Rankové statistiky jako aparát pro další použití. Spearmanův korelační koeficient, Kendalovo tau, elementy copul.

Pořadové statistiky

Mějme (iid) náhodný výběr $\{x_1, x_2, \dots, x_n\}$ rozsahu n náhodné proměnné ξ s distribuční funkcí $F(x)$ a hustotou $f(x)$, pak vzestupně seřazená pozorování $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$, kde $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$ budeme nazývat pořádkovými statistikami nad výběrem $\{x_1, x_2, \dots, x_n\}$ a $x_{(i)}$ i-tou pořádkovou statistikou. Budeme dále předpokládat, že jednotlivá pozorování jsou po dvou různá, proto $x_{(1)} < x_{(2)} < \dots < x_{(n)}$. Pak je
$x_{(1)} = min\{x_1, x_2, \dots, x_n\}$ a $x_{(n)} = max\{x_1, x_2, \dots, x_n\}$

Distribuční funkce i-té pořádkové statistiky: $F_{(i)}(x) = P\{x_{(i)} < x\}$ to je pravděpodobnost toho, že se v náhodném výběru nalezne alespoň *i* pozorování menších než *x*. Pravděpodobnost toho, že se v náhodném výběru nalezne právě *i* pozorování menších než *x* a právě *n-i* pozorování větších nebo rovno *x* je $P_i(x) = \binom{n}{i}(F(x))^i(1-F(x))^{n-i}$. $F_{(i)}(x) = P\{x_{(i)} < x\}$ je pak součet předchozích pravděpodobností od *i* až do *n* (jedná se o disjunktní jevy). $F_{(i)}(x) = \sum_{j=i}^{n}\binom{n}{j}(F(x))^j(1-F(x))^{n-j}$.

Tím je realizované zobrazení, které konkrétnímu prvku přiřazuje jeho pořadí (rank).

Tedy: $x_i = x_{(j)} \Rightarrow x_i \rightarrow j \quad \text{jinak} \quad rank(x_i) = r_i = j$.

Příklad: náhodný výběr $\{x_1 = 3.1, x_2 = 1.2, x_3 = 4.3, x_4 = 1.8, x_5 = 3.2\}$, tomu odpovídá uspořádaný náhodný výběr: $\{x_{(1)} = 1.2, x_{(2)} = 1.8, x_{(3)} = 3.1, x_{(4)} = 3.2, x_{(5)} = 4.3\}$ a přiřazení pořadí:

Pozorování	Jeho rank – pořadí
3.1	3
1.2	1
4.3	5
1.8	2
3.2	4

Aby zobrazení: pozorování $\rightarrow$ pořadí (pořádek)=rank bylo **jednoznačné,** je nezbytné, aby v definičním vztahu $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$, kde $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$ nenastávaly rovnosti (tj. ***aby byla vyloučena shodná pozorování***). K tomu stačí, aby byla distribuční funkce pozorování spojitá a rostoucí. Dále je nutné měřit (pozorovat) náhodnou proměnnou (její realizace) na plnou přesnost.

Námět: Je zapotřebí předpoklad toho, aby byla distribuční rostoucí, nestačí jen její spojitost?

Tím dostáváme tři „typy“ statistik:
$s = (x_1, x_2, x_3, x_4, \dots, x_n)$
$p = (x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, \dots, x_{(n-1)}, x_{(n)}),\quad x_{(1)} < x_{(2)} < x_{(3)} < x_{(4)}, \dots, x_{(n-1)} < x_{(n)}$
$r = (r_1, r_2, \dots, r_{n-1}, r_n)$.

Potom: $\overline{f}(x_{(1)}, \dots, x_{(n)}) = \sum_{r \in R_n} f(x_{r_1}, \dots, x_{r_n})$ a $P\{r / x_{(1)}, \dots, x_{(n)}\} = \dfrac{f(x_{(r_1)}, \dots, x_{(r_n)})}{\overline{f}(x_{(1)}, \dots, x_{(n)})}$. Důkaz viz Jurečková J.: Pořadové testy. Skripta MFF UK, Praha 1981.str. 28,29.

Námět: (i pro další) rozdělení náhodného výběru je **symetrické** vůči jakémukoliv uspořádání argumentů $f(x_1, x_2, x_3, x_4, \dots, x_{n-1}, x_n) = \prod_{i=1}^{n} f(x_i)$.

Věta 1.: Mějme náhodný výběr $X = (x_1, x_2, x_3, x_4, \dots, x_{n-1}, x_n)$ jemu přiřazený vektor pořadových statistik $X_{()} = (x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, \dots, x_{(n-1)}, x_{(n)})$ a vektor rankových statistik $r = (r_1, r_2, r_3, r_4, \dots, r_{n-1}, r_n)$.Potom: $P(r) = \dfrac{1}{n!}$ a $\overline{f}(x_{(1)}, \dots, x_{(n)}) = n!\, f(x_{(1)}, \dots, x_{(n)})$. Důkaz je zřejmý, pro pomoc viz Jurečková J.: Pořadové testy. Skripta MFF UK, Praha 1981., str. 29-31. Navíc vektory s a r jsou nezávislé.

Důkaz: $P\{r / x_{(1)}, \dots, x_{(n)}\} = \dfrac{f(x_{(r_1)}, \dots, x_{(r_n)})}{\overline{f}(x_{(1)}, \dots, x_{(n)})} = \dfrac{f(x_{(r_1)}, \dots, x_{(r_n)})}{n!\, f(x_{(r_1)}, \dots, x_{(r_n)})} = \dfrac{1}{n!} = P\{r\}$.

Věta 2.: Za předpokladů věty 1. platí:

$$P\{r_i = j\} = \frac{1}{n} \qquad i, j \in \{1, 2, \dots, n\}$$

$$P\{r_i = k, r_j = m\} = \frac{1}{n(n-1)} \qquad i, j, k, m \in \{1, 2, \dots, n\}; i \neq j; k \neq m$$

$$E\{r_i\} = \frac{n+1}{2} \qquad i \in \{1, 2, \dots, n\}$$

$$\sigma^2\{r_i\} = \frac{n^2 - 1}{12} \qquad i \in \{1, 2, \dots, n\}$$

$$Cov(r_i, r_j) = -\frac{n+1}{12} \qquad i, j \in \{1, 2, \dots, n\}; i \neq j$$

Důkaz:

$$P\{r_i = j\} = \sum_{(j_2, \dots, j_n)} P(r_1 = j_2, \dots, r_{i-1} = j_i, r_i = j, r_{i+1} = j_{i+1}, \dots, r_n = j_n) = \sum_{(j_2, \dots, j_n)} \frac{1}{n!} = \frac{(n-1)!}{n!} = \frac{1}{n}$$

Stejným způsobem se odvodí: $P\{r_i = k, r_j = m\} = \dfrac{1}{n(n-1)}$.Ostatní jsou již triviální důsledky předchozího. Viz též součtové vzorce konečných řad v příloze.

Párová pozorování

Mějme k dispozici náhodný výběr (***iid***):
$(X,Y)=\{(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)...,(x_{n-1},y_{n-1}),(x_n,y_n)\}$
a jemu odpovídající rankové statistiky $r(x_i)=r_i; s(y_i)=s_i$.

Potom ***Spearmanův koeficient korelace*** $r_S=Corr(r,s)$ - přesněji jeho statistický ekvivalent:

$$r_S=\frac{\frac{1}{n}\sum_{i=1}^{n} r_i s_i-\bar{r}\bar{s}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i-\bar{r})^2\ \frac{1}{n}\sum_{i=1}^{n}(s_i-\bar{s})^2}}$$. Protože:

$\bar{r}=\bar{s}=\frac{n+1}{2}$ a $\frac{1}{n}\sum_{i=1}^{n}(r_i-\bar{r})^2=\frac{1}{n}\sum_{i=1}^{n}(s_i-\bar{s})^2=\frac{1}{n}\sum_{i=1}^{n}(i)^2-\left(\frac{n+1}{2}\right)^2=\frac{n^2-1}{12}$, platí:

$$r_S=\frac{12}{n(n^2-1)}\sum_{i=1}^{n} r_i s_i-\frac{3(n+1)}{n-1}.$$

Nebo také ekvivalentně: $r_S=1-\frac{6}{n(n^2-1)}\sum_{i=1}^{n}(r_i-s_i)^2$.

Pokud jsou (X,Y) **nezávislé,** jsou i (r,s), pak $E\{r_S\}=\frac{12}{n(n^2-1)}\sum_{i=1}^{n}E\{r_i\}E\{s_i\}-\frac{3(n+1)}{n-1}=$ $=\frac{12}{n(n^2-1)}\sum_{i=1}^{n}\left(\frac{n+1}{2}\right)^2-\frac{3(n+1)}{n-1}=\frac{12n}{n(n^2-1)}\left(\frac{n+1}{2}\right)^2-\frac{3(n+1)}{n-1}=0$. Tedy $E\{r_S\}=0$ (ovšem za uvedené hypotézy nezávislosti), proto $\sigma^2\{r_S\}=E\{r_S^2\}$.

$E\{r_S\}=0$ implikuje $E\left\{\sum_{i=1}^{n} r_i s_i\right\}=\frac{n(n+1)^2}{4}$ a $\sigma^2\left\{\sum_{i=1}^{n} r_i s_i\right\}=\frac{n^2(n+1)^2(n-1)}{144}$. Dále platí (za hypotézy nezávislosti):

$$\sigma^2\{r_S\}=\left(\frac{12}{n(n^2-1)}\right)^2\sigma^2\left\{\sum_{i=1}^{n} r_i s_i\right\}=\left(\frac{12}{n(n^2-1)}\right)^2\frac{n^2(n+1)^2(n-1)}{144}=\frac{144}{n^2(n+1)^2(n-1)^2}\frac{n^2(n+1)^2(n-1)}{144}=\frac{1}{n-1}$$

Shrnutí za hypotézy nezávislosti: $E\{r_S\}=0$, $\sigma^2\{r_S\}=\frac{1}{n-1}$

Za hypotézy nezávislosti je navíc rozdělení asymptoticky normální.
Námět: použijte této skutečnosti pro test nezávislosti obou „párovaných" náhodných veličin.

Diskuse: Spearmanův koeficient korelace je „jistým zúžením" ***maximální korelace*** (viz Alfréd Rényi: Teorie pravděpodobnosti, ACADEMIA, Praha 1972, str. 243-249) $\psi(X,Y)=\sup_{u,v}[Corr(u(X),v(Y))]$, kde se supremum bere přes všechny možné *borelovsky měřitelné* funkce $u(X),v(Y)$, pro které existují

střední hodnoty a rozptyly. Spearmanův koeficient korelace je vlastně klasickou (Pearsonovou) korelací pořadí párovaných náhodných výběrů.
Pokud použijeme třídu transformačních funkcí $u(X), v(Y)$, které jsou obě rostoucí, jinak libovolné (s odpovídající podmínkou existence středních hodnot a rozptylů), evidentně se pro celou takovou třídu nezmění pořadí při srovnání podle velikosti. Obdobné, leč modifikované, tvrzení platí, pokud se zabýváme třídou dvojic transformačních funkcí, které jsou obě klesající.
Námět: Změní se, pokud ano, situace pokud jedna z funkcí bude rostoucí a druhá klesající?
Shrnutí: Spearmanův koeficient korelace (jeho statistické vyjádření) „měří sílu korelační" závislosti a je **invariantní** vůči **ryze monotónním transformacím**.

Kendalův koeficient korelace pořadí – Kendalovo τ

$$\tau_n = \frac{1}{n(n-1)} \sum_{i \neq j} sign(r_i - r_j) sign(s_i - s_j)$$

Evidentně $-1 \leq \tau_n \leq 1$. Kendalovo τ vlastně +1 čítá shody $(x_i < x_j) \wedge (y_i < y_j)$ a $(x_i > x_j) \wedge (y_i > y_j)$ (zachování monotonie = izotonie) a čítá -1 $(x_i < x_j) \wedge (y_i > y_j)$ a $(x_i > x_j) \wedge (y_i < y_j)$ (zachování proti-monotonie = antiizotonie). Všech možných porovnatelných párů je n^2 stejné páry se (pro trivialitu) nesrovnávají proto všech srovnávaných dvojic je $n^2 - n = n(n-1)$. Opět testuje „monotónní závislost" = testuje četnost jevů „když jedna z proměnných v páru se změní, změní se i druhá a to stejným směrem". Výše uvedený tvar je vhodný pro názorné pochopení smyslu testového kritéria, výpočetně jsou vhodnější jiná vyjádření, např.:

- $\tau_n = \frac{2}{n(n-1)} \sum_{j<i} sign(r_i - r_j) sign(s_i - s_j)$.
- Seřadíme-li pozorování ***párů*** podle pořadí r_i (nebo ekvivalentně podle x_i) vzestupně budeme mít nová pořadí r_i^0 a s_i^0, kde $r_i^0 = i$, pak ale $sign(r_i^0 - r_j^0) = 1 \text{ pro } j < i$. Odtud $\tau_n = \frac{2}{n(n-1)} \sum_{j<i} sign(s_i^0 - s_j^0)$.
- Používá se také statistiky $K_n^+ = \sum_{j<i} sign(s_i^0 - s_j^0)$ s lepší výpočetní strukturou nebo $K_n^+ = 2\sum_{j<i} \operatorname{sgn}(s_i^0 - s_j^0) - \frac{n(n-1)}{2}, \text{kde} \operatorname{sgn}(x) = 1 \Leftrightarrow x > 0, \operatorname{sgn}(x) = 0 \Leftrightarrow x \leq 0$.
  Námět: navrhněte efektivní algoritmus výpočtu $\sum_{j<i} \operatorname{sgn}(s_i^0 - s_j^0)$.
- … .

Pokud jsou X, Y nezávislé, pak platí: $E\{\tau_n\} = 0$ a $\sigma^2(\tau_n) = \frac{2(2n+5)}{9n(n-1)}$ a rozdělení je asymptoticky normální.
Námět: použijte této skutečnosti pro test nezávislosti obou „párovaných" náhodných veličin.
Shrnutí: Kendalovo τ (jeho statistické vyjádření) „měří sílu izotónní" závislosti a je **invariantní** vůči **ryze monotónním transformacím**.

Význam rankové statistiky:

$$(X,R)=\{(x_1,r_1),(x_2,r_2),(x_3,r_3),(x_4,r_4)\ldots,(x_{n-1},r_{n-1}),(x_n,r_n)\}$$

r_i je počet pozorování menších nebo rovno x_i. Potom poměr $\frac{r_i}{n}$ je odhadem pravděpodobnosti jevu: $\xi \le x_i$. Proto: $E\left\{\frac{r_i}{n}\right\}=\frac{E\{r_i\}}{n}=\frac{nP\{\xi \le x_i\}}{n}=P\{\xi \le x_i\}=F_\xi(x_i)$ pro absolutně spojitou náhodnou proměnnou (= se spojitou hustotou).

Copula – aneb **vydělení** závislostních vztahů

Theorem 1. [Sklar, 1959]

Mějme dvě náhodné proměnné ξ,η s marginálními distribučními funkcemi $F_\xi(x),F_\eta(y)$ a sdruženou distribuční funkcí $F_{\xi\eta}(x,y)$. Nechť existují inverzní funkce $F_\xi^{-1}(x),F_\eta^{-1}(y); x\in\{z:0<F_\xi(z)<1\}, y\in\{z:0<F_\eta(z)<1\}$ k marginálním distribučním funkcím $F_\xi(x),F_\eta(y)$. Pak existuje jediná funkce $C:\langle 0,1\rangle \mathrm{x} \langle 0,1\rangle \to \langle 0,1\rangle$ taková, že $F_{\xi\eta}(x,y)=C(F_\xi(x),F_\eta(x))$.

Důkaz (naznačení):

Nechť $u=F_\xi(x), v=F_\eta(y)$, pak $F_\xi^{-1}(u)=x, F_\eta^{-1}(v)=y$ a dále $F_{\xi\eta}(F_\xi^{-1}(u),F_\eta^{-1}(v))=C(u,v)$. To je vlastně předpis pro hledanou funkci $C(u,v)$.

Význam:

Zavedeme novou náhodnou proměnnou $\nu=F_\xi(\xi)$. $F_\nu(x)=P\{F_\xi(\xi)<x\}$. Vzhledem k předpokladu existence inverzí a k neklesajícímu charakteru $F_\xi(x)$ pak platí: $F_\nu(x)=P\{F_\xi(\xi)<x\}=P\{\xi<F_\xi^{-1}(x)\}=F_\xi(F_\xi^{-1}(x))=x \Leftrightarrow x\in\langle 0,1\rangle$. Tedy $\nu=F_\xi(\xi)$ je rovnoměrně rozdělená náhodná proměnná na intervalu $\langle 0,1\rangle$. Protože totéž můžeme udělat pro η je $C(u,v)$ sdruženou distribuční funkcí dvou rovnoměrně rozdělených náhodných proměnných na čtverci: $\langle 0,1\rangle \mathrm{x} \langle 0,1\rangle$.

Shrnuto:

$$F_{\xi\eta}(x,y)=C(F_\xi(x),F_\eta(x))$$

Příklad 1: $F_{\xi\eta}(x,y)=\min\left(1-e^{-\frac{x}{\tau_x}};1-e^{-\frac{y}{\tau_y}}\right);x,y\geq 0$ $\quad F_{\xi}(x)=1-e^{-\frac{x}{\tau_x}};x\geq 0$, $\quad F_{\eta}(y)=1-e^{-\frac{y}{\tau_y}};y\geq 0$.

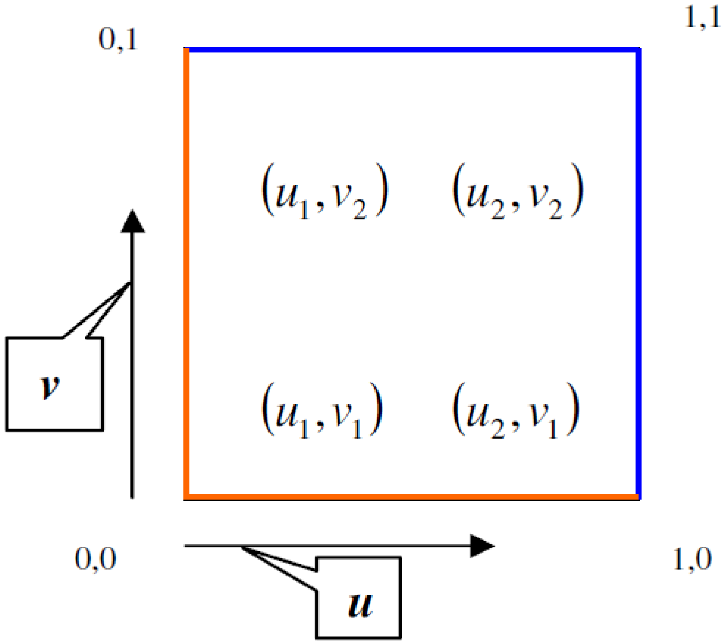
Potom: $F_{\xi\eta}(x,y)=\min\left(F_{\xi}(x);F_{\eta}(y)\right)\Rightarrow\ C(u,v)=\min(u;v)$.

Příklad 2: Dvourozměrné exponenciální rozdělení $F_{\xi\eta}(x,y)=\left(1-e^{-\frac{x}{\tau_x}}\right)\left(1-e^{-\frac{y}{\tau_y}}\right);x,y\geq 0$.

$F_{\xi}(x)=1-e^{-\frac{x}{\tau_x}};x\geq 0$, $F_{\eta}(y)=1-e^{-\frac{y}{\tau_y}};y\geq 0$. Potom: $F_{\xi\eta}(x,y)=F_{\xi}(x)F_{\eta}(y)\Rightarrow\ C(u,v)=uv$.

Vlastnosti copul (spon):

1. Pro $u,v\in\langle 0,1\rangle$ je $C(u,0)=C(0,v)=0$
2. Pro $u,v\in\langle 0,1\rangle$ je $C(u,1)=u;C(1,v)=v$
3. Pro $u_1,u_2,v_1,v_2\in\langle 0,1\rangle;u_1\leq u_2;v_1\leq v_2$ je $C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1)\geq 0$.
   $\left(\left[C(u_2,v_2)-C(u_2,v_1)\right]-\left[C(u_1,v_2)-C(u_1,v_1)\right]\geq 0\right)$



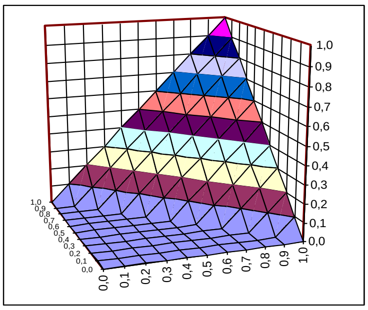
Velice volně řečeno funkce $C(u,v)$ je neklesající ve „ve všech směrech od bodu $(0,0)$ k bodu $(1,1)$“, a na „modrých“ hranicích je distribuční funkcí rovnoměrného rozdělení a na oranžových hranicích oblasti má hodnotu 0.

Další vlastnosti

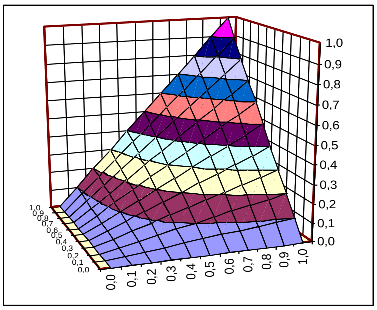
Frechetovy meze $\max(0,u+v-1)\le C(u,v)\le \min(u,v)$

Součinová copula $C(u,v)=uv$ pro případ nezávislosti.

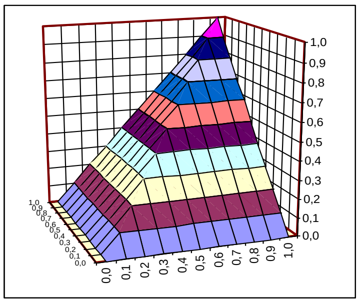
$C(u,v)$



Frechetova dolní mez

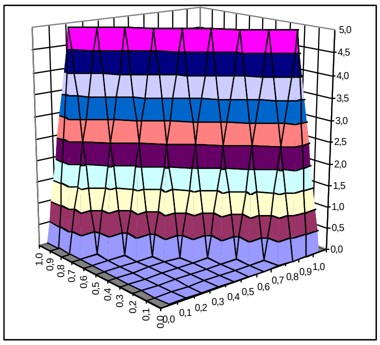


Součinová

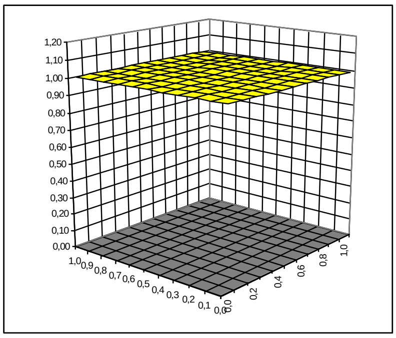


Frechetova horní mez

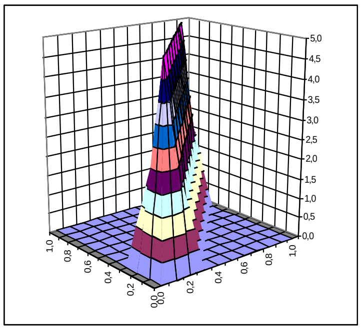
$c(u,v)$



Frechetova dolní mez



Součinová



Frechetova horní mez

Námět: Dokažte Frechetovu horní mez. Poznámka: meze platí nejen pro „copuly" ale pro jakékoliv distribuční funkce (marginály a sdružené). Pokuste se dokázat i dolní mez.

Odpovídající hustoty

$F_{\xi\eta}(x,y)=C\big(F_\xi(x),F_\eta(y)\big)$ standardní vyjádřená hustota: $f_{\xi\eta}(x,y)=\dfrac{\partial^2}{\partial x\partial y}F_{\xi\eta}(x,y)$. Označíme-li $c(u,v)=\dfrac{\partial^2}{\partial u\partial v}C(u,v)$, pak: $f_{\xi\eta}(x,y)=c\big(F_\xi(x),F_\eta(y)\big)f_\xi(x)\,f_\eta(x)$.

Test správnosti:

$f_\xi(x)=\int\limits_{-\infty}^{+\infty} f_{\xi\eta}(x,z)dz=\int\limits_{-\infty}^{+\infty} c\big(F_\xi(x),F_\eta(z)\big)f_\xi(x)\,f_\eta(z)dz=\int\limits_{0}^{+1} c\big(F_\xi(x),y\big)f_\xi(x)dy=Q$ po substituci

$y = F_\eta(z) \Rightarrow dy = f_\eta(z)dz$. Ale $Q = f_\xi(x)\int_0^{+1} c\left(F_\xi(x), y\right)dy = f_\xi(x)\int_0^{+1} \frac{\partial}{\partial u}\frac{\partial}{\partial v}C(u,v)\Big|_{\substack{u=F_\xi(x)\\ v=y}} dy =$

$$= f_\xi(x)\frac{\partial}{\partial F_\xi(x)}\int_0^{+1}\frac{\partial}{\partial y}C\left(F_\xi(x), y\right)dy = f_\xi(x)\frac{\partial}{\partial F_\xi(x)}\left[C\left(F_\xi(x),1\right) - C\left(F_\xi(x),0\right)\right] = f_\xi(x)\frac{\partial}{\partial F_\xi(x)}F_\xi(x) = f_\xi(x)$$

Námět: dokažte i druhou rovnost.

Konvexní kombinace dvou copul je také copula.

$0 \le \alpha \le 1; C(u,v) = \alpha\, C_1(u,v) + (1-\alpha)C_2(u,v)$, to bezprostředně vyplývá z vlastností distribučních funkcí.

Podmíněná distribuce.

$$F_{\xi\eta}(x/y) = \lim_{\delta\to 0}\frac{P\{\xi < x, \eta \in (y-\delta, y+\delta)\}}{P\{\eta \in (y-\delta, y+\delta)\}} = \lim_{\delta\to 0}\frac{F_{\xi\eta}(x, y+\delta) - F_{\xi\eta}(x, y-\delta)}{F_\eta(y+\delta) - F_\eta(y-\delta)} =$$

$$= \lim_{\delta\to 0}\frac{\frac{\partial}{\partial y}F_{\xi\eta}(x,y)(2\delta)}{\frac{\partial}{\partial y}F_\eta(y)(2\delta)} = \frac{\frac{\partial}{\partial y}F_{\xi\eta}(x,y)}{\frac{\partial}{\partial y}F_\eta(y)} = \frac{\frac{\partial}{\partial y}F_{\xi\eta}(x,y)}{f_\eta(y)}$$. Formulováno v termínech copul:

$$\frac{\frac{\partial}{\partial y}F_{\xi\eta}(x,y)}{f_\eta(y)} = \frac{\frac{\partial}{\partial y}C\left(F_\xi(x), F_\eta(y)\right)}{f_\eta(y)} = \frac{\frac{\partial}{\partial v}C\left(F_\xi(x), v\right)_{v=F_\eta(y)} f_\eta(y)}{f_\eta(y)} = \begin{cases}\frac{\partial}{\partial v}C\left(F_\xi(x), v\right)_{v=F_\eta(y)} \Leftrightarrow f_\eta(y) > 0\\ 0 \Leftrightarrow f_\eta(y) = 0\end{cases}$$

Shrnuto: $f_\eta(y) > 0 \Rightarrow F_{\xi\eta}(x/y) = \frac{\partial}{\partial v}C\left(F_\xi(x), v\right)_{v=F_\eta(y)}$

Podmíněná hustota.

$$f_{\xi\eta}(x/y) = \frac{\partial}{\partial x}F_{\xi\eta}(x/y) = \frac{\partial}{\partial u}\frac{\partial}{\partial v}C(u,v)_{v=F_\eta(y)}^{u=F_\xi(x)} f_\xi(x) = c\left(F_\xi(x), F_\eta(y)\right)f_\xi(x)$$

Další vztahy

$E\{x^n\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^n f_{\xi\eta}(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^n c\left(F_\xi(x), F_\eta(x)\right)f_\xi(x) f_\eta(x)dxdy =$ po substitucích $u = F_\xi(x) \Rightarrow du = f_\xi(x)dx,\ \ v = F_\eta(y) \Rightarrow dv = f_\eta(y)dy$ a $x = F_\xi^{-1}(u), x^n = \left(F_\xi^{-1}(u)\right)^n$

dostaneme $E\{x^n\} = \int_0^1\int_0^1 \left(F_\xi^{-1}(u)\right)^n c(u,v)dudv$. Obdobně:

$$E\{xy\} = \int_0^1\int_0^1 \left(F_\xi^{-1}(u)\right)\left(F_\eta^{-1}(v)\right)c(u,v)dudv.$$

Příklady copul

Table 1: One and Two Parameter Bivariate Copulas

Model Name	$C(u,v)$ where $(u,v) \in [0,1]^2$	parameters
Ali-Mikhail-Haq	$uv[1-\alpha(1-u)(1-v)]^{-1}$	$-1 < \alpha < 1$
Carrière	$(1-p)\,u\,v + p\big(u^{-\alpha} + v^{-\alpha} - 1\big)^{-1/\alpha}$	$0 \le p \le 1, \quad \alpha > 0$
Cook-Johnson	$[u^{-\alpha} + v^{-\alpha} - 1]^{-1/\alpha}$	$\alpha > 0$
Cuadras-Augé-1	$[\min(u,v)]^{\alpha}[u\,v]^{1-\alpha}$	$0 \le \alpha \le 1$
Cuadras-Augé-2	$u^{1-\alpha}v^{1-\beta}\min(u^{\alpha}, v^{\beta})$	$0 \le \alpha, \beta \le 1$
Frank	$\alpha^{-1}\ln[1 + (e^{\alpha u} - 1)(e^{\alpha v} - 1)(e^{\alpha} - 1)^{-1}]$	$\alpha \neq 0$
Fréchet-1	$p\max(0, u+v-1) + (1-p)\min(u,v)$	$0 \le p \le 1$
Fréchet-2	$p\max(0, u+v-1) + (1-p-q)u\,v + q\min(u,v)$	$0 \le p, q \le 1$
Gauss	$G\Big(\Phi^{-1}(u), \Phi^{-1}(v) \parallel \rho\Big)$	$-1 < \rho < 1$
Gumbel	$\exp\big\{-[(-\ln u)^{\alpha} + (-\ln v)^{\alpha}]^{1/\alpha}\big\}$	$\alpha > 0$
Morgenstern	$u\,v\,[1 + 3\rho\,(1-u)(1-v)]$	$-1/3 < \rho < 1/3$
Plackett	$\frac{1+(\alpha-1)(u+v)-\sqrt{1+(\alpha-1)(u+v)^2+4\alpha(1-\alpha)}}{1/2(\alpha-1)}$	$\alpha \ge 0$
t-copula	$C_{\mathcal{T}}(u, v \mid \rho, r)$	$-1 < \rho < 1, r > 0$
Yashin-Iachine	$(u\,v)^{1-p}\,(u^{-\alpha} + v^{-\alpha} - 1)^{-p/\alpha}$	$0 \le p \le 1, \quad \alpha > 0$

Příklady copul. Zdroj: J. F. CARRI`ERE: ***COPULAS***. Mathematical and Statistical Sciences University of Alberta, Edmonton, Canada, j.carriere@ualberta.ca

$$C_r^{Gauss}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-r^2}} e^{-\left(\frac{x^2-2rxy+y^2}{2(1-r^2)}\right)} dx\,dy; -1 \le r \le +1$$

Kendalovo τ a Spearmanův koeficient pořadové korelace pravděpodobnostně, vyjádřené pomocí aparátu copul.

Spearmanova korelace

Pravděpodobnostní ekvivalent k/ke statistickému Spearmanovu korelačnímu koeficientu pro dvojici náhodných proměnných (X,Y) je $r_S = Corr(F_X(x), F_Y(y))$ neboť $Corr(\alpha x+\beta, \gamma y+\delta) = Corr(x,y)$.

Mějme náhodnou proměnnou ξ popsanou hustotou a distribuční funkcí $f(x), F(x)$ - k $F(x)$ existuje její inverzní funkce $x = F^{-1}(y)\ \forall y \in (0,1)$, dále mějme transformovanou náhodnou proměnnou $\eta = F(\xi)$. Pak:
$F_\eta(x) = P(\eta < x) = P(F(\xi) < x) = P(\xi < F^{-1}(x)) = F(F^{-1}(x)) = x \Leftrightarrow x \in (0,1)$, tedy náhodná proměnná $\eta = F(\xi)$ má rovnoměrné rozdělení na intervalu $(0,1)$. Proto $E\{\eta\} = \frac{1}{2}$ a $\sigma^2\{\eta\} = \frac{1}{12}$.

Proto:

$$r_S = Corr(F_X(x), F_Y(x)) = \frac{Cov(F_X(x), F_Y(y))}{\sigma\{F_X(x)\}\sigma\{F_Y(y)\}} = \frac{Cov(F_X(x), F_Y(y))}{\frac{1}{\sqrt{12}}\frac{1}{\sqrt{12}}} = 12Cov(F_X(x), F_Y(y))$$

$$Cov(F_X(x), F_Y(y)) = E\{F_X(x)F_Y(y)\} - E\{F_X(x)\}E\{F_Y(y)\} = E\{F_X(x)F_Y(y)\} - \frac{1}{2}\frac{1}{2} =$$

$$= E\{F_X(x)F_Y(y)\} - \frac{1}{4}$$

$E\{F_X(x)F_Y(y)\} = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} F_X(x)F_Y(y)c(F_X(x), F_Y(y))f_X(x)f_Y(y)dxdy$, po substituci

$u = F_X(x); v = F_Y(y)$ dostaneme:

$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} F_X(x)F_Y(y)c(F_X(x), F_Y(y))f_X(x)f_Y(y)dxdy = \int\limits_0^1\int\limits_0^1 uv\,c(u,v)dudv$. Tedy:

$$r_S = Corr(F_X(x), F_Y(x)) = 12\left(\int\limits_0^1\int\limits_0^1 uv\,c(u,v)dudv - \frac{1}{4}\right) = 12\int\limits_0^1\int\limits_0^1 uv\,c(u,v)dudv - 3.$$

Protože ale: $\int\limits_0^1\int\limits_0^1 uv\,dudv = \frac{1}{4}$ a $\int\limits_0^1\int\limits_0^1 uv\,c(u,v)dudv = \int\limits_0^1\int\limits_0^1 C(u,v)dudv$, dostáváme:

$$r_S = 12\int\limits_0^1\int\limits_0^1 uv\,c(u,v)dudv - 3 \quad \text{nebo}$$

$$r_S = 12\int\limits_0^1\int\limits_0^1 uv\left(c(u,v) - 1\right)dudv \quad \text{nebo}$$

$$r_S = 12\int\limits_0^1\int\limits_0^1 C(u,v)dudv - 3 \quad \text{nebo}$$

$$r_S = 12\int\limits_0^1\int\limits_0^1 \left(C(u,v) - uv\right)dudv.$$

Kde: $C(u,v)$ je copula a $c(u,v)$ je její copulová hustota.

C(u,v)	
min(u,v)	$r_S = 12\int_0^1\int_0^1 C(u,v)dudv - 3 = 12\int_0^1\int_0^1 \min(u,v)dudv - 3 = 12\int_0^1\int_0^v udu dv + 12\int_0^1\int_v^1 vdudv - 3 = 12\frac{2}{6} - 3 = +1$
max(0,u+v-1)	$r_S = 12\int_0^1\int_0^1 C(u,v)dudv - 3 = 12\int_0^1\int_0^1 max(0,u+v-1)dudv - 3 = 12\int_0^1\int_{1-v}^1 (u+v-1)dudv - 3 = 12\frac{1}{6} - 3 = -1$
uv	$r_S = 12\int_0^1\int_0^1 uvdudv - 3 = 12\frac{1}{2}\frac{1}{2} - 3 = 0$
uv(1+α(1-u)(1-v)) ***-1≤α≤+1***	$r_S = 12\int_0^1\int_0^1 C(u,v)dudv - 3 = 12\int_0^1\int_0^1 uv(1+\alpha(1-u)(1-v))dudv - 3 = 12 = 12\frac{9-2\alpha}{36} - 3 = \frac{-2\alpha}{3}$ ***-1/3≤rS≤1/3***

Odtud se někdy pro Frechetovu dolní mez $\max(0, u+v-1)$ používá názvu maximální „negativní závislost“ a pro horní mez $\min(u,v)$ názvu maximální „pozitivní závislost“. A poloha konkrétní copuly mezi těmito mezemi „stupněm závislosti“.

Kendalovo τ

Obdobnými úvahami jako v případě Spearmanova koeficientu pořadové korelace lze odvodit pro Kendalovo τ:

$$\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1$$

, kde $dC(u,v) = c(u,v)dudv$, pokud copulová hustota existuje a $C(u,v)$ je copula a $c(u,v)$ je její copulová hustota.

Doporučená a zdrojová literatura:

Jaroslav Hátle, Jiří Likeš	Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974.
Alfréd Rényi	Teorie pravděpodobnosti, ACADEMIA, Praha 1972
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
Machek J.	Teorie odhadu, SPN Praha 1974, skripta MFFUK
Jurečková J.	Pořadové testy. Skripta MFF UK, Praha 1981.
Hájek J., Vorlíčková D.	Neparametrické metody. Skripta MFFUK, Praha 1974
J. F. CARRI`ERE	COPULAS. Mathematical and Statistical Science. University of Alberta, Edmonton, Canada. j.carriere@ualberta.ca
Tchilabalo Abozou Kpanzou	Copulas in Statistics. African Institute for Mathematical Sciences (AIMS). Supervised by Tertius De Wet, University of Stellenbosch, May 24, 2007.

Příloha – součtové vzorce konečných řad

$$\sum_{i=1}^{n} i = \quad 1 \quad + \quad 2 \quad + \quad 3 \quad + \quad \dots \quad +(n-1) \quad +n$$

$$\sum_{i=1}^{n} i = \quad n \quad + \quad (n-1) \quad + \quad (n-2) \quad + \quad \dots \quad +2 \quad +1$$

$$2\sum_{i=1}^{n} i = \quad (n+1) \quad + \quad (n+1) \quad + \quad (n+1) \quad + \quad \dots \quad +(n+1) \quad +(n+1)$$

$\Rightarrow$

$$2\sum_{i=1}^{n} i = n(n+1) \Rightarrow \qquad \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Odtud:

$$\left(\sum_{i=1}^{n} i\right)^2 = \left(\frac{n(n+1)}{2}\right)^2 = \left(\sum_{i=1}^{n} i\right)^2 = \sum_{i=1}^{n} i^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} ij = \sum_{i=1}^{n} i^2 + 2\sum_{i=2}^{n} i\sum_{j=1}^{i-1} j = \sum_{i=1}^{n} i^2 + 2\sum_{i=2}^{n} i\frac{i(i-1)}{2} =$$

$$= \sum_{i=1}^{n} i^2 + \sum_{i=2}^{n} i^2(i-1) = \sum_{i=1}^{n} i^2 + \sum_{i=1}^{n} i^3 - \sum_{i=1}^{n} i^2 = \sum_{i=1}^{n} i^3$$, a odtud:

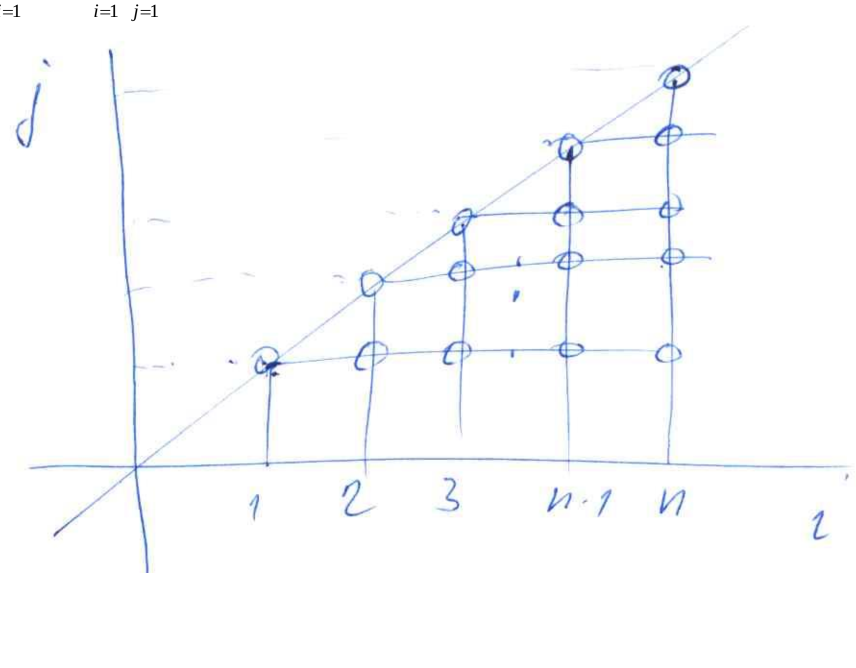
$$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}$$

Ze vztahu $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ bezprostředně plyne:

$$\sum_{i=1}^{n}(2i-1) = 2\sum_{i=1}^{n} i - n = 2\frac{n(n+1)}{2} - n = n^2 + n - n = n^2 \qquad \Rightarrow \qquad i^2 = \sum_{j=1}^{i}(2j-1)$$

a pak:

$\sum_{i=1}^{n} i^2 = \sum_{i=1}^{n}\sum_{j=1}^{i}(2j-1) = \mathrm{A}$. Záměnou sumačního pořadí:



$A=\sum_{j=1}^{n}\sum_{i=j}^{n}(2j-1)=\sum_{j=1}^{n}(n-j+1)(2j-1)=2n\sum_{j=1}^{n}j-n^2-2\sum_{j=1}^{n}j^2+\sum_{j=1}^{n}j+2\sum_{j=1}^{n}j-n=$

$=2n\frac{n(n+1)}{2}-n^2-2\sum_{j=1}^{n}j^2+3\frac{n(n+1)}{2}-n$. Shrnuto:

$\sum_{i=1}^{n}i^2=2n\frac{n(n+1)}{2}-n^2-2\sum_{i=1}^{n}i^2+3\frac{n(n+1)}{2}-n\Rightarrow 3\sum_{i=1}^{n}i^2=2n\frac{n(n+1)}{2}-n^2+3\frac{n(n+1)}{2}-n=$

$3\sum_{i=1}^{n}i^2=n^2(n+1)-n^2+3\frac{n(n+1)}{2}-n=n^3+n^2-n^2+\frac{3}{2}n^2+\frac{3}{2}n-n=\frac{2n^3+3n^2+n}{2}=$

$=\frac{n(2n^2+3n+1)}{2}$, ale $(2n^2+3n+1)=2\left(n+\frac{1}{2}\right)(n+1)=(2n+1)(n+1)$.

Proto: $\frac{n(2n^2+3n+1)}{2}=\frac{n(2n+1)(n+1)}{2}$. Opět shrnuto: $3\sum_{i=1}^{n}i^2=\frac{n(2n+1)(n+1)}{2}$ a proto:

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

Výsledky:

$$\sum_{i=1}^{n}i=\frac{n(n+1)}{2}$$

$$\sum_{i=1}^{n}i^2=\frac{n(2n+1)(n+1)}{6}$$

$$\sum_{i=1}^{n}i^3=\frac{n^2(n+1)^2}{4}$$

Námět: Na základě uvedeného odvoďte součty vyšších mocnin.