Závislost, nezávislost a vztahy náhodných (číselných, převážně spojitých) proměnných (nekategoriálních)[1]

Definice 3: Náhodné (číselné) proměnné X, Y [2] jsou nezávislé (značeno $X \perp Y$) právě když platí $F_{X,Y}(x,y) = F_X(x)F_Y(y); \forall x, y \in R_1^2$ [3].

Elementární vlastnosti:

- $X \perp Y \Rightarrow Y \perp X$.
- $X \perp Y \Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y); \forall x, y \in R_1^2$ [4], pokud všechny uvedené hustoty existují. Platí to i pro pravděpodobnosti (v případě diskrétních náhodných proměnných) důkaz analogický důkazu následujícího tvrzení (o kvantování).
- Pokud K je degenerovaná náhodná proměnná[5] (konstanta), pak pro libovolnou náhodnou proměnnou Y je $K \perp Y$.
  Dk.: $F_{K,}(x,y) = P(K \le x, Y \le y) = \begin{array}{ll} P(0, Y \le y) = 0 = 0 * P(Y \le y) = F_K(x)F_Y(y) & \Leftarrow x < k. \\ P(\Omega, Y \le y) = P(Y \le y) = 1 * P(Y \le y) = F_K(x)F_Y(y) & \Leftarrow x \ge k. \end{array}$
  Volněji: každá náhodná proměnná je nezávislá na konstantě.
- Pokud existují $E(X), E(Y)$ a $E(XY)$, pak $X \perp Y \Rightarrow E(XY) = E(X)E(Y)$.
- Pokud existují $E(X), E(Y)$ a $E(XY)$, pak $E(XY) \ne E(X)E(Y) \Rightarrow X \not\perp Y$
  Dk.: $[E(XY) \ne E(X)E(Y)] \wedge [X \perp Y]$ je spor.

Nezávislost a kvantování

Tvrzení: $X \perp Y$, $-\infty = x_0 < x_1 < x_2 < \ldots < x_n < x_{n+1} = +\infty$ dělící body definičního oboru X a $-\infty = y_0 < y_1 < y_2 < \ldots < y_m < y_{m+1} = +\infty$ dělící body definičního oboru Y a $1 \le n, m \in N$. Označíme-li $p_{ij} = P(x_{i-1} < X \le x_i, y_{j-1} < Y \le y_j)$, pak platí $\mathbf{p_{ij}} = \mathbf{P}(\mathbf{x_{i-1}} < \mathbf{X} \le \mathbf{x_i})\mathbf{P}(\mathbf{y_{j-1}} < \mathbf{Y} \le \mathbf{y_j}) =$ $= [F(x_i) - F(x_{i-1})][F(y_j) - F(y_{j-1})]$, pro $i, j \in N$.

Dk.: Obecně platí $p_{ij} = F(x_i, y_j) - F(x_{i-1}, y_j) - F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1})$, viz následující obrázek.

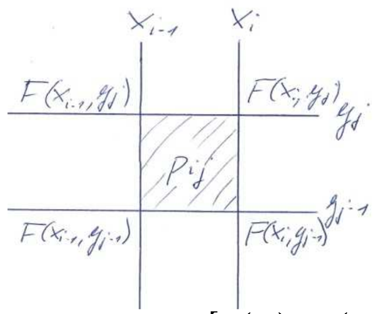
[1] Dependence relations between random variables is one of the most widely studied subjects in probability and statistics. The nature of the dependence can take a variety of forms and unless some specific assumptions are made about the dependence, no meaningful statistical model can be contemplated. Jogdeo K (1982) Concepts of dependence. In: Kotz S, Johnson NL (eds) Encyclopedia of Statistical Sciences, Vol 1. Wiley, New York, pp 324-334. Citováno z: Roger B. Nelsen: An Introduction to Copulas. Second Edition. Springer 2006. str. 171.

[2] Náhodná proměnná (jednorozměrná) X je měřitelná funkce na pravděpodobnostním prostoru $X : \omega \rightarrow R_1$.

[3] Distribuční funkce $F_{X,Y}(x,y) = P(X \le x, Y \le y)$ a $F_X(x) = P(X \le x)$. Navíc, v tomto textu budou analyzovány jen binární a z nich odvozené vztahy.

[4] $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$ a $f_X(x) = \frac{\partial}{\partial x} F_X(x)$ v bodech kde příslušné derivace existují.

[5] $K = k \in R_1$, kde $F_K(x) = 0 \Leftrightarrow x < k$ a $F_K(x) = 1 \Leftrightarrow x \ge k$.



Dosazením $F(x,y)=F(x)F(y)$, dostaneme $p_{ij}=[F(x_i)-F(x_{i-1})][F(y_j)-F(y_{j-1})]$.

Příklad: Opačná implikace nemusí platit. Vezměme dvě stejně rozdělené exponenciální proměnné X, Y s distribuční funkcí $F(x)=1-e^{-x}$ propojené copulou $C(u,v)=p\max(u+v-1,0)+(1-p)\min(u,v)$, tedy $F_{X,Y}(x,y)=p\max(F(x)+F(y)-1,0)+(1-p)\min(F(x),F(y))$, $p=\dfrac{1}{2}$ (existuje alespoň jedna dvojice $(x,y)\in(0,+\infty)$, pro kterou platí $F_{X,Y}(x,y)\neq F_X(x)F_Y(y)$) a dělení $0=x_0<x_1=\lg(2)<x_{n+1}=+\infty$, $0=y_0<y_1=\lg(2)<y_{n+1}=+\infty$. Je zřejmé, že $F(x_1)=F(y_1)=F(\lg(2))=1-e^{-\lg(2)}=\dfrac{1}{2}$ a

x – vodorovně, y – svisle v tabulce $F_{X,Y}(x,y)$	0	$\lg(2)$	$+\infty$
0	0	0	0
$\lg(2)$	0	$1/4$	$1/2$
$+\infty$	0	$1/2$	1

X – vodorovně, Y – svisle, v tabulce pravděpodobnosti oblastí	$(0,\lg(2))$	$\langle\lg(2),+\infty)$
$(0,\lg(2))$	$1/4$	$1/4$
$\langle\lg(2),+\infty)$	$1/4$	$1/4$

Tedy, ač se jedná o kvantování dvou závislých proměnných, pravděpodobnosti, pro toto kvantování platí $P(x_{i-1}<X\le x_i, y_{j-1}<Y\le y_j)=P(x_{i-1}<X\le x_i)P(y_{j-1}<Y\le y_j), i,j=1,2$, projevuje se nezávislost.

Nezávislost a korelovanost

Příklad: Nekorelované náhodné veličiny nemusí být nezávislé: $X\approx N(0,1)$, $Y=X^2$, pak $(x\in R_1, y\in R_{1+})\Rightarrow P(X\le x, Y\le y)=P(X\le x, X^2\le y)$.
$(X\le x, X^2\le y)\Rightarrow(X\le x, -\sqrt{y}\le X\le+\sqrt{y})$. V případě kdy $x<-\sqrt{y}$ je $(X\le x, X^2\le y)=\emptyset$ a proto $P(X\le x, Y\le y)=P(\emptyset)=0$. Ale $P(X\le x)>0$ a $P(Y\le y)>0$. Pak $P(X\le x, Y\le y)\neq P(X\le x)P(Y\le y)$. **Tedy X, Y NEJSOU nezávislé.**

$Cov(X,Y)=E(XY)-E(X)E(Y)=E(X^3)-E(X)E(Y)=0-0*1=0$, $\sigma(X)=1, \sigma(Y)=\sigma(X^2)=\sqrt{2}$ [6], proto $Corr(X,Y)=0$. **Tedy** X,Y **JSOU nekorelované.**

Námět: Lze předpoklad $X \approx N(0,1)$ zeslabit a jak?

Příklad: Buď $X \approx N(0,1)$ a $\{x_1, x_2, ..., x_n\}$ náhodný výběr realizací náhodné proměnné X, pak $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ a $x_i - \bar{x}$ jsou nezávislé $i = 1,...,n$.

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}\right] = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2\right)}.$$

Zavedeme transformaci:

$y_1 = \bar{x}$	$\bar{x} = y_1$	$\Rightarrow x_1 = y_1 - \sum_{i=2}^{n} y_i$ [7]
$y_2 = x_2 - \bar{x}$	$x_2 = y_2 + y_1$	
…		
$y_n = x_n - \bar{x}$	$x_n = y_n + y_1$	

Odtud: Jakobián této transformace $\det \mathbf{J} = \det\left[\left.\frac{\partial x_i}{\partial y_j}\right|_{i,j=1,...,n}\right]$ je nenulový a funkčně nezávislý na $\left. y_j \right|_{j=1,...,n}$. Tedy je vzhledem k proměnným $\left. y_j \right|_{j=1,...,n}$ konstantou.

Po této transformaci bude $\sum_{i=1}^{n} x_i^2 = x_1^2 + \sum_{i=2}^{n} x_i^2 = \left(y_1 - \sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n}(y_i + y_1)^2 =$

$$= \left[y_1^2 - 2y_1\sum_{i=2}^{n} y_i + \left(\sum_{i=2}^{n} y_i\right)^2\right] + \left[\sum_{i=2}^{n} y_i^2 + 2y_1\sum_{i=2}^{n} y_i + (n-1)y_1^2\right] = \left(\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2 + ny_1^2.$$

Proto $f(y_1, y_2, ..., y_n) = c(n) e^{-\frac{1}{2}\left(\left(\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2 + ny_1^2\right)} = c(n) e^{-\frac{1}{2}\left(\left(\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2\right)} e^{-\frac{n}{2}y_1^2} = f(y_2, ..., y_n) f(y_1)$ a $c(n) = (2\pi)^{-\frac{n}{2}} |\det \mathbf{J}|$.

Důsledek $f(y_j, y_1) = f(y_j) f(y_1); j = 2,...,n$. Tedy $y_j = x_j - \bar{x}$ a $y_1 = \bar{x}$ jsou nezávislé pro $j = 2,...,n$.

Abychom dokázali nezávislost pro $x_1 - \bar{x}$ a $\bar{x}$, stačí místo původní transformace

---

[6] χ^2 rozdělení s jedním stupněm volnosti.

[7] $y_1 = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(x_1 + \sum_{i=2}^{n} x_i\right)$, ale $\sum_{i=2}^{n} x_i = \sum_{i=2}^{n} y_i + (n-1)y_1$, proto

$$y_1 = \frac{1}{n}\left(x_1 + \sum_{i=2}^{n} x_i\right) = \frac{1}{n}\left(x_1 + \sum_{i=2}^{n} y_i + (n-1)y_1\right) \Rightarrow x_1 = y_1 - \sum_{i=2}^{n} y_i .$$

$y_1 = \bar{x}$	$\bar{x} = y_1$
$y_2 = x_2 - \bar{x}$	$x_2 = y_2 + y_1$
…	…
$y_n = x_n - \bar{x}$	$x_n = y_n + y_1$

použít např. transformaci:

$y_n = \bar{x}$	$\bar{x} = y_n$
$y_1 = x_1 - \bar{x}$	$x_1 = y_1 + y_n$
…	…
$y_{n-1} = x_{n-1} - \bar{x}$	$x_{n-1} = y_{n-1} + y_n$

Shrnuto $x_j - \bar{x}$ a $\bar{x}$ jsou nezávislé dvojice pro $j = 1,\dots,n$ pro $\{x_1, x_2, \dots, x_n\}$ náhodný výběr realizací náhodné proměnné $X \approx N(0,1)$.

Důsledek 1: $(x_j - \bar{x})^2$ a $\bar{x}$ jsou nezávislé pro $j = 1,\dots,n$ pro $\{x_1, x_2, \dots, x_n\}$ náhodný výběr realizací náhodné proměnné $X \approx N(0,1)$.

Důsledek 2: $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$ a $\bar{x}$ jsou nezávislé pro $\{x_1, x_2, \dots, x_n\}$ náhodný výběr realizací náhodné proměnné $X \approx N(0,1)$.

Důsledek 3: $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$ a $\bar{x}$ jsou nezávislé pro $\{x_1, x_2, \dots, x_n\}$ náhodný výběr realizací náhodné proměnné $X \approx N(\mu, \sigma^2)$.

Příklad: Buď $X_i \approx F(x), E(X_i) = 0, E(X_i^2) > 0, i = 1,\dots,n > 1$ (nemusí, ale mohou být nezávislé), pak pro $Y = \sum_{i=}^{n} X_i$ platí $Y \not\perp X_1$.

Dk.: $Cov(Y, X_1) = E(X_1 Y) = E\left(X_1 \sum_{i=1}^{n} X_i\right) = E(X_1^2) > 0$.

Důsledek: Buď $X_i \approx F(x), \sigma^2(X_i) > 0, i = 1,\dots,n > 1$, pak pro $Y = \sum_{i=}^{n} X_i$ platí $Y \not\perp X_1$. Dk.: $X_i - E(X_i) \approx F(x + E(X)), i = 1,\dots,n$. $E(X - E(X)) = 0, E\left((X - E(X))^2\right) = \sigma^2(X) > 0, i = 1,\dots,n$.

Příklad: Buď X, Y $X \perp Y$ číselné náhodné proměnné a jejich zobrazení $U = g_1(X), V = g_2(Y), g_1, g_2 : R_1 \to R_1$, spojité, pak $U \perp V$.

Dk.: $F_{U,V}(u,v) = P(U \le u, V \le v) = P(g_1(X) \le u, g_2(Y) \le v) = P(g_1(X) \le u, g_2(Y) \le v) =$

$$= P\left(\bigcup_i \left[g_{1i}^{-d}(u) \le X \le g_{1i}^{-h}(u)\right], \bigcup_j \left[g_{2j}^{-d}(v) \le Y \le g_{2j}^{-h}(v)\right]\right) = A,$$

kde $g_{1i}^{-d}(u) \le X \le g_{1i}^{-h}(u), i = 1,\dots,$ jsou všechna možná a disjunktní řešení nerovnosti $g_1(X) \le u$ vůči X a $g_{2j}^{-d}(v) \le Y \le g_{2j}^{-h}(v), j = 1,\dots,$ jsou všechna možná a disjunktní řešení nerovnosti $g_2(Y) \le v$ vůči Y [8]. Pak:

$$A = \sum_{\substack{i,j \\ g_{1i}^{-d}(u) < g_{1i}^{-h}(u) \\ g_{2j}^{-d}(v) < g_{2j}^{-h}(v)}} P\left(g_{1i}^{-d}(u) \le X \le g_{1i}^{-h}(u), g_{2j}^{-d}(v) \le Y \le g_{2j}^{-h}(v)\right) =$$

[8] $g_{1i}^{-d}(u)$, $g_{1i}^{-h}(u)$, $g_{2j}^{-d}(v)$, $g_{2j}^{-h}(v)$ mohou být i nevlastní.

$$= \sum_{\substack{i,j \\ g_{1i}^{-d}(u)<g_{1i}^{-h}(u) \\ g_{2j}^{-d}(v)<g_{2j}^{-h}(v)}} P\left(g_{1i}^{-d}(u)\le X\le g_{1i}^{-h}(u)\right)P\left(g_{2j}^{-d}(v)\le Y\le g_{2j}^{-h}(v)\right)=$$

$= \sum_{\substack{i \\ g_{1i}^{-d}(u)<g_{1i}^{-h}(u)}} P\left(g_{1i}^{-d}(u)\le X\le g_{1i}^{-h}(u)\right) \sum_{\substack{j \\ g_{2j}^{-d}(v)<g_{2j}^{-h}(v)}} P\left(g_{2j}^{-d}(v)\le Y\le g_{2j}^{-h}(v)\right)=H(u)K(v)$, kde

$\sum_{\substack{i \\ g_{1i}^{-d}(u)<g_{1i}^{-h}(u)}} P\left(g_{1i}^{-d}(u)\le X\le g_{1i}^{-h}(u)\right)=H(u)$ a $\sum_{\substack{j \\ g_{2j}^{-d}(v)<g_{2j}^{-h}(v)}} P\left(g_{2j}^{-d}(v)\le Y\le g_{2j}^{-h}(v)\right)=K(v)$, odtud už je jenom formulační a technický problém dokončit $F_{U,V}(u,v)=H(u)K(v)=F_U(u)F_V(v)$.

Příklad: Buď X,Y $X\perp Y$ nedegenerované číselné náhodné proměnné, mající hustotu a jejich zobrazení $U=g_1(X,Y), V=g_2(X,Y)$, $g_1,g_2 : R_2 \to R_1$, spojité, pak **nemusí** být $U\perp V$.

Př.: $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$, odtud $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} U \\ V \end{pmatrix}$, pak $f_{U,V}(u,v)=f_X(u)f_Y(v-u)|1|$. Odtud

$f_U(u)=f_X(u)\int_{-\infty}^{+\infty} f_Y(v-u)dv=f_X(u)$ a

$f_V(v)=\int_{-\infty}^{+\infty} f_X(u)f_Y(v-u)du$.

Pokud by $U\perp V$, muselo by platit $f_{U,V}(u,v)=f_U(u)f_V(v)=f_X(u)\int_{-\infty}^{+\infty} f_X(z)f_Y(v-z)dz$, tedy: $f_X(u)\int_{-\infty}^{+\infty} f_X(z)f_Y(v-z)dz \overset{by}{=} f_X(u)f_Y(v-u)$ a pro oblast kde $f_X(u)>0$ (a ta je neprázdná) $\int_{-\infty}^{+\infty} f_X(z)f_Y(v-z)dz \overset{by}{=} f_Y(v-u)$. Pak při volbě $u=0$ by $\int_{-\infty}^{+\infty} f_X(z)f_Y(v-z)dz = f_Y(v)$, tedy $Y \overset{by}{=} Y+X$ a to by byl spor (s předpokladem nedegenerovanosti X,Y). V případě, že existují $E(X), E(Y), E(XY)$ je důkaz jednodušší (přes kovariance, viz výše).

Tvrzení: $Corr(X,aY+b)=sign(a)Corr(X,Y)$. Dk.: dosazením do definice.

Tvrzení: $|Corr(X,Y)|=1 \Leftrightarrow Y=aX+b$ s pravděpodobností jedna.
Dk.[9]:

- $Y=aX+b$ pak $E(Y)=aE(X)+b$, $\sigma(Y)=|a|\sigma(X)$, $Cov(X,Y)=aE\left((X-E(X))^2\right)$, $Corr(X,Y)=\dfrac{aE\left((X-E(X))^2\right)}{|a|\sigma^2(X)}=sign(a)$
- $Corr(X,Y)=1$, pak, při označení $X'=\dfrac{X-E(X)}{\sigma(X)}$ a $Y'=\dfrac{Y-E(Y)}{\sigma(Y)}$, bude $Corr(X',Y')=1$, tj. $E(X'Y')=1$, proto $E\left((X'-Y')^2\right)=E(X'^2)-2E(X'Y')+E(Y'^2)=1-2+1=0$. Shrnuto $E\left((X'-Y')^2\right)=0$ a odtud

[9] Viz Alfréd Rényi: Teorie pravděpodobnosti, ACADEMIA, Praha 1972, str. 111-113.

$P(X' - Y' = 0) = 1$[10]. Při návratu k původnímu značení $P\left(\frac{Y-E(Y)}{\sigma(Y)} = \frac{X-E(X)}{\sigma(X)}\right) = 1$ a po úpravě $P\left(Y = \frac{\sigma(Y)}{\sigma(X)}(X - E(X)) + E(Y)\right) = 1$. $Corr(X,Y) = -1$, analogické předchozímu, při využití $E\left((X' + Y')^2\right) = 0$.

Spearmanova korelace

Každou z náhodných proměnných $(X,Y) \approx F_{X,Y}(x,y), F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x,y)$, $F_Y(y) = \lim_{x \to +\infty} F_{X,Y}(x,y)$ transformujeme vlastní distribuční funkcí na $R(0,1)$. Tedy $U = F_X(X)$, $V = F_Y(Y)$, proto $E(U) = E(V) = \frac{1}{2}$, $\sigma^2(U) = \sigma^2(V) = \frac{1}{12}$. Pro jednoduchost budeme předpokládat existenci obou $F_X^{-1}(u), F_Y^{-1}(v)$ všude tam, kde $0 < F_X(x), F_Y(y) < 1$.

$Cov(U,V) = E(UV) - E(U)E(V) = E(UV) - \frac{1}{4}$

V terminologii copul[11]:

$$E(UV) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F_X(x)F_Y(y)f_{X,Y}(x,y)dxdy = \int_0^1\int_0^1 uvc(u,v)dudv =$$

$$= \int_0^1\int_0^1 uv\frac{\partial^2}{\partial u \partial v}C(u,v)dudv = \int_0^1\int_0^1 C(u,v)dudv, \text{ pak}$$

$$Cov(U,V) = \int_0^1\int_0^1 C(u,v)dudv - \frac{1}{4} \Rightarrow Corr(U,V) = \frac{\int_0^1\int_0^1 C(u,v)dudv - \frac{1}{4}}{\sqrt{\frac{1}{12}}\sqrt{\frac{1}{12}}} = 12\int_0^1\int_0^1 C(u,v)dudv - 3,$$

shrnuto: $Corr(U,V) = Corr(F_X(X), F_Y(Y)) = 12\int_0^1\int_0^1 C(u,v)dudv - 3 \overset{ozn}{=} Corr_S(X,Y)$.

Důsledek: $X \perp Y \Rightarrow Corr_S(X,Y) = 0$. Dk.: Pokud $X \perp Y$ je $C(u,v) = uv$ a $12\int_0^1\int_0^1 uvdudv - 3 = 0$.

Důsledek: $F_{X,Y}(x,y) = \min(F_X(x), F_Y(y)) \Rightarrow Corr_S(X,Y) = 1$.

Dk.: $Corr_S(X,Y) = 12\int_0^1\int_0^1 C(u,v)dudv - 3 = 12\int_0^1\int_0^1 \min(u,v)dudv = 12\int_0^1\int_0^v ududv + 12\int_0^1\int_v^1 vdudv - 3 = 12\frac{2}{6} - 3 = +1$

---

10 $E\left((X' - Y')^2\right) = 0$ a $E(X' - Y') = 0$ dá $\sigma^2(X' - Y') = 0$. Pak z Čebyševovy nerovnosti $\forall \varepsilon > 0$ je $P(|Z - E(Z)| \le \varepsilon) \ge 1 - \frac{\sigma^2(Z)}{\varepsilon^2}$ v tomto případě dostaneme $\forall \varepsilon > 0$ je $P(|X' - Y'| \le \varepsilon) \ge 1$.

11 $F_{X,Y}(x,y) = C(F_X(x), F_Y(y))$, odtud $C(u,v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v))$ a $f_{X,Y}(x,y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y)$, kde $c(u,v) = \frac{\partial^2}{\partial u \partial v}C(u,v)$. Protože $C(u,v)$ je vlastně sdruženou distribuční funkcí dvou marginálů s rozdělením $R(0,1)$, platí pro ni: $\max(u + v - 1, 0) \le C(u,v) \le \min(u,v)$ (Fréchetovy meze).

Důsledek: $F_{X,Y}(x,y)=\max\left(F_X(x)+F_Y(y)-1,0\right)\Rightarrow Corr_S(X,Y)=-1$.

Dk.: $Corr_S(X,Y)=12\int_0^1\int_0^1 C(u,v)dudv-3=12\int_0^1\int_0^1 \max(0,u+v\text{-}1)dudv-3=12\int_0^1\int_{1-v}^1 (u+v-1)dudv-3=12\frac{1}{6}-3=-1$

Shrnuto: Dosažení Frechetových mezí znamená dosažení mezí pro Spearmanův korelační koeficient. Nezávislé náhodné proměnné mají Spearmanův korelační koeficient rovný nule.

Příklad: Ač platí $X\perp Y\Rightarrow Corr_S(X,Y)=0$ obrácená implikace $Corr_S(X,Y)=0\Rightarrow X\perp Y$ platit nemusí. Stačí vzít např. následující copulu: $C(u,v)=\frac{1}{2}\max(u+v-1,0)+\frac{1}{2}\min(u,v)$. Pak

$$Corr_S(X,Y)=12\int_0^1\int_0^1 C(u,v)dudv-3=12\int_0^1\int_0^1\left(\frac{1}{2}\max(u+v-1,0)+\frac{1}{2}\min(u,v)\right)dudv-3=$$

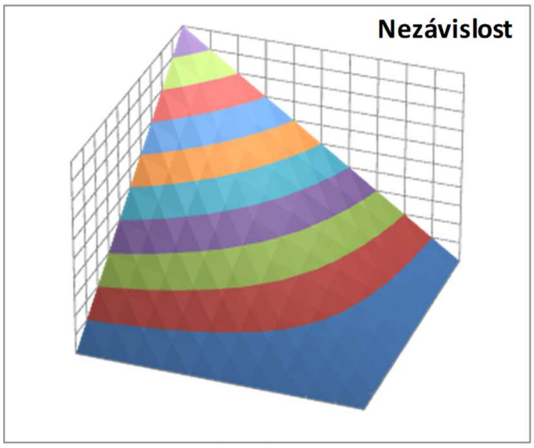
$$=\frac{1}{2}\frac{12}{6}+\frac{1}{2}\frac{24}{6}-3=0$$ (Viz předcházející důsledky).

Nezávislost

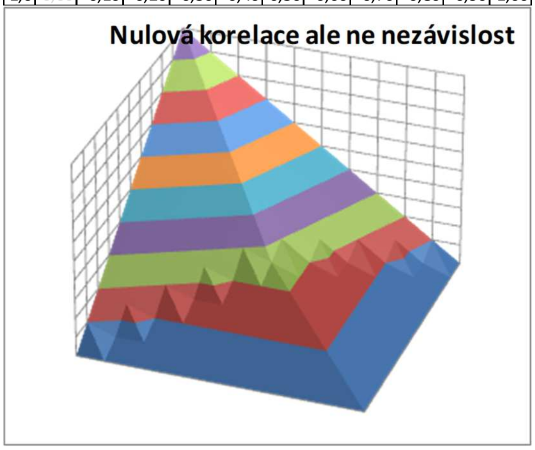
	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0,0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,1	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10
0,2	0,00	0,02	0,04	0,06	0,08	0,10	0,12	0,14	0,16	0,18	0,20
0,3	0,00	0,03	0,06	0,09	0,12	0,15	0,18	0,21	0,24	0,27	0,30
0,4	0,00	0,04	0,08	0,12	0,16	0,20	0,24	0,28	0,32	0,36	0,40
0,5	0,00	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0,6	0,00	0,06	0,12	0,18	0,24	0,30	0,36	0,42	0,48	0,54	0,60
0,7	0,00	0,07	0,14	0,21	0,28	0,35	0,42	0,49	0,56	0,63	0,70
0,8	0,00	0,08	0,16	0,24	0,32	0,40	0,48	0,56	0,64	0,72	0,80
0,9	0,00	0,09	0,18	0,27	0,36	0,45	0,54	0,63	0,72	0,81	0,90
1,0	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00

Nulová korelace ale ne nezávislost

	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0,0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,1	0,00	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,10
0,2	0,00	0,05	0,10	0,10	0,10	0,10	0,10	0,10	0,10	0,15	0,20
0,3	0,00	0,05	0,10	0,15	0,15	0,15	0,15	0,15	0,20	0,25	0,30
0,4	0,00	0,05	0,10	0,15	0,20	0,20	0,20	0,25	0,30	0,35	0,40
0,5	0,00	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0,6	0,00	0,05	0,10	0,15	0,20	0,30	0,40	0,45	0,50	0,55	0,60
0,7	0,00	0,05	0,10	0,15	0,25	0,35	0,45	0,55	0,60	0,65	0,70
0,8	0,00	0,05	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,75	0,80
0,9	0,00	0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,90
1,0	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00



$C(u,v)=uv$



$C(u,v)=\left(\frac{1}{2}\max(u+v-1,0)+\frac{1}{2}\min(u,v)\right)$

Tvrzení: Buď $g(x)$ neklesající funkce s jednoznačnou inverzí tam, kde $0<F_X(x)<1$ pak $Corr_S(X,Y)=Corr_S(g(X),Y)$. [12]

Dk.: Buď $F_{X,Y}(x,y)$ sdružená distribuční funkce, $f_{X,Y}(x,y)$ jí odpovídající hustota a $C_{X,Y}(u,v)=F_{X,Y}(F_X^{-1}(u),F_Y^{-1}(v))$. $Z=g(X)$ odtud $X=g^{-1}(Z)$, $F_{Z,Y}(z,y)=F_{X,Y}(g^{-1}(z),y)$ a

[12] Viz též Roger B. Nelsen: An Introduction to Copulas. Second Edition. Springer 2006. str. 25-26.

$F_Z(z) = F_X(g^{-1}(z))$[13]. Proto $C_{Z,Y}(u,v) = F_{Z,Y}(F_Z^{-1}(u), F_Y^{-1}(v)) = F_{Z,Y}(g(F_X^{-1}(u)), F_Y^{-1}(v)) =$ $= F_{X,Y}(g^{-1}(g(F_X^{-1}(u))), F_Y^{-1}(v)) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v))$.

Shrnuto: $C_{Z,Y}(u,v) = C_{X,Y}(u,v)$. Volněji: Jednoznačně invertovatelná (neklesající) transformace kterékoliv z proměnných mění pouze model marginálu a ne copulu. Protože $Corr_S(X,Y) = 12\int_0^1\int_0^1 C(u,v)dudv - 3$ nemění tedy ani hodnotu Spearmanova korelačního koeficientu. Totéž platí i pro rostoucí transformaci obou proměnných. Podobné je to i pro jednu (nebo obě) z obou transformací[12] klesající.

Kendalovo tau

Konkordance (Concordance)

Definice: Mějme náhodný výběr $\{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_j, y_j), \ldots, (x_n, y_n)\}$ z náhodných proměnných (X,Y) se spojitými distribučními funkcemi $F_X(x), F_Y(y)$ a sdruženou distribuční funkcí $F_{X,Y}(x,z)$. O páru $(x_i, y_i), (x_j, y_j);\ 1 \le i, j \le n;\ i \ne j$ řekneme, že je **konkordantní**[14] (concordant), pokud platí $(x_i < x_j \Rightarrow y_i < y_j) \vee (x_i > x_j \Rightarrow y_i > y_j)$. Obdobně, o páru $(x_i, y_i), (x_j, y_j);\ 1 \le i, j \le n;\ i \ne j$ řekneme, že je **diskordantní** (discordant), pokud platí $(x_i < x_j \Rightarrow y_i > y_j) \vee (x_i > x_j \Rightarrow y_i < y_j)$. Vzhledem ke spojitosti X a Y může nastat jen právě jeden z jevů $(x_i < x_j \Rightarrow y_i < y_j) \vee (x_i > x_j \Rightarrow y_i > y_j)$ a $(x_i < x_j \Rightarrow y_i > y_j) \vee (x_i > x_j \Rightarrow y_i < y_j)$. Konkordanci a diskordanci lze ekvivalentně vyjádřit $(x_i - x_j)(y_i - y_j) > 0$ a $(x_i - x_j)(y_i - y_j) < 0$.

Pravděpodobnostní formulace (tj. pravděpodobnost, že budeme v náhodném výběru pozorovat konkordantní pár):

Mějme dvě nezávislé ale stejně rozdělené dvojice náhodných proměnných (X_1, Y_1) a (X_2, Y_2)[15], pak pravděpodobnost, že budeme v náhodném výběru pozorovat konkordantní pár, bude $P((X_1 - X_2)(Y_1 - Y_2) > 0) = P(X_1 > X_2, Y_1 > Y_2) + P(X_1 < X_2, Y_1 < Y_2)$. Pak:

$$P(X_1 > X_2, Y_1 > Y_2) = P(X_2 < X_1, Y_2 < Y_1) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(X_2 < x, Y_2 < y) dF_{XY}(x,y) =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C(F_X(x), F_Y(y)) dC(F_X(x), F_Y(y)) = \int_0^1\int_0^1 C(u,v) dC(u,v)$$[16]. Podobně:

$P(X_1 < X_2, Y_1 < Y_2) = \int_0^1\int_0^1 C(u,v)dC(u,v)$ a proto: $\mathbf{P((X_1 - X_2)(Y_1 - Y_2) > 0) = 2\int_0^1\int_0^1 C(u,v)dC(u,v)}$.

---

[13] Zde je využit předpoklad, že $g(x)$ je rostoucí.

[14] Souhlasný, … .

[15] S distribucemi $F_{X_1}(x) = F_{X_2}(x) = F_X(x),\ F_{Y_1}(y) = F_{Y_2}(x) = F_Y(x),\ F_{XY}(x,y) = C(F_X(x), F_Y(y))$.

[16] S pomocí copulové hustoty $dC(F_X(x), F_Y(y)) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y) dxdy$ a $dC(u,v) = c(u,v)dudv$.

Kendalovo tau

Pro Kendalovo tau platí $\tau_{X,Y} = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0)$. Je to tedy rozdíl mezi pravděpodobností pozorování toho, že pozorovaný pár bude konkordátní a pravděpodobností toho, že pozovaný pár bude diskordantní.

Předpoklad spojitosti distribučních funkcí obou proměnných vylučuje shody a proto: $P((X_1 - X_2)(Y_1 - Y_2) < 0) + P((X_1 - X_2)(Y_1 - Y_2) > 0) = 1$ a proto $P((X_1 - X_2)(Y_1 - Y_2) < 0) = 1 - P((X_1 - X_2)(Y_1 - Y_2) > 0)$ a odtud $\tau_{X,Y} = 2P((X_1 - X_2)(Y_1 - Y_2) > 0) - 1$.

Dosazením výše odvozeného vztahu $P((X_1 - X_2)(Y_1 - Y_2) > 0) = 2\int_0^1\int_0^1 C(u,v)dC(u,v)$, dostaneme

$$\boldsymbol{\tau_{X,Y} = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1}$$

Příklad:

Buď $C(u,v) = uv$, pak $dC(u,v) = dudv$ a $\tau_{X,Y} = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 = 4\int_0^1\int_0^1 uvdudv - 1 = 0$.

Příklad: $C(u,v) = \min(u,v)$, pak $dC(u,v)$ je „nenulový“ v $\langle 0,1\rangle^2$ jen na podmnožině $u = v$, musí platit pro každou funkci $g(u,v)$, která má na $\langle 0,1\rangle^2$ integrál $\int_0^1\int_0^1 g(u,v)dC(u,v) = \int_0^1 g(u,u)du$. Pak $\tau_{X,Y} = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 = 4\int_0^1 udu - 1 = 1$.

Příklad: $C(u,v) = \max(0, u+v\text{-}1)$, pak $dC(u,v)$ je „nenulový“ v $\langle 0,1\rangle^2$ jen na podmnožině $u = 1 - v$, musí platit pro každou funkci $g(u,v)$, která má na $\langle 0,1\rangle^2$ integrál $\int_0^1\int_0^1 g(u,v)dC(u,v) = \int_0^1 g(u,1-u)du$. Pak

$$\tau_{X,Y} = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 = 4\int_0^1\int_0^1 \max(0,u+v\text{-}1)dC(u,v) - 1 = 4\int_0^1 \max(0,u+(1-u)\text{-}1)du - 1 =$$

$= 4\int_0^1 0du - 1 = -1$[17].

Poznámka: $\tau_{X,Y} = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 = 4E_{U,V}(C(u,v)) - 1$.

Tvrzení: $-1 \le \tau_{X,Y} \le 1$, přičemž $\tau_{X,Y}(\max(0,u+v\text{-}1)) = -1$, $\tau_{X,Y}(uv) = 0$ a $\tau_{X,Y}(\min(u,v)) = +1$.
Dk.: Jednoduché důsledky předchozího.

Dále bude užíváno značení[18]: $W = W(u,v) = \max(0,u+v\text{-}1)$, $\Pi = \Pi(u,v) = uv$ a $M = M(u,v) = \min(u,v)$.

---

[17] Detailní, obecnější a přesné odvození viz Roger B. Nelsen: An Introduction to Copulas. Second Edition. Springer 2006. str.161-162.
[18] Opět Nelsen.

Tvrzení: $X \perp Y \Rightarrow \tau_{X,Y} = \tau_{X,Y}(uv) = 0$. Nemusí však platit opačná implikace. Dk.: První implikace je zřejmá.

Dále sestrojíme copulu $C(u,v) = \alpha W(u,v) + (1-\alpha)M(u,v)$, pak $dC(u,v) = \alpha dW(u,v) + (1-\alpha)dM(u,v)$.

$$\tau_{X,Y} = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 = 4\int_0^1\int_0^1 (\alpha W(u,v) + (1-\alpha)M(u,v))(\alpha dW(u,v) + (1-\alpha)dM(u,v)) - 1 =$$

$$= 4\alpha^2\int_0^1\int_0^1 W(u,v)dW(u,v) + 4\alpha(1-\alpha)\int_0^1\int_0^1 W(u,v)dM(u,v) + 4\alpha(1-\alpha)\int_0^1\int_0^1 M(u,v)dW(u,v) +$$

$$+ 4(1-\alpha)^2\int_0^1\int_0^1 M(u,v)dM(u,v) - 1 = A.$$

………

$$\int_0^1\int_0^1 W(u,v)dW(u,v) = \int_0^1 W(u,1-u)du = \int_0^1 \max(0, u + (1-u) - 1)du = 0$$

$$\int_0^1\int_0^1 W(u,v)dM(u,v) = \int_0^1 W(u,u)du = \int_0^1 \max(0, 2u-1)du = \int_{1/2}^1 (2u-1)du = \frac{1}{4}$$

$$\int_0^1\int_0^1 M(u,v)dW(u,v) = \int_0^1 M(u,1-u)du = \int_0^1 \min(u, 1-u)du = \frac{1}{4}$$

$$\int_0^1\int_0^1 M(u,v)dM(u,v) = \int_0^1 M(u,u)du = \int_0^1 \min(u,u)du = \int_0^1 u\,du = \frac{1}{2}$$

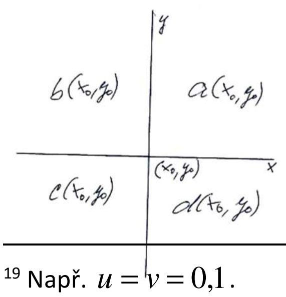
………

Pak $A = 2\alpha(1-\alpha) + 2(1-\alpha)^2 - 1 = 2(1-\alpha)(\alpha + 1 - \alpha) - 1 = 2(1-\alpha) - 1$ a stačí zvolit $\alpha = \frac{1}{2}$ aby $\tau_{X,Y} = 0$. Přitom existuje alespoň jedna dvojice (u,v) aby platilo: $\frac{1}{2}(W(u,v) + M(u,v)) \neq uv$ [19].

Shrnuto: pro $C(u,v) = \frac{1}{2}(W(u,v) + M(u,v))$ je $\tau_{X,Y} = 0$ a přitom $X \not\perp Y$ (obdobně jako pro Spearmanovu korelaci).

Kvadrantové míry konkordance[20]

Modelují a měří pravděpodobnost výskytu dvojice v jednom z kvadrantů daných bodem (x_0, y_0), viz následující obrázek. Bude značeno



$$a(x_0, y_0) = P(X > x_0, Y > y_0) = 1 - F_X(x_0) - F_Y(y_0) + F_{X,Y}(x_0, y_0)$$

$$b(x_0, y_0) = P(X \le x_0, Y > y_0) = F_X(x_0) - F_{X,Y}(x_0, y_0)$$

$$c(x_0, y_0) = P(X \le x_0, Y \le y_0) = F_{X,Y}(x_0, y_0)$$

[19] Např. $u = v = 0{,}1$.

[20] Značení, citace a zdroj dle Nils Blomqvist: On a Measure of Dependence Between two Random Variables. The Annals of Mathematical Statistics, Vol. 21, No. 4 (Dec., 1950), pp. 593-600. Dostupné na http://www.jstor.org/stable/2236609 nebo na http://projecteuclid.org/euclid.aoms/1177729754.

$d(x_0, y_0) = P(X > x_0, Y \le y_0) = F_Y(y_0) - F_{X,Y}(x_0, y_0)$

Označíme li: $u_0 = F_X(x_0), v_0 = F_Y(y_0), C(u_0, v_0) = F_{X,Y}(x_0, y_0)$, pak:

$a(x_0, y_0) = 1 - u_0 - v_0 + C(u_0, v_0)$,
$b(x_0, y_0) = u_0 - C(u_0, v_0)$,
$c(x_0, y_0) = C(u_0, v_0)$,
$d(x_0, y_0) = v_0 - C(u_0, v_0)$.

Pak lze zavést míru konkordance (za předpokladu $F_X(x), F_Y(y)$ spojité)

$\beta_0 = P((X - x_0)(Y - y_0) > 0) - P((X - x_0)(Y - y_0) < 0) =$
$= a(x_0, y_0) + c(x_0, y_0) - b(x_0, y_0) - d(x_0, y_0) = 1 - 2(u_0 + v_0) + 4C(u_0, v_0)$.

Shrnuto: $\mathbf{\beta_0 = 1 - 2(u_0 + v_0) + 4C(u_0, v_0)}$**, pro** $\mathbf{u_0 = F_X(x_0), v_0 = F_Y(y_0), C(u_0, v_0) = F_{X,Y}(x_0, y_0)}$.

Blomqvistovo beta

Pokud zvolíme (x_0, y_0) tak aby platilo $u_0 + v_0 = 1$, dostáváme $\beta_0 = 4C(u_0, v_0) - 1$, pokud vezmeme za (x_0, y_0) mediány obou marginálních rozdělení dostaneme $\beta_{X,Y} = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1$ a to bývá nazýváno Blomqvistovým beta.

Takové beta je pak rozdílem pravděpodobností obě náhodné proměnné X, Y leží ve stejném směru od svých mediánů a pravděpodobnosti toho, že obě náhodné proměnné X, Y leží v různých směrech od svých mediánů.

Příklad: Pokud $C(u,v) = W(u,v) = \max(0, u + v\text{-}1)$, pak $\beta_{X,Y} = 4W\left(\frac{1}{2}, \frac{1}{2}\right) - 1 = -1$.

Příklad: Pokud $C(u,v) = \Pi(u,v) = uv$, pak $\beta_{X,Y} = 4\Pi\left(\frac{1}{2}, \frac{1}{2}\right) - 1 = 0$.

Příklad: Pokud $C(u,v) = M(u,v) = \min(u,v)$, pak $\beta_{X,Y} = 4M\left(\frac{1}{2}, \frac{1}{2}\right) - 1 = +1$.

Poznámka: Z definice $\beta_0 = P((X - x_0)(Y - y_0) > 0) - P((X - x_0)(Y - y_0) < 0)$ bezprostředně plyne $-1 \le \beta_0 \le +1$ a tedy i $-1 \le \beta_{X,Y} \le +1$.

Poznámka: Pro nezávislé náhodné proměnné je $\beta_{X,Y} = 0$ opačně to platit nemusí. Stačí zvolit $C(u,v) = \frac{1}{2}(W(u,v) + M(u,v))$, pak je $\beta_{X,Y} = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1 = 2\left(W\left(\frac{1}{2}, \frac{1}{2}\right) + M\left(\frac{1}{2}, \frac{1}{2}\right)\right) - 1 = 2\left(0 + \frac{1}{2}\right) - 1 = 0$.

Poznámka: Blomqvistovo beta lze snadno určit bez explicitní znalosti copuly $\beta_{X,Y} = 4F_{X,Y}(x_{med}, y_{med}) - 1$.