



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

**BAKALÁŘSKÁ PRÁCE**

Alexander Terkovič

**Kendallovo tau pro diskrétní rozdělení**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Šárka Hudecová, Ph.D.

Studijní program: Obecná matematika

Praha 2024

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Na tomto mieste by som chcel vyjadriť vďaku mojej vedúcej práce, RNDr. Šárke Hudecovej, Ph.D., za jej neoceniteľnú pomoc a ochotu počas celého procesu písania. Taktiež by som chcel poďakovať Dominike Čižmárovej za podporu počas písania práce.

Název práce: Kendallové tau pro diskretní rozdělení

Autor: Alexander Terkovič

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Šárka Hudecová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Táto práca sa zaoberá Kendallovým tau ako metódou na meranie asociácie medzi dvomi diskretnými náhodnými veličinami. Prvá časť motivuje a definuje Kendallové tau, zavádza pojmy konkordancie, diskordancie a kopule, a ukazuje výpočet pre spojité rozdelenia. Druhá časť študuje zjednodušenie výpočtu pre diskretné rozdelenia pomocou spojitého rozšírenia a dokazuje, že miera asociácie sa aj po spojitom rozšírení zachová. V poslednej kapitole uvedieme ne-stranné odhady Kendallovho tau a pomocou nich porovnáme výsledky odhadov medzi diskretnými náhodnými veličinami a ich spojitými rozšíreniami.

Klíčová slova: Kendallové tau, konkordancia, diskordancia, spojité rozšírenie

Title: Kendall's tau for discrete distributions

Author: Alexander Terkovič

Department: Department of probability and mathematical statistics

Supervisor: RNDr. Šárka Hudecová, Ph.D., Department of probability and mathematical statistics

Abstract: This thesis focuses on Kendall's tau as a method for measuring association between two discrete random variables. The first part motivates and defines Kendall's tau, introduces the concepts of concordance, discordance, and copula, and demonstrates computation for continuous distributions. The second part investigates simplifying computations for discrete distributions through continuous extension and proves that the measure of association remains preserved after this extension. In the final chapter, unbiased estimates of Kendall's tau are presented and used to compare estimation results between discrete random variables and their continuous extensions.

Keywords: Kendall's tau, concordance, discordance, continuous extension

# Obsah

<b>Úvod</b>	<b>2</b>
<b>1 Kendallovo tau</b>	<b>3</b>
1.1 Pearsonov korelačný koeficient . . . . .	3
1.2 Kendallovo tau . . . . .	4
1.3 Kópula . . . . .	7
<b>2 Spojité rozšírenie diskkrétnej náhodnej veličiny</b>	<b>10</b>
2.1 Spojité rozšírenie . . . . .	10
2.2 Zachovanie usporiadania podľa konkordancie a PQD . . . . .	14
2.3 Zachovanie Kendallovho tau . . . . .	15
<b>3 Odhad Kendallovho tau</b>	<b>18</b>
3.1 Opis praktickej časti . . . . .	19
3.2 Nezávislé náhodné veličiny . . . . .	21
3.3 Závislé binárne veličiny . . . . .	22
<b>Záver</b>	<b>24</b>
<b>Zoznam použitej literatúry</b>	<b>25</b>

# Úvod

Táto práca je zameraná na skúmanie jedného z mnoha spôsobov merania asociácie dvoch náhodných veličín, konkrétne Kendalloveho tau. Kendalloveho tau je veľmi užitočný nástroj na meranie závislosti, pretože umožňuje vyjadriť mieru monotónnej závislosti medzi dvoma premennými. Často sa využíva v štatistike, ekonómii a pri analýze usporiadaných dát.

V prvej kapitole začneme krátkou motiváciou pre použitie Kendallovho tau a následne si ho zdefinujeme. Zavedieme základné pojmy, ktoré s Kendallovým tau úzko súvisia, ako napríklad konkordancia, diskordancia a kopula. K tomu vyslovíme a dokážeme užitočné tvrdenia, ktoré popisujú vlastnosti Kendallovho tau za konkrétnych podmienok. Ďalej ukážeme, ako sa dá Kendalloveho tau vypočítať pomocou kopuly v prípade, že náhodné veličiny majú spojitú rozdelenie.

V druhej kapitole sa budeme venovať jednému z riešení, ako zjednodušiť výpočet Kendallovho tau v prípade, že náhodné veličiny sú z diskrétného rozdelenia. Vysvetlíme, ako sa spojitú rozširuje diskrétna náhodná veličina a ilustrujeme to na príklade a obrázkoch. Nakoniec dokážeme, že toto rozšírenie zachováva základné vlastnosti a predovšetkým Kendalloveho tau.

Tretia kapitola bude zväčša praktická časť, ktorá je zameraná na skúmanie odhadov Kendallovho tau. Uvedieme postup, akým tieto odhady budeme počítat pre rôzne dvojice náhodných veličín. Tento postup ďalej aplikujeme na rôzne dvojice náhodných veličín a výsledky dôkladne analyzujeme a interpretujeme.

# 1. Kendallovo tau

Kendalovo tau je neparametrická miera asociácie a jeho cieľom je merať vzťah medzi dvomi náhodnými veličinami. Kendallovo tau nadobúda hodnoty z intervalu  $[-1,1]$ . Hovoríme, že dve náhodné veličiny sú negatívne korelované, ak Kendallovo nadobúda hodnoty z intervalu  $[-1, 0]$ , pozitívne korelované, ak Kendallovo nadobúda hodnoty z intervalu  $(0, 1]$ , a nekorelované, ak Kendallovo tau je rovné 0.

V tejto kapitole motivujeme skúmanie Kendallovho tau. Následne ho zdefinuujeme aj s pojmami, ktoré s ním úzko súvisia. Nakoniec ukážeme, ako sa dá jednoducho vyjadriť pomocou kopuly. V časti 1.1 sa zaoberáme predovšetkým materiálom z článku Trivedi a Zimmer (2007) a knihy Anděl (2005). Pre zvyšok kapitoly budeme používať najmä článok Denuit a Lambert (2005). V tejto kapitole je vlastným prínosom podrobné rozpísanie všetkých krokov v dôkaze vety 5 a tvrdenia 3, ako aj ilustrácia základných pojmov a vlastností prostredníctvom obrázkov a príkladov.

## 1.1 Pearsonov korelačný koeficient

Predtým, ako sa zoznámime s pojmom Kendallovo tau, ukážeme si najznámejší spôsob merania vzťahu medzi dvomi náhodnými veličinami. K tomu je potrebné zdefiniovať kovarianciu.

**Definícia 1.** *Nech  $(X,Y)^\top$  je náhodný vektor s konečnými druhými momentmi, potom kovariancia  $X$  a  $Y$  je definovaná ako*

$$\text{cov}(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Lahkou úpravou získame tvar  $\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$ , ktorý je praktickejší pre výpočty.

**Definícia 2.** *Nech  $(X,Y)^\top$  je náhodný vektor s konečnými druhými momentmi a nenulovými rozptylmi  $\sigma_X^2, \sigma_Y^2$ . Pearsonov korelačný koeficient  $X$  a  $Y$  definujeme ako*

$$\text{cor}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}.$$

Z definície môžeme vidieť, že Pearsonov korelačný koeficient nie je definovaný pre niektoré náhodné veličiny s rozdelením s ťažkými chvostmi, kde druhý moment neexistuje. V praktických aplikáciach to môže byť neprijemné obmedzenie.

**Veta 1.** *Budte  $X_1, X_2$  nezávislé náhodné veličiny a bud  $\mathbb{E}|X_i| < \infty, i = 1, 2$ . Potom  $\mathbb{E}|X_1 X_2| < \infty$  a platí  $\mathbb{E}[X_1 X_2] = \mathbb{E}X_1 \mathbb{E}X_2$ .*

*Dôkaz.* Dôkaz nájdeme na strane 33 v knihe Anděl (2005). □

Z práve spomenutej vety a definície kovariancie je vidieť, že pre dve nezávislé náhodné veličiny  $X_1, X_2$  s konečnými nenulovými rozptylmi, platí  $\text{cov}(X_1, X_2) = 0$  a teda aj  $\text{cor}(X_1, X_2) = 0$ . Opačná implikácia vo všeobecnosti neplatí, čo predvedieme na nasledujúcom príklade.

*Príklad 1.* Majme dve náhodné veličiny  $X_1 = Y$  a  $X_2 = Y^2$ , kde  $Y$  má rovnomerné rozdelenie na intervale  $[-1,1]$ . Náhodné veličiny  $X_1$  a  $X_2$  sú očividne na sebe závislé. Napriek tomu nám vyjde  $cor(X_1, X_2) = 0$ .

Nevýhodou Pearsonovho korelačného koeficientu je, že jeho hodnota závisí od marginálnych rozdelení  $X$  a  $Y$ . Rôzne vlastnosti tohto korelačného koeficientu sú skúmané v Embrechts a kol. (2002). Ako môžeme vidieť, Pearsonov korelačný koeficient čelí niekoľkým obmedzeniam. Tieto obmedzenia nás vedú k inému spôsobu merania závislosti, a to pomocou Kendallovho tau.

## 1.2 Kendallovo tau

Korelačný koeficient spomenutý vyššie meria lineárnu závislosť medzi náhodnými veličinami. Narozdiel od neho, Kendallovo tau meria monotónnu závislosť medzi usporiadanými náhodnými veličinami.

**Definícia 3.** *Nech  $(X, Y)^\top$  je náhodný vektor. Kendallovo tau definujeme ako*

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

kde  $(X_1, Y_1)^\top$  a  $(X_2, Y_2)^\top$  sú dva nezávislé náhodné vektory s rovnakým rozdelením ako  $(X, Y)^\top$ .

Terminológiu, ktorú budeme ďalej v práci používať, si teraz zavedieme. Dvojicu vektorov  $(X_1, Y_1)^\top$  a  $(X_2, Y_2)^\top$  splňujúcich  $X_1 \neq X_2 \vee Y_1 \neq Y_2$  nazývame konkordantnou, ak platí

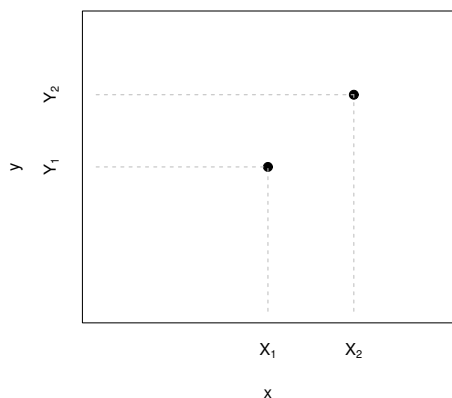
$$\text{sgn}(X_2 - X_1) = \text{sgn}(Y_2 - Y_1).$$

To znamená, že pozorovaniu  $X$  s väčšou hodnotou odpovedá  $Y$  s väčšou hodnotou, alebo pozorovaniu  $X$  s menšou hodnotou odpovedá  $Y$  s menšou hodnotou. Príklad konkordantnej dvojice vektorov môžeme vidieť na obrázku 1.1. Dvojicu vektorov  $(X_1, Y_1)^\top$  a  $(X_2, Y_2)^\top$  splňujúcich  $X_1 \neq X_2 \vee Y_1 \neq Y_2$  nazývame diskordantnou, ak platí

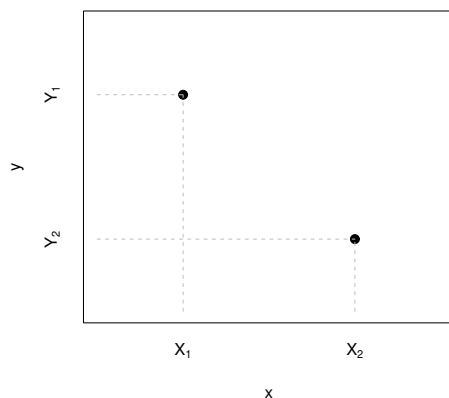
$$\text{sgn}(X_2 - X_1) = -\text{sgn}(Y_2 - Y_1).$$

V tomto prípade  $X$  s väčšou hodnotou odpovedá  $Y$  s menšou hodnotou, alebo  $X$  s menšou hodnotou odpovedá  $Y$  s väčšou hodnotou. Príklad diskordantnej dvojice vektorov môžeme vidieť na obrázku 1.2. V prípade, že dvojica vektorov  $(X_1, Y_1)^\top$  a  $(X_2, Y_2)^\top$  splňa  $X_1 = X_2 \vee Y_1 = Y_2$ , hovoríme, že nastala zhoda (tento jav budeme označovať  $Z$ ). Matematicky vyjadrujeme pravdepodobnosť, že nastane zhoda, ako

$$\mathbb{P}[Z] = \mathbb{P}(X_1 = X_2 \vee Y_1 = Y_2).$$



Obr. 1.1: Konkordantná dvojica vektorov.



Obr. 1.2: Diskordantná dvojica vektorov.

V nadchádzajúcom texte budeme označovať  $P[(X_1 - X_2)(Y_1 - Y_2) > 0]$  ako  $P[K]$  a  $P[(X_1 - X_2)(Y_1 - Y_2) < 0]$  ako  $P[D]$ . Použitím tohto značenia a definície 3 vieme Kendallovo tau  $X$  a  $Y$ , vyjadriť nasledovne

$$\tau(X, Y) = P[K] - P[D]. \quad (1.1)$$

Pri úpravách alebo výpočtoch je potrebné brať ohľad na to, či náhodné veličiny majú spojité rozdelenie, diskrétno alebo kombináciu predošlých. Kendallovo tau prichádza o niekoľko vlastností v prípade diskrétného rozdelenia náhodných veličín  $X$  a  $Y$ . Jedným z dôvodov je, že v diskrétnom prípade môžu vznikáť zhody. V tomto prípade platí

$$P[K] + P[D] + P[Z] = 1. \quad (1.2)$$

Z tejto rovnosti vyjadríme  $P[D]$  a dosadením do (1.1) dostaneme

$$\tau(X, Y) = 2P[K] - 1 + P[Z]. \quad (1.3)$$

Čím menej rôznych hodnôt nabývajú naše náhodné veličiny, tak tým je väčšia pradedobnosť, že dostaneme zhodu vo výbere  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$ .

**Tvrdenie 2.** *Nech  $(X, Y)^\top$  je náhodný vektor s diskrétnym rozdelením, potom  $P[Z] > 0$  a Kendallovo tau nemôže nadobúdať hodnôt  $-1$  a  $1$ .*

*Dôkaz.* Využitím toho, že  $P[Z] > 0$  a úpravou pravej strany rovnosti (1.3) dostaneme

$$\tau(X, Y) = 2P[K] - 1 + P[Z] \geq P[Z] - 1 > -1.$$

Podobne dokážeme, že  $\tau(X, Y) < 1$ . Z rovnosti (1.2) vyjadríme  $P[K]$  a dosadením do (1.1), dostaneme

$$\tau(X, Y) = 1 - P[Z] - 2P[D] \leq 1 - P[Z] < 1.$$

□

Toto tvrdenie demonštrujeme na príklade 2.

*Príklad 2.* Nech  $(X,Y)^\top$  je náhodný vektor s diskrétnym rozdelením na  $\{0,1\}^2$  takým, že marginálne rozdelenia  $X$  a  $Y$  sú  $Alt(p)$  a  $Alt(q)$  pre  $p,q \in (0,1)$  a platí  $P(X = 1, Y = 1) = a$ , kde  $a \in (\max\{p + q - 1, 0\}, \min\{p, q\})$ . Zvyšné združené pravdepodobnosti dopočítame a zapíšeme do tabuľky 1.1.

$X \setminus Y$	0	1
0	$1 - p - q + a$	$q - a$
1	$p - a$	$a$
	$1 - q$	$q$

Tabuľka 1.1: Združené rozdelenie vektoru  $(X,Y)^\top$ .

Výpočet pravdepodobnosti konkordancie a pravdepodobnosti diskordancie je jednoduchý. Nech  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$  sú nezávislé náhodné vektory s rovnakým rozdelením ako  $(X,Y)^\top$ . Upravíme pravdepodobnosť konkordancie nasledovne

$$P[K] = P[(X_1 - X_2)(Y_1 - Y_2) > 0] = P[(X_1 > X_2, Y_1 > Y_2) \vee (X_1 < X_2, Y_1 < Y_2)].$$

Z tohto rozpisu je vidieť, že jediné dvojice vektorov, ktoré sú konkordantné sú  $(0,0)^\top, (1,1)^\top$  a  $(1,1)^\top, (0,0)^\top$ . Takže pre pravdepodobnosť konkordancie platí

$$\begin{aligned} P[K] &= P[(X_1, Y_1) = (1,1)]P[(X_2, Y_2) = (0,0)] \\ &\quad + P[(X_1, Y_1) = (0,0)]P[(X_2, Y_2) = (1,1)] \\ &= a(1 - p - q + a) + (1 - p - q + a)a = 2(1 - p - q + a)a. \end{aligned}$$

Podobne vypočítame pravdepodobnosť pre diskordantné dvojice vektorov a dostaneme  $P[D] = 2(p - a)(q - a)$ . Vo všetkých zvyšných kombináciách nám nastáva zhoda. Dosadením  $P[K]$  a  $P[D]$  do rovnosti (1.1), dostaneme  $\tau(X,Y) = 2(a - pq)$ . Ak napríklad  $p = q = \frac{1}{2}$ , potom platí  $\tau(X,Y) \in [-\frac{1}{2}, \frac{1}{2}]$ , čo ukazuje, že  $\tau(X,Y)$  nenadobúda hodnoty  $-1$  ani  $1$ . Nech  $X$  a  $Y$  sú nezávislé, to nastáva práve vtedy, keď

$$a = P[(X,Y) = (1,1)] = P(X = 1)P(Y = 1) = pq.$$

Z toho dostávame, že  $a = pq$  a  $\tau(X,Y) = 0$ .

Fakt, že nezávislosť  $X$  a  $Y$  implikuje  $\tau(X,Y) = 0$ , neplatí len v práve spomenutom príklade, ale aj pre ľubovoľné rozdelenie  $X$  a  $Y$ . V nasledujúcom tvrdení si to dokážeme.

**Tvrdenie 3.** *Nech  $(X,Y)^\top$  je náhodný vektor s nezávislými zložkami. Potom platí*

$$\tau(X,Y) = 0.$$

*Dôkaz.* Predpokladajme, že  $(X_1, Y_1)^\top$  a  $(X_2, Y_2)^\top$  sú nezávislé vektory s rovnakým rozdelením ako  $(X,Y)^\top$ . Následne upravujeme rovnosť z definície 3

$$\begin{aligned} \tau(X,Y) &= P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= P[(X_1 - X_2 > 0, Y_1 - Y_2 > 0) \vee (X_1 - X_2 < 0, Y_1 - Y_2 < 0)] \\ &\quad - P[(X_1 - X_2 > 0, Y_1 - Y_2 < 0) \vee (X_1 - X_2 < 0, Y_1 - Y_2 > 0)] \\ &= P(X_1 - X_2 > 0, Y_1 - Y_2 > 0) + P(X_1 - X_2 < 0, Y_1 - Y_2 < 0) \\ &\quad - P(X_1 - X_2 > 0, Y_1 - Y_2 < 0) - P(X_1 - X_2 < 0, Y_1 - Y_2 > 0). \end{aligned}$$

Následne využijeme nezávislosť  $X$  a  $Y$

$$\begin{aligned}\tau(X,Y) &= \mathbf{P}(X_1 - X_2 > 0)\mathbf{P}(Y_1 - Y_2 > 0) + \mathbf{P}(X_1 - X_2 < 0)\mathbf{P}(Y_1 - Y_2 < 0) \\ &\quad - \mathbf{P}(X_1 - X_2 > 0)\mathbf{P}(Y_1 - Y_2 < 0) - \mathbf{P}(X_1 - X_2 < 0)\mathbf{P}(Y_1 - Y_2 > 0) \\ &= \mathbf{P}(X_1 - X_2 > 0)\left[\mathbf{P}(Y_1 - Y_2 > 0) - \mathbf{P}(Y_1 - Y_2 < 0)\right] \\ &\quad + \mathbf{P}(X_1 - X_2 < 0)\left[\mathbf{P}(Y_1 - Y_2 < 0) - \mathbf{P}(Y_1 - Y_2 > 0)\right].\end{aligned}$$

Z nezávislosti a identického rozdelenia  $Y_1$  a  $Y_2$  vyplýva, že

$$\mathbf{P}(Y_1 > Y_2) = \mathbf{P}(Y_1 < Y_2).$$

Z toho dostávame

$$\tau(X,Y) = \mathbf{P}(X_1 - X_2 > 0) \cdot 0 + \mathbf{P}(X_1 - X_2 < 0) \cdot 0 = 0.$$

□

V prípade, že náhodne veličiny majú spojité rozdelenie, platí  $\mathbf{P}[Z] = 0$  a teda

$$\mathbf{P}[K] + \mathbf{P}[D] = 1. \tag{1.4}$$

Z toho dostávame, že platí

$$\tau(X,Y) = 2\mathbf{P}[K] - 1. \tag{1.5}$$

V spojitom prípade Kendallovo tau vieme vyjadriť pomocou združenej distribučnej funkcie náhodných veličín. Ešte predtým ako si toto vyjadrenie ukážeme, je potrebné zdefinovať pojem kopula.

## 1.3 Kopula

Uvažujme  $d \in \mathbb{N}, d \geq 2$ . Nech  $(X_1, \dots, X_d)^\top$  je reálny náhodný vektor a  $F$  je jeho združená distribučná funkcia. Ďalej nech sú  $F_1, \dots, F_d$  marginálne distribučné funkcie náhodných veličín  $X_1, \dots, X_d$ .

**Definícia 4.** *Kopula je  $d$ -rozmerná distribučná funkcia  $C : [0,1]^d \rightarrow [0,1]$ , ktorej marginálne distribučné funkcie odpovedajú rovnomernému rozdeleniu na intervale  $[0,1]$ ,  $d \in \mathbb{N}$ .*

**Veta 4.** *Nech  $(X_1, \dots, X_d)^\top$  je reálny náhodný vektor a  $F(\mathbf{x})$  je jeho združená distribučná funkcia. Ďalej nech sú  $F_1, \dots, F_d$  marginálne distribučné funkcie náhodných veličín  $X_1, \dots, X_d$ . Potom existuje kopula  $C$ , že platí*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad \forall (x_1, \dots, x_d)^\top \in \mathbb{R}^d.$$

*Ak  $(X_1, \dots, X_d)^\top$  má spojité rozdelenie, potom kopula  $C$  je určená jednoznačne.*

*Dôkaz.* Dôkaz nájdeme v článku Sklar (1959) alebo aj v článku Durante a kol. (2013). □

Vďaka vete 4 vieme vyjadriť združenú distribučnú funkciu pomocou kopule a marginálnych distribučných funkcií. Túto vetu využijeme v dôkaze nasledujúcej vety.

**Veta 5.** *Nech  $(X,Y)^\top$  je náhodný vektor so spojitým rozdelením s distribučnou funkciou  $F_{X,Y}$ , s marginálnymi distribučnými funkciami  $F_X, F_Y$  a s jedinečnou kopulou  $C : [0,1]^2 \rightarrow [0,1]$ . Potom platí, že*

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

*Dôkaz.* Nech  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$  sú náhodné vektory, ktoré majú rovnaké rozdelenie ako  $(X,Y)^\top$ . Kvôli prehľadnosti budeme distribučné funkcie vektorov  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$  označovať  $F_{X_1, Y_1}$  a  $F_{X_2, Y_2}$ , a príslušné hustoty  $f_{X_1, Y_1}(x_1, y_1)$  a  $f_{X_2, Y_2}(x_2, y_2)$ . Potom  $\tau(X,Y)$  vieme upraviť nasledovne.

$$\begin{aligned} \tau(X,Y) &= \mathbf{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbf{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= 2\mathbf{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 \\ &= 2\mathbf{P}(X_2 < X_1, Y_2 < Y_1 \vee X_2 > X_1, Y_2 > Y_1) - 1. \end{aligned}$$

Vďaka spojitosti rozdelenia sme v druhej rovnosti využili vyjadrenie Kendallovho tau z rovnosti (1.5). Následne ukážeme, že platí

$$\mathbf{P}(X_2 < X_1, Y_2 < Y_1) = \mathbf{P}(X_2 > X_1, Y_2 > Y_1). \quad (1.6)$$

To platí, pretože vektory  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$  sú nezávislé a majú rovnaké rozdelenie.

Z rovnosti (1.6) dostávame, že platí

$$\tau(X,Y) = 4\mathbf{P}(X_2 \leq X_1, Y_2 \leq Y_1) - 1. \quad (1.7)$$

Teraz budeme upravovať  $\mathbf{P}(X_2 \leq X_1, Y_2 \leq Y_1)$ . Označme združenú hustotu vektoru  $(X_1, Y_1, X_2, Y_2)^\top$  ako  $f(x_1, y_1, x_2, y_2)$ ,  $(x_1, y_1, x_2, y_2) \in \mathbb{R}^4$ . Z nezávislosti  $(X_1, Y_1)^\top$  a  $(X_2, Y_2)^\top$  vieme  $f(x_1, y_1, x_2, y_2)$  vyjadriť ako súčin jednotlivých hustôt, t. j.

$$f(x_1, y_1, x_2, y_2) = f_{X_1, Y_1}(x_1, y_1) \cdot f_{X_2, Y_2}(x_2, y_2), \quad \forall (x_1, y_1, x_2, y_2) \in \mathbb{R}^4. \quad (1.8)$$

Rovnicu (1.8) využijeme v druhej rovnosti nasledujúceho výpočtu.

$$\begin{aligned}
& \mathbf{P}(X_2 \leq X_1, Y_2 \leq Y_1) \\
&= \int_{\mathbb{R}^2 \times ((-\infty, x_1) \times (-\infty, y_1))} f(x_1, y_1, x_2, y_2) d(x_1, y_1, x_2, y_2) \\
&= \int_{\mathbb{R}^2} \int_{(-\infty, x_1) \times (-\infty, y_1)} f_{X_1, Y_1}(x_1, y_1) \cdot f_{X_2, Y_2}(x_2, y_2) d(x_2, y_2) d(x_1, y_1) \\
&= \int_{\mathbb{R}^2} f_{X_1, Y_1}(x_1, y_1) \int_{(-\infty, x_1) \times (-\infty, y_1)} f_{X_2, Y_2}(x_2, y_2) d(x_2, y_2) d(x_1, y_1) \\
&= \int_{\mathbb{R}^2} f_{X_1, Y_1}(x_1, y_1) \cdot F_{X_2, Y_2}(x_1, y_1) d(x_1, y_1) \\
&= \int_{\mathbb{R}^2} F_{X_2, Y_2}(x_1, y_1) dF_{X_1, Y_1}(x_1, y_1).
\end{aligned}$$

V tejto rovnosti sme využili:  $\int_B f(x) d(x) = \int_B 1 dF_X(x)$ . Následne využijeme toho, že  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$  majú rovnaké rozdelenie, ktorého distribučnú funkciu označíme ako  $F_{X, Y}$ .

$$\mathbf{P}(X_2 \leq X_1, Y_2 \leq Y_1) = \int_{\mathbb{R}^2} F_{X, Y}(x_1, y_1) dF_{X, Y}(x_1, y_1). \quad (1.9)$$

Ďalej, využitím vety 4 dostaneme, že platí

$$\begin{aligned}
\mathbf{P}(X_2 \leq X_1, Y_2 \leq Y_1) &= \int_{\mathbb{R}^2} C(F_X(x_1), F_Y(y_1)) dC(F_X(x_1), F_Y(y_1)) \\
&= \int_0^1 \int_0^1 C(u, v) dC(u, v).
\end{aligned}$$

V poslednej rovnici sme previedli substitúciu  $(F_X(x_1), F_Y(y_1))$  za  $(u, v)$ . Nakoniec dosadením rovnosti

$$\mathbf{P}(X_2 \leq X_1, Y_2 \leq Y_1) = \int_0^1 \int_0^1 C(u, v) dC(u, v)$$

do (1.7) dostaneme

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

□

Toto vyjadrenie pre Kendallovo tau neplatí v prípade, že  $X$  a  $Y$  sú diskkrétne náhodné veličiny. Tento problém je podrobnejšie rozobraný v článku Genest a Nešlehová (2007).

## 2. Spojité rozšírenie diskkrétnej náhodnej veličiny

Predchádzajúcu kapitolu sme zakončili rovnosťou, ktorá nám hovorí, že Kendallovo tau náhodných veličín so spojitým rozdelením sa dá vyjadriť pomocou kopuly. Toto vyjadrenie neplatí pre diskkrétne náhodné veličiny, to vidieť pri rozpísaní Kendallovho tau. Vychádzame z rovnosti (1.3) a postupujeme podobne ako v dôkaze vety 5.

$$\tau(X,Y) = 2\mathbb{P}[K] - 1 + \mathbb{P}[Z] = 4\mathbb{P}(X_2 < X_1, Y_2 < Y_1) - 1 + \mathbb{P}[Z].$$

Z uvedeného dôvodu si v tejto kapitole ukážeme spojitú rozšírenie diskrétnych náhodných veličín, čo ako uvidíme, zachováva Kendallovo tau. V tejto kapitole čerpáme z článku Denuit a Lambert (2005). V sekcii 2.1 je vlastným prínosom podrobné prevedenie dôkazu vety 6, príklad 3 a k nemu patriace grafy. V sekcii 2.2 je vlastným prínosom podrobné dokázanie tvrdení. Nakoniec v sekcii 2.3 je vlastným prínosom podrobné prevedenie dôkazu tvrdenia 9.

### 2.1 Spojité rozšírenie

Nech  $X$  je diskrétna náhodná veličina, ktorá nabýva hodnoty na množine  $M \subseteq \mathbb{N}$  a jej rozdelenie zapíšeme ako

$$f_x = \mathbb{P}(X = x) \quad x \in M$$

a  $f_x = 0$  pre  $x \notin M$ . Potom definujeme spojitú rozšírenie  $X^*$  ako

$$X^* = X + (U - 1),$$

kde  $U$  je spojitá náhodná veličina s hodnotami v  $(0,1)$ , nezávislá od  $X$  a s rastúcou distribučnou funkciou  $G(u)$  na  $(0,1)$ .

Ďalej ukážeme ako vyzerá distribučná funkcia náhodnej veličiny  $X^*$ .

**Veta 6.** *Nech platia práve spomenuté predpoklady a  $F$  je distribučná funkcia náhodnej veličiny  $X$ . Potom  $X^*$  má distribučnú funkciu*

$$F^*(s) = F(\lfloor s \rfloor) + G(s - \lfloor s \rfloor)f_{\lfloor s \rfloor + 1}, \quad \forall s \in \mathbb{R}$$

kde  $\lfloor s \rfloor$  je dolná celá časť čísla  $s \in \mathbb{R}$ .

*Dôkaz.* Stačí nám počítať pre  $s \in \mathbb{R}^+$ , keďže  $X$  nabýva hodnoty v  $\mathbb{N}$  a teda aj  $X^*$  nabýva kladných hodnôt. Ak by  $s \in \mathbb{R}^-$ , potom  $F^*(s) = F(s) = 0$ . Začneme úpravou funkcie  $F^*(s)$

$$\begin{aligned} F^*(s) &= \mathbb{P}(X^* \leq s) = \mathbb{P}(X + U - 1 \leq s) = \mathbb{P}(U \leq s - X + 1) \\ &= \sum_{x \in M} \mathbb{P}(U \leq s - x + 1) \mathbb{P}(X = x). \end{aligned}$$

V poslednej rovnosti sme využili vetu o úplnej pravdepodobnosti. Preznačením  $s - x + 1$  na  $u_x$  dostaneme

$$\begin{aligned} F^*(s) &= \sum_{x \in M} \mathbf{P}(U \leq u_x) f_x \\ &= \sum_{x \leq s} \mathbf{P}(U \leq u_x) f_x + \sum_{x \in M \cap (s, s+1)} \mathbf{P}(U \leq u_x) f_x + \sum_{x \geq s+1} \mathbf{P}(U \leq u_x) f_x. \end{aligned}$$

Teraz využijeme toho, že ak  $x \leq s$  potom  $u_x \geq 1$ . Podobne pre  $x \geq s + 1$ , platí  $u_x \leq 0$  a teda

$$F^*(s) = \sum_{x \leq s} f_x + \sum_{x \in M \cap (s, s+1)} \mathbf{P}(U \leq u_x) f_x + 0. \quad (2.1)$$

Od tohto bodu berieme  $s$  reálne, také že  $s \notin \mathbb{N}$ . Na upravenie druhej sumy v (2.1) využijeme, že  $x$  je prirodzené číslo z intervalu  $(s, s + 1)$ , takže musí platiť  $x = \lfloor s + 1 \rfloor$ . Teda  $u_x = s - x + 1 = s - \lfloor s + 1 \rfloor + 1 = s - \lfloor s \rfloor$  a dostávame

$$F^*(s) = \sum_{x \leq s} f_x + G(s - \lfloor s \rfloor) f_{\lfloor s + 1 \rfloor}.$$

Znovu využijeme toho, že  $x$  je prirodzené číslo, a preto ak  $x \leq s$ , potom  $x \leq \lfloor s \rfloor$  a platí

$$\begin{aligned} F^*(s) &= \sum_{x \leq \lfloor s \rfloor} f_x + G(s - \lfloor s \rfloor) f_{\lfloor s + 1 \rfloor} \\ &= F(\lfloor s \rfloor) + G(s - \lfloor s \rfloor) f_{\lfloor s + 1 \rfloor}. \end{aligned}$$

V poslednej rovnosti sme len prepísali  $\sum_{x \leq \lfloor s \rfloor} f_x$  na  $F(\lfloor s \rfloor)$  pomocou definície distribučnej funkcie.

Ak by  $s \in \mathbb{N}$ , tak v (2.1) by sme v druhej sume sčítali cez prázdnu množinu, keďže interval  $(s, s + 1)$  je otvorený. V tomto prípade vieme, že  $s = \lfloor s \rfloor$  a postupujeme nasledovne

$$\sum_{x \leq s} f_x = \sum_{x \leq \lfloor s \rfloor} f_x = F(\lfloor s \rfloor).$$

Daný výsledok sa zhoduje s  $F(\lfloor s \rfloor) + G(s - \lfloor s \rfloor) f_{\lfloor s + 1 \rfloor}$  za predpokladu, že  $s \notin \mathbb{N}$ . Pretože  $s = \lfloor s \rfloor$  a teda  $G(s - \lfloor s \rfloor) = 0$ , pre  $s \in \mathbb{N}$ . □

*Poznámka 1.* Pôvodne sme zaviedli spojité rozšírenie pre  $X$  z množiny  $M \subseteq \mathbb{N}$ , lebo autori článku Denuit a Lambert (2005) chceli  $X^*$  nezáporné. Avšak, tento spôsob rozšírenia je aplikovateľný aj pre  $X$  z množiny  $M \subseteq \mathbb{N}_0$ .

Teraz si na jednoduchom príklade porovnáme distribučné funkcie náhodných veličín  $X$  a  $X^*$ . V nasledujúcom príklade budeme slovné spojenie, náhodná veličina, často označovať ako n.v..

*Príklad 3.* Nech  $X$  je diskrétna n.v., ktorá nadobúda hodnoty: 1,2,3 s pravdepodobnosťami  $\frac{1}{6}, \frac{2}{6}, \frac{3}{6}$ . Náhodnú veličinu  $U$ , ktorá je nezávislá od  $X$  zvolíme s rovnomerným rozdelením na intervale  $(0,1)$ ,  $G(u) = u \mathbb{I}_{(0,1)}(u)$ . Pomocou vzorca pre  $F^*(s)$  vyjadríme distribučnú funkciu náhodnej veličiny  $X^*$  ako

$$F^*(s) = \begin{cases} 0, & \text{ak } s \leq 0, \\ (s - \lfloor s \rfloor) \cdot \frac{1}{6}, & \text{ak } s \in (0,1), \\ \frac{1}{6} + (s - \lfloor s \rfloor) \cdot \frac{2}{6}, & \text{ak } s \in [1,2), \\ \frac{1}{2} + (s - \lfloor s \rfloor) \cdot \frac{3}{6}, & \text{ak } s \in [2,3), \\ 1, & \text{ak } s \in [3, \infty). \end{cases}$$

Potom distribučnú funkciu náhodnej veličiny  $X$  môžeme vidieť na obrázku 2.1. Následne môžeme vidieť graf 2.2 distribučnej funkcie spojitého rozšírenia  $X^*$  (modrá farba). V rámci porovnania je do grafu 2.2 pridaná aj empirická distribučná funkcia  $X^*$ , ktorá je spočítaná z 500 nezávislých opakovaní (červená farba).

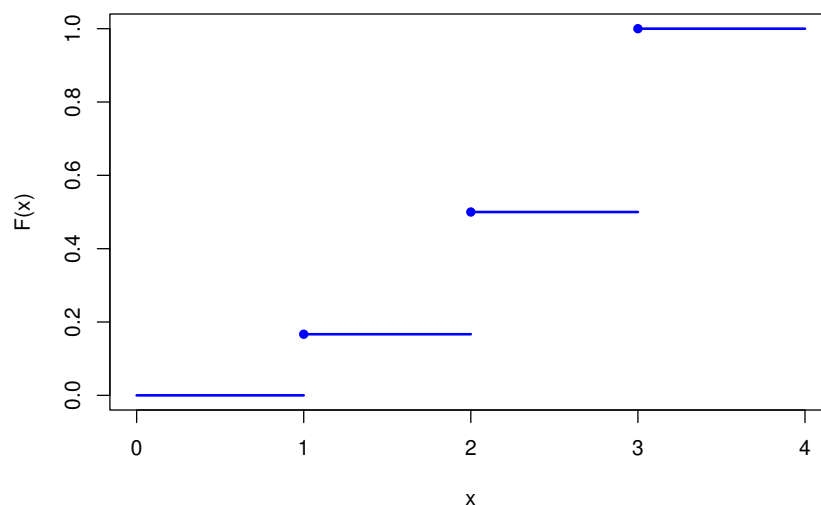
Pre porovnanie ešte spojito rozšírime  $X$ , náhodnou veličinou  $V$ , ktorá má distribučnú funkciu

$$H(v) = \begin{cases} 0, & \text{ak } v \leq 0, \\ v^2, & \text{ak } 0 < v < 1, \\ 1, & \text{ak } v \geq 1. \end{cases}$$

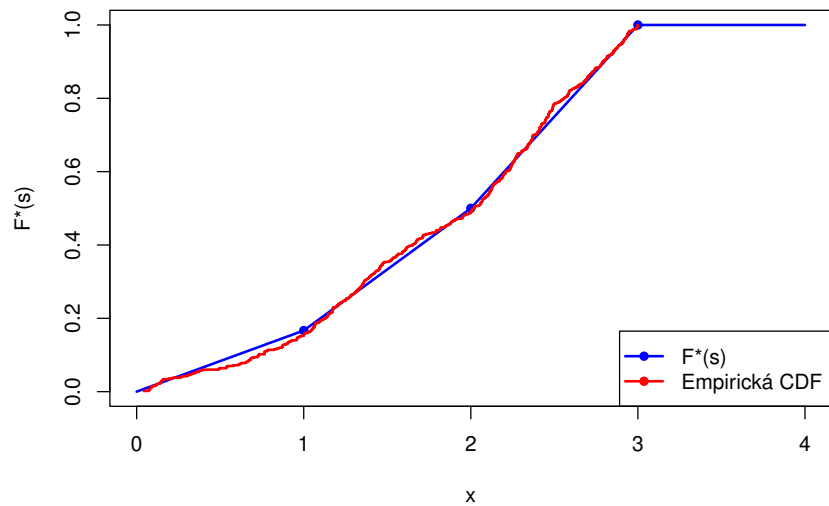
Potom distribučná funkcia spojito rozšírenej náhodnej veličiny  $X^\wedge = X + V - 1$ , bude vyzerat nasledovne

$$F^\wedge(s) = \begin{cases} 0, & \text{ak } s \leq 0, \\ (s - \lfloor s \rfloor)^2 \cdot \frac{1}{6}, & \text{ak } s \in (0,1), \\ \frac{1}{6} + (s - \lfloor s \rfloor)^2 \cdot \frac{2}{6}, & \text{ak } s \in [1,2), \\ \frac{1}{2} + (s - \lfloor s \rfloor)^2 \cdot \frac{3}{6}, & \text{ak } s \in [2,3), \\ 1, & \text{ak } s \in [3, \infty). \end{cases}$$

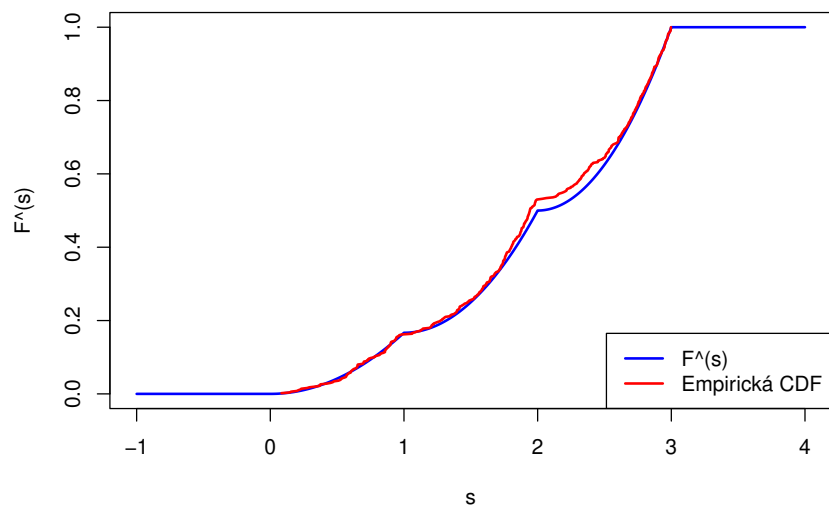
Graf tejto distribučnej funkcie v porovnaní s empirickou distribučnou funkciou môžeme vidieť na 2.3.



Obr. 2.1: Distribučná funkcia n.v.  $X$ .



Obr. 2.2: CDF a ECDF n.v.  $X^*$ .



Obr. 2.3: CDF a ECDF n.v.  $X^\wedge$ .

## 2.2 Zachovanie usporiadania podľa konkordancie a PQD

V tejto časti si zavedieme pojmy, kedy je vektor náhodných veličín viac konkordantný ako iný a kedy je pozitívne závislý podľa kvadrantu. Následne ukážeme, že ak jednotlivé veličiny vo vektoroch spojito rozšírime, tak výsledné dvojice zachovajú práve spomenuté vlastnosti. Počas celej sekcie 2.2 a 2.3 predpokladáme nasledovné. Náhodné veličiny, ktorými spojito rozširujeme, sú vždy nezávislé a spĺňajú vlastnosti zo začiatku sekcie 2.1. Ďalej predpokladáme, že náhodný vektor  $(X, Y)^\top$  s diskretným rozdelením nadobúda hodnoty v  $\mathbb{N}_0^2$ .

**Definícia 5.** *Nech  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$  sú dva náhodné vektory, kde  $X_1, X_2$  majú rovnaké rozdelenie a  $Y_1, Y_2$  majú rovnaké rozdelenie. Povieme, že  $(X_2, Y_2)^\top$  je viac konkordantný ako  $(X_1, Y_1)^\top$ , ak platí*

$$\mathbf{P}(X_1 \leq s, Y_1 \leq t) \leq \mathbf{P}(X_2 \leq s, Y_2 \leq t), \forall s, t \in \mathbb{R}.$$

Značíme  $(X_1, Y_1)^\top \prec_c (X_2, Y_2)^\top$ .

Ďalej povieme, že  $(X, Y)^\top$  je pozitívne závislý podľa kvadrantu (značíme PQD), ak platí

$$F(s)G(t) \leq \mathbf{P}(X \leq s, Y \leq t), \forall s, t \in \mathbb{R},$$

kde  $F$  a  $G$  sú marginálne distribučné funkcie  $X$  a  $Y$  v tomto poradí.

Teraz vyslovíme a dokážeme tvrdenia, ktoré nám hovoria, že pojmy z definície 5 zostanú zachované aj pre spojité roširenia náhodných veličín.

**Tvrdenie 7.** *Nech  $(X_1, Y_1)^\top, (X_2, Y_2)^\top$  sú dvojice diskretných náhodných veličín, pre ktoré platí, že  $(X_1, Y_1)^\top \prec_c (X_2, Y_2)^\top$ . Ďalej nech  $(X_1^*, Y_1^*)^\top, (X_2^*, Y_2^*)^\top$  sú dvojice spojito rozšírených náhodných veličín, kde  $X_1^*, X_2^*$  sú rozšírené rovnakou nahodnou veličinou a  $Y_1^*, Y_2^*$  sú rozšírené rovnakou nahodnou veličinou. Potom platí  $(X_1^*, Y_1^*)^\top \prec_c (X_2^*, Y_2^*)^\top$ .*

*Dôkaz.* Náhodnú veličinu, ktorá spojito rozširuje  $X_1$  aj  $X_2$  označíme  $U$ , náhodnou veličinou, ktorá spojito rozširuje  $Y_1, Y_2$  označíme  $V$ . Ďalej označíme hustoty náhodných veličín  $U, V$  ako  $f(u), u \in (0, 1), g(v), v \in (0, 1)$ . Potom máme

$$\begin{aligned} \mathbf{P}(X_1^* \leq s, Y_1^* \leq t) &= \mathbf{P}(X_1 + U - 1 \leq s, Y_1 + V - 1 \leq t) \\ &= \int_0^1 \int_0^1 \mathbf{P}(X_1 \leq s - u + 1, Y_1 \leq t - v + 1) f(u) g(v) du dv. \end{aligned}$$

Použitím predpokladu  $(X_1, Y_1)^\top \prec_c (X_2, Y_2)^\top$  dostávame

$$\begin{aligned} \mathbf{P}(X_1^* \leq s, Y_1^* \leq t) &\leq \int_0^1 \int_0^1 \mathbf{P}(X_2 \leq s - u + 1, Y_2 \leq t - v + 1) f(u) g(v) du dv \\ &= \mathbf{P}(X_2^* \leq s, Y_2^* \leq t). \end{aligned}$$

Týmto sme dokázali, že  $(X_1^*, Y_1^*)^\top \prec_c (X_2^*, Y_2^*)^\top$ . □

Podobne overíme, že ak  $(X, Y)^\top$  je PQD, tak aj  $(X^*, Y^*)^\top$  je PQD.

**Tvrdenie 8.** *Nech  $(X, Y)^\top$  je dvojica diskretných náhodných veličín, ktorá je PQD. Ďalej nech  $(X^*, Y^*)^\top$  je dvojica spojito rozšírených náhodných veličín. Potom  $(X^*, Y^*)^\top$  je PQD.*

*Dôkaz.* Náhodné veličiny, ktoré spojito rozširujú  $X, Y$  označíme  $U, V$ , v tomto poradí. Taktiež označíme hustoty  $U, V$  ako  $f, g$ . Ďalej označíme distribučné funkcie náhodných veličín  $X, Y$  ako  $F, G$  a distribučné funkcie náhodných veličín  $X^*, Y^*$  ako  $F^*, G^*$ . Následne počítame

$$\begin{aligned} F^*(s)G^*(t) &= \mathbf{P}(X + U - 1 \leq s)\mathbf{P}(Y + V - 1 \leq t) \\ &= \int_0^1 \mathbf{P}(X \leq s - u + 1)f(u) du \int_0^1 \mathbf{P}(Y \leq t - v + 1)g(v) dv \\ &= \int_0^1 \mathbf{P}(X \leq s - u + 1)f(u) \int_0^1 \mathbf{P}(Y \leq t - v + 1)g(v) dv du. \end{aligned}$$

V poslednej rovnici sme presunuli druhý integrál dovnútra prvého, keďže

$$\int_0^1 \mathbf{P}(Y \leq t - v + 1)g(v) dv$$

nezávisí od premennej  $u$ . Podobne presunieme  $\mathbf{P}(X \leq s - u + 1)f(u)$  dovnútra druhého integrálu

$$\begin{aligned} F^*(s)G^*(t) &= \int_0^1 \int_0^1 \mathbf{P}(X \leq s - u + 1)\mathbf{P}(Y \leq t - v + 1)f(u)g(v) dv du \\ &= \int_0^1 \int_0^1 F(s - u + 1)G(t - v + 1)f(u)g(v) dv du. \end{aligned}$$

Následne použijeme predpoklad  $F(s)G(t) \leq \mathbf{P}(X \leq s, Y \leq t)$  a dostaneme

$$\begin{aligned} F^*(s)G^*(t) &\leq \int_0^1 \int_0^1 \mathbf{P}(X \leq s - u + 1, Y \leq t - v + 1)f(u)g(v) dv du \\ &= \mathbf{P}(X + U - 1 \leq s, Y + V - 1 \leq t) \\ &= \mathbf{P}(X^* \leq s, Y^* \leq t). \end{aligned}$$

V predposlednej rovnosti sme využili to, že náhodné veličiny  $U$  a  $V$  sú nezávislé, takže ich združenú hustotu vieme vyjadriť ako súčin jednotlivých hustôt. Tým sme dokázali, že  $(X^*, Y^*)^\top$  je PQD. □

## 2.3 Zachovanie Kendallovho tau

V tejto sekcii si ukážeme, že spojité rozšírenie diskretných náhodných veličín zachováva Kendallovo tau. Kondkordanciu spojito rozšírených náhodných veličín budeme značiť  $K^*$

**Tvrdenie 9.** *Nech  $(X, Y)^\top$  je dvojica diskretných náhodných veličín a  $(X^*, Y^*)^\top$  je jej príslušná dvojica spojito rozšírených náhodných veličín. Potom platí, že*

$$\tau(X, Y) = \tau(X^*, Y^*). \tag{2.2}$$

*Dôkaz.* Majme dva nezávislé vektory  $(X_1, Y_1)^\top$ ,  $(X_2, Y_2)^\top$  s rovnakým rozdelením ako  $(X, Y)^\top$ . Ďalej určíme

$$X_1^* = X_1 + U_1 - 1, X_2^* = X_2 + U_2 - 1, Y_1^* = Y_1 + V_1 - 1, Y_2^* = Y_2 + V_2 - 1.$$

Pričom  $U_1$  má rovnaké rozdelenie ako  $U_2$ , podobne  $V_1$  má rovnaké rozdelenie ako  $V_2$ . V tomto prípade sú všetky štyri náhodné veličiny  $U_1, U_2, V_1, V_2$  nezávislé. Začneme úpravou

$$P(K^*) = P[(X_1^* - X_2^*)(Y_1^* - Y_2^*) > 0].$$

Výraz rozpíšeme ako

$$P(K^*) = P[(X_1 + U_1 - X_2 - U_2)(Y_1 + V_1 - Y_2 - V_2) > 0]. \quad (2.3)$$

Teraz rovnosť z (2.3) rozpíšeme pomocou vety o úplnej pravdepodobnosti

$$\begin{aligned} P(K^*) &= P[X_1 = X_2, Y_1 = Y_2] P[(U_1 - U_2)(V_1 - V_2) > 0] \\ &+ P[X_1 = X_2, Y_1 > Y_2] P[(U_1 - U_2)(Y_1 + V_1 - Y_2 - V_2) > 0] \\ &+ P[X_1 = X_2, Y_1 < Y_2] P[(U_1 - U_2)(Y_1 + V_1 - Y_2 - V_2) > 0] \\ &+ P[X_1 > X_2, Y_1 = Y_2] P[(X_1 + U_1 - X_2 - U_2)(V_1 - V_2) > 0] \\ &+ P[X_1 < X_2, Y_1 = Y_2] P[(X_1 + U_1 - X_2 - U_2)(V_1 - V_2) > 0] \\ &+ P[(X_1 - X_2)(Y_1 - Y_2) > 0] P[(X_1 + U_1 - X_2 - U_2)(Y_1 + V_1 - Y_2 - V_2) > 0]. \end{aligned}$$

Ďalej vieme, že  $X_1, X_2$  sú diskrétné náhodné veličiny z množiny  $\mathbb{N}_0$ , preto ak  $X_1 > X_2$ , potom  $X_1 - X_2 \geq 1$ . Náhodné veličiny  $U_1$  a  $U_2$  majú hodnoty na intervale  $(0, 1)$ , preto  $U_1 - U_2 > -1$ . Z tohto vieme, že ak  $X_1 > X_2$  potom aj  $(X_1 + U_1 - X_2 - U_2) > 0$ . Vďaka tomuto, môžeme upraviť predošlú rovnosť na

$$\begin{aligned} P(K^*) &= P[X_1 = X_2, Y_1 = Y_2] P[(U_1 - U_2)(V_1 - V_2) > 0] \\ &+ P[X_1 = X_2, Y_1 > Y_2] P[(U_1 - U_2) > 0] \\ &+ P[X_1 = X_2, Y_1 < Y_2] P[(U_1 - U_2) < 0] \\ &+ P[X_1 > X_2, Y_1 = Y_2] P[(V_1 - V_2) > 0] \\ &+ P[X_1 < X_2, Y_1 = Y_2] P[(V_1 - V_2) < 0] \\ &+ P[(X_1 - X_2)(Y_1 - Y_2) > 0]. \end{aligned}$$

Ďalej vieme, že platí

$$P(U_1 - U_2 > 0) = P(U_1 - U_2 < 0),$$

keďže to je jednorozmerný prípad (1.6). Takže  $P(K^*)$  vieme upraviť nasledovne

$$\begin{aligned} P(K^*) &= P[X_1 = X_2, Y_1 = Y_2] P[(U_1 - U_2)(V_1 - V_2) > 0] \\ &+ P[X_1 = X_2, Y_1 > Y_2] P[(U_1 - U_2) > 0] \\ &+ P[X_1 = X_2, Y_1 < Y_2] P[(U_1 - U_2) > 0] \\ &+ P[X_1 > X_2, Y_1 = Y_2] P[(V_1 - V_2) > 0] \\ &+ P[X_1 < X_2, Y_1 = Y_2] P[(V_1 - V_2) > 0] \\ &+ P[(X_1 - X_2)(Y_1 - Y_2) > 0]. \end{aligned}$$

Posledná časť súčtu je  $P[K]$  a zvyšná časť dáva dokopy  $\frac{1}{2}P(Z)$ . Takže celú rovnicu vieme prepísať ako

$$P(K^*) = P(K) + \frac{1}{2}P(Z). \quad (2.4)$$

Čo prepíšeme ako

$$P(K) = P(K^*) - \frac{1}{2}P(Z). \quad (2.5)$$

Z rovnosti (1.3) vieme, že pre diskkrétne náhodné veličiny platí

$$\tau(X,Y) = 2P(K) - 1 + P(Z). \quad (2.6)$$

Po dosadení rovnosti (2.5) do (2.6), dostaneme

$$\tau(X,Y) = 2P(K^*) - 1 = \tau(X^*,Y^*). \quad (2.7)$$

□

*Poznámka 2.* Tvrdenie 9 platí aj vtedy, keď náhodné veličiny  $X$  a  $Y$  spojito rozširujeme bez odčítania 1, t. j.

$$X_1^* = X_1 + U_1, X_2^* = X_2 + U_2, Y_1^* = Y_1 + V_1, Y_2^* = Y_2 + V_2.$$

Platnosť vzťahu (2.2) je očividná, keďže  $P(K^*)$  by sme rozpísali na rovnaký tvar ako výraz v rovnosti (2.3).

### 3. Odhad Kendallovho tau

V tejto kapitole budeme porovnávať odhad Kendallovho tau náhodných veličín s diskretným rozdelením s odhadmi Kendallových tau náhodných veličín, ktoré sú spojito rozšírené rôznymi náhodnými veličinami. Vlastným prínosom sú nasledovné kapitoly a podrobný dôkaz tvrdenia 10. Čerpať budeme z knihy Rémillard (2013).

Nech  $(X_1, Y_1)^\top, \dots, (X_n, Y_n)^\top$  je náhodný výber z dvojrozmerného diskretného rozdelenia. Odhad Kendallovho tau náhodných veličín  $X$  a  $Y$  s diskretným rozdelením vypočítame ako

$$\hat{\tau} = \frac{\#K}{\binom{n}{2}} - \frac{\#D}{\binom{n}{2}}, \quad (3.1)$$

kde  $\#K$  označuje počet konkordantných párov z daného náhodného výberu, čo vieme matematicky vyjadriť ako

$$\#K = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{I}[(X_i < X_j \wedge Y_i < Y_j) \vee (X_i > X_j \wedge Y_i > Y_j)].$$

Podobne  $\#D$  počet diskordantných párov vyjadríme ako

$$\#D = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{I}[(X_i < X_j \wedge Y_i > Y_j) \vee (X_i > X_j \wedge Y_i < Y_j)].$$

Odhad Kendallovho tau spojito rozšírených náhodných veličín  $X^*$  a  $Y^*$  vieme vďaka rovnosti (1.4) vyjadriť nasledovne

$$\tilde{\tau} = \frac{2 \cdot \#K^*}{\binom{n}{2}} - 1. \quad (3.2)$$

Teraz  $\#K^*$  označuje počet konkordantných dvojíc z náhodného výberu  $(X_1^*, Y_1^*)^\top, \dots, (X_n^*, Y_n^*)^\top$ . Postup ako tento náhodný výber získať, vysvetlíme v nasledujúcej sekcii.

**Tvrdenie 10.** *Nech  $(X, Y)^\top$  je náhodný vektor z dvojrozmerného diskretného rozdelenia a  $(X^*, Y^*)^\top$  je jemu príslušný vektor spojito rozšírených náhodných veličín, potom odhady*

$$\hat{\tau} = \frac{\#K}{\binom{n}{2}} - \frac{\#D}{\binom{n}{2}}, \quad \tilde{\tau} = \frac{2 \cdot \#K^*}{\binom{n}{2}} - 1$$

sú nestranné a konzistentné odhady  $\tau(X, Y)$ .

*Dôkaz.* Začneme dôkazom nestrannosti odhadu Kendallovho tau pre náhodné veličiny s diskretným rozdelením.

$$\begin{aligned}
\mathbb{E}\hat{\tau} &= \mathbb{E} \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{I}[(X_i < X_j \wedge Y_i < Y_j) \vee (X_i > X_j \wedge Y_i > Y_j)] \\
&\quad - \mathbb{E} \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{I}[(X_i < X_j \wedge Y_i > Y_j) \vee (X_i > X_j \wedge Y_i < Y_j)] \\
&= \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E} \mathbb{I}[(X_i < X_j \wedge Y_i < Y_j) \vee (X_i > X_j \wedge Y_i > Y_j)] \\
&\quad - \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E} \mathbb{I}[(X_i < X_j \wedge Y_i > Y_j) \vee (X_i > X_j \wedge Y_i < Y_j)] \\
&= \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{P}[(X_i < X_j \wedge Y_i < Y_j) \vee (X_i > X_j \wedge Y_i > Y_j)] \\
&\quad - \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{P}[(X_i < X_j \wedge Y_i > Y_j) \vee (X_i > X_j \wedge Y_i < Y_j)] \\
&= \mathbb{P}[K] - \mathbb{P}[D] = \tau(X, Y).
\end{aligned}$$

Teraz budeme dokazovať nestrannosť  $\tilde{\tau}$ ,

$$\begin{aligned}
\mathbb{E}\tilde{\tau} &= \mathbb{E} \frac{2 \cdot \#K^*}{\binom{n}{2}} - 1 \\
&= \frac{2}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E} \mathbb{I}[(X_i^* < X_j^* \wedge Y_i^* < Y_j^*) \vee (X_i^* > X_j^* \wedge Y_i^* > Y_j^*)] - 1 \\
&= \frac{2}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{P}[(X_i^* < X_j^* \wedge Y_i^* < Y_j^*) \vee (X_i^* > X_j^* \wedge Y_i^* > Y_j^*)] - 1 \\
&= 2\mathbb{P}[K^*] - 1 \\
&= 2\mathbb{P}[K] + \mathbb{P}[Z] - 1 = \tau(X, Y).
\end{aligned}$$

Predposledná rovnosť plynie z (2.4) a posledná rovnosť plynie z (1.3).

Navyše, sú oba odhady konzistentné, čo vyplýva z nestrannosti a toho, že ich rozptyly konvergujú k 0, pre  $n \rightarrow \infty$  viď str. v 14 Lee (1990). □

Ide aj ukázať, že  $\tilde{\tau}$  je asymptoticky normálne, viď Rémillard (2013).

V zvyšnej časti práce budeme skúmať vlastnosti a správanie práve spomenutých odhadov Kendallovho tau pomocou simulácií v praktickej časti.

### 3.1 Opis praktickej časti

V praktickej časti budeme porovnávať odhady Kendallovho tau pôvodných náhodných veličín a tých spojito rozšírených, pričom budeme uvažovať rôzne rozdelenia náhodných veličín, ktorými spojito rozširujeme. V tejto kapitole generujeme dáta a vykonávame výpočty pomocou kódu v programe R, R Core Team (2023).

Najskôr budeme generovať náhodný výber  $(X_1, Y_1)^\top, \dots, (X_n, Y_n)^\top$  z nejakého dvojrozmerného diskretného rozdelenia s daným Kendallovým tau. Pomocou jednoduchého cyklu spočítame všetky konkordantné a diskordantné dvojice a dosadením do rovnosti (3.1), vypočítame odhad Kendallovho tau (označíme  $\hat{\tau}$ ). Tento postup zopakujeme  $N$ -krát. Takto získame realizácie  $\hat{\tau}_1, \dots, \hat{\tau}_N$  a pomocou nich vypočítame odhad vychýlenia  $\hat{\tau}$  (označujeme  $bias(\hat{\tau})$ ) ako

$$bias(\hat{\tau}) = \frac{1}{N} \sum_{i=1}^N \hat{\tau}_i - \tau(X, Y).$$

Ďalej vypočítame smerodajnú odchýlku  $\hat{\tau}$  (označujeme  $sd(\hat{\tau})$ ) ako

$$sd(\hat{\tau}) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\hat{\tau}_i - \frac{1}{N} \sum_{i=1}^N \hat{\tau}_i)^2}.$$

Tieto hodnoty budeme počítat pre rôzne veľkosti náhodného výberu (t. j. rôzne  $n$ ).

Pri počítaní odhadu Kendallovho tau pre spojito rozšírené náhodné veličiny budeme postupovať nasledovne. Najskôr určíme náhodné veličiny  $U$  a  $V$ , ktoré majú rovnakú distribučnú funkciu  $G$  a vyhovujú vlastnostiam zo začiatku kapitoly 2.1. Týmito veličinami spojito rozšírime náhodné veličiny  $X$  a  $Y$ . Následne vygenerujeme náhodné výbery  $U_1, \dots, U_n$  a  $V_1, \dots, V_n$  z rozdelenia s distribučnou funkciou  $G$ . Tentokrát náhodný výber, s ktorým budeme pracovať, budú tvoriť dvojice  $(X_1^*, Y_1^*)^\top, \dots, (X_n^*, Y_n^*)^\top$ , kde

$$X_i^* = X_i + U_i \text{ a } Y_i^* = Y_i + V_i.$$

Teraz odhad Kendallovho tau budeme počítat pomocou (3.2). Ďalej budeme pokračovať rovnako ako v predošlom prípade. Tento postup zopakujeme pre rôzne distribučné funkcie  $G$ . Pre generovanie náhodného výberu zo zadanou distribučnou funkciou  $G$ , využijeme nasledujúce tvrdenie.

**Tvrdenie 11.** *Nech  $U$  je náhodná veličina s rovnomerným rozdelením na  $(0,1)$  a  $F$  je spojitá a rastúca distribučná funkcia na  $(0,1)$ . Potom  $F^{-1}(U)$  má rozdelenie s distribučnou funkciou  $F$ .*

*Dôkaz.* Pre distribučnú funkciu  $G$  náhodnej veličiny  $F^{-1}(U)$  platí

$$G(u) = \mathbb{P}(F^{-1}(U) \leq u) = \mathbb{P}(U \leq F(u)) = F(u), \forall u \in \mathbb{R}.$$

V poslednej rovnosti sme využili, že  $U$  je náhodná veličina s rovnomerným rozdelením na  $(0,1)$ . □

V prípade, že  $G$  nie je bežná distribučná funkcia, použijeme tvrdenie 11 na vygenerovanie náhodného výberu z rozdelenia s distribučnou funkciou  $G$ . Najskôr vygenerujeme náhodné výbery  $U_1^*, \dots, U_n^*, V_1^*, \dots, V_n^*$  z rovnomerného rozdelenia na intervale  $(0,1)$ . Potom ich dosadíme do  $G^{-1}$ . Podľa tvrdenia 11 dostaneme náhodné výbery  $U_1, \dots, U_n, V_1, \dots, V_n$  z rozdelenia s distribučnou funkciou  $G$ .

## 3.2 Nezávislé náhodné veličiny

V tejto sekcii sa budeme zaoberať prípadom, kde náhodné veličiny  $X$  a  $Y$  sú nezávislé a  $X, Y$  majú Poissonovo rozdelenie s parametrom 1, značíme  $Po(1)$ . Zvolíme počet opakovaní  $N = 1000$ . Z tvrdenia 3 vieme, že pre nezávislé náhodné veličiny platí  $\tau(X, Y) = 0$ . Hodnoty  $bias(\hat{\tau})$  a  $sd(\hat{\tau})$  môžeme vidieť v druhom a treťom stĺpci tabuľky 3.1.

V prípade spojitého rozšírenia, najskôr volíme  $G$  ako distribučnú funkciu rovnomerného rozdelenia t. j.

$$G_1(u) = \begin{cases} 0, & \text{ak } u \leq 0, \\ u, & \text{ak } 0 < u < 1, \\ 1, & \text{ak } u \geq 1. \end{cases}$$

Z tvrdenia 9 vieme, že aj v tomto prípade platí  $\tau(X^*, Y^*) = 0$ . Odhad Kendallovho tau budeme značiť  $\tilde{\tau}_1$  a hodnoty  $bias(\tilde{\tau}_1)$ ,  $sd(\tilde{\tau}_1)$  zapíšeme do tabuľky 3.1. Podobne postupujeme pre distribučné funkcie  $G_2$  a  $G_3$

$$G_2(u) = \begin{cases} 0, & \text{ak } u \leq 0, \\ u^2, & \text{ak } 0 < u < 1, \\ 1, & \text{ak } u \geq 1, \end{cases} \quad G_3(u) = \begin{cases} 0, & \text{ak } u \leq 0, \\ u^3, & \text{ak } 0 < u < 1, \\ 1, & \text{ak } u \geq 1. \end{cases}$$

Jediný rozdiel je, že pri generovaní náhodných výberov  $U_1, \dots, U_n, V_1, \dots, V_n$  postupujeme podľa odseku na konci kapitoly 3.1. Odhady postupne označíme  $\tilde{\tau}_2$ ,  $\tilde{\tau}_3$  a pridáme do tabuľky.

$n$	$bias(\hat{\tau})$	$sd(\hat{\tau})$	$bias(\tilde{\tau}_1)$	$sd(\tilde{\tau}_1)$	$bias(\tilde{\tau}_2)$	$sd(\tilde{\tau}_2)$	$bias(\tilde{\tau}_3)$	$sd(\tilde{\tau}_3)$
50	-0,003	0,086	-0,001	0,098	-0,004	0,099	-0,003	0,098
100	-0,002	0,061	-0,003	0,069	-0,003	0,069	-0,003	0,068
200	0,000	0,043	-0,001	0,048	-0,001	0,047	0,000	0,047
500	-0,001	0,027	-0,001	0,030	-0,001	0,030	-0,001	0,031
1000	-0,001	0,016	0,000	0,019	0,000	0,019	0,000	0,020

Tabuľka 3.1: Odhad vychýlenia a smerodajná odchýlka odhadov Kendallovho tau pre nezávislé  $X$  a  $Y$  s Poissonovým rozdelením s parametrom 1.

Teraz skúmame, či sa  $bias$  alebo  $sd$  odhadov Kendallovho tau zmenia, ak zmeníme distribučnú funkciu náhodných veličín, ktorými spojitاً rozširujeme  $X$  a  $Y$ . Vidíme, že  $bias$  odhadov sú veľmi blízke 0, čo by sme od nestranných odhadov čakali. U smerodajných odchýliek sa najviac líši smerodajná odchýlka u odhadu  $\hat{\tau}$ . Táto odchýlka je pre každý počet pozorovaní trochu menšia, ako všetky zvyšné odchýlky. Tento rozdiel nie je príliš zásadný pre vyšší počet pozorovaní (t. j.  $n$ ). Z nazbieraných dát sa nedá povedať, že by bola nejaká náhodná veličina, ktorou spojitاً rozširujeme, lepšia ako iná. Pri výpočte  $\tilde{\tau}$  nám stačí počet konkordantných párov, pričom výpočet  $\hat{\tau}$  vyžaduje aj počet diskordantných, ako je vidieť vo vzorcoch (3.2) a (3.1). Výpočet sa môže časovo natiahnuť alebo skrátiť, v závislosti od metódy, ktorou odhady počítame.

Tento postup zopakujeme pre nezávislé náhodné veličiny  $X$  a  $Y$ , ktoré majú Poissonovo rozdelenie s parametrom 0,1. V tomto prípade náhodné veličiny  $X$  a  $Y$  nadobúdajú hodnoty 0 a 1 s väčšou pravdepodobnosťou ako v predošlom prípade.

$n$	$bias(\hat{\tau})$	$sd(\hat{\tau})$	$bias(\tilde{\tau}_1)$	$sd(\tilde{\tau}_1)$	$bias(\tilde{\tau}_2)$	$sd(\tilde{\tau}_2)$	$bias(\tilde{\tau}_3)$	$sd(\tilde{\tau}_3)$
50	0,000	0,025	-0,004	0,099	-0,003	0,100	-0,002	0,097
100	0,000	0,017	-0,003	0,069	-0,001	0,067	-0,002	0,069
200	0,001	0,012	0,002	0,046	-0,001	0,047	0,004	0,047
500	0,000	0,008	-0,001	0,029	-0,001	0,030	0,000	0,031
1000	0,000	0,005	0,001	0,021	0,001	0,021	-0,001	0,021

Tabuľka 3.2: Odhad vychýlenia a smerodajná odchýlka odhadov Kendallovho tau pre nezávislé  $X$  a  $Y$  s Poissonovým rozdelením s parametrom 0,1.

Vidíme, že v prípade spojito rozšírených náhodných veličín,  $bias$  a  $sd$  odhadov Kendallovho tau sú skoro rovnaké ako v prípade vyššie. Výrazná zmena nastala u odhadu, kde sme nerozširovali. Smerodajná odchýlka je niekoľkokrát menšia ako v predošlom prípade. To je zapríčinené malou variabilitou hodnôt v rozdelení  $Po(0,1)$ . V tomto prípade sú odhady Kendallovho tau rozšírených veličín výrazne nepresnejšie, ako odhady z pôvodných hodnôt.

### 3.3 Závislé binárne veličiny

V tejto sekcii sa budeme zaoberať náhodným vektorom  $(X,Y)^\top$ , ktorého rozdelenie a Kendallovo tau sme rozoberali v príklade 2. Rovnako, ako v predošlej sekcii, bude počet opakovaní  $N = 1000$ . Pracovať budeme s hodnotami  $p = \frac{1}{2}$ ,  $q = \frac{1}{3}$ ,  $a = \frac{1}{4}$ , z toho vyrátame  $\tau(X,Y) = \frac{1}{6}$  a rozdelenie, ktoré zapíšeme do tabuľky nižšie.

$X \setminus Y$	0	1	
0	5/12	1/12	1/2
1	3/12	3/12	1/2
	2/3	1/3	

Tabuľka 3.3: Združené rozdelenie vektoru  $(X,Y)^\top$ .

Pri počítaní  $bias$  a smerodajných odchýlok odhadov Kendallovho tau, budeme postupovať skoro rovnako, ako v predošlej sekcii. Budeme spojito rozširovať rovnakými náhodnými veličinami, ako v predošlej sekcii. Rozdiel bude pri generovaní náhodného výberu  $(X_1, Y_1)^\top, \dots, (X_n, Y_n)^\top$ , ten budeme generovať nasledujúcim spôsobom. Najskôr vygenerujeme náhodný výber  $Z_1, \dots, Z_n$  z rovnomerného rozdelenia na intervale  $(0,1)$ , potom prechádzame každou hodnotou a určíme dvojice  $(X_i, Y_i)^\top$  nasledovným spôsobom. Ak je  $Z_i < \frac{5}{12}$ , volíme bod  $(0,0)^\top$ . Pre  $\frac{5}{12} < Z_i < \frac{6}{12}$  volíme bod  $(0,1)^\top$ . Pre  $\frac{6}{12} < Z_i < \frac{9}{12}$  volíme bod  $(1,0)^\top$ . Pre  $\frac{9}{12} < Z_i < 1$  volíme bod  $(1,1)^\top$ . Po vygenerovaní  $(X_1, Y_1)^\top, \dots, (X_n, Y_n)^\top$  a prevedení rovnakého postupu ako v predošlej sekcii dostávame tabuľku 3.4.

$n$	$bias(\hat{\tau})$	$sd(\hat{\tau})$	$bias(\tilde{\tau}_1)$	$sd(\tilde{\tau}_1)$	$bias(\tilde{\tau}_2)$	$sd(\tilde{\tau}_2)$	$bias(\tilde{\tau}_3)$	$sd(\tilde{\tau}_3)$
50	-0,002	0,064	-0,001	0,095	-0,004	0,095	0,000	0,093
100	-0,001	0,043	-0,003	0,064	-0,002	0,064	-0,001	0,064
200	0,000	0,031	-0,002	0,046	-0,001	0,045	0,001	0,046
500	0,000	0,020	0,000	0,029	-0,001	0,028	-0,001	0,028
1000	0,000	0,014	0,000	0,020	-0,001	0,021	-0,001	0,019

Tabuľka 3.4: Odhad vychýlenia a smerodajná odchýlka odhadov Kendallovho tau pre závislé  $X$  a  $Y$  s alternatívnym rozdelením.

Opäť skúmame, či sa  $bias$  alebo  $sd$  odhadov Kendallovho tau zmenia, ak zmeníme distribučnú funkciu náhodných veličín, ktorými spojitاً rozširujeme  $X$  a  $Y$ . V tomto prípade sa  $sd(\hat{\tau})$  výraznejšie líši od ostatných  $sd$ , v porovnaní s Poissonovým rozdelením s parametrom 1. Tento rozdiel stále nie je až tak výrazný ako  $Po(0,1)$ . Ostatne, záver zostáva rovnaký ako v prípade pre  $Po(1)$ .

# Záver

V prvej kapitole sme si motivovali a zadefinovali Kendallove tau. Uviedli sme pár základných pojmov, ktoré s ním úzko súvisia a tie sme ilustrovali na obrázkoch. Zúženie hraníc Kendallovho tau pre diskkrétne náhodné veličiny sme dokázali v tvrdení a demonštrovali na príklade. Pre náhodné veličiny so spojitým rozdelením sme Kendallove tau vyjadrili pomocou kopuly. V druhej kapitole sme zaviedli spojitú rozšírenie pre diskkrétne náhodné veličiny. Ďalje sme dokázali, že Kendallove tau a ďalšie vlastnosti sú zachované aj po spojitom rozšírení.

V poslednej časti sme zaviedli odhady Kendallovho tau a skúmali sme, či pri výpočte odhadu Kendallovho tau záleží na náhodnej veličine, ktorou spojitou rozširujeme. V našej voľbe náhodných veličín, nezáležalo, ktorou náhodnou veličinou sme rozširovali. Záležalo iba na tom, či vôbec sme spojitou rozširovali. Smerodajné odchýlky boli vždy vo všetkých príkladoch menšie u odhadu Kendallovho tau nerozšírených náhodných veličín. Avšak treba brať ohľad na to, že v prípade, keď spojitou nerozširujeme, potrebujeme pri výpočte odhadu Kendallovho tau okrem počtu konkordantných dvojíc, počítať aj počet diskordantných dvojíc. To nám vie pri veľkom množstve dát predĺžiť čas výpočtu.

# Zoznam použitej literatúry

- ANDĚL, J. (2005). *Základy matematické statistiky*. Matfyzpress, Praha.
- DENUIT, M. a LAMBERT, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, **93**(1), 40–57.
- DURANTE, F., FERNANDEZ-SANCHEZ, J. a SEMPI, C. (2013). A topological proof of Sklar’s theorem. *Applied Mathematics Letters*, **26**(9), 945–948.
- EMBRECHTS, P., KLÜPPELBERG, C. a MIKOSCH, T. (2002). Modelling extremal events: for insurance and finance. *Springer-Verlag*.
- GENEST, C. a NEŠLEHOVÁ, J. (2007). A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA*, **37**(2), 475–515.
- LEE, A. J. (1990). *U-Statistics: Theory and Practice*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis. ISBN 9780824782535.
- R CORE TEAM (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RÉMILLARD, B. (2013). *Statistical methods for financial engineering*. CRC press.
- SKLAR, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Annales de l’ISUP*, **8**(3), 229–231.
- TRIVEDI, P. K. a ZIMMER, D. M. (2007). Copula modeling: an introduction for practitioners. *Foundations and Trends® in Econometrics*, **1**(1), 1–111.