), G(y)) \leq z)$. Tolimesni veiksmai susideda iš etapų:

1. įvertinamas Kendalo koreliacijos koeficientas naudojant neparametrinį įvertį:
$$\tau_n = \left(C_n^2\right)^{-1} \sum_{i<j} sign\left[(X_i - X_j)(Y_i - Y_j)\right]. \tag{5}$$

2. Konstruojamas funkcijos K neparametrinis įvertis:

   - Apibrėžiami pseudo stebėjimai
$$Z_i = \frac{\#\left\{(X_j, Y_j) : X_j < X_i, Y_j < Y_i\right\}}{(n-1)}, \qquad i = 1, ..., n. \tag{6}$$

   - Randamas funkcijos K neparametrinis įvertis
$$K_n(z) = \frac{\#\{i : 1 \leq i \leq n, Z_i \leq z\}}{n}. \tag{7}$$

3. Ieškomas parametrinis funkcijos K įvertis, naudojantis sąryšiu
$$K_\varphi(z) = t - \frac{\varphi(z)}{\varphi'(z)}.$$

Parametrinis funkcijos K įvertis ieškomas keletui pasirinktų Archimedo jungčių šeimų. Pirmiausia, naudojantis Kendalo τ įverčiu, surandamas jungties parametro α įvertis $\hat{\alpha}$, po to parametro įvertis panaudojamas jungties generatoriaus φ įvertinimui. Turint generatoriaus įvertį φ_n, galima surasti funkcijos $K_\varphi(z)$ įvertį $K_{\varphi_n}(z)$.

Pakartojus 3-iąjį žingsnį keletui jungčių šeimų, kiekvienas gautas parametrinis funkcijos K įvertis palyginamas su 2-ąjame žingsnyje sukonstruota neparametrine K funkcija. Pasirenkamas generatorius φ toks, kad parametrinis įvertis $K_{\varphi_n}(z)$ labiausiai atitiktų neparametrinį įvertį $K_n(z)$. Įverčiai lyginami grafiškai. Tuo tikslu brėžiami kvantilių-kvantilių grafikai (angl. *Q-Q plot*). Jei parametras yra "geras", grafike kvantiliai turėtų būti išsidėstę ant įstrižos tiesios linijos.

Įvertintų jungties parametrų įverčių atitikimą taip pat galima patikrinti ir minimizuojant atstumą $\int (K_{\varphi_n}(z) - K_n(z))^2 dK_n(z)$. Šį būdą pasiūlė Freez and Valdez [2].

6.2 Didžiausio tikėtinumo metodas

Kita jungties parinkimo procedūra yra parametrinė. Ji grindžiama maksimalaus tikėtinumo įverčiu. Priežingai nei Genest ir Rivest pasiūlytoje procedūroje, čia tinkamos jungties parinkimas priklauso ir nuo marginalinių skirstinių. Nei vienas iš šių metodų nelaikomas geresniu už kitą. Yra žinoma, kad jei duomenyse yra išskirčių arba marginaliniai skirstiniai turi sunkias uodegas, tai geriau rinktis Genest ir Rivest procedūrą, nes šiuo būdu rasti įverčiai stabilesni ir nėra keliami reikalavimai marginalinėms pasiskirstymo funkcijoms $F(x)$ ir $G(y)$. Tačiau jei turimas didelis duomenų rinkinys, tikslesni įverčiai randami didžiausio tikėtinumo metodu.
Nagrinėjame n dvimačių stebėjimų (X_1, Y_1), (X_2, Y_2),..., (X_n, Y_n). Jie pasiskirstę pagal nežinomą simultaninę pasiskirstymo funkciją $H(x, y)$ su marginaliniais skirstiniais $F(x)$ ir $G(y)$ ir Archimedo jungtimi $C(u, v)$.
Pirmiausia daromos prielaidos apie marginalinius skirstinius ir įvertinami marginalinių pasiskirstymo funkcijų parametrai.
Žinoma, kad $H(x, y) = C(F(x), G(y))$, tuomet simultaninė tankio funkcija yra

$$h(x, y) = f(x)g(y)C_{12}(F(x), G(y)),$$

kur

$$C_{12}(u, v) = \frac{\partial}{\partial u}\frac{\partial}{\partial v}C(u, v).$$

Kadangi laikoma, jog marginalinių pasiskirstymo funkcijų parametrai yra jau įvertinti, tai funkcijų F ir G reikšmės stebėjimų taškuose $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ yra žinomos ($U_i = F(X_i)$ ir $V_i = G(Y_i)$) ir pasiskirsčiusios tolygiai. Tuomet tikėtinumo funkcijoje marginalinių skirstinių tankio funkcijos tampa konstantomis ir tikėtinumo funkciją galima užrašyti supaprastinta forma:

$$L(\alpha; U, V) = \prod_{i=1}^{n} C_{12}(U_i, V_i).$$

Maksimumo patogiau ieškoti logaritmavus tikėtinumo funkciją. Galutinė įverčio išraiška lygi

$$\hat{\alpha} = argmax_{\alpha \in A} \sum_{i=1}^{n} \log L(\alpha; U_i, V_i),$$

kur A yra galimų parametro α reikšmių aibė (Galiojančios α reikšmės pateiktos 1-oje Archimedo jungčių lentelėje).

Tolimesniame skyriuje atliekama empirinių duomenų analizė. Archimedo jungtys pritaikomos remiantis Genest ir Rivest procedūra. Tokį pasirinkimą lėmė tai, kad marginaliniai skirstiniai neatitiko tiksliai nei vieno parametrinio skirstinio iš žinomų skirstinių klasės. Todėl paprastumo dėlei pasirinktas "Amerikietiškas" Pareto skirstinys (taip pat žinant, jog praktikoje žalos dažnai modeliuojamos šio skirstinio pagalba). Be to, stebėjimų imtis nėra didelė. Dėl šių priežasčių laikoma, kad neparametrinis įvertis duos tikslesnius rezultatus.

7 Duomenų analizė

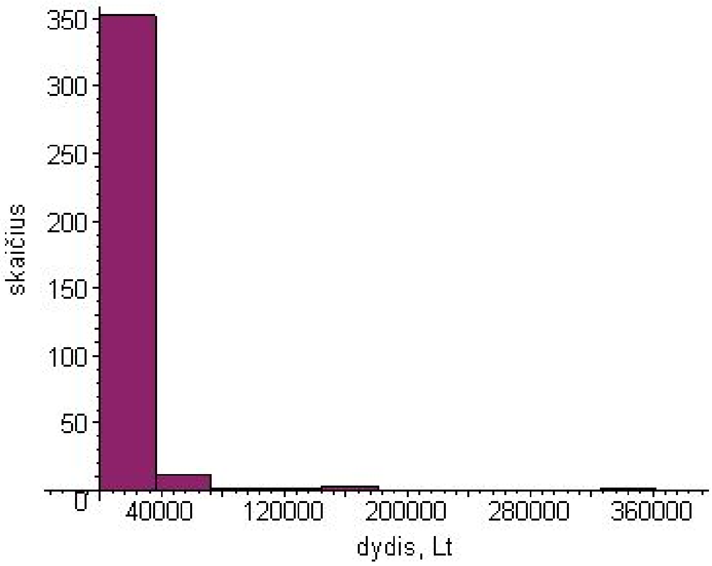
Aukščiau aprašytai teorijai iliustruoti atliksime draudimo žalų duomenų analizę. Turimi duomenys sudaryti iš 371 Transporto priemonių valdytojų civilinės atsakomybės privalomojo draudimo (toliau - TPVCAPD) žalos, atsiradusios dėl ne Lietuvoje įvykusių eismo įvykių. Kiekviena žala susideda iš nuostolių (žala, X) ir žalos admnistravimo išlaidų (išlaidos, Y). Į išlaidas žaloms administruoti įeina teisinių paslaugų mokesčiai, turto vertintojo išlaidos, banko išlaidos, žalos nagrinėjimo išlaidos ir pan. Mūsų tikslas yra nustatyti bendrą žalų ir išlaidų žaloms administruoti pasiskirstymą. Tam, kad galėtume daryti prielaidas apie marginalinių skirstinių tolydumą, iš duomenų pašalinti atvejai, kai $X = 0$ arba $Y = 0$. Žalų ir išlaidų dydžiai matuojami litais. Taip pat atkreipiamas dėmesys, jog duomenys yra transformuoti ir realios išvados iš gautų rezultatų negali būti daromos.
2-oje lentelėje pateikiamas duomenų apibendrinimas.

2 lentelė: Žalų ir išlaidų aprašomoji statistika

	ŽALOS	IŠLAIDOS
Skaičius	371	371
Vidurkis	11316,72	1539,09
Mediana	4555,49	791,64
Moda	9322,56	690,56
Standartinis nuokrypis	26750,64	1986,449
Minimali reikšmė	216,94	25,83
Maksimali reikšmė	361726,65	16288,45
0,25-asis kvantilis	1842,45	690,56
0,75-asis kvantilis	10527,44	1437,498

Taip pat pateikiame žalų (14 pav.) ir išlaidų žaloms admministruoti (15 pav.) histogramas bei žalų lyginant su išlaidomis grafiką (16 pav).

Žalų lyginant su išlaidomis grafike matoma, kad tarp šių dydžių yra stipri



14 pav.: Žalų histograma

tiesinė priklausomybė. Tai patvirtina ir atitinkamas koreliacijos koeficientas, kuris lygus 0,743.

7.1 Marginalinių skirstinių parinkimas

Pradinis žingsnis modelio konstravime yra tinkamų marginalinių skirstinių parinkimas. Nagrinėjamus žalų ir išlaidų duomenis modeliuosime "Amerikietiškojo" Pareto skirstinio pagalba. Šio skirstinio su parametrais x_0 ir k pasiskirstymo funkcijos išraiška yra

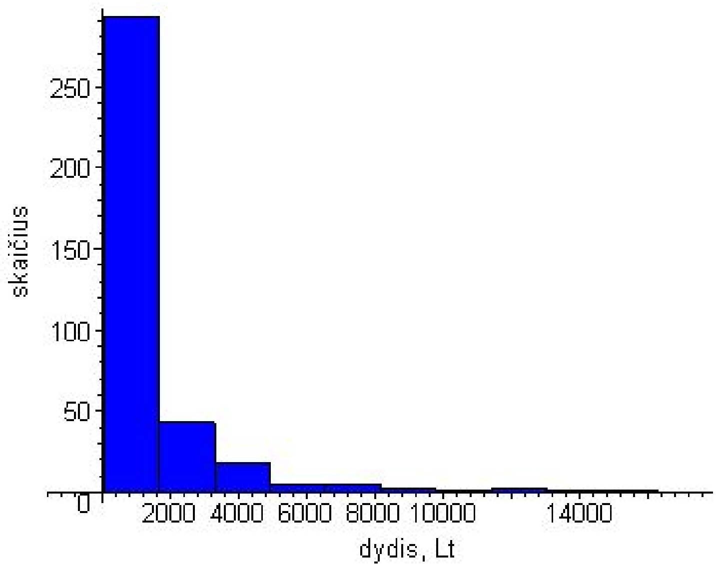
$$F(x) = 1 - \left(\frac{x_0}{x_0 + x}\right)^k, \qquad x > 0 \tag{8}$$

Tankio funkcija lygi

$$f(x;k;x_0) = \frac{k x_0^k}{(x + x_0)^{k+1}}, \qquad x \geq x_0.$$

Parametrai vertinami didžiausio tikėtinumo metodu. Tikėtinumo funkcija Pareto skirstinio atveju

$$L(k, x_0) = \prod_{i=1}^{n} \frac{k x_0^k}{(x_0 + x_i)^{k+1}} = k^n x_0^{nk} \prod_{i=1}^{n} \frac{1}{(x_0 + x_i)^{k+1}}.$$



15 pav.: Išlaidų histograma

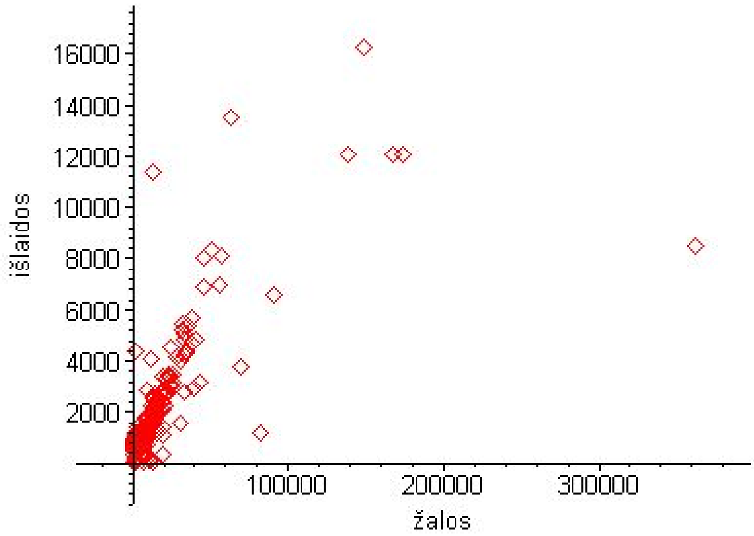
Logaritmavus tikėtinumo funkciją ir ją maksimizavus, gauname parametro k įverčio išraišką:

$$\hat{k} = \frac{n}{\sum_{i=1}^{n} \ln(x_0 + x_i) - \sum_{i=1}^{n} \ln(x_0)}.$$

Šioje išraiškoje esantis parametras x_0 laikomas žinomu ir lygus minimaliai stebėjimų reikšmei, t.y., $x_0 = \min_i x_i$.
Išlaidų duomenų atveju šį parametrą x_0 ir parinkome lygų minimaliai išlaidų reikšmei, o parametrą k įvertinome. Tačiau žalų duomenų atveju, atlikus papildomą analizę, pasiskirstymas gaunamas tikslesnis, kai x_0 laikomas nežinomu ir taip pat kaip ir k įvertinamas.
Nagrinėjamu atveju, įvertinę "Amerikietiško" Pareto skirstinio parametrus žaloms ir išlaidoms, atitinkamai turime tokias marginalinių funkcijų išraiškas:

$$F(x) = 1 - \left(\frac{9916,8}{x + 9916,8} \right)^{1,8909} \tag{9}$$

$$G(y) = 1 - \left(\frac{25,83}{y + 25,83} \right)^{0,2716} \tag{10}$$



16 pav.: Žalų ir išlaidų grafikas

7.2 Tinkamos jungties pritaikymas

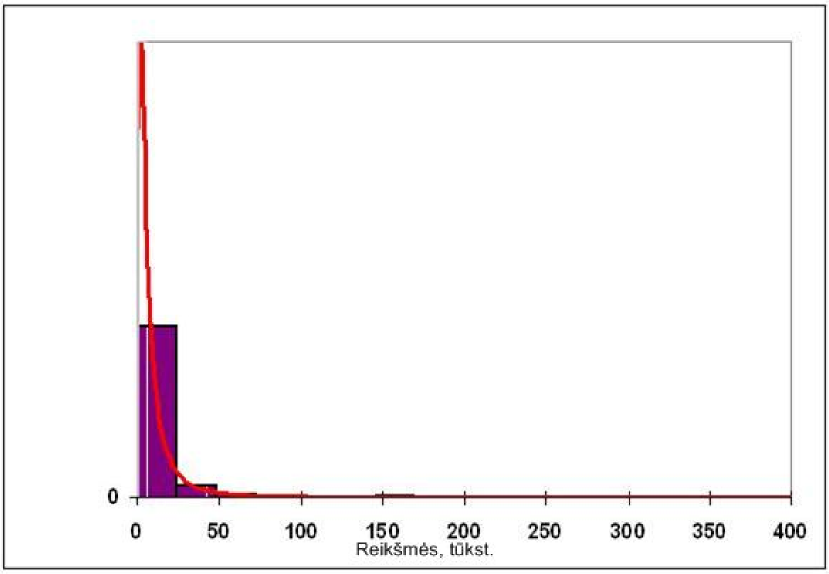
Tinkamai jungčiai nustatyti naudosime 6.1 skyrelyje aprašytą procedūrą. Vadovaudamiesi Genest ir Rivest procedūra atlikome tokius veiksmus:

- Iš (5) lygybės įvertinome Kendalo koreliacijos koeficientą, kurį gavome lygų $\hat{\tau}_C = 0,59628469$;

- Remdamiesi (2) išraiška, radome 1-oje lentelėje išvardintų jungčių Kendalo tau koreliacijos koeficientų išraiškas (pateikiamos A priede). Po to įvertinome Archimedo jungčių šeimų parametrus α ir patikrinome, ar jie priklauso reikiamam intervalui. Parametrų vertinimas atliktas toms 1-oje lentelėje išvardintoms jungtims, kurioms suskaičiavus (2) integralą, gautos tikslios Kendalo τ išraiškos. Tokiu būdu į nagrinėjimą įtrauktos 1, 2, 3, 4, 5, 7, 8, 9, 12, 14, 15, 18 numeriais pažymėtos jungtys. Po parametrų α įverčių galiojimo patikrinimo, iš tolesnio tyrimo pašalintos 7, 8 ir 9 jungtys.

- Turėdami parametro α įverčius $\hat{\alpha}$, įvertinome atitinkamų jungčių generatorius φ ir funkcijas K_{φ}.

- Iš nagrinėjamų duomenų sukonstravome pseudo stebėjimus Z_i, $i = 1..n$ (žr., (6)) ir jų neparametrinę pasiskirstymo funkciją $K_n(z)$ (žr., (7)) taip, kaip aprašyta 6.1 skyrelyje.



17 pav.: Žalų histograma ir Pareto pasiskirstymas su parametrais $x_0 = 9916,8$ ir $k = 1,8909$

- Gautą neparametrinį $K_n(z)$ įvertį lyginome su keletos nagrinėtų jungčių parametriniais funkcijos $K_{\varphi_n}(z)$ įverčiais. Palyginimui nubraižyti neparametrinės funkcijos K_n ir parametrinių įverčių K_{φ_n} kvantilių grafikai (19 - 27 pav.) bei suskaičiuoti atstumai $\int (K_{\varphi_n} - K_n)^2 dK_n$.

Skaičiavimų rezultatai pateikiami 3-oje lentelėje:

3 lentelė: Genest ir Rivest procedūros rezultatai

Nr.	$\hat{\alpha}$	Ar galioja	K_{φ_n}	$\int (K_{\varphi_n} - K_n)^2 dK_n$
1	2,9547013	Taip	$1,3384437t - 0,33844369t^{3,9547013}$	0,014874341
2	4,9547013	Taip	$0,79817149t + 0,20182851$	0,0042798458
3	0,50813238	Taip	$-16,727534t - 18,313737t^2 + \ln(\frac{(24593381+2540661949t)}{t})t + 1,0330674\ln(\frac{(24593381+25406619t)}{t})t^2$	0,013353830
4	2,4773506	Taip	$t - 0,40365704\ln(t)t$	0,0066529509
5	-2,3621293	Taip	$t - \frac{0,42334684\ln(e^{2,3621293t}-1)(e^{2,3621293t}-1)}{e^{2,3621293t}}$	0,17143648
6	-	-	-	-
7	0	Ne	-	-

3 lentelė: Genest ir Rivest procedūros rezultatai

8	-5,0695227	Ne	-	-
9	-	-	-	-
10	-	-	-	-
11	-	-	-	-
12	1,6515671	Taip	$1,6054855t - 0,60548554t^2$	0,011158189
13	-	-	-	-
14	1,9773506	Taip	$t + t^{0,99999999} - t^{1,5057272}$	0,0093074836
15	2,9773506	Taip	$t^{0,66413092}$	0,0044591873
16	-	-	-	-
17	-	-	-	-
18	3,3031342	Taip	$0,39451446t + 0,30274277 + 0,30274277t^2$	0,010148066
19	-	-	-	-
20	-	-	-	-
21	-	-	-	-
22	-	-	-	-

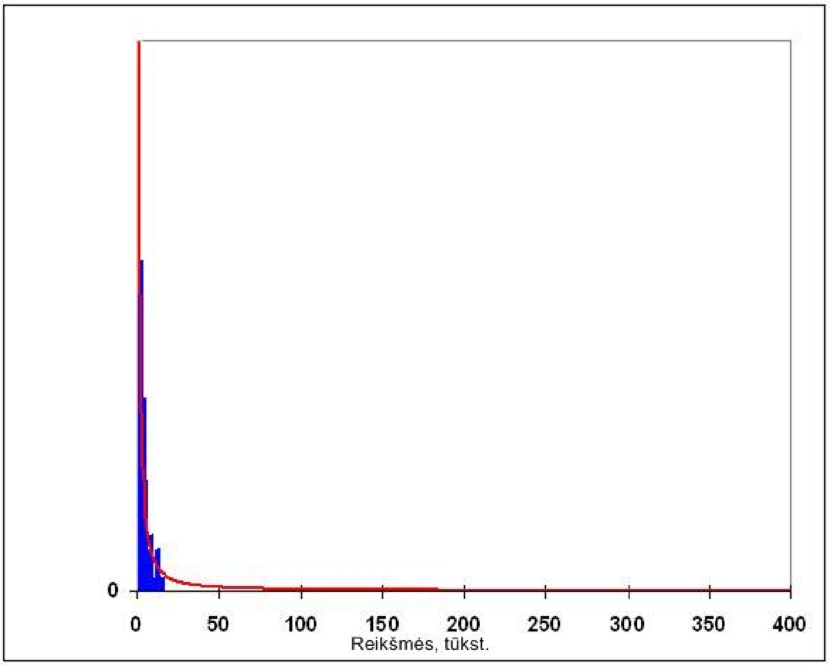
Q-Q grafikai pateikti 19 - 27 brėžiniuose. Juose horizantalioje ašyje atidėti parametriškai įvertintų funkcijų K_{φ_n} kvantiliai, o vertikalioje - neparametrinio įverčio kvantiliai.

Iš gautų kvantilių-kvantilių grafikų negalima vienareikšmiškai nuspręsti, kuri jungtis tinka labiausiai. Akivaizdžiai galime atmesti 4-osios jungties (Frank'o jungtis) variantą. 2-oji ir 18-oji jungtis gerai atitinka didelius kvantilius, tačiau matomas skirtumas tarp mažų kvantilių. 15-osios bei 14-os jungties kvantiliai vizualiai atrodo artimiausi neparametrinės funkcijos kvantiliams. Analitiškai suskaičiavus, mažiausi atstumai tarp parametrinės funkcijos įverčio ir neparametrinės funkcijos įverčio gaunami 2-os ir 15-os jungties atveju. Jei tariame, kad būtent šios dvi jungtys geriausiai nusako priklausomumą tarp nagrinėjamų žalų ir išlaidų duomenų, tai galime užrašyti ir šių dydžių sumultaninę pasiskirstymo funkciją. Paėmę 1-oje lentelėje esančią 2-os jungties išraišką ir įvertintą parametrą, turime:

$$C_{4,9547013} = \max\left(1 - \left[(1-u)^{4,9547013} + (1-v)^{4,9547013}\right]^{0,2018295}, 0\right).$$

Jei pasirenkama 15-oji jungtis, tai išraiška bus:

$$C_{2,9773506} = \max\left(0, \left\{1 - \left[(1-u^{0,335869})^{2,9773506} + (1-v^{0,335869})^{2,9773506}\right]^{0,335869}\right\}^{2,9773506}\right),$$

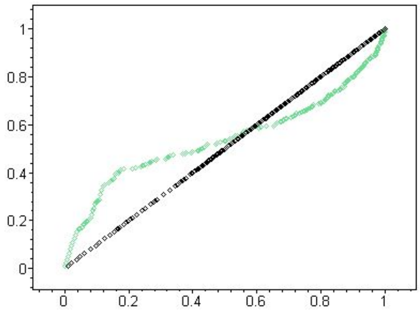


18 pav.: Išlaidų histograma ir Pareto pasiskirstymas su parametrais $y_0 = 25,83$ ir $k = 0,2716$

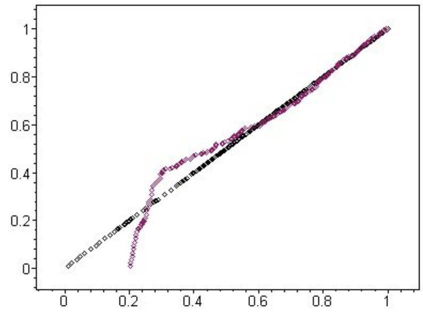
kur abiem atvejais vietoj u ir v įstatome marginalinių skirstinių išraiškas:

$$u = F(x) = 1 - \left(\frac{9916,8}{x + 9916,8)}\right)^{1,8909}$$

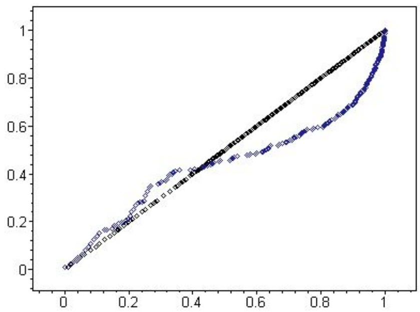
$$v = G(y) = 1 - \left(\frac{25,83}{y + 25,83}\right)^{0,2716}.$$



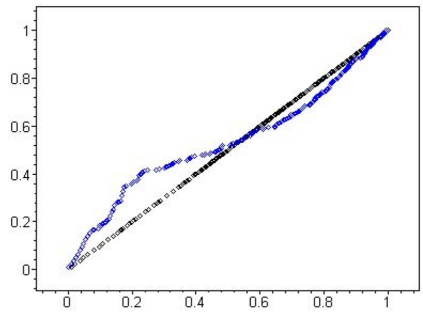
19 pav.: 1-os jungties ir neparametrinis įvertis



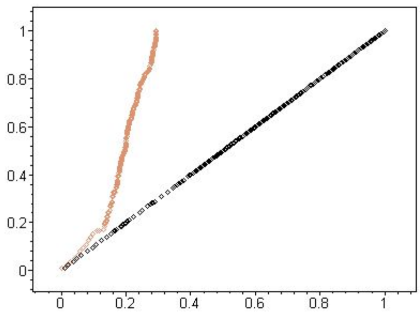
20 pav.: 2-os jungties ir neparametrinis



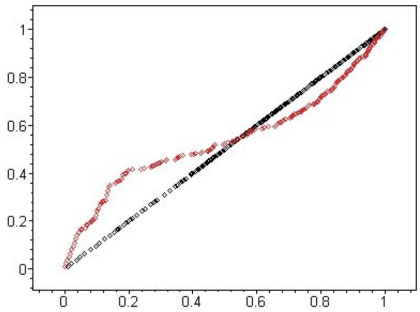
21 pav.: 3-os jungties ir neparametrinis įvertis



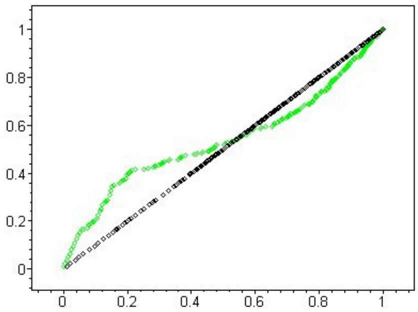
22 pav.: 4-os jungties ir neparametrinis įvertis



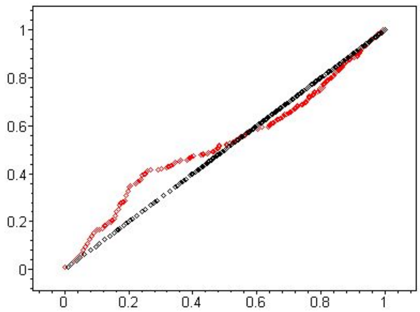
23 pav.: 5-os jungties ir neparametrinis įvertis



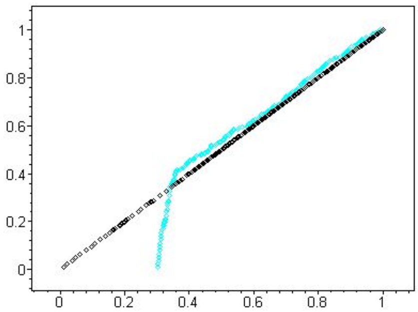
24 pav.: 12-os jungties ir neparametrinis įvertis



25 pav.: 14-os jungties ir neparametrinis įvertis



26 pav.: 15-os jungties ir neparametrinis įvertis



27 pav.: 18-os jungties ir neparametrinis įvertis

8 Išvados

Šiame darbe buvo įvesta jungties, kaip tinkamos priemonės priklausomų dydžių modeliavimui, sąvoka ir jos taikymas draudimo srityje. Nagrinėjimui pasirinkti transporto priemonių valdytojų privalomojo draudimo žalų duomenys. Pagrindinis uždavinys buvo surasti žalų ir išlaidų žaloms administruoti bendrą simultaninę pasiskirstymo funkciją. Vadovaujantis neparametrine Genest ir Rivest [1] procedūra parinktos tinkamos Archimedo jungtys. Įverčiai palyginti grafiškai Q-Q diagramose. Gauti rezultatai parodė, kad analizuojamų duomenų atveju negalima vienareikšmiškai nustatyti labiausiai tinkančios vienos jungties. Tačiau iš žinomų Archimedo jungčių šeimų atrinktos kelios, kurios galėtų būti naudojamos tolimesniuose skaičiavimuose. Turint simultaninių pasiskirstymo funkcijų išraiškas, galime spręsti įvairius aktualius draudimo uždavinius: skaičiuoti perdraudimo kainą, įvertinti ir prognuozuoti galimas išlaidas, duotiems žalų dydžiams, modeliuoti tikėtinas žalas ir išlaidas, kai žinomas jų tarpusavio priklausomumo ryšys ir pan.
Pastaruoju metu jungčių teorija populiarėja, literatūroje aprašomi nauji būdai jų parinkimui. Todėl tolimesniuose tyrimuose, tikslesnių įverčių radimui, reikėtų naudoti keletą metodų, jais gautus rezultatus palyginti ne tik grafiškai, bet ir konstruojant formalius testus.

Literatūra

[1] C. Genest ir L.-P. Rivest. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, **88(423)**: 1034-1043, 1993.

[2] E. W. Frees ir Valdez. Understanding Relationships using copulas. *North American Actuarial Journal*, **2(1)**:1-25, 1998.

[3] R. De Matteis. *Fitting Copulas to data*, 2001. *http : //www.math.ethz.ch/ degiorgi/CreditRisk/ftp/de_mattheis_fitting_copula.pdf*

[4] Mario R. Melchiori. *Which Archimedean Copula is the right one?*, 2003. *http : //www.riskglossary.com/papers/Copula_carta.PDF*

[5] R. B. Nelson. *An Introduction to Copula*. Springer, New York, 1999.

[6] R. B. Nelson. *Properties and applications of copulas: A brief survey*, *http : //www.lclark.edu/ mathsci/brazil.pdf*

A Priedas: Archimedo jungčių Kendalo τ išraiškos

Eil. Nr.	$\varphi_\alpha(t)$	$\varphi'_\alpha(t)$	$\hat{\tau}_C$
1	$\frac{1}{\alpha}\left(t^{-\alpha}-1\right)$	$-t^{(-1-\alpha)}$	$\frac{\alpha}{\alpha+2}$
2	$(1-t)^\alpha$	$-\alpha(1-t)^{(-1+\alpha)}$	$1-\frac{2}{\alpha}$
3	$\ln\frac{1-\alpha(1-t)}{t}$	$\frac{-1+\alpha}{t-\alpha t-t^2}$	$1-\frac{2}{3}\frac{\ln(1-\alpha)\alpha^2+\alpha-2\ln(1-\alpha)\alpha+\ln(1-\alpha)}{\alpha^2}$
4	$(-\ln t)^\alpha$	$\frac{\alpha(-\ln(t)^\alpha)}{t\ln(t)}$	$1-\alpha^{-1}$
5	$-\ln\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}$	$\alpha/(1-\exp(\alpha t))$	$1-\frac{4}{\alpha}[D_1(-\alpha)-1]$, čia $D_1(\alpha):=\frac{1}{\alpha}\int_0^\alpha\frac{t}{e^t-1}dt$ ir $D_1(-\alpha)=D_1(\alpha)+\frac{\alpha}{2}$
6	$-\ln[1-(1-t)^\alpha]$	$\frac{\alpha(1-t)^{-1+\alpha}}{-1+(1-t)^\alpha}$	tikslios išraiškos nėra
7	$-\ln[\alpha t+(1-\alpha)]$	$\frac{\alpha}{-1+\alpha-\alpha t}$	$\frac{2}{\alpha^2}(-\ln(1-\alpha)+\alpha^2\ln(1-\alpha)+2\alpha\ln(1-\alpha)+\alpha)$
8	$\frac{1-t}{1+(\alpha-1)t}$	$-\frac{\alpha}{(1-t+\alpha t)^2}$	$\frac{\alpha-4}{3\alpha}$
9	$\ln(1-\alpha\ln t)$	$\frac{-a}{t(1-\alpha\ln(t))}$	tikslios išraiškos nėra
10	$\ln(2t^{-\alpha}-1)$	$\frac{-2\alpha t^{-\alpha}}{t(2t^{-\alpha}-1)}$	tikslios išraiškos nėra
11	$\ln(2-t^\alpha)$	$\frac{\alpha t^{-1+\alpha}}{-2+t^\alpha}$	tikslios išraiškos nėra
12	$\left(\frac{1}{t}-1\right)^\alpha$	$-\frac{\alpha(\frac{1}{t}-1)^{-1+\alpha}}{t^2}$	$1-\frac{2}{3\alpha}$
13	$(1-\ln t)^\alpha-1$	$-\frac{\alpha(1-\ln(t))^{-1+\alpha}}{t}$	tikslios išraiškos nėra
14	$(t^{-1/\alpha}-1)^\alpha$	$-\frac{(t^{\frac{-1}{\alpha}}-1)^\alpha}{t(1-t^{\frac{1}{\alpha}})}$	$\frac{2\alpha-1}{2\alpha+1}$
15	$(1-t^{1/\alpha})^\alpha$	$-(t^{-1+\frac{1}{\alpha}}(1-t^{\frac{1}{\alpha}})^{-1+\alpha}$	$\frac{3-2\alpha}{1-2\alpha}$
16	$\left(\frac{\alpha}{t}+1\right)(1-t)$	$-1-\frac{\alpha}{t^2}$	tikslios išraiškos nėra
17	$-\ln\frac{(1+t)^{-\alpha}-1}{2^{-\alpha}-1}$	$\frac{\alpha(1+t)^{-\alpha-1}}{(1+t)^{-\alpha}-1}$	tikslios išraiškos nėra
18	$e^{\alpha/(t-1)}$	$-\frac{\alpha\exp(\frac{\alpha}{t-1})}{(t-1)^2}$	$1-\frac{4}{3\alpha}$
19	$e^{\alpha/t}-e^\alpha$	$-\frac{\alpha\exp(\frac{\alpha}{t})}{t^2}$	tikslios išraiškos nėra
20	$\exp(t^{-\alpha})-e$	$-\alpha t^{-\alpha-1}\exp(t^{-\alpha})$	tikslios išraiškos nėra
21	$1-[1-(1-t)^\alpha]^{1/\alpha}$	$-(1-t)^{\alpha-1}(1-(1-t)^\alpha)^{\frac{1}{\alpha}-1}$	tikslios išraiškos nėra
22	$\arcsin(1-t^\alpha)$	$-\frac{\alpha t^{\alpha-1}}{\sqrt{1-(-1+t^\alpha)^2}}$	tikslios išraiškos nėra

B Priedas: Skaičiavimų algoritmai

Kendalo tau skaičiavimas Archimedo jungtims

```
restart;
kendall:=proc(phi)
local f,df,rho;
description "Kendalo tau";
f:=phi;
df:=diff(f,t);
rho:=1+4*int(f/df,t=0..1);
end proc;
> #1.
phi1:=(1/a)*(t^(-a)-1);
kendall(phi1);
> #2. jungtis
phi2:=(1-t)^a;
kendall(phi2);
> #3.
phi3:=ln((1-a*(1-t))/t);
kendall(phi3);
simplify(%);
> #4. Gumbel jungtis
phi4:=-ln(t)^a;
kendall(phi4);
> #5.
phi5:=-ln((exp(-a*t)-1)/(exp(-a)-1));
kendall(phi5);
> #6.
phi6:=-ln(1-(1-t)^a);
kendall(phi6);
> #7.
phi7:=-ln(a*t+(1-a));
kendall(phi7);
> #8.
phi8:=(1-t)/(1+(a-1)*t);
kendall(phi8);
simplify(%);
> #9.
phi9:=ln(1-a*ln(t));
kendall(phi9);
> #10.
phi10:=ln(2*(t^(-a))-1);
kendall(phi10);

> #11.
phi11:=ln(2-t^a);
kendall(phi11);
> #12.
phi12:=((1/t)-1)^a;
kendall(phi12);
> #13.
phi13:=(1-ln(t))^a-1;
kendall(phi13);
```

```
> #14.
phi14:=(t^(-1/a)-1)^a;
kendall(phi14);
> #15.
phi15:=(1-t^(1/a))^a;
kendall(phi15);
> #16.
phi16:=(a/t+1)*(1-t);
kendall(phi16);
> #17.
phi17:=-ln(((1+t)^(-a)-1)/(2^(-a)-1));
kendall(phi17);
> #18.
phi18:=exp(a/(t-1));
kendall(phi18);
> #19.
phi19:=exp(a/t)-exp(a);
kendall(phi19);
> #20.
phi20:=exp(t^(-a))-exp(1);
kendall(phi20);
> #21.
phi21:=1-(1-(1-t)^a)^(1/a);
kendall(phi21);
> #22.
phi22:=arcsin(1-t^a);
kendall(phi22);
```

```
#Kendalo tau neparametrinis įvertis
n:=describe[count](data);
Digits:=8;
s:=0;
for i from 1 to n do
    for j from 1 to n do
         if j>i then s:=s+signum((X[i]-X[j])*(Y[i]-Y[j]));
         end if;
     end do;
end do;
print(s);
with(combinat, numbcomb):
C:=numbcomb(n,2);
K_tau:=(C^(-1))*s:

#Pseudo stebėjimai
Digits:=8;
n:=describe[count](data);
V:=array(1..n):
for i from 1 to n do
S:=0:
for j from 1 to n do
        if (X[j]<X[i]) and (Y[j]<Y[i]) then S:=S+1:
        end if;
    end do;
V[i]:=evalf(S/(n-1)):
end do:
```

```
print(V);
```

#-----Parametriniai K įverčiai----#
#1. Clayton copula

```
Digits:=8;
l[1]:=alpha[1]/(alpha[1]+2)=K_tau;
a[1]:=solve(l[1],alpha[1]);
if (a[1]>=-1) and (a[1]<infinity) and (a[1]<>0) then "PARAMETRAS GALIOJA";
phi[1]:=(1/a[1])*(t^(-a[1])-1);
dphi[1]:=diff(phi[1], t);
K[iv1]:=simplify(t-(phi[1]/dphi[1]));
else "PARAMETRAS NEGALIOJA!";
end if;
```

#4. Gumbel - Hougaard copula

```
Digits:=8;
l[4]:=1-alpha[4]^(-1)=K_tau;
a[4]:=solve(l[4],alpha[4]);
if (a[4]>=1) and (a[4]<infinity) then "PARAMETRAS GALIOJA";
phi[4]:=(-ln(t))^a[4];
dphi[4]:=diff(phi[4],t);
K[iv4]:=simplify(t-(phi[4]/dphi[4]));
else "PARAMETRAS NEGALIOJA";
end if;
```

#5. Frank copula

```
Digits:=8;
with(student):
Da:=value(rightsum(t/(a*(exp(t)-1)),t=0..a));
Dminus:=Da+a/2;
tau:=1-(4/a)*(Dminus-1)=.59634297;
Digits:=8;
sprend:=fsolve(tau,a);
a[5]:=-2.3621293;
if a[5]<>0 then "PARAMETRAS GALIOJA";
phi[5]:=-ln(exp(-a[5]*t)-1)/(exp(-a[5])-1);
dphi[5]:=diff(phi[5],t);
K[iv5]:=t-(phi[5]/dphi[5]);
else "PARAMETRAS NEGALIOJA!";
end if;
```

#2. copula

```
Digits:=8;
l[2]:=1-(2/alpha[2])=K_tau;
a[2]:=solve(l[2],alpha[2]);
if a[2]>=1 then "PARAMETRAS GALIOJA";
phi[2]:=(1-t)^a[2];
dphi[2]:=diff(phi[2],t);
K[iv2]:=simplify(t-(phi[2]/dphi[2]));
else "PARAMETRAS NEGALIOJA!";
end if;
```

#3. AMH copula

```
Digits:=8;
l[3]:=1+(4*(-1/(6*alpha[3]))-4*((-1+alpha[3])^2)*log(1-
alpha[3]))/(6*(alpha[3]^2))=K_tau;
```

```
a[3]:=solve(l[3],alpha[3]);
if (a[3]>=-1) and (a[3]<1) then "PARAMETRAS GALIOJA";
phi[3]:=ln((1-a[3]*(1-t))/t);
dphi[3]:=diff(phi[3],t);
K[iv3]:=simplify(t-(phi[3]/dphi[3]));
else "PARAMETRAS NEGALIOJA!";
end if;
```

#7.copula

```
Digits:=8;
l[7]:=(2*(-1+alpha[7])*(alpha[7]+log(1-alpha[7])-alpha[7]*log(1-
alpha[7])))/(alpha[7]^2)=K_tau;
a[7]:=solve(l[7], alpha[7]);
if (a[7]>0) and (a[7]<=1) then "PARAMETRAS GALIOJA";
phi[7]:=-ln(a[7]*t+(1-a[7]));
dphi[7]:=diff(phi[7], t);
K[iv7]:=t-(phi[7]/dphi[7]);
else "PARAMETRAS NEGALIOJA!";
end if;
```

#8.copula

```
Digits:=8;
l[8]:=(-4+alpha[8])/(3*alpha[8])=K_tau;
a[8]:=solve(l[8],alpha[8]);
if a[8]>=1 then "PARAMETRAS GALIOJA";
phi[8]:=(1-t)/(1+(a[8]-1)*t);
dphi[8]:=diff(phi[8],t);
K[iv8]:=simplify(t-(%%/%));
else "PARAMETRAS NEGALIOJA!";
end if;
```

#12. copula

```
Digits:=8;
l[12]:=1-(2/(3*alpha[12]))=K_tau;
a[12]:=solve(l[12], alpha[12]);
if a[12]>=1 then "PARAMETRAS GALIOJA";
phi[12]:=((1/t)-1)^a[12];
dphi[12]:=diff(phi[12],t);
K[iv12]:=simplify(t-(%%/%));
else "PARAMETRAS NEGALIOJA!";
end if;
```

#14. copula

```
Digits:=8;
l[14]:=1-(4/(2+4*alpha[14]))=K_tau;
a[14]:=solve(l[14],alpha[14]);
if a[14]>=1 then "PARAMETRAS GALIOJA";
phi[14]:=(t^(-1/a[14])-1)^a[14];
dphi[14]:=diff(phi[14],t);
K[iv14]:=simplify(t-(%%/%));
else "PARAMETRAS NEGALIOJA!";
end if;
```

#15.copula

```
Digits:=8;
l[15]:=1+(4/(2-4*alpha[15]))=K_tau;
a[15]:=solve(l[15],alpha[15]);
```

```
if a[15]>=1 then "PARAMETRAS GALIOJA";
phi[15]:=(1-t^(1/a[15]))^a[15];
dphi[15]:=diff(phi[15],t);
K[iv15]:=simplify(evalf(t-(%%/%)));
else "PARAMETRAS NEGALIOJA!";
end if;
```

 # 18.copula

```
Digits:=8;
l[18]:=1-4/(3*alpha[18])=K_tau;
a[18]:=solve(l[18],alpha[18]);
if a[18]>=2 then "PARAMETRAS GALIOJA";
phi[18]:=exp(a[18]/(t-1));
dphi[18]:=diff(phi[18],t);
K[iv18]:=simplify(t-(%%/%));
else "PARAMETRAS NEGALIOJA!";
end if;
```

#empirinė K[n] pasiskirstymo funkcija

```
data:=sort(Z);
d:=array(1..n):
for j from 1 to n do
    s:=0;
    for i from 1 to n do
    if data[i]<=data[j]
    then s:=s+1;
    end if;
    end do;
    d[j]:=evalf(s/n):
end do:
```

#Atstumo tarp K[n] ir K skaičiavimas";

```
Minimumas:=proc(KC,KN,n)
local kc,kn,m,sum,p,sub;
m:=n;
kc:=array(1..m);
kn:=array(1..m);
kc:=KC;
kn:=KN;
sum:=0;
for p from 1 to m do
sub[p]:=eval((kc[p]-kn[p])^2);
sum:=sum+sub[p];
end do;
sum:=sum/m;
print(sum);
end proc;
```