UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MATEMATICKÉ ZAUJÍMAVOSTI

BAKALÁRSKA PRÁCA

2014 Michaela ILIŤOVÁ

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MATEMATICKÉ ZAUJÍMAVOSTI

BAKALÁRSKA PRÁCA

Študijný program:	Ekonomická a finančná matematika
Študijný odbor:	1114 Aplikovaná matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	**RNDr. Beáta Stehlíková, PhD.**

Bratislava 2014

Michaela ILIŤOVÁ

14112131

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Michaela Ilit'ová
Študijný program:	ekonomická a finančná matematika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor:	9.1.9. aplikovaná matematika
Typ záverečnej práce:	bakalárska
Jazyk záverečnej práce:	slovenský

Názov: Matematické zaujímavosti
Mathematical curiosities

Cieľ: Bakalárske práce z predchádzajúcich rokov [Ďuratná, 2011], [Trajová, 2012], [Mészárosová, 2013] obsahovali motivačné príklady založené na ukážkach z filmov a románov. Táto práca bude sledovať podobný cieľ – vzbudiť záujem o vysokoškolskú matematiku, ukázať, že je zaujímavá. Spracované témy však nebudú vychádzať z filmov a kníh, ale budú motivované rôznym spôsobom, podobne, ako je tomu v populárnych matematických knihách. Výsledkom bude teda text podobný týmto knihám.

Vedúci:	RNDr. Mgr. Beáta Stehlíková, PhD.
Katedra:	FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry:	prof. RNDr. Daniel Ševčovič, CSc.
Dátum zadania:	18.10.2013

Dátum schválenia: 14.11.2013

doc. RNDr. Margaréta Halická, CSc.
garant študijného programu

.... študent

.... vedúci práce

Poďakovanie

Touto cestou sa chcem poďakovať svojej vedúcej bakalárskej práce RNDr. Beáte Stehlíkovej, PhD. za jej ochotu, pomoc, odborné rady a podnetné pripomienky, ktoré mi pomohli pri písaní tejto práce. Ďakujem aj svojej rodine, priateľovi a priateľom za ich trpezlivosť a podporu.

Abstrakt

ILIŤOVÁ, Michaela: *Matematické zaujímavosti* [Bakalárska práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: RNDr. Beáta Stehlíková, PhD. Bratislava, 2014, 52 s.

V našej práci sa zaoberáme tromi pravdepodobnostnými paradoxmi a to paradoxom kravát, problémom dvoch obálok a Simpsonovým paradoxom. Pri každom uvádzame jeho zadanie, riešenie, prípadne riešenia a zovšeobecnenia. Naším cieľom je vzbudiť záujem o štúdium vysokoškolskej matematiky a pokračovanie v ňom, preto sú riešenia formulované čo možno najzrozumiteľnejšie. Diskrétne riešenie paradoxu kravát z podkapitoly 1.1 ako aj výpočty v reálnych príkladoch výskytu Simpsonovho paradoxu z 3.1 sme formulovali zrozumiteľne aj pre študentov posledných ročníkov gymnázií a stredných škôl. V práci sa venujeme už spomínanému diskrétnemu a aj spojitému riešeniu paradoxu kravát, riešeniu problému dvoch obálok aj pri formulácii s dodatočnou informáciou, hľadaniu podmienky pre výhodnosť výmeny obálky, reálnym výskytom Simpsonovho paradoxu a otázke pravdepodobnosti jeho výskytu. Práca obsahuje zaujímavé matematické problémy, ktoré môžu poslúžiť na propagáciu vysokoškolského štúdia matematiky ako aj na rozšírenie obzoru čitateľa.

Kľúčové slová: Paradox kravát, Problém dvoch obálok, Simpsonov paradox

Abstract

ILIŤOVÁ, Michaela: *Mathematical curiosities* [Bachelor Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: RNDr. Beáta Stehlíková, PhD. Bratislava, 2014, 52 p.

In our thesis we discuss three probability paradoxes, namely the necktie paradox, the two envelopes problem and the Simpson's paradox. For each we present its formulation, solution, or solutions and expansions. Our objective is to rouse interest in university mathematical studies and their continuation; therefore the solutions are drawn up in a most understandable way. The discrete solution of the two envelopes paradox from subsection 1.1 as well as the calculations in the examples of real occurrence of the Simpson's paradox in 3.1 are made to be understandable even for high school students. In the thesis we address the aforementioned discrete, as well as the continuous solution of the necktie paradox, the solution of the two envelopes paradox, even of the formulation with additional information, the search for the condition of the advantageous envelopes switch, the real occurrence of Simpson's paradox and the question of the probability of its occurrence. The thesis contains interesting mathematical problems, which can serve to propagate university studies of mathematics, as well as to widen our reader's horizon.
Keywords: Necktie Paradox, Two Envelopes Problem, Simpson's Paradox

Predhovor

Ako to iste vieš aj ty, milý čitateľ, matematika je pre veľkú časť populácie nezaujímavá, ba je až postrachom. Jej využitie v reálnom svete si ľudia viac či menej uvedomujú, avšak prívlastkom pútavá či príťažlivá alebo nebodaj zaujímavá by ju len ťažko označili. Preto je málo tých, ktorí sa odvážia preniknúť do nej hlbšie a ešte menej tých, ktorí sa rozhodnú zasvätiť jej svoj život. Veď na prvý pohľad vyzerá tak nelákavo, samé čísla, počty, problémy, koho by už len niečo také mohlo osloviť? Dokonca aj tí, ktorí sa jej sprvu nezľaknú, sa často utopia v mori nudných definícii, viet a dôkazov. Nadobudnú presvedčenie, že matematika nemôže byť zaujímavá, že je na to príliš skostnatená, že záľubu v nej môžu mať len čudáci, jednoducho to nie je nič, čo by upútalo moderného človeka.

Napriek tomu má matematika aj inú tvár, ktorá je tak zarážajúca, až z nej človeku padne sánka od úžasu a ktorej pochopenie ho naplní slastným pocitom uspokojenia. Presne túto jej stránku sa ti na nasledujúcich stranách pokúsime priblížiť. Samozrejme len jej malú časť. Matematických zaujímavostí je totižto neúrekom a stále pribúdajú nové. Preto nie je v našich silách ťa oboznámiť so všetkými. My ti len ponúkneme pohľad cez kľúčovú dierku, malý náznak toho, čo sa nachádza za dverami. Otvoriť ich a vstúpiť do jej divotvorného sveta je už ale na tebe. Jedno ti však vieme povedať s istotou. Neoľutuješ to.

Obsah

Úvod

Matematika je už od dávna predmetom ľudského skúmania. Cez stáročia v nej pribudlo nespočetne veľa poznatkov, z ktorých mnohé sú pre ľudstvo životne dôležité, nenahraditeľné ale aj fascinujúce. Táto práca bude zameraná na posledné z menovaných, fascinujúce, zaujímavé, pozoruhodné. O mnohých takýchto poznatkoch píšu autori populárnych matematických kníh, ako sú [1] alebo [15]. My sme sa rozhodli v tejto práci priblížiť čitateľovi zopár paradoxov, ktoré ako dúfame, ho upútajú a vzbudia v ňom záujem o matematiku ako celok.

Cieľom tejto práce je poskytnúť čitateľovi náhľad na vybrané matematické problémy, konkrétne tri pravdepodobnostné paradoxy a týmto spôsobom v ňom podnietiť záujem o ďalšie štúdium matematiky. Vzhľadom k tomu, že toto štúdium je spoločnosťou vnímané skôr v negatívnom svetle, ako náročné a nezaujímavé, pokúsime sa poukázaním na príťažlivú stránku matematiky povzbudiť čitateľa k zmene tohto názoru.

Dôvodom pre spracovanie tejto témy je na jednej strane snaha poskytnúť pútavý obraz matematiky čitateľovi a na druhej strane úsilie autora o prehĺbenie svojich poznatkov z rôznych oblastí matematiky.

V jednotlivých kapitolách sa budeme venovať už spomínaným trom paradoxom. Prvá kapitola sa zameriava na paradox kravát. Vychádzame hlavne z [9], [12] a [6]. V druhej kapitole sa budeme venovať problému dvoch obálok ako ho popisuje [3]. Simpsonov paradox, jeho výskyt v reálnom svete ako aj pravdepodobnosť jeho vzniku je obsahom tretej kapitoly. Čerpať budeme hlavne z [15] a [14]. V celej práci využívame základné poznatky z analýzy, ktoré sú čitateľovi k dispozícii v [5], a pravdepodobnosti, ktoré nájde napríklad v [4].

1 Paradox kravát

Paradox kravát je relatívne známy a rozšírený. Jeho zadanie, ktoré teraz uvedieme, sme čerpali z [13]. Predstavme si nasledovnú situáciu: Dvaja muži dostali od svojich manželiek na Vianoce kravaty. Začnú sa hádať, koho kravata je drahšia, až ich hádka vyústi do stávky. Opýtajú sa manželiek koľko stála ich kravata a ten s drahšou kravatou ju dá tomu s lacnejšou. Predpokladajme, že obe manželky sú s rovnakou pravdepodobnosťou ochotné zaplatiť za tú ktorú kravatu príslušnú sumu a kupovali ich nezávisle od seba, bez toho, aby vedeli, akú kravatu kúpila tá druhá. Obom sa zdá, že stávka je v ich prospech. Veď možnosť, že prehrajú alebo vyhrajú je rovnako pravdepodobná. Ak je ich kravata drahšia, tak prídu len o *hodnotu svojej kravaty*, ak je však lacnejšia, tak získajú *viac ako hodnotu svojej kravaty*. Avšak aby takáto stávka bola v prospech oboch nie je možné.

Paradox kravát má aj inú obmenu, ktorá sa veľmi často vyskytuje v literatúre, napr. v [9], odkiaľ vychádzame. Je ňou paradox peňaženiek občas nazývaný aj paradox hry s peňaženkami. Problém v tomto paradoxe zostáva rovnaký ako pri paradoxe kravát, mení sa len situácia, v ktorej sa uskutočňuje stávka, respektíve hra. Dvaja hráči položia svoje peňaženky na stôl. Následne sa spočítajú peniaze v každej peňaženke a hráč s menšou sumou v peňaženke vyhráva všetky peniaze, ktoré mal súper v tej jeho. Aby sa dosiahla istá „férovosť" tejto hry musíme predpokladať, že ani jeden z hráčov nenosí vždy viac peňazí v peňaženke, lebo v takom prípade by bol samozrejme stále v nevýhode a pravdepodobne by sa do tejto hry nezapojil. Ďalej s ohľadom na „férovosť" budeme predpokladať, že sumy v peňaženkách (teda aj ceny kravát) sú nezávislé náhodné premenné z rovnakého rozdelenia, čo považujú autori [9] za najprirodzenejší predpoklad.

Aj v paradoxe peňaženiek predpokladajú obaja hráči, že hra je v ich prospech, nakoľko ak prehrajú, tak stratia len sumu v svojej peňaženke, ale ak vyhrajú, tak získajú viac ako sumu vo svojej peňaženke. Avšak nie je možné, aby bola táto hra výhodná pre oboch. Dochádza teda k rovnakému omylu ako pri paradoxe kravát (samozrejme za predpokladu, že v peňaženkách nemôže byť dlžobný úpis). Analýza paradoxu peňaženiek je vo svojej podstate rovnaká ako paradoxu kravát, preto sa mu nebudeme venovať samostatne. Riešenia, ktoré uvedieme pre paradox kravát je teda možné aplikovať aj na paradox peňaženiek.

1.1 Diskrétne riešenie paradoxu kravát

V nasledujúcich úvahách vychádzame z [12]. Uvažujme najskôr jednoduchú situáciu, že sa kravaty predávajú len za dve ceny a to 10€ a 15€. V Tab. 1.1 sú uvedené všetky možné situácie, pričom zo zadania úlohy vyplýva (vzhľadom na nezávislosť a rovnakú ochotu u oboch manželiek kúpiť kravatu za 10€, alebo 15€), že pravdepodobnosť každej zo situácii je rovnaká.

Cena kravaty prvého muža	Cena kravaty druhého muža	Zisk prvého muža	Pravdepodobnosť zisku prvého muža
10€	10€	0	50%
10€	15€	Získa 15€	25%
15€	10€	Stratí 15€	25%
15€	15€	0	50%

Tab. 1.1 Tabuľka cien kravát a zisku prvého muža [12]

Ako vidíme v Tab. 1.1, pravdepodobnosť, že ani jeden nič nezíska ani nestratí, je 0,5. Šanca, že prvý muž získa 15€ je rovnaká, ako že 15€ stratí, konkrétne 0,25. Pozrime sa teda ešte raz na to, čo sme nazvali *hodnotou jeho kravaty* a *viac ako hodnotou jeho kravaty* a všimnime si, že ide o rovnakú sumu, konkrétne 15€. Zároveň je pravdou, že v prvom prípade je to skutočne *viac ako hodnota jeho kravaty*, keďže tá stála 10€ a v druhom len *hodnota jeho kravaty*.

V [6] zovšeobecnili prípad s dvoma cenami kravát na rovnomerné rozdelenie cien s hodnotami 0,1,...,100. My budeme uvažovať n cien. Stále predpokladáme rovnomerné rozdelenie pravdepodobnosti kúpy kravaty medzi cenami od x_1 po x_n, pričom $x_1 < x_2 < \cdots < x_n$. Pravdepodobnosť kúpy každej z n kravát je potom rovná $1/n$. Tab. 1.2 zobrazuje zisk prvého muža pri všetkých kombináciách cien kravát.

Cena kravaty prvého muža	x_1	x_1	...	x_1	x_2	x_2	...	x_2	...
Cena kravaty druhého muža	x_1	x_2	...	x_n	x_1	x_2	...	x_n	...
Zisk prvého muža	0	x_2	...	x_n	$-x_2$	0	...	x_n	...

...	x_i	...	x_i	x_i	...	x_i	...	x_n	...	x_n
...	x_1	...	x_i	x_{i+1}	...	x_n	...	x_1	...	x_n
...	$-x_i$	...	0	x_{i+1}	...	x_n	...	$-x_n$	...	0

Tab. 1.2 Zisk prvého muža pri n rôznych cenách kravát

Podľa Tab. 1.2 vieme určiť pravdepodobnosť zisku pre každú z cien x_j pre $j = 1, \dots, n$. Tieto pravdepodobnosti zachytáva Tab. 1.3.

Zisk	$-x_n$	...	$-x_i$	...	$-x_2$	$-x_1$	0	x_1	x_2	...	x_i	...	x_n
P(Zisk)	$\frac{n-1}{n^2}$	...	$\frac{i-1}{n^2}$	...	$\frac{1}{n^2}$	0	$\frac{1}{n}$	0	$\frac{1}{n^2}$	...	$\frac{i-1}{n^2}$	...	$\frac{n-1}{n^2}$

Tab. 1.3 Pravdepodobnosti zisku prvého muža pri n rôznych cenách kravát

Strednú hodnotu zisku pre diskrétne rozdelenie potom vypočítame zo vzorca

$$E(Zisk) = \sum_{j=1}^{2n+1} Zisk_j \cdot P\left(Zisk_j\right),$$

kde $Zisk_j$ je j-ta možná hodnota dosiahnutého zisku ($Zisk_1 = -x_n$, $Zisk_2 = -x_{n-1}$,..., $Zisk_{2n+1} = x_n$).

V našom prípade

$$E(Zisk) = (-x_n)\frac{n-1}{n^2} + \cdots + (-x_i)\frac{i-1}{n^2} + \cdots + (-x_2)\frac{1}{n^2} + (-x_1) \cdot 0$$

$$+0 \cdot \frac{1}{n} + x_1 \cdot 0 + x_2\frac{1}{n^2} + \cdots + x_i\frac{i-1}{n^2} + \cdots + x_n\frac{n-1}{n^2},$$

čo môžeme napísať ako

$$E(Zisk) = \sum_{i=1}^{n} (-x_i)\frac{i-1}{n^2} + x_i\frac{i-1}{n^2},$$

a teda

$$E(Zisk) = 0.$$

Vidíme, že v konečnom dôsledku je stredná hodnota zisku prvého a analogicky aj druhého muža nulová. Stávka preto nie je v prospech žiadneho z nich.

1.2 Spojité riešenie paradoxu kravát

Teraz prejdeme k všeobecnejšiemu, spojitému riešeniu, pričom zovšeobecníme naše úvodné predpoklady. Pri spojitom riešení, na rozdiel od diskrétneho, berieme do úvahy stávky pri všetkých (aj reálne neexistujúcich ako napríklad π eur) možnostiach cien kravát prvého a druhého muža. Takáto aproximácia sa bežne využíva pri veľkom množstve možných výsledkov. Hodnoty kravát sú teda náhodné premenné, ktoré označíme X u prvého muža a Y u druhého. X a Y sú kladné, iným slovom vylučujeme možnosť, že manželky dostali kravatu zadarmo, alebo že by im dokonca obchody za kúpu kravaty zaplatili. Stále predpokladáme, že manželky kupovali kravaty nezávisle od seba a že obe boli, voľne povedané[1], ochotné zaplatiť za kravatu konkrétnu sumu s rovnakou pravdepodobnosťou. Matematicky vyjadrené, X a Y sú nezávislé s rovnakým rozdelením. Hustota ceny, ktorú označíme $f(x)$ ako aj distribučná funkcia $F(x)$, je teda rovnaká pre kravaty oboch mužov. Obe funkcie sú nulové pre nekladné argumenty, a teda aj

$$F(0) = 0. \tag{1.1}$$

Bez ujmy na všeobecnosti budeme riešiť tento problém z pohľadu prvého muža. Zisk prvého muža, ktorý je tiež náhodnou premennou označíme Z, jeho hustotu $f_Z(z)$ a distribučnú funkciu $F_Z(z)$. Zisk Z môže nadobúdať tri hodnoty. Ak je $X < Y, Z = Y$, iným slovom ak je kravata prvého muža lacnejšia tak získa kravatu druhého. Ak je naopak jeho kravata drahšia, musí ju druhému mužovi dať a teda ak $X > Y, Z = -X$. V prípade rovnosti $X = Y$ je $Z = 0$, žiaden z nich nič nezíska ani nestratí. Teda

$$Z = \begin{cases} Y & pre \quad X < Y, \\ -X & pre \quad X > Y, \\ 0 & pre \quad X = Y. \end{cases} \tag{1.2}$$

V nasledujúcom texte vypočítame hustotu zisku Z a pomocou nej vyjadríme hodnotu nepodmienenej strednej hodnoty zisku. Náš prístup je teda iný ako autorov [9], ktorý nepočítali hustotu zisku a uviedli len strednú hodnotu podmienenú cenou kravaty prvého muža.

Teraz uvedieme výpočet $f_Z(z)$, hustoty náhodnej premennej Z, z jej distribučnej funkcie $F_Z(z)$. Budeme pri tom využívať základné vlastnosti náhodných premenných a vektorov, ktoré s popísané napríklad v [4].

[1] V skutočnosti je pravdepodobnosť kúpy kravaty za jednu konkrétnu sumu pri spojitom rozdelení nulová. Táto formulácia je teda len intuitívnym vyjadrením.

Podľa definície distribučnej funkcie

$$F_Z(z) = P(Z < z). \tag{1.3}$$

Pravdepodobnosť $P(Z < z)$ vyjadríme podľa vety o úplnej pravdepodobnosti

$$\begin{gathered} F_Z(z) = P\big((Z < z) \wedge (X < Y)\big) \vee P\big((Z < z) \wedge (X > Y)\big) \\ \vee P\big((Z < z) \wedge (X = Y)\big). \end{gathered} \tag{1.4}$$

Z disjunktnosti udalostí v (1.4) vieme, že

$$\begin{gathered} F_Z(z) = P\big((Z < z) \wedge (X < Y)\big) + P\big((Z < z) \wedge (X > Y)\big) \\ +P\big((Z < z) \wedge (X = Y)\big). \end{gathered} \tag{1.5}$$

Keďže uvažujeme spojité rozdelenie,

$$P(X = Y) = 0, \tag{1.6}$$

a preto posledný člen v (1.5) je rovný nule a môžeme túto rovnosť napísať ako

$$F_Z(z) = P\big((Z < z) \wedge (X < Y)\big) + P\big((Z < z) \wedge (X > Y)\big). \tag{1.7}$$

Rozlíšime tri prípady a to $z = 0$, $z > 0$ a $z < 0$.

<u>Prípad 1: $z = 0$</u>

V prvom prípade pre $z = 0$ je výsledok zrejmý priamo z definície distribučnej funkcie (1.3) po dosadení

$$F_Z(0) = P(Z < 0).$$

Pravdepodobnosť, že zisk Z je záporný, je vzhľadom na nezávislosť X, Ya ich rovnaké rozdelenie rovný[2] 0,5, a teda

$$F_Z(z) = \frac{1}{2} \quad pre \quad z = 0. \tag{1.8}$$

<u>Prípad 2: $z > 0$</u>

Pre $z > 0$ vyjadríme hodnoty oboch sčítancov v (1.7) samostatne. Začneme s $P\big((Z < z) \wedge (X > Y)\big)$. Ak je $X > Y$, potom podľa (1.2) je $Z = -X$. Keďže $X > 0$, Z je v tomto prípade záporné a keďže z je kladné, určite platí $Z < z$, a preto

---

[2] Presný dôkaz tohto intuitívne jasného tvrdenia spravíme nasledovným výpočtom:

$$\begin{gathered} P(Z < 0) = P(X > Y) = \int_0^\infty \int_x^x f(x) f(y) dx\, dy \\ = \int_0^\infty f(x) F(x) dx = \int_0^1 u du = \frac{1}{2}, \end{gathered}$$

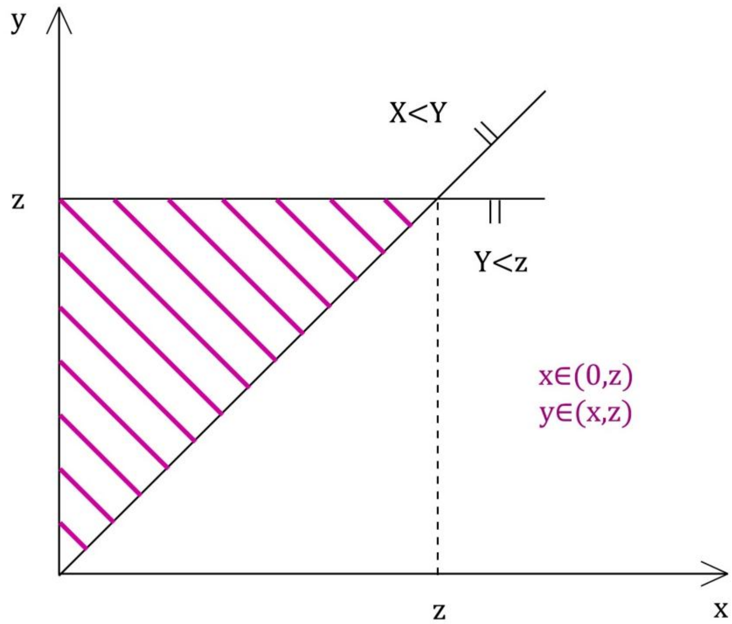
kde sme pri výpočte využili substitúciu $F(x) = u$. Zdôvodnenie jednotlivých krokov je podobné ako pri výpočtoch, ktoré uvádzame v hlavnom texte, preto ich v tejto poznámke vynechávame.

$$P\big((Z < z) \wedge (X > Y)\big) = P(X > Y) = \frac{1}{2}. \tag{1.9}$$

Ďalej upravíme $P\big((Z < z) \wedge (X < Y)\big)$ podľa (1.2)

$$P\big((Z < z) \wedge (X < Y)\big) = P\big((Y < z) \wedge (X < Y)\big) \tag{1.10}$$

a vyjadríme (1.10) pomocou integrálu zo združenej hustoty vektora (X, Y), ktorého integračnú množinu vidno na Obr. 1.1.



Obr. 1.1 Integračná množina pre $P\big((Y < z) \wedge (X < Y)\big)$, kde $z > 0$

Z nezávislosti X a Y vieme, že združená hustota je súčin hustôt, a preto

$$\begin{aligned} P\big((Y < z) \wedge (X < Y)\big) &= \int_0^z \int_x^z f(x) f(y) dy\, dx \\ &= \int_0^z f(x) \int_x^z f(y) dy\, dx. \end{aligned} \tag{1.11}$$

Keďže

$$\int_x^z f(y) dy = P\big(Y \in (x, z)\big) = F(z) - F(x), \tag{1.12}$$

pravdepodobnosť v (1.11) nadobudne tvar

$$P\big((Y < z) \wedge (X < Y)\big) = \int_0^z f(x)[F(z) - F(x)] dx \tag{1.13}$$

$$= F(z)\int_0^z f(x)dx - \int_0^z f(x)F(x)dx.$$

Obidva integrály v (1.13) vyjadríme samostatne. Prvý z nich podobne ako (1.12)

$$\int_0^z f(x)dx = P\big(Y \in (0,z)\big) = F(z) - F(0),$$

čo využitím (1.1) môžeme napísať ako

$$\int_0^z f(x)dx = F(z). \tag{1.14}$$

V druhom integráli použijeme nasledovnú substitúciu:

$$F(x) = u,$$

$$f(x)dx = du,$$

a teda

$$\int_0^z f(x)F(x)dx = \int_{F(0)}^{F(z)} u\,du. \tag{1.15}$$

Po integrovaní a s využitím (1.1) vieme (1.15) upraviť na

$$\int_0^z f(x)F(x)dx = \left[\frac{u^2}{2}\right]_0^{F(z)} = \frac{F(z)^2}{2}. \tag{1.16}$$

Dosadením (1.14) a (1.16) do (1.13) získavame

$$P\big((Y < z) \wedge (X < Y)\big) = F(z)^2 - \frac{F(z)^2}{2} = \frac{F(z)^2}{2}. \tag{1.17}$$

Distribučnú funkciu (1.7) pre z kladné vyjadríme pomocou (1.9) a (1.17) ako

$$F_Z(z) = \frac{1}{2} + \frac{F(z)^2}{2} \quad pre \quad z > 0. \tag{1.18}$$

<u>Prípad 3: $z < 0$</u>

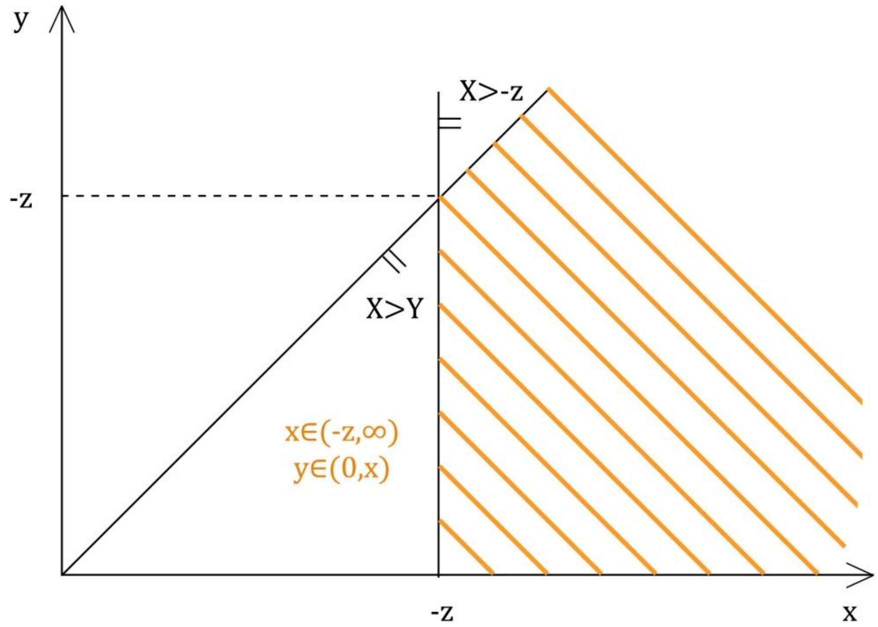
Zostáva nám posledný prípad a to $z < 0$. Budeme postupovať obdobne ako v predchádzajúcom prípade. Najskôr vyjadríme z (1.7) pravdepodobnosť $P\big((Z < z) \wedge (X < Y)\big)$. Podľa (1.2) vieme, že pre $X < Y$ je $Z = Y$. Avšak keďže $Y > 0$, nie je možné aby platilo $Z < z$ pre z záporné. Preto

$$P\big((Z < z) \wedge (X < Y)\big) = 0. \tag{1.19}$$

Prejdeme k druhému sčítancu v (1.7). Pre $X > Y$ podľa (1.2) platí $Z = -X$ a pravdepodobnosť $P\big((Z < z) \wedge (X > Y)\big)$ vieme potom prepísať takto

$$P\big((Z < z) \wedge (X > Y)\big) = P\big((X > -z) \wedge (X > Y)\big), \tag{1.20}$$

kde $-z > 0$. Rovnako ako v predošlom prípade budeme pravdepodobnosť (1.20) vyjadrovať zo združenej hustoty vektora (X, Y). Integračnú množinu zobrazuje Obr. 1.2.



Obr. 1.2 Integračná množina pre $P\big((X > -z) \wedge (X > Y)\big)$, kde $z < 0$

Znovu využijeme nezávislosť X a Y a vyjardíme pravdepodobnosť (1.20) ako integrál zo súčinu hustôt

$$\begin{aligned} P\big((X > -z) \wedge (X > Y)\big) &= \int_{-z}^{\infty} \int_{0}^{x} f(x) f(y) dy\, dx \\ &= \int_{-z}^{\infty} f(x) \int_{0}^{x} f(y) dy\, dx. \end{aligned} \tag{1.21}$$

Vnútorný integrál vyjadríme cez distribučné funkcie tak, ako sme to spravili aj v (1.12) a zároveň využijeme na úpravu (1.1)

$$\int_{0}^{x} f(y) dy = P\big(Y \in (0, x)\big) = F(x) - F(0) = F(x),$$

a teda (1.13) môžeme napísať ako

$$P\big((X > -z) \wedge (X > Y)\big) = \int_{-z}^{\infty} f(x)F(x)dx. \tag{1.22}$$

Opäť využijeme substitúciu

$$F(x) = u,$$

$$f(x)dx = du,$$

pomocou ktorej vieme integrál v (1.22) vyjadriť takto:

$$\begin{aligned} P\big((X > -z) \wedge (X > Y)\big) &= \int_{F(-z)}^{F(\infty)} u\,du = \left[\frac{u^2}{2}\right]_{F(-z)}^{F(\infty)} \\ &= \frac{F(\infty)^2 - F(-z)^2}{2}, \end{aligned} \tag{1.23}$$

kde sme limitu $\lim_{s\to\infty} F(s)$ označili $F(\infty)$. Keďže distribučná funkcia v nekonečne dosahuje limitne hodnotu jedna, (1.23) je rovné

$$P\big((X > -z) \wedge (X > Y)\big) = \frac{1}{2} - \frac{F(-z)^2}{2} \tag{1.24}$$

Využitím (1.19) a (1.24) vieme potom distribučnú funkciu (1.7) pre z záporné upraviť nasledovne

$$F_Z(z) = \frac{1}{2} - \frac{F(-z)^2}{2} \quad pre \quad z < 0. \tag{1.25}$$

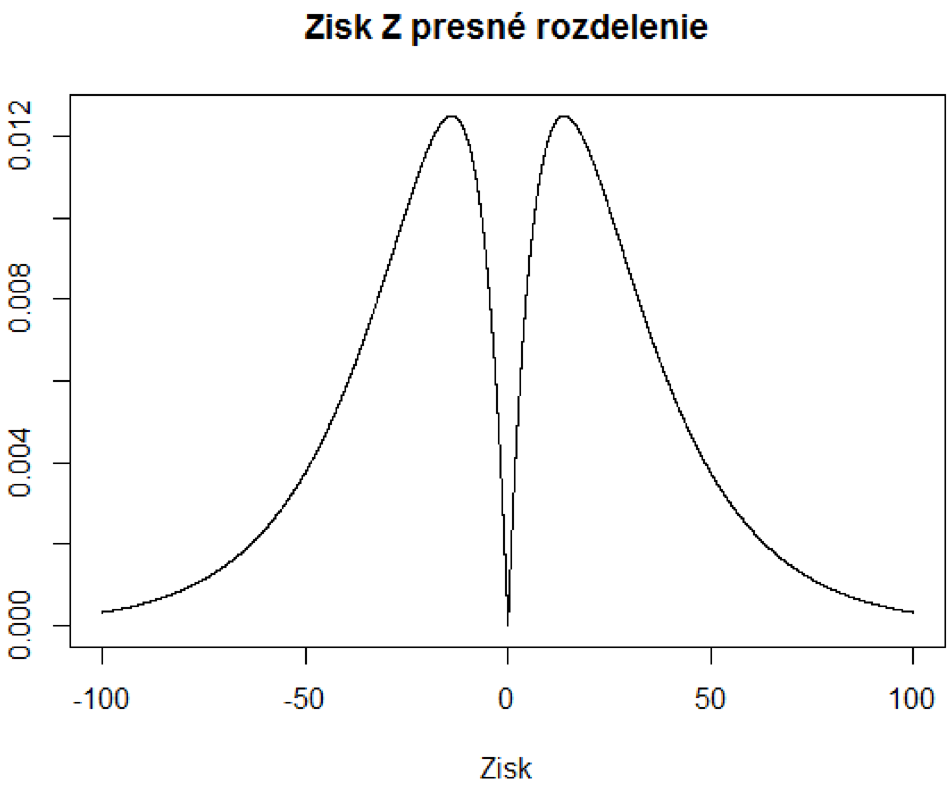
Teraz vyjadríme $F_Z(z)$ v (1.7) pre ľubovoľné z spojením (1.8), (1.18) a (1.25)

$$F_Z(z) = \begin{cases} \dfrac{1}{2} + \dfrac{F(z)^2}{2} & pre \quad z > 0, \\ \dfrac{1}{2} & pre \quad z = 0, \\ \dfrac{1}{2} - \dfrac{F(-z)^2}{2} & pre \quad z < 0. \end{cases} \tag{1.26}$$

Hustota zisku Z je deriváciou jeho distribučnej funkcie (1.26), pričom hodnotu pre $z = 0$ si zvolíme[3] tak, aby bola $f_Z(z)$ spojitá v prípade, že $f(0) = 0$, alebo aspoň $F(z)f(z) \to 0$ pre $z \to 0^+$, čo je pre mnohé pravdepodobnostné rozdelenia splnené:

$$f_Z(z) = \begin{cases} F(z)f(z) & pre \quad z > 0, \\ 0 & pre \quad z = 0, \\ F(-z)f(-z) & pre \quad z < 0. \end{cases} \tag{1.27}$$

[3] Ak zmeníme hodnotu hustoty v jednom bode, stále zostáva hustotou tej istej náhodnej premennej.



Obr. 1.3 Hustota zisku Z pre X, Y z exponenciálneho rozdelenia s $\lambda = 0{,}05$

Hustotu zisku, t.j. náhodnej premennej Z, ak X, Y sú náhodné premenné s exponenciálnym rozdelením s $\lambda = 0{,}05$ ilustruje Obr. 1.3. Pri definovaní exponenciálneho rozdelenia nie je v literatúre jednoznačná parametrizácia. My uvažujeme hustotu

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & inak. \end{cases}$$

To znamená, že stredná hodnota takejto náhodnej premennej je $\frac{1}{\lambda}$, v našom prípade 20.

Ako vidíme Obr. 1.3 je symetrický, teda očakávame že stredná hodnota zisku prvého muža, ktorú označíme $E(Z)$ je rovná 0. Overíme to nasledovným výpočtom, v ktorom využijeme rovnosť $f_Z(z) = f_Z(-z)$, vyplývajúcu z (1.27), matematicky vyjadrujúcu symetriu, ktorú môžeme pozorovať na Obr. 1.3. Budeme pritom predpokladať, že stredná hodnota náhodnej premennej X existuje.

Do všeobecného vzorca pre strednú hodnotu

$$E(Z) = \int_{-\infty}^{\infty} t f_Z(t) dt$$

dosadíme hustotu (1.27), čím dostaneme

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} t\left(\frac{1}{2}F(t)f(t) + \frac{1}{2}F(-t)f(-t)\right)dt \\ &= \frac{1}{2}\int_{-\infty}^{\infty} tF(t)f(t)dt + \frac{1}{2}\int_{-\infty}^{\infty} tF(-t)f(-t)dt. \end{aligned} \tag{1.28}$$

Existencia týchto integrálov vyplýva z predpokladu existencie strednej hodnoty náhodnej premennej[4] X. Teraz uskutočníme substitúciu $-t = s$ v druhom sčítanci (1.28):

$$E(Z) = \frac{1}{2}\int_{-\infty}^{\infty} tF(t)f(t)dt + \frac{1}{2}\int_{\infty}^{-\infty} (-s)F(s)f(s)(-ds).$$

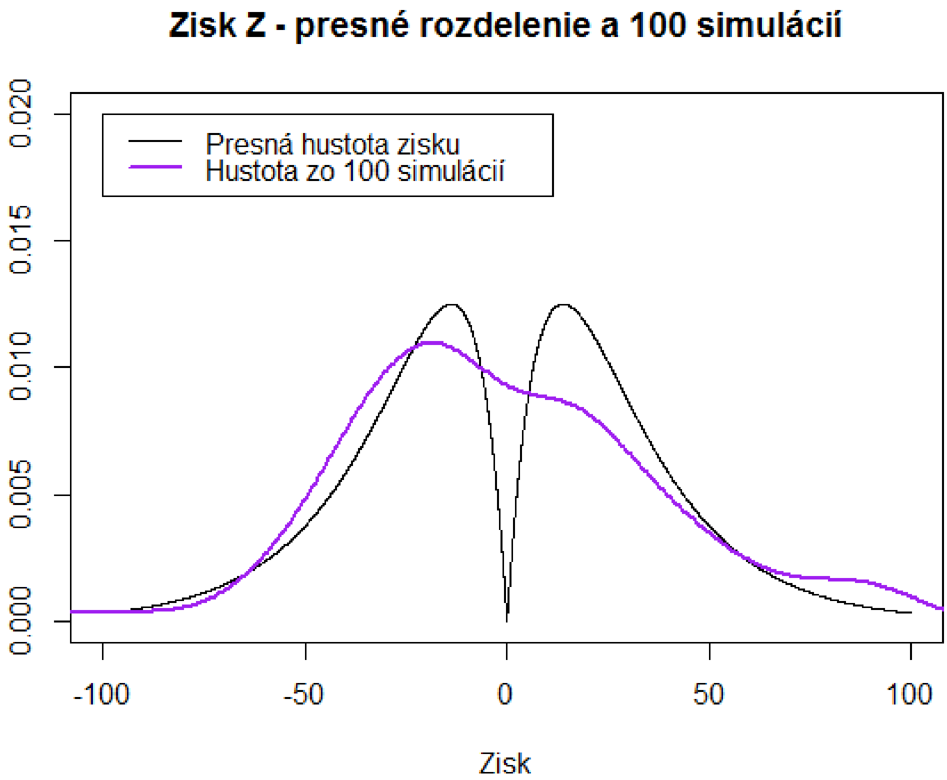
Nakoniec otočíme interval integrácie druhého sčítanca, čím sa zmení aj jeho znamienko:

$$E(Z) = \frac{1}{2}\int_{-\infty}^{\infty} tF(t)f(t)dt - \frac{1}{2}\int_{-\infty}^{\infty} sF(s)f(s)ds = 0. \tag{1.29}$$

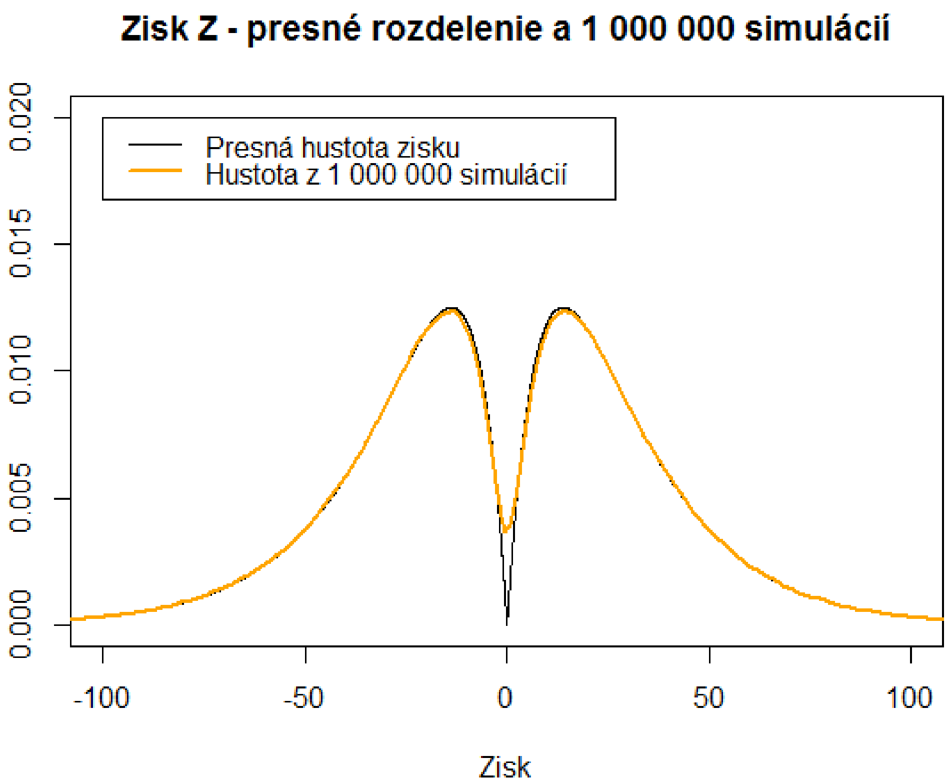
Podľa (1.29) ako aj podľa symetrickosti funkcie hustoty (1.27) vidíme, že v zovšeobecnenom modeli je stredná hodnota zisku prvého aj druhého muža rovná nule, a teda vzhľadom na zisk daného muža tento model funguje rovnako ako ilustračný model s dvomi cenami.

Pridávame možnosť pozrieť si simulácie stávok zisku Z pre náhodné hodnoty X, Y z exponenciálneho rozdelenia s $\lambda = 0{,}05$ na Obr. 1.4 a Obr. 1.5. Ako vidíme, so stúpajúcim počtom simulácii sa hustota stále viac približuje presnej hustote zisku Z pre takéto X, Y.

[4] Podľa tohto predpokladu existuje integrál $\int_{-\infty}^{\infty}|x|f(x)\,dx$. Keďže distribučná funkcia F má obor hodnôt [0,1], platí $|tF(t)f(t)| \leq |t|f(t)$ a tiež $|tF(-t)f(-t)| \leq |t|f(-t)$.



Obr. 1.4 Sto simulácii zisku Z pre X, Y z exponenciálneho rozdelenia s $\lambda = 0{,}05$



Obr. 1.5 Milión simulácii zisku Z pre X, Y z exponenciálneho rozdelenia s $\lambda = 0{,}05$

Na generovanie obrázkov Obr. 1.4 a Obr. 1.5 sme použili program RStudio. Kód, ktorým sme generovali Obr. 1.4 zobrazuje Obr. 1.6. V prípade Obr. 1.5 sme použili ten istý kód (až na popisy a farbu grafu) avšak pre $N = 10^6$.

```
osX1 <- seq(-100,0,by=.01)
osX2 <- seq(0,100,by=.01)
osX3 <- seq(100,0,by=-.01)

N=10^2
x <- rexp(N, 0.05)   # cena kravaty prvého muža
y <- rexp(N, 0.05)   # cena kravaty druhého muža
zisk=rep(0,N)        # zisk prvého muža
for (i in 1:N){
  if (x[i]>y[i]) zisk[i]=-x[i]   # príde o svoju kravatu
  if (x[i]<y[i]) zisk[i]=y[i]    # získa kravatu druhého muža
}

plot(c(osX1,osX2) , c(dexp(osX3, 0.05)*pexp(osX3, 0.05)
                      ,dexp(osX2, 0.05)*pexp(osX2, 0.05)),
     type="l", ylim=c(0,0.02), xlab="Zisk", ylab="",
     main="Zisk Z - presné rozdelenie a 100 simulácií"
    )
lines(density(zisk) ,col="purple", lwd=2)
legend(x=-100,
       y=0.02,
       legend=c("Presná hustota zisku","Hustota zo 100 simulácií"),
       lty=c(1,1),
       lwd=c(1,2),
       col=c("black","purple")
      )
```

Obr. 1.6 Zdrojový kód použitý na vygenerovanie Obr. 1.4 v programe RStudio

Pri generovaní Obr. 1.4 a Obr. 1.5 sme využili funkcie density, dexp a pexp, ktoré teraz vysvetlíme. Funkcia density(x) podľa [16] odhaduje funkciu hustoty pre x, kde x je vektor hodnôt. V kóde sme ju použili na výpočet hustoty zisku Z zo simulácii. Funkcie dexp$(y, rate)$ a pexp$(y, rate)$, ako to uvádza [17], sa obe viažu na exponenciálne rozdelenie, kde y je vektor hodnot a $rate$ je rovné prevrátenej hodnote zo strednej hodnoty daného exponenciálneho rozdelenia, pričom dexp$(y, rate)$ počíta hodnotu hustoty a pexp$(y, rate)$ hodnotu distribučnej funkcie v y. Pomocou nich sme počítali hustotu zisku Z, ktorú sme funkciou plot následne vykreslili.

Dokázali sme teda matematicky, že ani jeden z mužov nemá v takto formulovanom prípade oproti tomu druhému výhodu. Uvedieme ešte jednoduchú intuitívnu úvahu, ktorou sa pokúsime úplne ozrejmiť, prečo je tomu tak.

V obchodoch sa nachádzajú kravaty s istými cenami. Obe manželky sú rovnako ochotné kúpiť kravatu pri danej cene a navzájom sa pri kúpe nijako neovplyvňujú, t.j. pravdepodobnosť, že kúpia kravatu za istú cenu, napr. 10€, je rovnaká u oboch, ale napríklad väčšia ako pravdepodobnosť, že kúpia kravatu za 100€. Cena kravaty ktorú kúpia sa stáva ziskom (alebo stratou) v stávke ich manželov, a teda pravdepodobnosť zisku oboch mužov pri každej cene je pre oboch rovnaká, čo znamená, že ani jeden nemôže s istotou povedať, že je preňho stávka výhodnejšia.

Pozrieme sa ešte na možný dôvod vzniku paradoxu. Budeme uvažovať len zisk Z za podmienky, že je $Z > 0$. Hustota v tomto prípade nadobudne tvar

$$f_Z(z) = \begin{cases} 2F(z)f(z) & z > 0, \\ \\ 0 & inak. \end{cases} \tag{1.30}$$

Teraz vypočítame strednú hodnotu $E(Z|Z > 0)$ pre náhodné premenné X, Y z exponenciálneho rozdelenia a porovnáme ju so strednou hodnotou $E(X)$. Použitím (1.30) vieme podmienenú strednú hodnotu zisku Z za podmienky, že je kladný vyjadriť nasledovne

$$\begin{aligned} E(Z|Z > 0) &= \int_0^\infty x \cdot 2\left(1 - e^{-\lambda x}\right)\lambda e^{-\lambda x}\, dx \\ &= 2\int_0^\infty x\lambda e^{-\lambda x}\, dx + \int_0^\infty x(-2\lambda)e^{-2\lambda x}\, dx. \end{aligned} \tag{1.31}$$

Oba integrály z (1.31) upravíme použitím per partes a základných poznatkov o určitých integráloch, ktoré sú uvedené v [5]. Potom

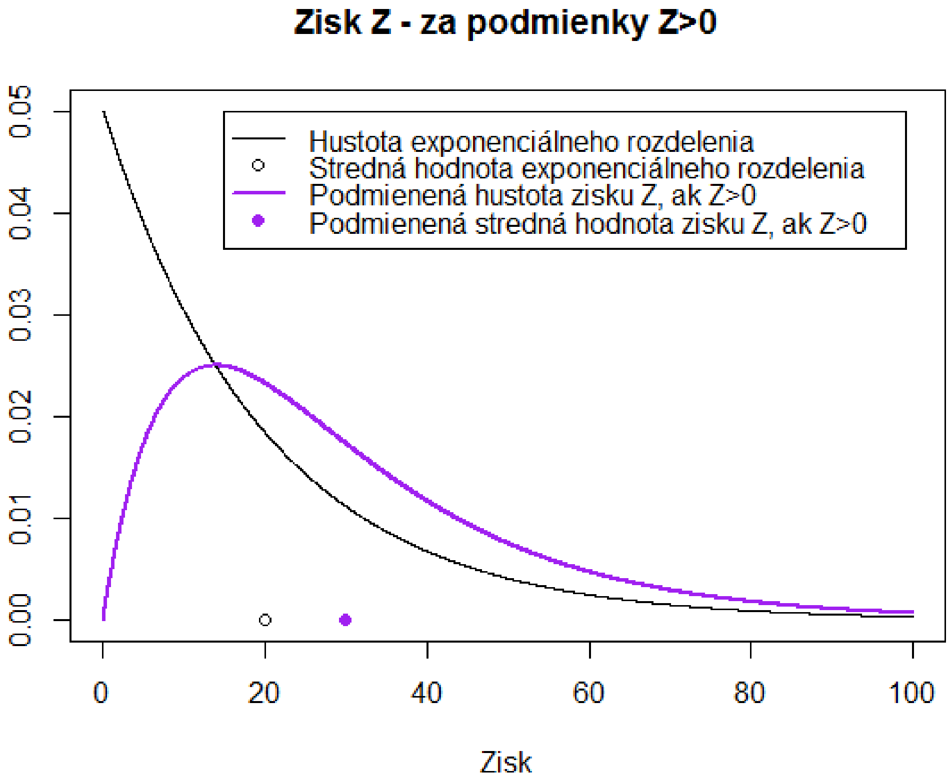
$$\begin{aligned} E(Z|Z > 0) &= 2\left(\left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\, dx\right) + \left[xe^{-2\lambda x}\right]_0^\infty - \int_0^\infty e^{-2\lambda x}\, dx \\ &= 2\left(0 + \left[\frac{-e^{-\lambda x}}{\lambda}\right]_0^\infty\right) + 0 - \left[\frac{-e^{-2\lambda x}}{2\lambda}\right]_0^\infty = \frac{2}{\lambda} - \frac{1}{2\lambda}, \end{aligned}$$

a teda

$$E(Z|Z > 0) = \frac{3}{2} \cdot \frac{1}{\lambda}. \tag{1.32}$$

Podmienená stredná hodnota zisku Z, ak $Z > 0$ ktorú vyjadruje (1.32) je väčšia ako stredná hodnota náhodnej premennej X, keďže $E(X) = \frac{1}{\lambda}$. Je možné, že práve táto skutočnosť vzbudzuje dojem výhodnosti stávky z pohľadu oboch mužov.

Obr. 1.7 zobrazuje podmienenú hustotu zisku (1.30) ako aj podmienenú strednú hodnotu zisku (1.32) pre X, Y z exponenciálneho rozdelenia s $\lambda = 0{,}05$ v porovnaní s hustotou a strednou hodnotou tohto exponenciálneho rozdelenia. Vidíme, že podmienená stredná hodnota $E(Z|Z > 0) = 30$ je väčšia ako stredná hodnota $E(X) = 20$.



Obr. 1.7 Hustota exponenciálneho rozdelenia s $\lambda = 0{,}05$ a zisku Z za podmienky $Z > 0$ pre X, Y z exponenciálneho rozdelenia s $\lambda = 0{,}05$

2 Problém dvoch obálok

V tomto probléme je hráčovi ponúknutá na výber jedna z dvoch nerozlíšiteľných obálok s peniazmi. V jednej je pri tom dvakrát viac peňazí ako v druhej. Hráč si vyberie jednu obálku a rozhoduje sa, či si ju nechá, alebo vymení za druhú.

Predpokladajme, že v obálke ktorú sme si zvolili je 20€. To znamená, že v druhej obálke môže byť buď 10€, alebo 40€. Pravdepodobnosť, že sme si zvolili obálku s menším obnosom peňazí je rovnaká ako že sme si zvolili tú s väčším a to 0,5. Môžeme sa domnievať, že stredná hodnota očakávanej sumy peňazí v druhej obálke bude rovná 25€, keďže

$$\frac{1}{2}10€+\frac{1}{2}40€=25€. \tag{2.1}$$

Vzhľadom k tomu, že tento obnos je vyšší ako obnos v pôvodne zvolenej obálke, zdá sa, že je pre nás vždy výhodné obálku vymeniť. Čo však ale na to hovorí matematika? Jednoducho, ako to často žartovne nazývajú profesori a prednášajúci, že spočítavame jablká s hruškami. V skutočnosti pri takomto zadaní problému vôbec nezáleží na tom aká suma sa nachádza v otvorenej obálke. Vždy sa v jednej nachádza x a v druhej $2x$ peňazí, ale ktorá z týchto súm je v našej obálke nevieme. To, či sa v nami zvolenej obálke nachádza 20€, či hoci aj 2000€ nedáva žiadnu ďalšiu informáciu pre výpočet strednej hodnoty, keďže skutočná stredná hodnota našej výhry sa vypočíta ako

$$\frac{1}{2}x+\frac{1}{2}2x=1{,}5x. \tag{2.2}$$

V celej druhej kapitole a jej podkapitolách vychádzame z [3].

2.1 Riešenie problému dvoch obálok

Prečo sa skutočná stredná hodnota vypočíta podľa (2.2) a nie podľa (2.1) a prečo hovoríme o spočítavaní jabĺk s hruškami? To si ukážeme v nasledujúcich výpočtoch.

Prvotne zvolenú obálku označíme $O1$ a jej obsah X. Druhú obálku označíme $O2$, sumu v nej Y. Množstvá peňazí v obálkach označíme M menšie z nich a V to väčšie. Konkrétne realizácie M, resp. V označíme x, resp. y. Vzťah medzi nimi udáva (2.3)

$$y=2x, \quad x>0. \tag{2.3}$$

Pravdepodobnosť, že v obálke, ktorú si zvolíme je menšia suma, je rovnaká ako pravdepodobnosť zvolenia obálky s väčšou sumou, a teda

$$P(X = x|M = x) = P(X = y|M = x) = \frac{1}{2}\cdot \qquad (2.4)$$

Ďalej zo zadania vieme, že

$$Y = \begin{cases} \frac{X}{2} & pre \quad Y < X, \\ 2X & pre \quad Y > X, \end{cases} \qquad (2.5)$$

čo platí analogicky, t.j. ak vymeníme všetky X a Y, aj pre X.

Počítajme strednú hodnotu pre sumu v $O2$ z vety o úplnej strednej hodnote

$$E(Y|M = x) = E(Y|Y > X, M = x)P(Y > X|M = x)$$
$$+E(Y|Y < X, M = x)P(Y < X|M = x). \qquad (2.6)$$

Využitím (2.4) a (2.5) môžeme strednú hodnotu (2.6) upraviť nasledovne

$$E(Y \,|\, M = x) = \frac{1}{2}E(2X \,|\, Y > X, M = x) + \frac{1}{2}E(X/2 \,|\, Y < X, M = x)$$
$$= E(X|Y > X, M = x) + \frac{1}{4}E(X|Y < X, M = x).$$

Keďže obsah obálky $O1$ už poznáme, strednú hodnotu Y vieme priamo vyjadriť využitím (2.3) a (2.5) nasledovne

$$E(Y|M = x) = x + \frac{1}{4}2x = \frac{3}{2}x.$$

Vidíme teda, že z pravdepodobnostného hľadiska je jedno, ktorú obálku si vyberieme, keďže nám ani jedna z možností, vymeniť či ponechať, neprinesie s vyššou pravdepodobnosťou väčší zisk.

2.2 Iný variant problému dvoch obálok s dodatočnou informáciou

Pozrime sa ešte na iný variant nášho problému, konkrétne, že sme dostali dodatočnú informáciu o podmienenej pravdepodobnosti toho, ktorá zo súm, väčšia, alebo menšia, je vo zvolenej obálke. Za podmienku budeme v tomto prípade pokladať zistenie výšky peňažného obnosu v nami zvolenej obálke. Prečo by ale takáto informácia mala zmeniť naše rozhodnutie? Najjednoduchšie si to predstavíme na situácii, kedy s istotou vieme, aká suma sa má nachádzať v obálke s väčším finančným obnosom. Ak sa v tomto prípade pozrieme do prvotne zvolenej obálky, hneď sa vieme rozhodnúť, či si ju ponechať, alebo vymeniť, a teda informácia o podmienenej pravdepodobnosti úplne zmenila situáciu.

Ako ale správne narábať s takouto dodatočnou informáciou v náročnejších prípadoch, keď už nie je na prvý pohľad zrejmé, či sa viac oplatí ponechať alebo vymeniť obálku? To sa pokúsime v nasledujúcej časti objasniť, pričom sa zameriame na prípady diskrétnych rozdelení, vzhľadom k tomu, že konečný počet možností pre konkrétne sumy v obálkach je podľa nás názornejší.

2.2.1 Podmienka výhodnosti výmeny obálky

V nasledujúcom texte budeme používať už zavedené označenia. Náhodne si zvolíme obálku, označíme ju $O1$, pozrieme sa do nej a zistíme, že jej obsah X sa rovná istej sume, ktorú označíme s. Budeme uvažovať možnosť, že vymeniť ju za $O2$ je pre nás výhodné len vtedy, ak stredná hodnota peňažného obnosu v obálke $O2$ je väčšia ako s, teda ak platí

$$E(Y|X=s) = \frac{s}{2}P(X=V|X=s) + 2sP(X=M|X=s) > s. \tag{2.7}$$

Keďže s je kladné, predelíme ním nerovnosť (2.7):

$$\frac{1}{2}P(X=V|X=s) + 2P(X=M|X=s) > 1. \tag{2.8}$$

Ak obsah $O1$ nie je V, teda väčší, musí byť M, preto platí

$$P(X=M|X=s) = 1 - P(X=V|X=s). \tag{2.9}$$

Dosadením (2.9) do (2.8) dostávame podmienku pre výhodnosť v tvare

$$\frac{1}{2}P(X=V|X=s) + 2[1 - P(X=V|X=s)] > 1,$$

$$\frac{3}{2}P(X=V|X=s) < 1.$$

Výmena $O1$ za $O2$ je preto výhodná len ak

$$P(X=V|X=s) < \frac{2}{3}, \tag{2.10}$$

teda ak podmienená pravdepodobnosť, že sme si zvolili obálku s väčším množstvom peňazí je po zistení sumy v $O1$ menšia ako $2/3$, pričom podmienkou je, že sme zistili sumu v $O1$.

Nerovnosť (2.10) nám udáva všeobecnú podmienku výhodnosti výmeny obálky. My ju ešte upravíme pre diskrétne rozdelenia. Predpokladajme, že existuje fixné m kladné také, že suma v akejkoľvek obálke sa musí rovnať $2^k m$€ pre nejaké $k \in \mathbb{Z}$. Člen

2^k odzrkadľuje fakt, že najbližšie menšie ako aj najbližšie väčšie sumy sú polovicou, respektíve dvojnásobkom obnosu v pôvodne zvolenej obálke. Definujeme

$$P(X = 2^k m | X = V) = p_k \qquad pre\ k = \cdots, -1, 0, 1, \ldots, \tag{2.11}$$

čo je pre prehľadnosť názornejšie uvedené v tabuľke Tab. 2.1. Požadujeme, aby $\{\ldots, p_{-1,}, p_0, p_1, \ldots\}$ definovali pravdepodobnostné rozdelenie, takže predpokladáme, že $p_k \geq 0$ a $\sum_{k=-\infty}^{\infty} p_k = 1$.

...	$2^{-k}m$	...	$\frac{1}{8}m$	$\frac{1}{4}m$	$\frac{1}{2}m$	m	$2m$	$4m$	$8m$	...	$2^k m$	...
...	p_{-k}	...	p_{-3}	p_{-2}	p_{-1}	p_0	p_1	p_2	p_3	...	p_k	...

Tab. 2.1 Definovanie pravdepodobnosti $P(X = 2^k m | X = V)$ pri rôznych hodnotách k

Pre lepšie pochopenie uvádzame údaje z Tab. 2.1 ešte raz, tento krát konkrétne pre $m = 1€$ v Tab. 2.2.

...	$2^{-k}€$	...	$12{,}5c$	$25c$	$50c$	1€	2€	4€	8€	...	$2^k€$	...
...	p_{-k}	...	p_{-3}	p_{-2}	p_{-1}	p_0	p_1	p_2	p_3	...	p_k	...

Tab. 2.2 Príklad súm v obálkach pre rôzne pravdepodobnosti

Všimnime si, že

$$P(X = 2^k m | X = M) = P(X = 2^{k+1} m | X = V) = p_{k+1}. \tag{2.12}$$

Predpokladajme, že sme zistili, že hodnota vo zvolenej obálke X je rovné $2^k m$ pre nejaké k, potom aplikáciou Bayesovej vety na $P(X = V | X = 2^k m)$ dostávame

$$P(X = V | X = 2^k m) = \frac{P(X = 2^k m | X = V) P(X = V)}{P(X = 2^k m)}. \tag{2.13}$$

(2.9) upravíme pre $s = 2^k m$ nasledovne

$$P(X = M | X = 2^k m) + P(X = V | X = 2^k m) = 1,$$

potom pravdepodobnosť, že $X = 2^k m$ vieme napísať ako

$$P(X = 2^k m) = P(X = 2^k m)\Big(P(X = M | X = 2^k m) + P(X = V | X = 2^k m)\Big),$$

čo po roznásobení a použití Bayesovej vety nadobudne tvar

$$P(X = 2^k m) = P(X = 2^k m | X = M) P(X = M) + P(X = 2^k m | X = V) P(X = V).$$

Potom (2.13) vieme prepísať ako

$$P(X = V|X = 2^k m)$$

$$= \frac{P(X = 2^k m|X = V)P(X = V)}{P(X = 2^k m|X = M)P(X = M) + P(X = 2^k m|X = V)P(X = V)}. \qquad (2.14)$$

a keďže pravdepodobnosť, že vo zvolenej obálke je menší obnos je rovnaká, ako pravdepodobnosť, že sa v nej nachádza obnos väčší, t.j.

$$P(X = M) = P(X = V),$$

môžeme upraviť (2.14) na tvar

$$P(X = V|X = 2^k m) = \frac{P(X = 2^k m|X = V)}{P(V = 2^k m|X = V) + P(M = 2^k m|X = M)}. \qquad (2.15)$$

Po dosadení (2.11) a (2.12) do (2.15) dostávame

$$P(X = V|X = 2^k m) = \frac{p_k}{p_k + p_{k+1}}. \qquad (2.16)$$

Všeobecná podmienka výhodnosti výmeny (2.10) využitím (2.16) nadobudne pre takéto diskrétne rozdelenia tvar

$$\frac{p_k}{p_k + p_{k+1}} < \frac{2}{3},$$

t.j. vzhľadom na kladnosť súčtu $p_k + p_{k+1}$

$$p_k < 2p_{k+1}, \qquad (2.17)$$

teda pri akejkoľvek sume v nami zvolenej obálke $O1$ by sme ju mali vymeniť len vtedy, ak pravdepodobnosť, že sme si zvolili obálku s väčším obnosom V je menej ako dvojnásobok pravdepodobnosti, že sme si zvolili tú s obnosom menším, teda s M. Inak povedané, ak pravdepodobnosť, že sme si zvolili obálku s väčšou čiastkou v porovnaní s pravdepodobnosťou zvolenia tej s menšou je dostatočne veľká (väčšia nanajvýš rovná dvojnásobku pravdepodobnosti zvolenia obálky s menšou čiastkou) ponecháme si zvolenú obálku $O1$.

2.2.2 Príklady splnenia a nesplnenia podmienky o výhodnosti výmeny obálky

Pozrieme sa ešte na dva príklady, v ktorých je podmienka (2.17) obvykle splnená, resp. nesplnená.

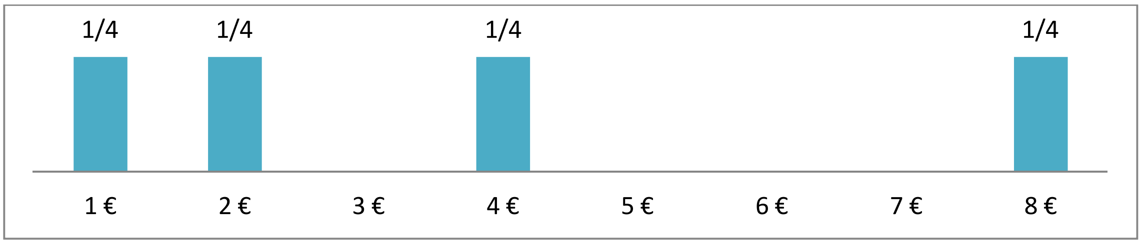
Jedným z príkladov, kedy je (2.17) obvykle splnená je nasledujúci. Nech $m = 1$,

$$p_k = \begin{cases} \frac{1}{4} & pre \quad k \in \{0,1,2,3\}, \\ 0 & pre \quad k \in \mathbb{Z} \setminus \{0,1,2,3\}, \end{cases}$$

ako je to uvedené v Tab. 2.3. V s rovnakou pravdepodobnosťou obsahuje 1€, 2€, 4€, alebo 8€. Podmienka (2.17) je splnená pre s rovné 50c, 1€, 2€ a 4€ ale nie pre s rovné 8€. Zrejme je teda splnená s pravdepodobnosťou $7/8$, pretože aby nebola splnená muselo by súčasne platiť $X = L$ aj $L = 8$, čo má pravdepodobnosť rovnú $1/2 \cdot 1/4 = 1/8$.

...	2^{-k}€	...	$\frac{1}{4}$€	$\frac{1}{2}$€	1€	2€	4€	8€	16€	...	2^k€	...
...	0	...	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	...	0	...

Tab. 2.3 Pravdepodobnosti $P(X = 2^k€|X = V)$ z prvého príkladu pri rôznych hodnotách k



Obr. 2.1 Zobrazenie vybraných pravdepodobností z Tab. 2.3

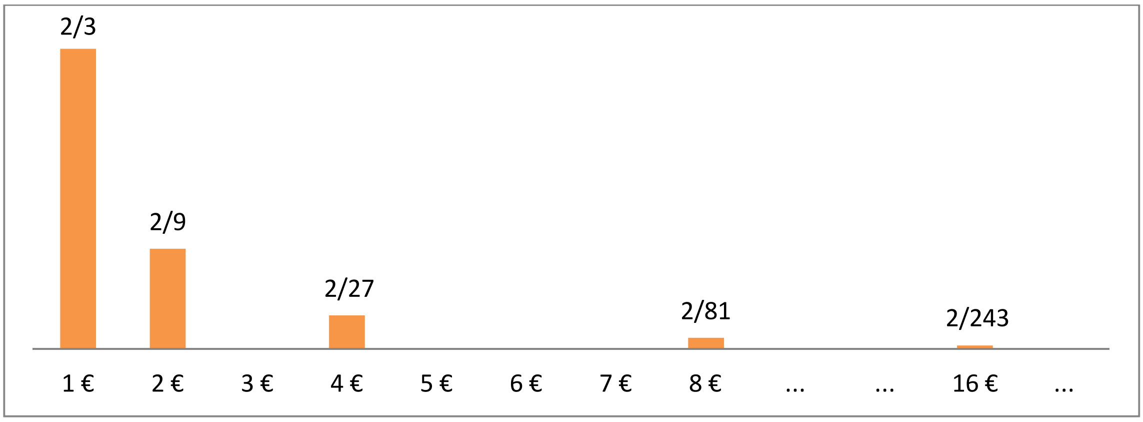
Pozrime sa na príklad, kedy je podmienka (2.17) obvykle nesplnená. Ako v predchádzajúcom príklade, nech $m = 1$,

$$p_k = \begin{cases} \frac{2}{3^k} & pre \quad k \in \mathbb{N}, \\ 0 & pre \quad k \in \mathbb{Z} \setminus \mathbb{N}. \end{cases}$$

Ako si môžeme všimnúť v Tab. 2.4, obálka s väčšou sumou obsahuje 2€ s pravdepodobnosťou $2/3$, 4€ s pravdepodobnosťou $2/9$, 8€ s pravdepodobnosťou $2/27$, atď. Podmienka (2.17) je splnená, iba ak $s = 1$, čo platí len ak $X = M$ a $M = 1$ a má pravdepodobnosť $1/2 \cdot 2/3 = 1/3$.

...	2^{-k}€	...	$\frac{1}{8}$€	$\frac{1}{4}$€	$\frac{1}{2}$€	1€	2€	4€	8€	...	2^k€	...
...	0	...	0	0	0	$\frac{2}{3}$	$\frac{2}{9}$	$\frac{2}{27}$	$\frac{2}{81}$	...	$\frac{2}{3^k}$	...

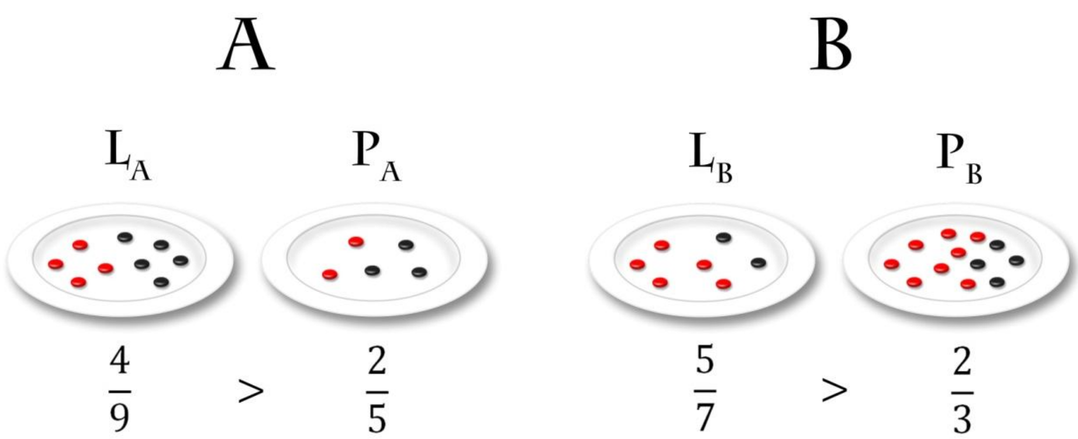
Tab. 2.4 Pravdepodobnosti $P(X = 2^k€|X = V)$ z druhého príkladu pri rôznych hodnotách k



Obr. 2.2 Zobrazenie vybraných pravdepodobností z Tab. 2.4

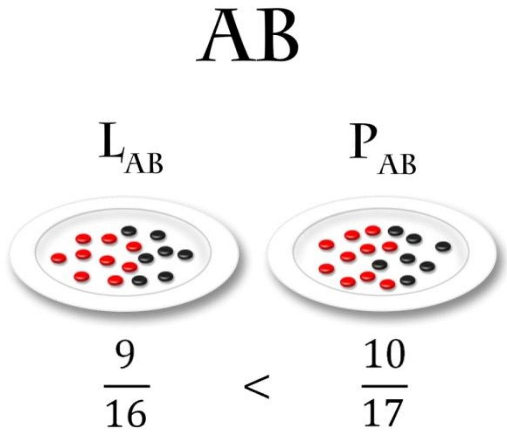
3 Simpsonov paradox

Vysvetlenie Simpsonovho paradoxu, ktoré nasleduje sme voľne prevzali z [15]. Majme dva stoly A a B. Na každý stôl umiestnime dva taniere, pravý a ľavý, v ktorých sa nachádzajú červené a čierne lentilky. Na stole A sa na jeho ľavom tanieri nachádzajú 4 červené a 5 čiernych lentiliek. Nazveme ho L_A. Pravý tanier na stole A, ktorý označíme P_A obsahuje 2 červené a 3 čierne lentilky. Podobne pomenujeme aj taniere na stole B, ľavý L_B ktorý obsahuje 5 červených a 2 čierne lentilky a pravý P_B v ktorom je 8 červených a 4 čierne lentilky. Chceme zistiť s akou pravdepodobnosťou vyberieme so zavretými očami červenú lentilku z toho ktorého taniera. Skúmajme teda na každom stole zvlášť jednotlivé pravdepodobnosti, a porovnajme, z ktorého taniera sa nám podarí červenú lentilku vybrať s väčšou pravdepodobnosťou. Pre lepšiu predstavu si celú situáciu, aj s danými pravdepodobnosťami znázorníme na obrázku Obr. 3.1.



Obr. 3.1 Stoly A a B s taniermi a lentilkami

Ako vidíme, na oboch stoloch je z hľadiska pravdepodobnosti vybratia červenej lentilky „lepším" ľavý tanier. Prisunieme teraz tretí stôl AB, ktorý bude tiež obsahovať dva taniere, do ktorých zosypeme lentilky zo súborov A a B nasledovne: ľavé taniere zo stolov A a B do ľavého taniera AB a pravé do pravého. Z L_A a L_B teda vznikne L_{AB} ako označíme ľavý tanier stola AB, ktorý obsahuje 9 červených a 7 čiernych lentilky a z P_A a P_B vznikne P_{AB} pravý tanier AB obsahujúci 10 červených a 7 čiernych lentiliek. Pozrime sa na obrázok Obr. 3.2 ako to s pravdepodobnosťami vybratia červenej lentilky vyzerá teraz.



Obr. 3.2 Stôl AB s taniermi a lentilkami

Všimnime si, že v prípade stolov A a B vyšiel vždy ako „lepší“ ľavý tanier, avšak po zosypaní lentiliek, t.j. na stole AB, je paradoxne „lepším“ pravý tanier.

Na prvý pohľad teda vnímame tento príklad ako paradox, avšak podľa [1] na to má významný vplyv práve jeho formulácia. Ak by sme k problému pristupovali matematicky, dajú sa jednotlivé pravdepodobnosti zapísať ako

$$\frac{a_1}{b_1} < \frac{c_1}{d_1}, \tag{3.1}$$

$$\frac{a_2}{b_2} < \frac{c_2}{d_2}, \tag{3.2}$$

z čoho vidno, že nevyplýva

$$\frac{a_1 + a_2}{b_1 + b_2} < \frac{c_1 + c_2}{d_1 + d_2}. \tag{3.3}$$

3.1 Príklady výskytu Simpsonovho paradoxu

Situácie pri ktorých dochádza k vzniku Simpsonovho paradoxu vznikajú aj v praxi. V nasledujúcom texte uvedieme tri z mnohých reálnych príkladov výskytu.

3.1.1 Simpsonov paradox pri prijímacom konaní na univerzitu Berkeley

Jedným z najznámejších takýchto príkladov je jesenné prijímacie konanie na univerzitu Berkeley v Kalifornii, ktoré sa uskutočnilo v roku 1973. Zo všetkých uchádzačov bolo prijatých približne 44% mužov a iba 35% žien, z čoho prirodzene vzišla otázka, či rozhodnutie o prijatí alebo neprijatí bolo ovplyvnené pohlavím uchádzača. Presné počty prijatých ako aj neprijatých uchádzačov sú uvedené v nasledujúcej tabuľke Tab. 3.1.

	Počet prijatých	Počet neprijatých	Percento prijatých
Ženy	1494	2827	34,6%
Muži	3738	4704	44,3%

Tab. 3.1 Počty a percentá prijatých a neprijatých uchádzačov na Berkeley [2]

Výskum [2], ktorý viedli Bickel, Hammel a O'Connell však dokázal, že po rozdelení údajov o uchádzačoch podľa katedier na ktoré sa hlásili existovala malá, avšak štatisticky významná odchýlka v prospech žien. Ak by sme sa vrátili k úvodnému príkladu s lentilkami, celkové počty uchádzačov uvedené v Tab. 3.1 môžeme chápať ako lentilky zosypané z tanierov, ktorými boli jednotlivé katedry. Vidíme teda, že správne spätné rozdelenie uchádzačov je dôležité pre vyvodzovanie záverov o pravdepodobnosti prijatia a celkové počty sú v tomto prípade nedostačujúce. Pre lepšiu predstavu rozdelenia, ktorým sa zmení pomer prijatých mužov a žien na jednotlivé katedry oproti pomeru celkového počtu prijatých udáva tabuľka Tab. 3.2.

	Počet prijatých	Počet neprijatých	Percento prijatých
Katedra matematiky			
Ženy	100	100	50%
Muži	200	200	50%
Katedra sociálnej starostlivosti			
Ženy	150	300	33,3%
Muži	50	100	33,3%
Spolu			
Ženy	250	400	38,5%
Muži	250	300	45,5%

Tab. 3.2 Hypotetické rozdelenie uchádzačov do katedier prezentujúce dôležitosť správneho rozdelenia údajov [2]

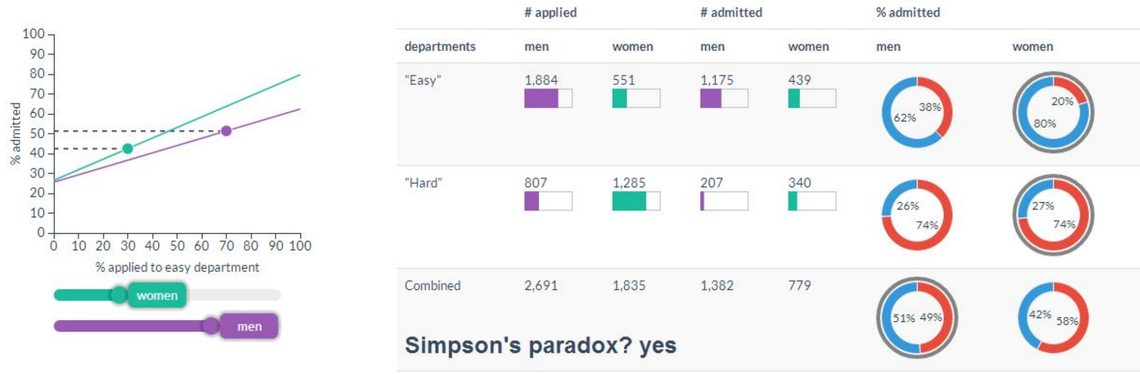
Ďalej výskum uvádza, že táto odchýlka nebola spôsobená diskrimináciou zo strany univerzity pri výberovom konaní, ale tendenciou žien hlásiť sa na niektoré študijné programy viac ako na iné.

Obr. 3.3 Počet prijatých a neprijatých uchádzačov vzhľadom na jednotlivé katedry [8]

Lepšiu ilustráciu vplyvu rozdelenia uchádzačov do katedier poskytuje [8] odkiaľ uvádzame obrázky Obr. 3.3 a Obr. 3.4. Guľôčky na nich predstavujú každá desať uchádzačov, modré prijatých a červené neprijatých. V prvom riadku sú údaje o ženách, v druhom o mužoch. Druhý až siedmy stĺpec označené A-F predstavujú jednotlivé katedry, stĺpec combined obsahuje súčty prijatých a neprijatých uchádzačov.

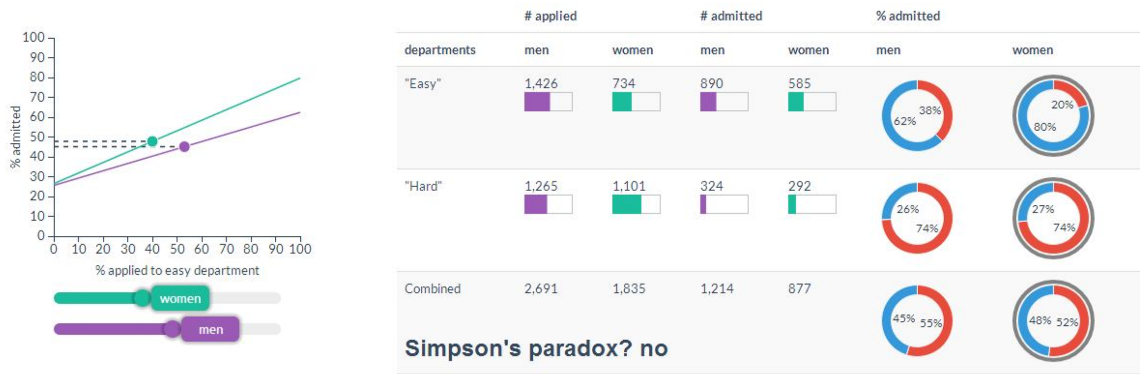
Obr. 3.4 Počet prijatých a neprijatých uchádzačov spolu [8]

Do pozornosti dávame aj interaktívnu ilustráciu Simpsonovho paradoxu, ktorú čitateľ môže nájsť a vyskúšať si na spomínanej internetovej stránke [8].



Obr. 3.5 Ilustrácia vzniku Simpsonovho paradoxu vzhľadom na počet mužov a žien uchádzajúcich sa o štúdium na ľahších odboroch – paradox vzniká [8]

Zobrazuje či dochádza, ako vidno na Obr. 3.5, alebo nedochádza, čo si môžeme všimnúť na Obr. 3.6, k Simsponovmu paradoxu vzhľadom na počet mužov a žien hlásiacich sa na ľahšie odbory. Percento prijatých mužov a žien samostatne na ľahšie a ťažšie odbory zostáva rovnaké, avšak mení sa percento prijatých mužov a žien na univerzitu ako celok.



Obr. 3.6 Ilustrácia vzniku Simpsonovho paradoxu vzhľadom na počet mužov a žien uchádzajúcich sa o štúdium na ľahších odboroch – paradox nevzniká [8]

3.1.2 Simpsonov paradox vo výskume rastu detí v Afrike

Druhý, novší prípad výskytu tohto paradoxu sa nachádza vo výskume rastu detí v Južnej Afrike [11] uskutočnenom v rokoch 1990-95. Zo 4092 detí, ktoré boli pri narodení zapojené do výskumu sa po piatich rokoch dostavilo na interview, ktoré bolo súčasťou výskumu len 964. Výskumníci teda čelili problému, či vzorka detí, ktoré prišli na interview môže poslúžiť na vyvodenie záverov o počiatočnej vzorke, a teda či sú jej charakteristiky podobné skupine detí, ktoré sa výskumníkom nepodarilo vystopovať. Z tohto dôvodu skúmali podobnosť týchto dvoch skupín vzhľadom na viacero faktorov, jedným z ktorých bolo aj či mala matka lekársku starostlivosť v čase narodenia dieťaťa. Ako si môžeme všimnúť v tabuľke Tab. 3.3, lekárska starostlivosť bola poskytnutá matkám nevystopovaných detí v 16,6% prípadov, pričom u tých, ktoré sa dostavili len v 11,1%.

	Deti, ktoré sa nedostavili		Deti, ktoré sa dostavili	
	Počet	Percento	Počet	Percento
Matka mala lek. pomoc	195	16,6%	46	11,1%
Matka nemala lek. pomoc	979	83,4%	370	88,9%

Tab. 3.3 Porovnanie detí vzhľadom na lekársku pomoc matkám pri ich narodení [11]

Ak ale rozdelíme deti do skupín podľa rasovej príslušnosti na belochov a černochov a opäť porovnáme skupinu nevystopovaných detí so skupinou tých, ktoré sa dostavili, zistíme, že v oboch týchto skupinách je percento matiek, ktorým bola poskytnutá lekárska pomoc vyššie u detí, ktoré sa dostavili, čo vidno v tabuľke Tab. 3.4.

	Belosi				Černosi			
	Deti, ktoré sa nedostavili		Deti, ktoré sa dostavili		Deti, ktoré sa nedostavili		Deti, ktoré sa dostavili	
	Počet	Percento	Počet	Percento	Počet	Percento	Počet	Percento
Matka mala lek. pomoc	104	82,5%	10	83,3%	91	8,7%	36	8,9%
Matka nemala lek. pomoc	22	17,5%	2	16,7%	957	91,3%	368	91,1%

Tab. 3.4 Porovnanie detí, rozdelených podľa rasovej príslušnosti, vzhľadom na lekársku pomoc matkám pri ich narodení [11]

3.1.3 Simpsonov paradox v bejzbale

Výskyt Simpsonovho paradoxu nie je niečím neobvyklým ani v športoch. [18] uvádza príklad jeho výskytu v bejzbale v rokoch 1995-96, kde porovnáva úspešnosť dvoch vtedajších hráčov Dereka Jetera a Davida Justicea. U oboch hráčov sleduje počet úspešných zásahov, ktoré označíme H (hits) a autov O (outs), ako to udáva Tab. 3.5. Presné pravidlá bejzbalu sú k dispozícii aj na internetovej stránke [13].

1995			1996		
	H	O		H	O
Justice	104	307	Justice	45	95
Jeter	12	36	Jeter	183	399

Tab. 3.5 Údaje o H a O Justicea a Jetera v rokoch 1995 a 1996 [18]

Z údajov o H a O pre oboch hráčov vypočítame úspešnosti u postupne pre 1995, 1996 a 1995-1996. V roku 1995 boli podľa Tab. 3.5 nasledovné

$$Justice: \quad u = \frac{104}{104+307} = 0{,}253,$$
$$Jeter: \quad u = \frac{12}{12+36} = 0{,}250.$$

V roku 1996 sa zopakovala situácia z predchádzajúceho roku a opäť bol Justice úspešnejší. Vychádzajúc z údajov v Tab. 3.5

$$Justice: \quad u = \frac{45}{45+95} = 0{,}321,$$
$$Jeter: \quad u = \frac{183}{183+399} = 0{,}314.$$

1995-1996		
	H	O
Justice	149	402
Jeter	195	435

Tab. 3.6 Údaje o H a O Justicea a Jetera za obdobie 1995-96 [18]

Teraz sa však pozrieme na celkové údaje v rokoch 1995-96, ako ich zachytáva Tab. 3.6. Úspešnosť pre toto obdobie je

$$Justice: \quad u = \frac{149}{149+402} = 0{,}270,$$
$$Jeter: \quad u = \frac{195}{195+435} = 0{,}310.$$

Ako vidíme, dochádza k Simpsonovmu paradoxu, nakoľko za obdobie 1995-1996 mal vyššiu úspešnosť Jeter, napriek tomu, že v ročných porovnaniach vyhrával Justice.

Podľa [18] sa v bejzbale u niektorých zaujímavých dvojíc hráčov[5] vyskytuje Simpsonov paradox približne raz do roka. Vynára sa teda otázka, ako pravdepodobný je jeho výskyt? Na túto otázku odpovedá nasledujúca podkapitola.

[5] Autor bližšie nešpecifikuje, ktorých hráčov zaradil medzi zaujímavých.

3.2 Pravdepodobnosť výskytu Simpsonovho paradoxu

Výskyt Simpsonovho paradoxu je podľa [11] zriedkavý. Na otázku ako veľmi zriedkavý sa teraz pokúsime odpovedať, pričom budeme vychádzať z [14]. Na rozdiel od autorov článku [14] nebudeme riešiť všeobecný problém výskytu v tabuľke 2x2xl, kde $l \in \{2,3,\dots\}$, ale zameriame sa na konkrétny prípad $l = 2$. V [14] je tento prípad uvedený len okrajovo na zhruba 2,5 strany aj s predpokladmi, z ktorých vychádza, ktoré sú formulované len všeobecne. V tejto podkapitole ich uvedieme konkrétne pre $l = 2$ a jednotlivé kroky vysvetlíme podrobnejšie.

Majme tri náhodné korelované udalosti A, B a C. Simpsonov paradox potom nastáva ak buď

$$P(A \cap B|C) \geq P(A|C)P(B|C), \tag{3.4}$$

$$P(A \cap B|C^C) \geq P(A|C^C)P(B|C^C), \tag{3.5}$$

$$P(A \cap B) \leq P(A)P(B), \tag{3.6}$$

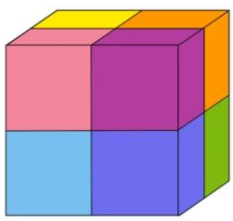
pričom aspoň jedna z nerovností musí byť neostrá, alebo musia byť nahradené opačnými nasledovne

$$P(A \cap B|C) < P(A|C)P(B|C), \tag{3.7}$$

$$P(A \cap B|C^C) < P(A|C^C)P(B|C^C), \tag{3.8}$$

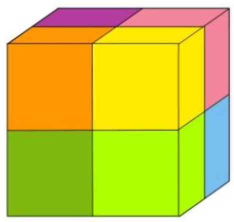
$$P(A \cap B) > P(A)P(B). \tag{3.9}$$

Nerovnosti (3.4)-(3.6) nazveme *pozitívny Simpsonov zvrat* a nerovnosti k nim opačné (3.7)-(3.9) *negatívny Simpsonov zvrat.*

	C			
	$A \rightarrow$	p_1	p_3	
	$A^C \rightarrow$	p_2	p_4	
		$\uparrow$ B	$\uparrow$ B^C	

Tab. 3.7 Tabuľka pravdepodobností 2x2x2 spredu

Predstavme si tabuľku 2x2x2 v ktorej sa v každej kolónke nachádza jedna z pravdepodobností $p_1, \dots, p_8$ korešpondujúcich s dichotómiami A, alebo A^C, B, alebo B^C, C, alebo C^C. Keďže sa v tomto prípadne jedná o 3D tabuľku, ilustratívne je zobrazená v Tab. 3.7 pri pohľade spredu a v Tab. 3.8 pri pohľade zozadu.

<table>
<tr><td rowspan="4"></td><td colspan="4">C^C</td></tr>
<tr><td>$A \rightarrow$</td><td>p_7</td><td>p_5</td><td></td></tr>
<tr><td>$A^C \rightarrow$</td><td>p_8</td><td>p_6</td><td></td></tr>
<tr><td></td><td>$\uparrow$
B^C</td><td>$\uparrow$
B</td><td></td></tr>
</table>

Tab. 3.8 Tabuľka pravdepodobností 2x2x2 zozadu

Aby sme mohli vypočítať pravdepodobnosti udalostí závisiacich od A, B, C, potrebujeme poznať pravdepodobnostné rozdelenie $p_1, \dots, p_8$. Predpokladajme, že pravdepodobnosti $\boldsymbol{p} \equiv \{p_1, \dots, p_8\}$ sú náhodné a rovnomerne rozdelené na pravdepodobnostnom priestore

$$S_8 = \{\boldsymbol{p} | p_1 \geq 0, \dots, p_8 \geq 0,\ p_1 + \cdots + p_8 = 1\} \tag{3.10}$$

čo je 7-dimenzionálna množina v $\mathbb{R}^8$. Súčet pravdepodobností $\sum_{k=1}^{8} p_k = 1$. Teraz vyjadríme *pozitívny Simpsonov zvrat* (3.4)-(3.6) pomocou pravdepodobností p_k. Pre lepšie pochopenie úprav rovníc odporúčame čitateľovi sledovať Tab. 3.7 a Tab. 3.8. Najskôr vyjadríme poslednú nerovnosť *pozitívneho Simpsonovho zvratu* (3.6). Prepísaním pravej strany nerovnosti pomocou pravdepodobností p_k dostávame

$$P(A \cap B) = p_1 + p_5. \tag{3.11}$$

Ľavú stranu prepíšeme nasledovne

$$P(A) = p_1 + p_3 + p_5 + p_7, \tag{3.12}$$

$$P(B) = p_1 + p_2 + p_5 + p_6. \tag{3.13}$$

(3.11), (3.12) a (3.13) dosadíme do (3.6)

$$p_1 + p_5 \leq (p_1 + p_3 + p_5 + p_7)(p_1 + p_2 + p_5 + p_6),$$

po roznásobení čoho získame

$$p_1 + p_5 - {p_1}^2 - p_1p_2 - p_1p_3 - 2p_1p_5 - p_1p_6 - p_1p_7 - p_2p_5 - p_3p_5 - {p_5}^2 - p_5p_6$$
$$-p_5p_7 \leq p_2p_3 + p_2p_7 + p_3p_6 + p_6p_7,$$

čo ďalej upravíme na

$$p_1(1 - p_1 - p_2 - p_3 - p_5 - p_6 - p_7) + p_5(1 - p_1 - p_2 - p_3 - p_5 - p_6 - p_7)$$
$$\leq p_3(p_2 + p_6) + p_7(p_2 + p_6). \tag{3.14}$$

Využijeme, že $\sum_{k=1}^{8} p_k = 1$, a teda (3.14) vieme prepísať nasledovne

$$p_1(p_4 + p_8) + p_5(p_4 + p_8) \leq p_3(p_2 + p_6) + p_7(p_2 + p_6).$$

Konečne

$$(p_1 + p_5)(p_4 + p_8) \leq (p_2 + p_6)(p_3 + p_7).$$

Využitím základných poznatkov o združenej pravdepodobnosti prepíšeme pravú stranu (3.4) nasledovne

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{p_1}{p_1 + p_2 + p_3 + p_4}.$$

Rovnako upravíme aj ľavú stranu a to

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{p_1 + p_3}{p_1 + p_2 + p_3 + p_4},$$

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4}.$$

Dosadením do nerovnosti (3.4) dostávame

$$\frac{p_1}{p_1 + p_2 + p_3 + p_4} \geq \frac{(p_1 + p_3)(p_1 + p_2)}{(p_1 + p_2 + p_3 + p_4)^2},$$

čo upravíme

$$p_1(p_1 + p_2 + p_3 + p_4) \geq (p_1 + p_3)(p_1 + p_2),$$

$${p_1}^2 + p_1p_2 + p_1p_3 + p_1p_4 \geq {p_1}^2 + p_1p_2 + p_1p_3 + p_2p_3,$$

a teda

$$p_1p_4 \geq p_2p_3.$$

Nerovnosť (3.5) sa dá ekvivalentnými úpravami prepísať na

$$p_5p_8 \geq p_6p_7,$$

a teda *pozitívny Simpsonov zvrat* môžeme ekvivalentne vyjadriť nasledujúcimi nerovnosťami

$$p_1 p_4 \geq p_2 p_3, \tag{3.15}$$

$$p_5 p_8 \geq p_6 p_7, \tag{3.16}$$

$$(p_1 + p_5)(p_4 + p_8) \leq (p_2 + p_6)(p_3 + p_7), \tag{3.17}$$

pričom aspoň jedna z nich musí byť neostrá.

V nasledujúcom texte ukážeme, aká je pravdepodobnosť, že $\boldsymbol{p}$ splní nerovnosti (3.15), (3.16) a (3.17), ktoré označíme R_1^+, R_2^+ a R_3^+. Nerovnosti k nim opačné označíme R_1^-, R_2^- a R_3^-, a teda

$$\{pozitívny\ Simpsonov\ zvrat\} = \bigcap_{k=1}^{3} R_k^+,$$

$$\{negatívny\ Simpsonov\ zvrat\} = \bigcap_{k=1}^{3} R_k^-,$$

$$\{Simpsonov\ paradox\} = \left(\bigcap_{k=1}^{3} R_k^+\right) \cup \left(\bigcap_{k=1}^{3} R_k^-\right).$$

Keďže $\bigcap_{k=1}^{3} R_k^+$ a $\bigcap_{k=1}^{3} R_k^-$ sú disjunktné udalosti na S_8,

$$\begin{aligned} \pi &:= P(Simpsonov\ paradox\ pri\ D_8(1)) \\ &= P\left(\bigcap_{k=1}^{3} R_k^+\right) + P\left(\bigcap_{k=1}^{3} R_k^-\right) \qquad (3.18) \\ &=: \pi^+ + \pi^-. \end{aligned}$$

Na výpočet týchto pravdepodobností použijeme Dirichletovo rozdelenie. Vo všeobecnosti má podľa [7] n-rozmerné Dirichletovo rozdelenie s parametrami $\alpha = (\alpha_1, \dots, \alpha_n)$ hustotu

$$f(p,\alpha) = \frac{\Gamma(\sum_{i=1}^{n} \alpha_i)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} \prod_{i=1}^{n-1} p_i^{\alpha_i - 1} \left(1 - \sum_{i=1}^{n-1} p_i\right)^{\alpha_n - 1}, p_i \geq 0, i = 1, \dots, n-1, \sum_{i=1}^{n-1} p_i \leq 1.$$

Ak označíme $p_n = 1 - \sum_{i=1}^{n-1} p_i$, tak dostávame pravdepodobnostné rozdelenie na množine $\{(p_1, \dots, p_n) | p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}$, t.j. na pravdepodobnostnom simplexe S_n. Označíme $D_n(1)$ špeciálny prípad Dirichletovho rozdelenia, v ktorom $\alpha = (1, \dots, 1)$. Potom hustota $D_n(1)$ je

$$f(p,\alpha) = \frac{\Gamma(n)}{\prod_{i=1}^{n} \Gamma(1)} \prod_{i=1}^{n-1} p_i^0 \left(1 - \sum_{i=1}^{n-1} p_i\right)^0 = \Gamma(n) = (n-1)!.$$

Hustota je konštantná, čo korešponduje s rovnomerným rozdelením na S_n.

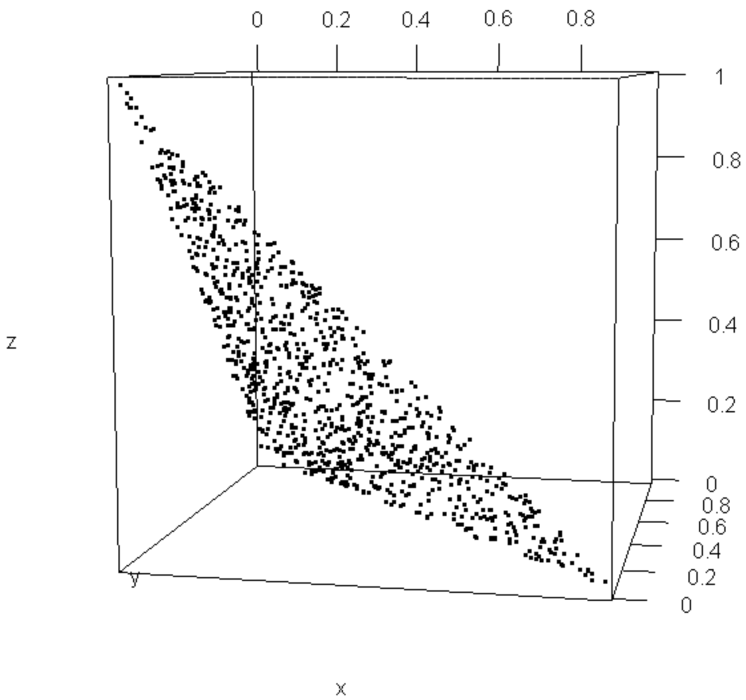
Podľa algoritmu 4.67 v [7] vieme generovať hodnoty Dirichletovho rozdelenia pomocou nezávislých náhodných premenných s gama rozdelením. Prepíšeme tento algoritmus pre náš špeciálny prípad, kedy všetky α_i sú rovné 1. Nech $V_1, \dots, V_n$ sú nezávislé náhodné premenné s $V_i \sim Gama(1,1)$ pre $i = 1, \dots, n$ (všimnime si, že $Gama(1,1)$ je vlastne exponenciálne rozdelenie s parametrom 1). Potom:

$$\boldsymbol{p} := \left(\frac{V_1}{\sum_{i=1}^{n} V_i}, \dots, \frac{V_n}{\sum_{i=1}^{n} V_i} \right) \sim D_n(1). \tag{3.19}$$

Uvedieme názornú ukážku pre $n+1=3$, kedy vieme získané hodnoty p graficky znázorniť. Simuláciu na S_3, zobrazuje Obr. 3.8, ktorý sme generovali v programe RStudio nasledovne:

```
re1=rexp(1000, rate=1);   # náhodné číslo z Exp(1) rozdelenia
re2=rexp(1000, rate=1);
re3=rexp(1000, rate=1);
x=re1/(re1+re2+re3);   # p1
y=re2/(re1+re2+re3);   # p2
z=re3/(re1+re2+re3);   # p3
library(rgl)
plot3d(x,y,z)
```

Obr. 3.7 Zdrojový kód k vygenerovaniu Obr. 3.8 v programe RStudio



Obr. 3.8 Simulácia rovnomerného rozdelenia na S_3 pomocou nezávislých náhodných premenných s exponenciálnym rozdelením v programe RStudio

V našej analýze Simpsonovho paradoxu uvažujeme rovnomerné rozdelenie na S_8. Ďalej teda označíme $V_1, \dots, V_8$ nezávislé náhodné premenné s exponenciálnym rozdelením so strednou hodnotou 1. Potom využitím (3.19) teda R_1^+, R_2^+ a R_3^+ môžeme vyjadriť pomocou $V_1, \dots, V_8$ nasledovne

$$V_1 V_4 \geq V_2 V_3 \tag{3.20}$$

$$V_5 V_8 \geq V_6 V_7 \tag{3.21}$$

$$(V_1 + V_5)(V_4 + V_8) \leq (V_2 + V_6)(V_3 + V_7) \tag{3.22}$$

kde (3.20), (3.21) a (3.22) zodpovedajú R_1^+, R_2^+ a R_3^+ v uvedenom poradí.

Rovnako môžeme vyjadriť aj R_1^-, R_2^- a R_3^- v danom poradí takto:

$$V_1 V_4 < V_2 V_3 \tag{3.23}$$

$$V_5 V_8 < V_6 V_7 \tag{3.24}$$

$$(V_1 + V_5)(V_4 + V_8) > (V_2 + V_6)(V_3 + V_7) \tag{3.25}$$

Označme

$$\tilde{V}_1 = V_2, \quad \tilde{V}_2 = V_1, \quad \tilde{V}_3 = V_4, \quad \tilde{V}_4 = V_3, \quad \tilde{V}_5 = V_6, \quad \tilde{V}_6 = V_5, \quad \tilde{V}_7 = V_8, \quad \tilde{V}_8 = V_7,$$

potom $\tilde{V}_1, \dots, \tilde{V}_8$ sú nezávislé náhodné premenné s exponenciálnym rozdelením so strednou hodnotou 1. Nerovnosti (3.23), (3.24) a (3.25), vyjadrené pomocou $\tilde{V}_i$, vyzerajú v danom poradí nasledovne:

$$\tilde{V}_2 \tilde{V}_3 < \tilde{V}_1 \tilde{V}_4 \Longrightarrow \tilde{V}_1 \tilde{V}_4 > \tilde{V}_2 \tilde{V}_3,$$

$$\tilde{V}_6 \tilde{V}_7 < \tilde{V}_5 \tilde{V}_8 \Longrightarrow \tilde{V}_5 \tilde{V}_8 > \tilde{V}_6 \tilde{V}_7,$$

$$\left(\tilde{V}_2 + \tilde{V}_6\right)\left(\tilde{V}_3 + \tilde{V}_7\right) > \left(\tilde{V}_1 + \tilde{V}_5\right)\left(\tilde{V}_4 + \tilde{V}_8\right) \Longrightarrow$$

$$\left(\tilde{V}_1 + \tilde{V}_5\right)\left(\tilde{V}_4 + \tilde{V}_8\right) < \left(\tilde{V}_2 + \tilde{V}_6\right)\left(\tilde{V}_3 + \tilde{V}_7\right),$$

Ak tieto nerovnosti porovnáme s (3.20), (3.21) a (3.22) vidíme, že sa líšia len vlnovkami a možnosťou rovnosti. Rozdelenie $(V_1, \dots, V_8)$ a $\left(\tilde{V}_1, \dots, \tilde{V}_8\right)$ je však rovnaké a keďže je spojité, pravdepodobnosť sa nezmení, ak v (3.23), (3.24) a (3.25) použijeme neostré nerovnosti. Preto môžeme (3.18) prepísať na

$$\begin{aligned} \pi &= 2P\left(\bigcap_{k=1}^{3} R_k^+\right) \\ &= 2\pi^+. \end{aligned} \tag{3.26}$$

Teraz vyjadríme hodnotu π. Pravdepodobnosť prieniku R_1^+, R_2^+ a R_3^+ pri Dirichletovom rozdelení $D_8(1)$, t.j. $P(\bigcap_{k=1}^{3} R_k^+)$, vyjadríme pomocou integrálu, pričom budeme integrovať združenú hustotu vektora $(V_1, \dots, V_8)$ a podmienky (3.20), (3.21) a (3.22) sa premietnu do integračnej množiny. Každé V_i má exponenciálne rozdelenie s parametrom 1, takže jeho hustota je

$$f_{V_i}(x) = \begin{cases} e^{-x} & x > 0, \\ 0 & inak. \end{cases}$$

Náhodné premenné $V_1, \dots, V_8$ sú nezávislé, takže združená hustota vektora $(V_1, \dots, V_8)$ je súčin hustôt V_i:

$$f_{V_1,\dots,V_8}(v_1, \dots, v_8) = e^{-v_1} \dots e^{-v_8} = e^{-(v_1+\dots+v_8)}.$$

Rovnosť (3.26) potom nadobudne tvar

$$\pi = 2 \int_{K_1} e^{-(v_1+\dots+v_8)} dv_1 \dots dv_8, \tag{3.27}$$

kde

$$K_1 = \{(v_1, \dots, v_8) \in (0, \infty)^8 | v_1 v_4 \geq v_2 v_3, v_5 v_8 \geq v_6 v_7,$$
$$ale\ (v_1 + v_5)(v_4 + v_8) \leq (v_2 + v_6)(v_3 + v_7)\}.$$

Výpočet integrálu uskutočníme pomocou dvoch transformácii.

Najprv uvažujme transformáciu $T_1: (v_1, \dots, v_8) \mapsto (\omega_1, \dots, \omega_4, \delta_1, \dots, \delta_4)$ kde $\omega_j \coloneqq v_j/(v_j + v_{j+4})$ a $\delta_j \coloneqq v_j + v_{j+4}$ pre $j \in \{1,2,3,4\}$. Konkrétne

$$\omega_1 = \frac{v_1}{v_1 + v_5}, \quad \omega_2 = \frac{v_2}{v_2 + v_6}, \quad \omega_3 = \frac{v_3}{v_3 + v_7}, \quad \omega_4 = \frac{v_4}{v_4 + v_8},$$

$$\delta_1 = v_1 + v_5, \quad \delta_2 = v_2 + v_6, \quad \delta_3 = v_3 + v_7, \quad \delta_4 = v_4 + v_8.$$

Vidíme, že definičný obor premenných δ_j zostáva $(0, \infty)$, ale pri premenných ω_j sa mení na (0,1).

Zobrazenie k T_1 inverzné, ktoré označíme T_1^{-1} je potom definované

$$v_1 = \omega_1 \delta_1, \qquad v_2 = \omega_2 \delta_2, \qquad v_3 = \omega_3 \delta_3, \qquad v_4 = \omega_4 \delta_4,$$

$$v_5 = (1 - \omega_1)\delta_1, \ v_6 = (1 - \omega_2)\delta_2, \ v_7 = (1 - \omega_3)\delta_3, \ v_8 = (1 - \omega_4)\delta_4.$$

Nerovnosti z množiny K_1 vyjadríme pomocou nových premenných. Prvú nerovnosť $v_1 v_4 \geq v_2 v_3$ môžeme upraviť nasledovne, pričom využijeme kladnosť pôvodných premenných v_i

$$\frac{v_1 v_4}{v_2 v_3} \geq 1. \tag{3.28}$$

Nerovnosť (3.28) upravíme tak, aby bolo možné nahradiť pôvodné premenné novými, čo dosiahneme nasledujúcim násobením

$$\frac{\frac{v_1}{v_1 + v_5} \cdot \frac{v_4}{v_4 + v_8}}{\frac{v_2}{v_2 + v_6} \cdot \frac{v_3}{v_3 + v_7}} \geq \frac{(v_2 + v_6)(v_3 + v_7)}{(v_1 + v_5)(v_4 + v_8)}.$$

Substitúciou nových premenných dostávame

$$\frac{\omega_1 \omega_4}{\omega_2 \omega_3} \geq \frac{\delta_2 \delta_3}{\delta_1 \delta_4}. \tag{3.29}$$

Druhú nerovnosť $v_5 v_8 \geq v_6 v_7$ môžeme upraviť obdobne ako prvú na

$$\frac{v_5 v_8}{v_6 v_7} \geq 1,$$

čo po dosadení nových premenných nadobudne tvar

$$\frac{(1-\omega_1)\delta_1(1-\omega_4)\delta_4}{(1-\omega_2)\delta_2(1-\omega_3)\delta_3} \geq 1,$$

ktorý už len upravíme, aby sa nám s nerovnosťou neskôr lepšie pracovalo, na

$$\frac{(1-\omega_1)(1-\omega_4)}{(1-\omega_2)(1-\omega_3)} \geq \frac{\delta_2 \delta_3}{\delta_1 \delta_4}. \tag{3.30}$$

V tretej nerovnosti $(v_1 + v_5)(v_4 + v_8) \leq (v_2 + v_6)(v_3 + v_7)$ postupujeme rovnako ako u prvých dvoch, pričom stále využívame kladnosť premenných v_i

$$\frac{(v_1 + v_5)(v_4 + v_8)}{(v_2 + v_6)(v_3 + v_7)} \leq 1,$$

po substitúcii

$$\frac{\delta_1 \delta_4}{\delta_2 \delta_3} \leq 1. \tag{3.31}$$

Pomocou (3.29), (3.30) a (3.31) vieme množinu K_1 napísať v zobrazení T_1 takto

$$\begin{gathered} T_1(K_1) = \{(\omega_1, .., \omega_4, \delta_1, \dots, \delta_4) \in (0,1)^4 \times (0,\infty)^4 | \\ t := \frac{\delta_1 \delta_4}{\delta_2 \delta_3} \leq 1, \frac{\omega_1 \omega_4}{\omega_2 \omega_3} \geq \frac{1}{t}, \frac{(1-\omega_1)(1-\omega_4)}{(1-\omega_2)(1-\omega_3)} \geq \frac{1}{t} \Big\}. \end{gathered} \tag{3.32}$$

Kvôli transformácii integrálu v (3.27) potrebujeme ešte vyjadriť jakobián zobrazenia T_1, ktorý označíme J_1.

$$J_1 = \begin{vmatrix} \delta_1 & 0 & 0 & 0 & \omega_1 & 0 & 0 & 0 \\ 0 & \delta_2 & 0 & 0 & 0 & \omega_2 & 0 & 0 \\ 0 & 0 & \delta_3 & 0 & 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \delta_4 & 0 & 0 & 0 & \omega_4 \\ -\delta_1 & 0 & 0 & 0 & 1-\omega_1 & 0 & 0 & 0 \\ 0 & -\delta_2 & 0 & 0 & 0 & 1-\omega_2 & 0 & 0 \\ 0 & 0 & -\delta_3 & 0 & 0 & 0 & 1-\omega_3 & 0 \\ 0 & 0 & 0 & -\delta_4 & 0 & 0 & 0 & 1-\omega_4 \end{vmatrix}. \tag{3.33}$$

Keďže Jacobiho matica prvých derivácii je po blokoch diagonálna, využijeme pri jej počítaní poznatky o determinante blokových matíc[10]

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| \cdot \left|D - CA^{-1}B\right|.$$

Jacobiho maticu J_1 vieme napísať ako

$$J_1 = \begin{vmatrix} A & B \\ C & D \end{vmatrix},$$

kde každý z blokov má rozmer 4x4.

K výpočtu J_1 vyjadríme matice A^{-1} a $CA^{-1}B$ nasledovne

$$A^{-1} = \begin{pmatrix} \frac{1}{\delta_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\delta_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\delta_3} & 0 \\ 0 & 0 & 0 & \frac{1}{\delta_4} \end{pmatrix},$$

$$CA^{-1}B = \begin{pmatrix} -\delta_1 \frac{1}{\delta_1} \omega_1 & 0 & 0 & 0 \\ 0 & -\delta_2 \frac{1}{\delta_2} \omega_2 & 0 & 0 \\ 0 & 0 & -\delta_3 \frac{1}{\delta_3} \omega_3 & 0 \\ 0 & 0 & 0 & -\delta_4 \frac{1}{\delta_4} \omega_4 \end{pmatrix}.$$

Jakobián z (3.33) teda môžeme prepísať takto:

$$J_1 = \begin{vmatrix} \delta_1 & 0 & 0 & 0 \\ 0 & \delta_2 & 0 & 0 \\ 0 & 0 & \delta_3 & 0 \\ 0 & 0 & 0 & \delta_4 \end{vmatrix} \cdot \left| \begin{pmatrix} 1-\omega_1 & 0 & 0 & 0 \\ 0 & 1-\omega_2 & 0 & 0 \\ 0 & 0 & 1-\omega_3 & 0 \\ 0 & 0 & 0 & 1-\omega_4 \end{pmatrix} - \begin{pmatrix} -\omega_1 & 0 & 0 & 0 \\ 0 & -\omega_2 & 0 & 0 \\ 0 & 0 & -\omega_3 & 0 \\ 0 & 0 & 0 & -\omega_4 \end{pmatrix} \right|,$$

t.j. po odpočítaní

$$J_1 = \begin{vmatrix} \delta_1 & 0 & 0 & 0 \\ 0 & \delta_2 & 0 & 0 \\ 0 & 0 & \delta_3 & 0 \\ 0 & 0 & 0 & \delta_4 \end{vmatrix} \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}. \tag{3.34}$$

Z (3.34) vidíme, že J_1 je súčinom determinantov dvoch diagonálnych matíc, a teda

$$J_1 = \delta_1\delta_2\delta_3\delta_4. \tag{3.35}$$

Nakoniec vyjadríme $\sum_{i=1}^{8} v_i$ pomocou nových premenných

$$\begin{gathered} \sum_{i=1}^{8} v_i = \omega_1\delta_1 + \omega_2\delta_2 + \omega_3\delta_3 + \omega_4\delta_4 + \delta_1 - \omega_1\delta_1 + \delta_2 - \omega_2\delta_2 + \\ \delta_3 - \omega_3\delta_3 + \delta_4 - \omega_4\delta_4 = \sum_{j=1}^{4} \delta_j. \end{gathered} \tag{3.36}$$

Na (3.27) aplikujeme transformáciu T_1, pričom využijeme (3.32), (3.35) a (3.36)

$$\pi = 2 \int_{T_1(K_1)} e^{-(\delta_1+\delta_2+\delta_3+\delta_4)} \delta_1\delta_2\delta_3\delta_4 \, d\delta_1 \dots d\omega_4. \tag{3.37}$$

Ako si môžeme všimnúť v (3.37), v integráli vystupujú už len premenné δ_j. Z tohto dôvodu rozdelíme integračnú množinu nasledovne

$$\pi = 2 \int_{K_2} e^{-(\delta_1+\delta_2+\delta_3+\delta_4)} \delta_1\delta_2\delta_3\delta_4 \int_{S\left(\frac{\delta_1\delta_4}{\delta_2\delta_3}\right)} d\omega_1 \dots d\omega_4 \, d\delta_1 \dots d\delta_4, \tag{3.38}$$

kde

$$K_2 = \left\{ (\delta_1, \dots, \delta_4) \in (0,\infty)^4 \middle| \frac{\delta_1\delta_4}{\delta_2\delta_3} \le 1 \right\},$$

$$S(t) = \left\{ (\omega_1, \dots, \omega_4) \in (0,1)^4 \middle| \frac{\omega_1\omega_4}{\omega_2\omega_3} \ge \frac{1}{t}, \frac{(1-\omega_1)(1-\omega_4)}{(1-\omega_2)(1-\omega_3)} \ge \frac{1}{t}, kde\ t = \frac{\delta_1\delta_4}{\delta_2\delta_3} \right\}.$$

Podľa [14] je pre $0 < t < 1$ integrál

$$q(t) := \int_{S(t)} d\omega_1 \dots d\omega_4 = -\frac{(1-t)^2}{6t} \ln(1-t) + \frac{3t-2}{12}. \tag{3.39}$$

Tento výsledok neskôr využijeme pri dopočítaní presnej hodnoty $\pi_2(1)$. Podľa zavedeného značenia teda môžeme (3.38) prepísať ako

$$\pi = 2 \int_{K_2} e^{-(\delta_1+\delta_2+\delta_3+\delta_4)} \delta_1\delta_2\delta_3\delta_4 \cdot q\left(\frac{\delta_1\delta_4}{\delta_2\delta_3}\right) d\delta_1 \dots d\omega_4. \tag{3.40}$$

Teraz využijeme ešte jednu transformáciu T_2: $(\delta_1, \delta_2, \delta_3, \delta_4) \longmapsto (x, y, z, t)$ definovanú

$$x = \delta_1, \qquad y = \delta_2, \qquad z = \delta_3, \qquad t = \frac{\delta_1\delta_4}{\delta_2\delta_3}.$$

Transformácia k nej inverzná je potom

$$\delta_1 = x, \qquad \delta_2 = y, \qquad \delta_3 = z, \qquad \delta_4 = \frac{tyz}{x}.$$

Vyjadríme transformáciu množiny K_2 v zobrazení T_2

$$T_2(K_2) = \{(x, y, z, t) \in (0, \infty)^3 \times (0,1] | t \leq 1\},$$

kde podmienka $t \leq 1$ je vždy splnená, vzhľadom na jeho definičný obor, a teda

$$T_2(K_2) = \{(x, y, z, t) \in (0, \infty)^3 \times (0,1]\}. \tag{3.41}$$

Jakobián zobrazenia T_2, ktorý označíme J_2 je

$$J_2 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{tyz}{x^2} & \frac{tz}{x} & \frac{ty}{x} & \frac{yz}{x} \end{vmatrix}.$$

Keďže J_2 je dolná trojuholníková matica, jej determinant je súčin jej diagonálnych prvkov

$$J_2 = 1 \cdot 1 \cdot 1 \cdot \frac{yz}{x}. \tag{3.42}$$

Teraz transformujeme integrál (3.40) v zobrazení T_2, pričom využijeme (3.41) a (3.42)

$$\begin{aligned} \pi &= 2 \int_{T_2(K_2)} e^{-x-y-z-\frac{tyz}{x}} \cdot \left(xyz\frac{tyz}{x}\right) \cdot J_2 \cdot q(t) dxdydzdt \\ &= 2 \int_{T_2(K_2)} e^{-\left(x+y+z+\frac{tyz}{x}\right)} \cdot \left(t\frac{(yz)^3}{x}\right) \cdot q(t) dxdydzdt. \end{aligned} \tag{3.43}$$

Keďže množina $T_2(K_2)$ už neobsahuje podmienky na premenné, len ich definičné obory, využitím Fubíniho vety upravíme (3.43) na

$$\pi = 2 \int_0^1 t \cdot q(t) \left(\int_0^\infty \int_0^\infty \int_0^\infty e^{-\left(x+y+z+\frac{tyz}{x}\right)} \cdot \frac{(yz)^3}{x} dx\, dy\, dz \right) dt \tag{3.44}$$

$$= 2\int_0^1 t \cdot q(t) \cdot I(t)\, dt.$$

Integrál $I(t)$vypočítame samostatne pomocou trial verzie programu Mathematica pre $0 < t < 1$, ako to môžeme vidieť na Obr. 3.9

$$I(t) = \frac{12\left[3ln\left(\frac{1}{t}\right)(1+t)(t^2+8t+1) - (1-t)(11t^2+38t+11)\right]}{(1-t)^7}. \tag{3.45}$$

```
In[1]:= Assuming[0 < t && t < 1, Integrate[E^(-(x + y + z + (t*y*z)/x))*((y*z)^3/x),
        {x, 0, Infinity}, {y, 0, Infinity}, {z, 0, Infinity}]]

Out[1]= - (12 ((-1 + t) (11 + t (38 + 11 t)) - 3 (1 + t) (1 + t (8 + t)) Log[t])) / (-1 + t)^7

In[2]:= 2*Integrate[(-((1 - t)^2)*Log[1 - t]/(6 t) + (3 t - 2)/12)*t*
        ((-12 ((-1 + t) (11 + t (38 + 11 t)) - 3 (1 + t) (1 + t (8 + t)) Log[t]))/
        ((-1 + t)^7)), {t, 0, 1}]

Out[2]= 1/60
```

Obr. 3.9 Zdrojový kód na výpočet integrálov $I(t)$ a (3.44) v programe Mathematica

Na záver dosadíme (3.39) a (3.45) do (3.44) a vypočítame výsledný integrál ako to zobrazuje Obr. 3.9. Pravdepodobnosť vzniku Simpsonovho paradoxu v tabuľke 2x2x2 je teda

$$\pi = \frac{1}{60}.$$

Záver

V tejto práci sme postupne prešli tri vybrané paradoxy. Venovali sme sa paradoxu kravát, problému dvoch obálok a Simpsonovmu paradoxu. Spracovali sme ich tak, aby boli zrozumiteľné študentom vysokých škôl po absolvovaní základného kurzu pravdepodobnosti a matematickej analýzy v rozsahu [4] a [5]. Podkapitola 1.1 o diskrétnom riešení paradoxu kravát ako aj výpočty v reálnych príkladoch výskytu Simpsonovho paradoxu z 3.1 sú formulované tak, aby im porozumeli aj študenti posledného ročníka gymnázií a stredných škôl, a teda aby sme touto prácou vzbudili záujem o vysokoškolské štúdium matematiky, prípadne pokračovanie v ňom..

V snahe zmeniť negatívny pohľad na štúdium matematiky sme oboznámili čitateľa s tromi pozoruhodnými pravdepodobnostnými paradoxmi, no neobišli sme pri tom ani analýzu a algebru. V práci sme využili poznatky o náhodných premenných, Riemanových integráloch ale aj o determinantoch blokových matíc. Uviedli sme riešenia každého z paradoxov. Pri paradoxe kravát sme sa venovali jeho diskrétnemu aj spojitému riešeniu. V probléme dvoch obálok sme okrem riešenia uviedli aj rozšírenie na variant s dodatočnou informáciou, vyjadrili sme podmienku výhodnosti výmeny obálky ako aj príklady jej splnenia a nesplnenia. V tretej kapitole sme sa venovali Simpsonovmu paradoxu, trom príkladom jeho reálneho výskytu a na záver sme vyjadrili pravdepodobnosť jeho vzniku v tabuľke 2x2x2.

Ako prípadné rozšírenia uvádzame možnosť použitia iného rozdelenia ceny kravaty u každého z mužov v podkapitolách 1.1 a 1.2, ako aj porovnanie stredných hodnôt z podkapitoly 1.2 pre ďalšie rozdelenia, napríklad gama rozdelenie (exponenciálne rozdelenie je jeho špeciálnym prípadom), lognormálne rozdelenie a pod.

Táto práca bola prínosom aj pre autorku, keďže mala možnosť lepšie sa zorientovať hlavne v problematike pravdepodobnosti a analýzy. Okrem toho jej poskytla príležitosť na rozvoj schopností ako je tvorivé písanie, samostatná práca a riešenie problémov. Dúfame, že bude prínosná aj pre čitateľov a vzbudí v nich túžbu po hlbšom spoznaní zaujímavostí matematiky.

Zoznam použitej literatúry

[1] Anděl, J.: *Matematika náhody*. Matfyz Press, Praha, 2007

[2] Bickel, P.J., O'Connel, E. A., Hammel J. W.: *Sex Bias in Graduate Admissions: Data from Berkeley*, Science 187 (1975), 398-404

[3] Brams, S. J., Kilgour, D. M.: *The Box Problem: To Switch or Not to Switch*, Mathematics Magazine 68 (1995), 27-34

[4] Janková, K., Pázman, A.: *Pravdepodobnosť a štatistika*, Vydavateľstvo UK, 2011

[5] Kollár, M., Kossaczká, Ľ., Ševčovič, D.: *Diferenciálny a integrálny počet funkcií viac premenných v príkladoch*, Knižničné a edičné centrum FMFI UK, Bratislava, 2012

[6] Kraitchik, M.: *Mathematical Recreations*, Dover, New York, 1953

[7] Kroese, D. P., Taimre, T., Botev, Z. I.: *Handbook of Monte Carlo Methods*, John Wiley & Sons, 2013

[8] Lehe, L., Powell, V.: *Simpson's Paradox*, dostupné na internete (21.5.2014): http://vudlab.com/simpsons/

[9] Merryfield, K. G., Viet, N., Watson, S.: *The wallet paradox*, American Mathematical Monthly 104 (1997), 647-649, dostupné na internete (21.5.2014): http://www.jstor.org/discover/10.2307/2975058?uid=3739024&uid=2&uid=4&sid=21104052014467

[10] Meyer, C. D.: *Matrix analysis and applied linear algebra*, Siam, Philadelphia, 2000, dostupné na internete (21.5.2014): http://www.matrixanalysis.com/

[11] Morrell, Ch. H.: *Simpson's Paradox: An Example From a Longitudinal Study in South Africa*, Journal of Statistics Education 7 (1999), dostupné na internete (21.5.2014): http://www.amstat.org/publications/jse/secure/v7n3/datasets.morrell.cfm

[12] Necktie paradox, dostupné na internete (21.5.2014): http://en.wikipedia.org/wiki/Necktie_paradox

[13] Oficiálne pravidlá bejzbalu, dostupné na internete (21.5.2014):

http://www.bkzvolen.sk/data/baseball_pravidla.pdf

[14] Pavlides, M. G., Perlman, M. D.: *How Likely Is Simpson's Paradox?*, The American Statistician 63 (2009), 226-233

[15] Plocki, A., Tlustý, P.: *Pravděpodobnost a statistika pro začátečníky a mírně pokročilé*, Prometheus, Praha, 2007

[16] R: Kernel Density Estimation, dostupné na internete (21.5.2014): http://stat.ethz.ch/R-manual/R-patched/library/stats/html/density.html

[17] R: The Exponential Distribution, dostupné na internete (21.5.2014): http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Exponential.html

[18] Ross, K.: *A Mathematician at the Ballpark: Odds and Probabilities for Baseball Fans*, Pi Press, New York, 2004