

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

## DIPLOMOVÁ PRÁCE

Modelování závislostí mezi rizikovými faktory  
v simulaci Monte Carlo



Vedoucí diplomové práce: **RNDr. Ondřej Pavlačka, Ph.D.**  
Vypracovala: **Bc. Barbora Sedláčková**  
Studijní program: B1103 Aplikovaná matematika  
Studijní obor Aplikace matematiky v ekonomii  
Forma studia: prezenční  
Rok odevzdání: 2017

## BIBLIOGRAFICKÁ IDENTIFIKACE

**Autor:** Bc. Barbora Sedláčková

**Název práce:** Modelování závislostí mezi rizikovými faktory v simulaci Monte Carlo

**Typ práce:** Diplomová práce

**Pracoviště:** Katedra matematické analýzy a aplikací matematiky

**Vedoucí práce:** RNDr. Ondřej Pavlačka, Ph.D.

**Rok obhajoby práce:** 2017

**Abstrakt:** Tato diplomová práce se zabývá modelováním závislostí mezi rizikovými faktory v simulaci Monte Carlo. V první části této práce je představena problematika simulace Monte Carlo. Další část je věnována metodám pro modelování závislostí mezi rizikovými faktory. Je v ní popsáno modelování korelací, a to pomocí Spearmanova korelačního koeficientu pořadové korelace a pomocí kopulí, dále obálková metoda a alternativní přístupy k modelování závislostí. V poslední části práce jsou jednotlivé metody ilustrovány na numerických příkladech.

**Klíčová slova:** simulace Monte Carlo, analýza rizika, rizikové faktory, závislost, korelace, Spearmanův koeficient pořadové korelace, kopuly, rozdělení pravděpodobnosti, obálková metoda

**Počet stran:** 66

**Počet příloh:** 1

**Jazyk:** český

## BIBLIOGRAPHICAL IDENTIFICATION

**Author:** Bc. Barbora Sedláčková

**Title:** Modelling of dependencies between risk factors in Monte Carlo simulations

**Type of thesis:** Diploma's

**Department:** Department of Mathematical Analysis and Application of Mathematics

**Supervisor:** RNDr. Ondřej Pavlačka, Ph.D.

**The year of presentation:** 2017

**Abstract:** This thesis deals with the modelling of dependencies between risk factors in Monte Carlo simulations. The first part of the thesis concentrates on the problem of Monte Carlo simulations. The next part is dedicated to methods for modelling of dependencies between risk factors. There are described modelling of correlations, namely by Spearman's rank order correlation coefficient and by copulas, further the envelope method and the alternative approaches for modelling dependencies. The last part of the thesis is focused on illustration of these methods on the numerical examples.

**Key words:** Monte Carlo simulation, risk analysis, risk factors, dependence, correlation, Spearman's rank order correlation coefficient, copulas, probability distribution, the envelope method

**Number of pages:** 66

**Number of appendices:** 1

**Language:** Czech

### **Prohlášení**

Prohlašuji, že jsem tuto diplomovou práci zpracovala samostatně pod odborným vedením pana RNDr. Ondřeje Pavlačky, Ph.D. a že jsem v seznamu literatury uvedla všechny použité zdroje.

V Olomouci dne .....  
.....  
podpis

# Obsah

<b>Úvod</b>	<b>7</b>
<b>1 Simulace Monte Carlo</b>	<b>8</b>
1.1 Úvod do simulace Monte Carlo . . . . .	8
1.2 Popis simulace Monte Carlo . . . . .	9
1.3 Přednosti a nedostatky simulace Monte Carlo . . . . .	12
<b>2 Modelování závislostí mezi rizikovými faktory</b>	<b>13</b>
2.1 Modelování korelací . . . . .	13
2.1.1 Pořadová korelace . . . . .	16
2.1.2 Modelování závislostí pomocí kopulí . . . . .	21
2.2 Obálková metoda . . . . .	39
2.2.1 Odvození parametrů rozdělení pravděpodobnosti rizikového faktoru $Y$ na základě realizace rizikového faktoru $X$ . . . . .	39
2.2.2 Modelování závislostí mezi rizikovými faktory $X$ a $Y$ pomocí regresní analýzy . . . . .	42
2.3 Alternativní přístupy k modelování závislostí . . . . .	45
2.3.1 Závislost definovaná pomocí vyhledávacích tabulek . . . . .	45
2.3.2 Závislost definovaná pomocí logických podmínek . . . . .	45
<b>3 Ilustrační příklady</b>	<b>46</b>
3.1 Příklad využívající kopulí pro modelování expertně stanovené korelace . . . . .	46
3.2 Příklad využívající kopulí pro modelování korelace odhadnuté z dat	50
3.2.1 Vytvoření matematického modelu . . . . .	51
3.2.2 Určení marginálních rozdělení rizikových faktorů . . . . .	52
3.2.3 Stanovení statistické závislosti . . . . .	55
3.2.4 Výsledky modelu . . . . .	58
3.3 Příklad využívající obálkové metody pro modelování závislosti mezi rizikovými faktory na základě expertních odhadů . . . . .	60
3.4 Příklad využívající obálkové metody pro modelování závislosti mezi rizikovými faktory na základě historických dat . . . . .	62
<b>Závěr</b>	<b>65</b>
<b>Literatura</b>	<b>66</b>

## **Poděkování**

Na tomto místě bych ráda poděkovala vedoucímu mé diplomové práce panu RNDr. Ondřeji Pavlačkovi, Ph.D. za obětavou spolupráci, cenné rady, trpělivost a za čas, který mi věnoval na konzultacích. Ráda bych také poděkovala mé rodině a mému partnerovi za velkou podporu při studiu.

# Úvod

V této práci se budeme zabývat metodami, pomocí kterých můžeme modelovat závislost mezi rizikovými faktory v simulaci Monte Carlo. Simulace Monte Carlo představuje analytickou metodu, jejíž princip je založen na využívání náhodných čísel. Podstata této metody spočívá ve vygenerování velkého počtu možných kombinací vstupních veličin, kdy pro každou takto vygenerovanou kombinaci je vypočtena hodnota výstupní veličiny.

Poměrně běžnou situací, se kterou se v praxi často setkáváme, je provázanost rizikových faktorů. Hodnoty jednotlivých faktorů rizika mohou záviset na hodnotách některých jiných faktorů, přičemž závislost mezi takovými rizikovými faktory může nabývat různých podob. V případě, kdy nezahrneme závislost do modelu, se výsledky simulace mohou výrazně lišit od výsledků, kdy je závislost mezi jednotlivými rizikovými faktory v modelu respektována. V důsledku toho je velmi důležité při sestavování modelu věnovat této problematice značnou pozornost.

Cílem diplomové práce je seznámit čtenáře s problematikou stanovení závislostí mezi rizikovými faktory, jeho významu pro simulaci, a především představit jednotlivé metody, na jejichž základě lze závislost mezi rizikovými faktory v simulaci Monte Carlo modelovat.

V první kapitole této práce se budeme věnovat metodě Monte Carlo. Nejdříve si představíme něco z historie této metody, poté si ukážeme, jak postupovat při simulaci, a nakonec zmíníme její přednosti a nedostatky. V druhé kapitole se zaměříme na jednotlivé metody pro modelování, a to na korelace, které můžeme modelovat pomocí Spearmanova korelačního koeficientu pořadové korelace nebo pomocí kopulí, dále na obálkovou metodu a alternativní přístupy. Ve třetí kapitole si názorně představíme popsání metody na ilustračních příkladech, pro jejichž zpracování použijeme software Matlab a MS Excel.

# 1. Simulace Monte Carlo

V této kapitole se seznámíme se statistickou metodou simulace Monte Carlo. Nejdříve si představíme, v čem tato metoda spočívá a proč je vhodné s ní pracovat a poté si popíšeme, jak postupovat při konstrukci modelů, založených na této metodě. Ke zpracování této kapitoly byla využita literatura [3], [4] a [10].

## 1.1. Úvod do simulace Monte Carlo

Simulace Monte Carlo je statistická metoda, která má široké použití v mnoha vědních oborech. V případě, kdy máme více významných rizikových faktorů, které ovlivňují výsledky analýzy rizika objektu (investičních projektů, firmy aj.), mnohdy nelze vzhledem k možnému počtu kombinací potenciálních stavů využít běžné nástroje analýzy rizika, jako jsou například pravděpodobnosti, kvantitativní scénáře apod. Výsledkem této situace je právě využití simulace Monte Carlo. Podstata simulace Monte Carlo spočívá v generování velkého počtu scénářů, řádově stovek až desítek tisíců, a v následném propočtu hodnot zvolených kvantitativních kritérií (např. finančních) pro každý tento scénář. Výstup simulace je v podobě grafického zobrazení rozdělení pravděpodobnosti těchto kvantitativních kritérií a odhadů jejich číselných charakteristik.

První simulace, které se podobají simulacím Monte Carlo, jsou již známy z počátků teorie pravděpodobnosti. S prvním skutečným využitím simulace Monte Carlo, jakožto výpočetního postupu, se však můžeme setkat až ve čtyřicátých letech 20. století. Zakladateli této simulace byli vědečtí pracovníci Stanislav Ulam a John von Neumann. Zabývali se zkoumáním chování neutronů a především je zajímalo, jaké množství neutronů projde různými materiály. Tento úkol bohužel nemohli vyřešit teoreticky ani prakticky, a to i napříč tomu, že disponovali s důležitými údaji, jako s pravděpodobností srážky neutronu s atomem vodíku nebo kyslíku, s množstvím energie, kterou ztrácí neutron během srážky atd. I přes znalost těchto informací nemohli nalézt metodu, která by jim pomohla předpovědět další chování neutronu (předpovědět „historii jeho života“). K nalezení vý-



sledku jim dopomohla právě až simulace Monte Carlo, kdy pro modelování předpovědi „historie života neutronu“ autoři využili techniku kola rulety. Simulace Monte Carlo jim následně pomohla předpovědět trajektorii každého neutronu.

Metoda simulace Monte Carlo je dnes využívána nejčastěji v situacích, kdy nelze předem požadovaný výsledek spočítat přímo (např. ve formě vzorce), a proto je nezbytné postup simulovat s pomocí výpočetní techniky.

## 1.2. Popis simulace Monte Carlo

Simulaci Monte Carlo můžeme rozdělit do několika následujících kroků:

1. Stanovení celkového **kontextu** modelu, tedy určení **kriteriálních veličin** (výstupů) a dále identifikace **rizikových faktorů** (vstupů).

V tomto kroku je nejdříve nezbytné si uvědomit, čeho má být dosaženo, co pomocí modelu chceme získat. Je nutné stanovit si kriteriální veličiny, kterými mohou být například čistá současná hodnota, zisk apod. Také je nezbytné si uvědomit, v jakém prostředí se budou účelové aktivity pro dosažení cílů realizovat. Následně pak lze mluvit o možných náhodných vlivech, které by mohly požadovaný či předpokládaný stav ohrozit nebo naopak, které by mohly postup k požadovanému cíli usnadnit. Teprve potom je možné provést analýzu rizika.

2. **Tvorba matematického modelu** objektu analýzy rizika a **zpracování** jeho **programu** v MS Excel či jiném matematickém softwaru.

Tvorba matematického modelu je založena na zkonstruování určitých matematických vzorců, kdy z jednotlivých rizikových faktorů (vstupních veličin) vypočteme hodnotu kvantitativního kritéria (výstupní veličiny). Na základě takto zkonstruovaných vzorců poté zpracujeme model například v softwaru MS Excel.

Nejčastější podobou tohoto modelu, při analýze rizika firmy či určitého finančního projektu, může být výkaz zisku a ztráty nebo rozvaha peněžních toků. Může však být i ve formě vztahů pro výpočet jednotlivých fi-

nančních kritérií či ukazatelů pro hodnocení firmy nebo investičního projektu (např. ukazatel rentability, čistá současná hodnota, doba návratnosti apod.).

### 3. Vytýčení klíčových rizikových faktorů.

Pod pojmem klíčové rizikové faktory rozumíme takové vstupní veličiny modelu, které zásadně ovlivňují nejistotu výstupů simulace (uvažované finanční ukazatele nebo kvantitativní kritéria). Nejistota těchto rizikových faktorů se bude v simulaci respektovat a současně budeme ostatní vstupní veličiny modelu uvažovat jako konstanty ve formě jejich bodových odhadů. Mezi klíčové faktory rizika budeme především řadit ta rizika, jejichž změny budou velmi citlivé na výstupy simulace. Jako užitečný nástroj pro výběr těchto klíčových faktorů můžeme využít *analýzu citlivosti* (viz [4]).

### 4. Stanovení rozdělení pravděpodobnosti klíčových rizikových faktorů.

V případě diskrétních rizikových faktorů můžeme rozdělení pravděpodobnosti stanovit přímo (vypíšeme hodnoty a jejich pravděpodobnosti) nebo můžeme využít některé z teoretických diskrétních rozdělení pravděpodobnosti (Poissonovo, binomické atd.). Naopak v případě spojitých rizikových faktorů se volí některé z teoretických spojitých rozdělení pravděpodobnosti (normální, exponenciální atd.) a zadávají se jeho parametry.

Stanovení rozdělení pravděpodobnosti bývá značně náročné a máme k užití dvě možnosti.

- Jestliže máme k dispozici historická data určitého rizikového faktoru (např. měnového kurzu, prodeje aj.), můžeme tato data aproximovat některým teoretickým rozdělením.
- Pokud nemáme historická data k dispozici, musíme vycházet ze znalostí a zkušeností expertů z oblastí, ke kterým se dané rizikové faktory vztahují.

## 5. Stanovení statistické závislosti rizikových faktorů.

Hodnoty určitých rizikových faktorů, které modelujeme, mohou na sobě různě záviset (například můžeme mít závislost mezi poptávkou po určitém produktu a jeho prodejní cenou, jelikož růst prodejní ceny ovlivní pokles poptávky a naopak). Důsledkem toho pak nelze při simulaci generovat nezávisle na sobě tyto hodnoty rizikových faktorů, ale je třeba jejich větší či menší závislost respektovat, jinak bychom dostali nepřesné výsledky. Můžeme rozlišit dva typy statistické závislosti rizikových faktorů:

- párová závislost = vzájemná závislost dvou rizikových faktorů ve stejném období (příkladem může být závislost poptávky na prodejní ceně),
- časová závislost = závislost téhož rizikového faktoru, avšak ve dvou časových obdobích (například prodej určitého nového produktu v roce jeho nastoupení na trh ovlivňuje prodej tohoto výrobku v letech dalších).

## 6. Vlastní proces simulace s užitím počítačového programu.

Vlastní proces simulace probíhá pomocí speciálního softwaru, kterým může být např. Crystal Ball, @RISK, MS Excel či Matlab, kdy vygenerujeme možné kombinace hodnot rizikových faktorů z jejich rozdělení pravděpodobnosti, tzn., že vytvoříme určitý scénář, a pro každý takto vygenerovaný scénář propočteme výstupní hodnotu uvažovaného kritéria modelu. Následně po dostatečně velkém počtu iterací získáváme soubor nasimulovaných dat  $x_1, \dots, x_n$  daného kritéria, který můžeme dále statisticky analyzovat:

- graficky - rozdělení pravděpodobnosti výstupní veličiny (zvoleného kritéria), příkladem může být empirická distribuční funkce či histogram,
- spočítat odhady číselných charakteristiky rizika, které mohou být v podobě průměru, rozptylu, směrodatné odchylky, variačního koeficientu, koeficientu šikmosti, koeficientu špičatosti nebo také pravděpodobnosti překročení či nedosažení požadované hodnoty kritéria aj.

### 1.3. Přednosti a nedostatky simulace Monte Carlo

Simulace prováděné metodou Monte Carlo představují užitečný nástroj analýzy rizika, která zvyšuje kvalitu významných rozhodnutí za rizika a nejistoty v různých oblastech managementu. Dle [4] můžeme mezi přednosti simulace Monte Carlo zařadit především to, že dokáže přinutit analytiku či rozhodovatele, kteří analyzují a hodnotí určité objekty jako např. investiční projekty, aby hlouběji přemýšleli a analyzovali tyto objekty z hlediska jednotlivých rizikových faktorů, jejich nejistot a vzájemných závislostí. To vše poté vede k hlubšímu poznání rizikové stránky objektů a také k lépe podloženému rozhodování souvisejícího například s přijetím či zamítnutím určitých investičních projektů.

Na druhou stranu má metoda Monte Carlo i své nedostatky. Jedním z problémů, které mohou nastat, je značná obtížnost týkající se stanovení rozdělení pravděpodobnosti vstupních rizikových faktorů a respektování jejich závislosti. Další nevýhodou této metody je poskytnutí jen přibližného výsledku. Avšak je teoreticky dokázané, že prostým zvýšením počtu iterací můžeme dosáhnout libovolné přesnosti výsledku. Zde však můžeme narazit na problém týkající se výpočetních možností počítačů (např. nepříjemně vzroste doba výpočtů). Největší nedostatek simulace představuje skutečnost, že nejvýznamnější rizikové faktory (ty, které nejvíce ovlivňují výsledky analýzy rizika), jsou často na základě hodnocení současnosti a minulosti nepředvídatelné. Jelikož se obvykle vychází pouze ze známých, v minulosti a přítomnosti vystupujících, faktorů rizika a nedochází k opakované identifikaci rizikových faktorů, může to mít za následek při uplatnění simulace tzv. *tunelový efekt*. To vše pak může směřovat ke kvantifikaci nesprávných rizik, což představuje hlavní riziko simulace.

## 2. Modelování závislostí mezi rizikovými faktory

V této části diplomové práce se budeme podrobněji zabývat již výše zmíněným krokem simulace Monte Carlo, a to stanovením statistické závislosti rizikových faktorů. Pro zpracování této kapitoly byla využita literatura [2], [4], [6], [7] a [10].

Rizikové faktory, které modelujeme s využitím určitého rozdělení pravděpodobnosti, mohou být na sobě různě závislé. V této kapitole se seznámíme vedle klasické lineární závislosti, která je definovaná přes korelace, také s jinými možnostmi modelování závislostí, jako např. *modelování závislostí pomocí kopulí, obálková metoda, závislost definovaná pomocí vyhledávacích tabulek* a také *závislost definovaná pomocí logických podmínek*.

**Zlaté pravidlo analýzy rizika** nám říká: „**Každý vygenerovaný scénář musí být pozorovatelný v reálném životě.**“ (Dle D. Voseho, [10]). Toto pravidlo nám vlastně sděluje, že model musí být sestaven tak, aby nemohl vytvořit scénář, který by v reálném životě nemohl nastat. Na základě toho bychom měli do modelu zavést možné závislosti mezi jednotlivými rizikovými faktory. Velmi důležité je uvědomit si, že nebudeme-li definovat závislost mezi rizikovými faktory, implicitně předpokládáme, že jsou rizikové faktory vzájemně nezávislé. Tento předpoklad je však v mnoha případech zcela nerealistický a oslabující hodnotu simulační analýzy, jelikož v rámci jednoho scénáře může být vytvořena jakákoliv kombinace hodnot jednotlivých rizikových faktorů.

### 2.1. Modelování korelací

Pomocí korelací můžeme modelovat *lineární typ* závislostí. Při modelování jsou však vhodné pouze tehdy, existuje-li předpoklad, že závislosti jsou „přibližně“ lineární. V případech, kdy tento předpoklad neexistuje, korelace jako způsob analýzy nebo modelování závislostí selhává. Korelace může být *pozitivní*, kdy vysoké hodnoty jednoho rizikového faktoru odpovídají vysokým hodnotám druhého rizikového faktoru a *negativní*, kdy nízké hodnoty jednoho rizikového faktoru odpovídají vysokým hodnotám druhého rizikového faktoru (či naopak).

Korelaci mezi rizikovými faktory  $X$  a  $Y$  můžeme zjistit pomocí *korelačního koeficientu*  $\rho(X, Y)$  označovaného také jako *lineární korelační koeficient*, který je definován vztahem:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y},$$

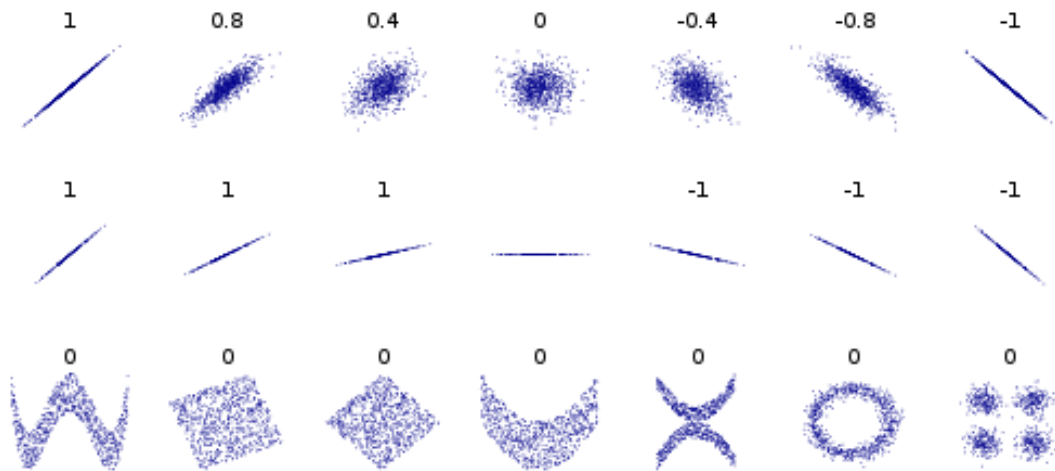
kde  $\text{cov}(X, Y)$  představuje kovarianci mezi rizikovými faktory  $X$  a  $Y$ , která je definovaná vztahem [5]

$$\text{cov}(X, Y) = E[X - E(X)][X - E(Y)],$$

a  $\sigma_X, \sigma_Y$  jsou výběrové směrodatné odchylky rizikových faktorů  $X$  a  $Y$ .

Korelační koeficient nabývá hodnot z intervalu  $\langle -1, 1 \rangle$ . Pokud je hodnota korelačního koeficientu rovna  $-1$ , jde o *úplnou nepřímou závislost* rizikových faktorů  $X$  a  $Y$ . Jestliže naopak hodnota korelačního koeficientu nabývá hodnoty  $1$ , jedná se o *úplnou přímou závislost* rizikových faktorů  $X$  a  $Y$ . Pokud je však korelační koeficient roven  $0$ , rizikové faktory jsou *nekorelované*, tzn., že mezi nimi neexistuje lineární závislost. Je však důležité si uvědomit, že to neznamená, že by mezi faktory nebyla žádná závislost.

Využitím následujícího funkčního vztahu můžeme velmi názorně demonstrovat, že korelace nejsou ve všech situacích vhodným způsobem pro modelování závislosti. Máme náhodnou veličinu  $X$ , která pochází z rovnoměrného rozdělení  $Ro(-1; 1)$  a také náhodnou veličinu  $Y$ , která bude definovaná jako druhá mocnina veličiny  $X$ , tedy  $Y = X^2$ . Přestože se jedná zcela o funkční, tj. deterministický vztah, tak je korelace mezi náhodnou veličinou  $X$  a náhodnou veličinou  $Y$  *nulová*, viz příložený soubor *příklad.m*. Další ukázky korelací a závislostí mezi náhodnými veličinami můžeme vidět na obrázku 1. Jsou zde znázorněny korelační koeficienty a také různé typy závislostí mezi náhodnými veličinami  $X$  a  $Y$ . Na prvním řádku tohoto obrázku můžeme vidět, že korelační koeficienty zachycují pouze „sílu“ lineárního vztahu a jeho směr, nikoliv však sklon, který je naopak vidět na druhém řádku. Třetí řádek nám pak ukazuje celou řadu nelineárních závislostí, které korelační koeficient modelovat nedokáže.



Obrázek 1: Příklady korelací a závislostí mezi náhodnými veličinami X a Y (obrázek je převzat z [11])

Obecně existují 3 základní důvody, **proč mohou být data korelovaná**:

- existuje určitý *logický vztah mezi dvěma (či více) rizikovými faktory*, příkladem může být úroková míra, která statisticky určuje hypoteční úrokovou míru nebo také výška a váha člověka,
- existuje nějaký *externí faktor, který ovlivňuje oba rizikové faktory*, příkladem může být počasí, které během výstavby budovy může ovlivnit dobu trvání výkopu základů a také dobu trvání výstavby základů,
- *pozorovaná korelace mezi rizikovými faktory má čistě náhodný charakter*<sup>1</sup>, ve skutečnosti jsou rizikové faktory na sobě nezávislé, tedy neexistuje mezi nimi žádný logicky odůvodnitelný vztah, příkladem zde může být spotřeba objemu mobilních dat v České republice a populace v Asii (jelikož hodnoty obou faktorů neustále rostou, mohli bychom zde pozorovat silně korelovaný vztah).

V prvním případě lze při tvorbě modelu využít dále popsané přístupy pro modelování korelací mezi rizikovými faktory. V případě existence externího faktoru

<sup>1</sup>Korelace popsaná v tomto případě by se měla v modelech „ignorovat“.

je vhodné řešit situaci modelováním závislosti obou rizikových faktorů na daném externím faktoru.

Pro modelování korelací v případě, kdy existuje logický vztah mezi dvěma rizikovými faktory, se v simulačních modelech využívají **2 přístupy**:

- využití tzv. *Spearmanova koeficientu pořadové korelace*,
- modelování pomocí tzv. „kopulí“ (z anglického Copulas).

Nejdříve se zaměříme na pořadové korelace, se kterými se můžeme často setkat v mnoha softwarových produktech. Následně se budeme podrobněji věnovat modelování korelací pomocí tzv. *kopulí*, které představují flexibilnější metodu.

### 2.1.1. Pořadová korelace

Pořadová korelace je založena na podobném principu jako klasická korelace. Opět nabývá hodnot z intervalu  $\langle -1, 1 \rangle$ , ale místo původních hodnot se pracuje s jejich pořadím. Nyní budeme brát v potaz místo lineárního korelačního koeficientu Spearmanův koeficient pořadové korelace, jehož hodnota vychází většinou hodně podobně jako hodnota lineárního korelačního koeficientu.

Při výpočtu tohoto koeficientu nepracujeme přímo s hodnotami  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , ale s jejich pořadím  $(u_i, v_i)$  v rámci skupin  $x_1, \dots, x_n$  a  $y_1, \dots, y_n$ , kde  $n$  značí rozsah souboru. To znamená, že  $u_i$  nám představuje pořadí hodnoty  $x_i$  a  $v_i$  nám představuje pořadí hodnoty  $y_i$ .

V případě, že se v datovém souboru nevyskytují stejné hodnoty faktoru  $X$  nebo  $Y$ , je Spearmanův koeficient pořadové korelace definován následovně:

$$\rho_s(X, Y) = 1 - \frac{6 \sum_{i=1}^n (u_i - v_i)^2}{n(n^2 - 1)},$$

kde  $u_i$  představuje pořadí hodnoty  $x_i$  rizikového faktoru  $X$  a  $v_i$  představuje pořadí hodnoty  $y_i$  rizikového faktoru  $Y$ . Jinak se vypočítá pomocí vzorce:

$$\rho_s(X, Y) = \frac{\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^n (u_i - \bar{u})^2 \sum_{i=1}^n (v_i - \bar{v})^2}},$$

kde  $\bar{u}$  a  $\bar{v}$  jsou průměry z hodnot pořadí ve skupinách  $x_1, \dots, x_n$  a  $y_1, \dots, y_n$ .



**Příklad 2.1** Pro ukázkou výpočtu lineárního korelačního koeficientu a také Spearmanova koeficientu pořadové korelace můžeme využít hodnoty výšek (faktor  $X$ ) a hmotností (faktor  $Y$ ) 10 studentů, které jsou zobrazeny v tabulce 1. Zajímá nás tedy, zda existuje závislost mezi faktorem  $X$  a faktorem  $Y$ .

Student	$x_i$	$y_i$	$u_i$	$v_i$	$(u_i - v_i)^2$
1	182	67	10	9	1
2	173	58	6	4	4
3	165	56	1	3	4
4	171	61	4	7	9
5	180	68	9	10	1
6	177	65	8	8	0
7	167	60	2	6	16
8	175	59	7	5	4
9	169	51	3	1	4
10	172	54	5	2	9

Tabulka 1: Tabulka hodnot výšek a hmotností 10 studentů

Koeficient	Hodnota
Lineární korelační koeficient	0,763
Spearmanův koeficient pořadové korelace	0,685

Tabulka 2: Výsledné hodnoty jednotlivých koeficientů

Z tabulky 2 je zřejmé, že mezi faktory  $X$  a  $Y$  existuje pozitivní korelace.

Postup simulace Monte Carlo, který se používá ve speciálních simulačních softwarech (např. Crystal Ball či @Risk) poté probíhá v následujících 5 krocích:

1. nejdříve se zadá hodnota korelačního koeficientu  $\rho(X, Y)$ ,
2. následně pro zadaný počet scénářů simulace ( $n$ ) vygeneruje speciální simulační software  $n$ -tici uspořádaných dvojic pořadí takovou, že korespondující hodnota Spearmanova korelačního koeficientu je rovna požadované korelaci  $\rho(X, Y)$  (např. (1, 21), (2, 12), (3, 22), ...),

3. dále se nezávisle na sobě vygenerují  $n$ -tice hodnot  $x_1, \dots, x_n$  a  $y_1, \dots, y_n$  z příslušných pravděpodobnostních rozdělení jednotlivých faktorů rizika,
4. poté se vygenerované hodnoty uspořádají podle velikosti, tzn.  $x_{(1)}, \dots, x_{(n)}$  a  $y_{(1)}, \dots, y_{(n)}$  a následně se spárují podle 2. kroku, například tedy dostaneme  $(x_{(1)}, y_{(21)}), (x_{(2)}, y_{(12)}), (x_{(3)}, y_{(22)}), \dots,$
5. nakonec pak vstupují do jednotlivých scénářů pouze spárované uspořádané dvojice  $(x_i, y_i)$  ze 4. kroku.

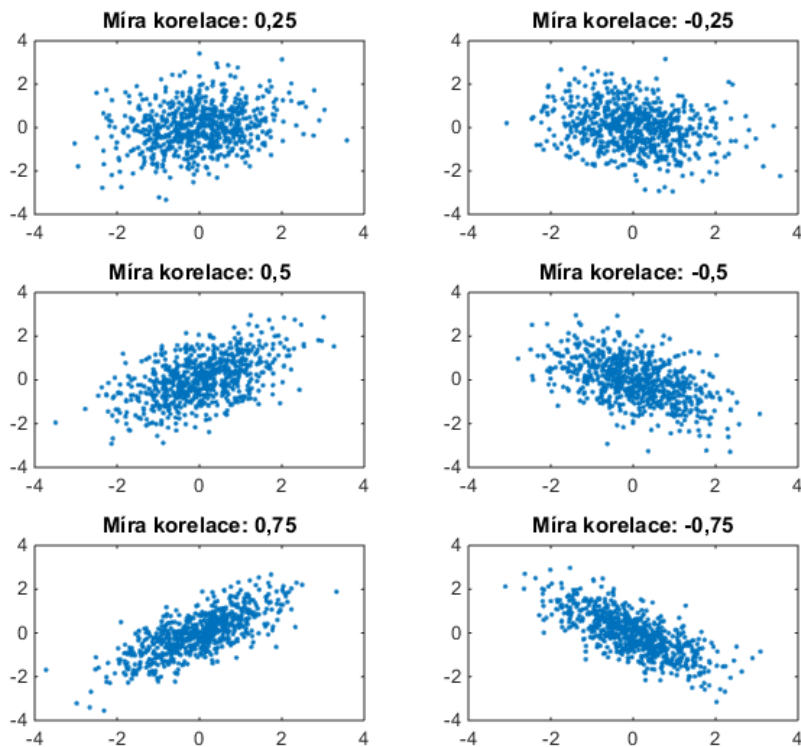
My však v této práci budeme k tomuto využívat postupu s využitím tzv. kopulí, který si představíme v následující kapitole.

Jako výhodu pořadových korelací můžeme zmínit, že na rozdíl od klasických korelací nijak neovlivňují tvar korelovaných rozdělení. Díky tomu jsou pořadové korelace vhodné pro modelování korelací mezi různými typy pravděpodobnostních rozdělení.

Hlavní nevýhodu pořadových korelací podle [10] představuje problém, který je spojen s volbou konkrétních hodnot korelačních koeficientů, jelikož pro jeho stanovení neexistuje žádné jednoznačné doporučení. Nejvýhodnější je případ, kdy máme k dispozici určitá historická data, o které se můžeme opřít či z nich přímo hodnotu korelačního koeficientu vypočítat. Bohužel se však nejčastěji setkáváme s projekty, které jsou jedinečné a nezbyvá nám než se spolehnout na expertní odhady.

Při určení odhadu korelačního koeficientu  $\rho(X, Y)$  expertně je vhodné využít vizuální pomůcky, jako jsou např. bodové grafy, které mohou napomoci k odhadu korelačního koeficientu (příkladem je obrázek 2). Můžeme zde pozorovat, že korelace je vizuálně viditelná přibližně od úrovně 0,5 a vyšší či  $-0,5$  a nižší.

Na základě takto vytvořených obrázků můžeme poté diskutovat nad možnou mírou závislosti mezi příslušnými rizikovými faktory. Nejčastěji se expertně definuje pouze jedna z hodnot 0,  $\pm 0,25$  pro značně slabou míru závislosti,  $\pm 0,5$  pro silnou míru závislosti a  $\pm 0,75$  pro značně silnou míru závislosti.

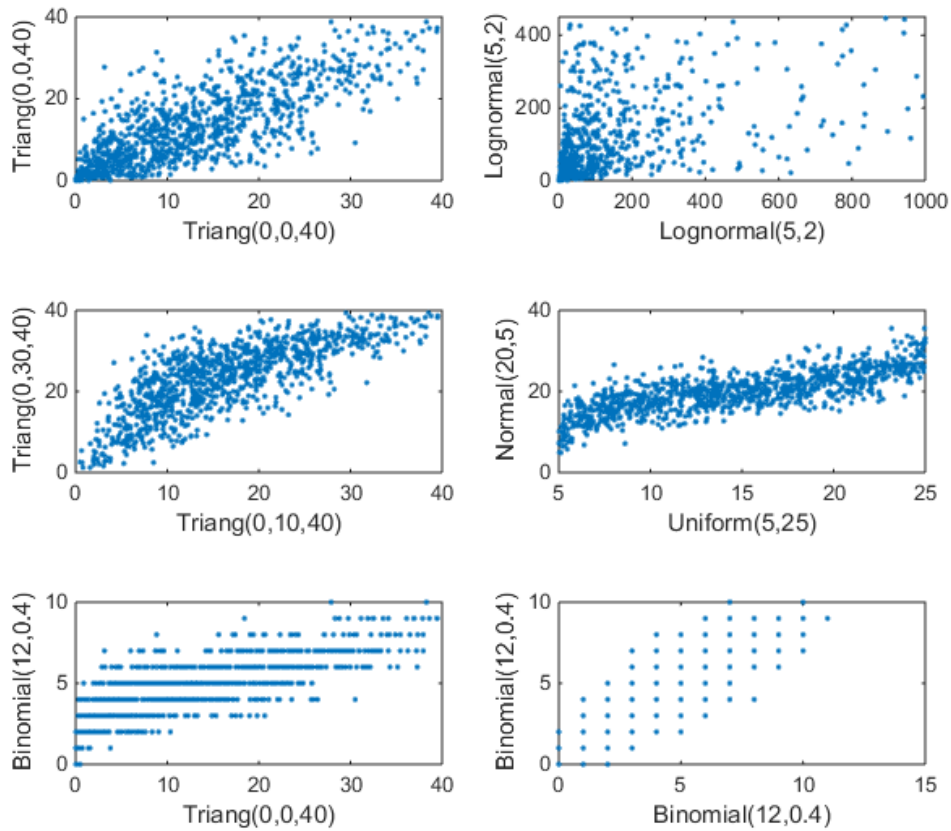


Obrázek 2: Míry korelace v grafickém vyjádření

V tomto postupu však nastává podle D. Vose [10] problém v tom, že vizuální vzhled výsledku generování silně závisí také na pravděpodobnostním rozdělení jednotlivých rizikových faktorů. Tato problematika je znázorněna na obrázku 3. Na tomto obrázku můžeme vidět výsledky vygenerovaných hodnot náhodných veličin s různými typy pravděpodobnostních rozdělení a korelačním koeficientem, který je pro všechna tato rozdělení roven hodnotě 0,8.

Navzdory neodmyslitelným nevýhodám spojených s určením pořadových korelací je tato technika velmi praktická díky svému snadnému použití a rychlosti realizace. Následující shrnutí, týkající se použití pořadových korelací, nám pomůže zajistit vyhnutí se jakémukoliv problému [10]:

- Pořadové korelace by se měly využívat k modelování závislostí mezi proměnnými pouze v případě, pokud mají malý vliv na výsledky našeho mo-



Obrázek 3: Vizuální srovnání vygenerovaných hodnot náhodných veličin s různými rozděleními pravděpodobnosti a s korelačním koeficientem 0,8

delu. Pokud si nejsme jisti jeho vlivem, můžeme si pomoci vytvořením dvou simulací, kde v jedné budeme uvažovat závislost mezi rizikovými faktory s určitým korelačním koeficientem a v druhé naopak nebudeme uvažovat závislost mezi rizikovými faktory (tedy simulaci s nulovým korelačním koeficientem). Pokud zjistíme, že mezi konečnými výsledky tohoto modelu je podstatný rozdíl, tak raději využijeme jednu z přesnějších technik, které budou popsány později v této práci.

- Všude tam, kde to bude možné, by se mělo použití pořadových korelací omezit pouze na modelování závislostí dvojic rizikových faktorů, které mají

podobně tvarované rozdělení pravděpodobnosti.

- Před zavedením závislosti do modelu je velmi důležité zobrazit si bodový graf jako vizuální pomůcku, a to hlavně v případě, kdy mají korelované rizikové faktory rozdílně tvarované rozdělení pravděpodobnosti.
- Při využití expertních názorů pro odhady korelací je užitečné používat grafy s různými úrovněmi korelace, které jim usnadní zvolit vhodnou míru korelace.
- Měli bychom se vyvarovat modelování korelací tam, kde neexistuje žádný logický vztah mezi rizikovými faktory ani důkaz pro jeho existenci.
- Měli bychom zvážit použití kopulí, pokud korelace v našem modelu hrají důležitou roli.

### 2.1.2. Modelování závislostí pomocí kopulí

Dle [6] se v dřívějších dobách většinou vycházelo ze značně zjednodušených předpokladů a využívalo se k modelování vícerozměrného rozdělení hlavně více-rozměrné normální rozdělení, kde nám stačila jen znalost korelačního koeficientu pro popis závislosti mezi dvěma náhodnými veličinami. Avšak s postupem času vyvstala potřeba používat i jiná marginální rozdělení (vedle normálního) a také další míry asociace, které by kromě lineární závislosti určitým způsobem charakterizovaly i jiné typy závislostí mezi náhodnými veličinami. Užitečným nástrojem pro modelování závislosti se proto stal přístup pomocí kopulí, který nám dává možnost zabývat se marginálními rozděleními a závislostní strukturou odděleně.

Nejdříve si v této kapitole nadefinujeme, co to vlastně kopuly jsou, a poté si představíme základní větu teorie kopulí, kterou je Sklarova věta. Následně si ukážeme jejich základní vlastnosti a nakonec si představíme klasifikaci kopulí na tři základní skupiny. V práci se budeme zaměřovat především na dvourozměrný případ kopulí, ale ukážeme si i jejich zobecnění. K vypracování této části byla použita literatura [2], [6], [7] a [10].

**Definice 2.1** *Dvourozměrná kopula  $C : \langle 0, 1 \rangle^2 \rightarrow \langle 0, 1 \rangle$  je neklesající zprava spojitá funkce, pro kterou platí*

- $C(u_1, 0) = 0, C(0, u_2) = 0$ , pro všechna  $u_1, u_2 \in \langle 0, 1 \rangle$ ,
- $C(u_1, 1) = u_1, C(1, u_2) = u_2$ , pro všechna  $u_1, u_2 \in \langle 0, 1 \rangle$ ,
- a  $\forall u_1, u_2, v_1, v_2 \in \langle 0, 1 \rangle, u_1 \leq u_2, v_1 \leq v_2$  platí

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Z definice vzhledem k posledním dvěma vlastnostem můžeme vidět, že kopula je vlastně sdružená distribuční funkce náhodného vektoru, jejíž marginální rozdělení jsou rovnoměrně rozdělena na intervalu  $\langle 0, 1 \rangle$ . V obecném případě můžeme poté  $d$ -rozměrnou kopulu definovat zjednodušeně jako sdruženou distribuční funkci náhodného vektoru  $(U_1, \dots, U_d)$ , kde náhodná veličina  $U_i$  pochází z rovnoměrného rozdělení na intervalu  $\langle 0, 1 \rangle$ ,  $i = 1, \dots, d$ , tj.

$$C(u_1, \dots, u_d) = P(U_1 \leq u_1, \dots, U_d \leq u_d), \quad u_1, \dots, u_d \in \langle 0, 1 \rangle.$$

Kopula nám určitým způsobem spojuje sdruženou distribuční funkci s jedno-rozměrnými marginálními distribučními funkcemi, což nám ukazuje následující věta, kterou publikoval Abe Sklar v roce 1959 [9].

**Věta 2.1** *(Sklarova věta) Nechť  $F$  je dvourozměrná sdružená distribuční funkce se spojitými marginálními distribučními funkcemi  $F_1, F_2$ . Potom existuje jediná kopula  $C$  taková, že*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \quad x_1, x_2 \in R.$$

*Obráceně, je-li  $C$  kopula a  $F_1, F_2$  jsou spojitě distribuční funkce, potom*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \quad x_1, x_2 \in R$$

*je sdružená distribuční funkce s marginálními distribučními funkcemi  $F_1, F_2$ .*

**Myšlenka důkazu ([6]).** Necht'  $(X_1, X_2)$  má spojité marginální distribuční funkce, potom

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) =$$

$$P(F_1(X_1) \leq F_1(x_1), F_2(X_2) \leq F_2(x_2)) = C(F_1(x_1), F_2(x_2)), \quad x_1, x_2 \in R,$$

kde  $C$  je sdružená distribuční funkce náhodného vektoru  $(F_1(X_1), F_2(X_2))$ , kde  $F_i(X_i)$  jsou rovnoměrně rozdělené na intervalu  $\langle 0, 1 \rangle$ ,  $i = 1, 2$ .

Sklarova věta je pokládána za hlavní větu teorie kopulí. Ukazuje nám, že každou sdruženou distribuční funkci můžeme zapsat pomocí kopuly a marginálních distribučních funkcí vzorcem  $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ ,  $x_1, x_2 \in R$ . A také naopak, že každou kopulu je možné vyjádřit pomocí distribučních funkcí jako

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), \quad u_1, u_2 \in \langle 0, 1 \rangle,$$

přičemž toto vyjádření představuje návod, jak lze konstruovat sdružená rozdělení s libovolnými marginálními distribučními funkcemi.

**Poznámka 2.1** *Je důležité si uvědomit, že Sklarova věta platí i za předpokladu diskrétních marginálních distribučních funkcí  $F_1, F_2$ . V takovém případě je kopula  $C$  určena jednoznačně pro každé  $x_1, x_2$  z oboru hodnot marginálních distribučních funkcí, tj. pro  $\forall x_1, x_2 \in (\text{Range}(F_1) \times \text{Range}(F_2))$ . Dále budeme v práci pracovat pouze se spojitým případem.*

### Vlastnosti kopulí obecně pro $d$ -rozměrné kopuly

- *Propojení marginálních rozdělení pomocí kopulí:* V rámci této vlastnosti jsou důležité dva explicitní vzorce. Prvním je vzorec

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

který nám obecně znázorňuje, jak daná marginální rozdělení můžeme propojit do sdruženého rozdělení pravděpodobnosti pomocí dané kopuly. Druhým je vzorec

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)),$$

který nám obecně znázorňuje, jak můžeme z daného sdruženého rozdělení náhodného vektoru  $(X_1, \dots, X_d)$  při spojitých marginálních rozdělení extrahovat kopulu. Jelikož je tato kopula určena jednoznačně, lze na základě toho použít pro její označení pojmenování *kopula vícerozměrného pravděpodobnostního rozdělení* či *kopula náhodného vektoru*.

- *Invariance kopulí vůči rostoucím transformacím:* V případě, že náhodný vektor  $(X_1, \dots, X_d)$  se spojitými marginálními rozděleními má kopulu  $C$ , pak platí, že  $C$  je také kopula náhodného vektoru  $(T_1(X_1), \dots, T_d(X_d))$ , pro libovolné rostoucí transformace  $T_1(\cdot), \dots, T_d(\cdot)$ .
- *Fréchetovy meze:* Jako pro každou vícerozměrnou distribuční funkci existují Fréchetovy meze, které ji ohraničují zdola i shora, tak i pro libovolnou kopulu  $C(u_1, \dots, u_d)$  existují Fréchetovy meze, které jsou definovány vztahem:

$$\max\left(\sum_{i=1}^d u_i + 1 - d, 0\right) \leq C(u_1, \dots, u_d) \leq \min(u_1, \dots, u_d). \quad (1)$$

- *Hustota pravděpodobnosti kopulí:* V některých případech může být užitečné mít k dispozici pravděpodobnostní hustotu kopuly (např. pro získání maximálně věrohodných odhadů). Za předpokladu, že dané vícerozměrné rozdělení má sdruženou hustotu  $f(x_1, \dots, x_d)$  a rostoucí marginální distribuční funkce  $F_1(x_1), \dots, F_d(x_d)$  s hustotami  $f_1(x_1), \dots, f_d(x_d)$ , tak poté příslušná kopula  $C(u_1, \dots, u_d)$  má hustotu pravděpodobnosti:

$$c(u_1, \dots, u_d) = \frac{\partial C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d} = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdot \dots \cdot f_d(F_d^{-1}(u_d))}.$$

Jelikož kopula plně vysvětluje závislostní strukturu mezi náhodnými veličinami, a to bez ohledu na jejich marginálním rozdělení, nazývá se někdy jako *závislostní funkce*. Na základě toho můžeme využít *míry závislosti*, které jsou na kopulích založeny.



### Kendallův koeficient (Kendalovo tau)

Jedná se o míru závislosti, která je založena na výskytu konkordancí a diskordancí mezi náhodnými veličinami  $X_1$  a  $X_2$ .

**Definice 2.2** *Dvě pozorování  $(x_1, x_2)$ ,  $(\tilde{x}_1, \tilde{x}_2)$  z náhodného vektoru  $(X_1, X_2)$  nazveme konkordantní, jestliže platí  $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0$ . Platí-li opačná nerovnost  $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0$ , tak tato dvě pozorování nazveme diskordantní.*

V případě náhodných veličin se pojmy konkordantní (diskordantní) používají ve smyslu, že velké (malé) hodnoty jedné proměnné se zpravidla objevují s velkými (malými) hodnotami druhé proměnné.

*Kendallův koeficient* je definován jako rozdíl pravděpodobnosti konkordance a pravděpodobnosti diskordance:

$$\tau(X_1, X_2) = P[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0] - P[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0],$$

kde  $(\tilde{X}_1, \tilde{X}_2)$  je náhodný vektor se stejným rozdělením jako  $(X_1, X_2)$ , ale nezávislý na  $(X_1, X_2)$ .

Kendallův koeficient patří mezi míry závislost, které nezávisí na marginálním rozdělení příslušných náhodných veličin, ale pouze na jejich kopuly. Nechť  $X_1$  a  $X_2$  jsou náhodné veličiny, které mají spojitá marginální rozdělení, tedy jednoznačně určenou kopulu  $C$ , pak je Kendallův koeficient dán následujícím vztahem:

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1. \quad (2)$$

Mezi další vlastnosti tohoto koeficientu patří:

- symetrie, tj.  $\tau(X_1, X_2) = \tau(X_2, X_1)$ ,
- $-1 \leq \tau(X_1, X_2) \leq 1$ ,
- $\tau(X_1, X_2) = 1$  právě tehdy, když náhodné veličiny  $X_1$  a  $X_2$  jsou komonotónní (tj. pokud se náhodná veličina  $X_1$  rovná  $T(X_2)$ , kde  $T$  je rostoucí funkce),

- $\tau(X_1, X_2) = -1$  právě tehdy, když náhodné veličiny  $X_1$  a  $X_2$  jsou kontramontónní (tj. pokud se náhodná veličina  $X_1$  rovná  $T(X_2)$ , kde  $T$  je klesající funkce),
- pokud jsou náhodné veličiny  $X_1$  a  $X_2$  nezávislé, pak  $\tau(X_1, X_2) = 0$ ,
- $\tau$  je invariantní vůči striktně rostoucím transformacím.

### Spearmanovo rho (Spearmanův koeficient)

Stejně jako Kendallův koeficient lze i Spearmanův koeficient vyjádřit pomocí výskytu konkordancí a diskordancí mezi náhodnými veličinami  $X_1$  a  $X_2$ :

$$\rho_s(X_1, X_2) = 3 \left[ P((X_1 - \tilde{X}_1)(X_2 - X_2^*) > 0) - P((X_1 - \tilde{X}_1)(X_2 - X_2^*) < 0) \right],$$

kde  $(X_1^*, X_2^*)$  je náhodný vektor se stejným rozdělením jako  $(X_1, X_2)$ , ale nezávislý na  $(X_1, X_2)$  a  $(\tilde{X}_1, \tilde{X}_2)$ .

Také můžeme tento koeficient vyjádřit pomocí lineárního korelačního koeficientu a distribučních funkcí  $F_1, F_2$  jako

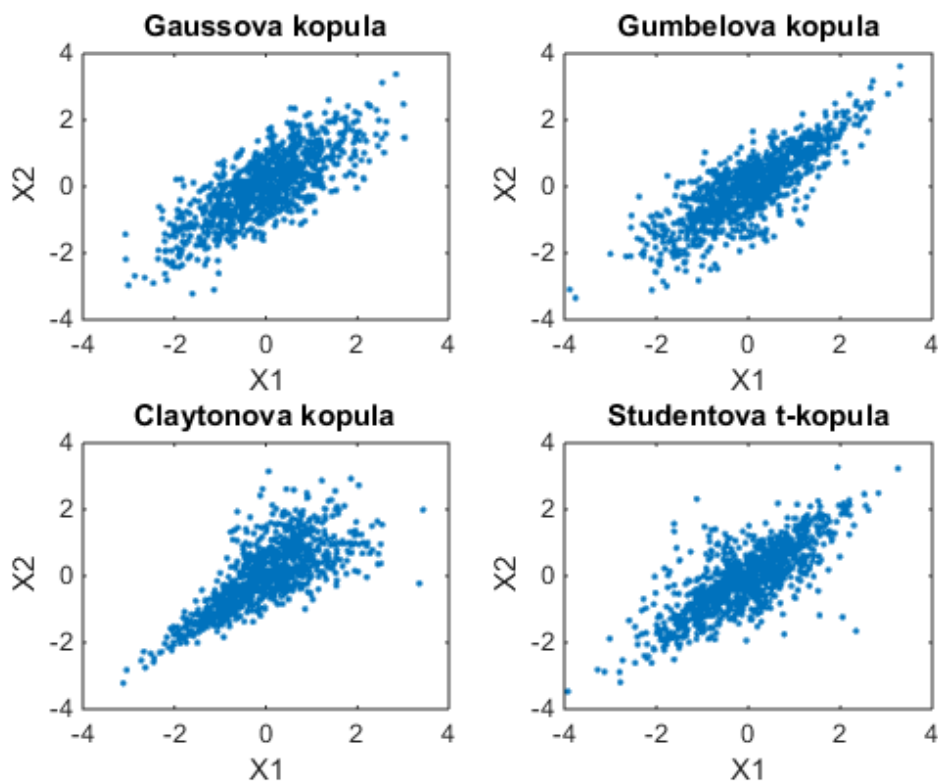
$$\rho_s(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)).$$

Spearmanův koeficient má stejné vlastnosti jako Kendallův koeficient. Nyní však tento koeficient můžeme vyjádřit pomocí kopuly následně

$$\rho_s(X_1, X_2) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2.$$

### Koncová závislost neboli závislost chvostů

Jedná se o závislost mezi extrémními hodnotami náhodných veličin. Rozlišujeme *koeficienty horní a dolní závislosti*, které představují symetrické míry závislosti. Jelikož jsou tyto míry nezávislé na marginálních rozdělení příslušných náhodných veličin, můžeme je opět vyjádřit pomocí jejich kopuly. Pomocí těchto koeficientů můžeme měřit sílu závislosti na horním nebo dolním konci dvourozměrných rozdělení, tedy například na obrázku 4 v horním pravém nebo dolním levém rohu jednotlivých obrázků kopulí.



Obrázek 4: Příklady jednotlivých typů kopulí, které mají různé závislosti chvostů

*Koeficient horní koncové závislosti*  $\lambda_U$  náhodných veličin  $X_1, X_2$  je definován jako

$$\lambda_U(X_1, X_2) = \lim_{u \rightarrow 1^-} P(X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$

a *koeficient dolní koncové závislosti*  $\lambda_L$  náhodných veličin  $X_1, X_2$  je definován jako

$$\lambda_L(X_1, X_2) = \lim_{u \rightarrow 0^+} P(X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u},$$

pokud existují uvedené limity.

Poslední rovnosti v těchto vzorcích jsou platné v případě spojitých marginálních distribučních funkcí, kdy je jednoznačně dána kopula  $C$  náhodných veličin  $X_1, X_2$ .

Je-li  $\lambda_U(X_1, X_2) \in (0, 1)$ , pak řekneme, že náhodné veličiny  $X_1, X_2$  vykazují závislost horních chvostů. Naopak, je-li  $\lambda_U(X_1, X_2) = 0$ , tak jsou náhodné veličiny  $X_1, X_2$  asymptoticky nezávislé v horních chvostech.

Analogicky je-li  $\lambda_L(X_1, X_2) \in (0, 1)$ , pak řekneme, že náhodné veličiny  $X_1, X_2$  vykazují závislost dolních chvostů. Naopak, je-li  $\lambda_L(X_1, X_2) = 0$ , tak jsou náhodné veličiny  $X_1, X_2$  asymptoticky nezávislé v dolních chvostech.

Nyní se podíváme na kopuly různého typu, které je možné aplikovat v praxi. Základní dělení kopulí je na *elementární kopuly*, *implicitní neboli také eliptické kopuly* a *Archimédovské kopuly*.

### Elementární kopuly

Elementární kopuly nám představují zvláštní druh závislostí náhodných veličin. Mezi základní reprezentanty elementárních kopulí můžeme zařadit tyto kopuly:

NEZÁVISLOSTNÍ KOPULA  $C_I$

Tato kopula má jednoduchý tvar:

$$C_I(u_1, \dots, u_d) = u_1 \cdot \dots \cdot u_d.$$

V případě, že ji zapíšeme ve tvaru, v jakém se obvykle používá

$$C_I(F(x_1), \dots, F(x_d)) = F(x_1) \cdot \dots \cdot F(x_d), \quad (3)$$

je její název vzhledem ke Sklarově větě logický. Podle [2] tedy platí, že spojitě složky náhodného vektoru jsou nezávislé právě tehdy, když kopula tohoto náhodného vektoru je nezávislostní kopula (3).

KOMONOTÓNÍ KOPULA  $C_{min}$

Tato kopula je také někdy označována jako *minimální kopula* a má tvar

$$C_{min}(u_1, \dots, u_d) = \min(u_1 \cdot \dots \cdot u_d)$$

či

$$C_{min}(F(x_1), \dots, F(x_d)) = \min(F(x_1), \dots, F(x_d)).$$

Na rozdíl od nezávislostní kopuly nám komonotónní kopula naopak představuje *perfektní pozitivní závislost*, a to ve smyslu, že můžeme například složky  $X_2, \dots, X_d$  náhodného vektoru  $X_1, \dots, X_d$  s touto kopulou vyjádřit jako rostoucí transformace první složky  $X_1$ , tj.  $X_i = T_i(X_1)$ ,  $i = 2, \dots, d$ . Samozřejmě můžeme místo první složky zvolit jinou libovolnou složku tohoto náhodného vektoru.

Komonotónní kopula v každém případě představuje horní Fréchetovu mez (vztah (1)).

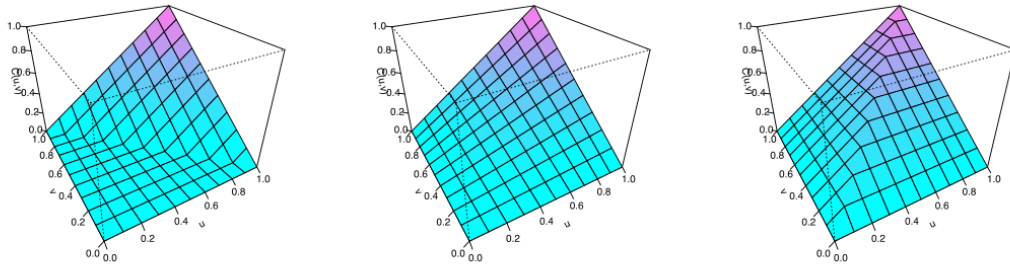
#### KONTRAMONOTÓNŇÍ KOPULA $C_{max}$

Tuto kopulu můžeme také pojmenovat jako *maximální*. Zavádí se pouze pro hodnotu  $d = 2$ , tedy dvourozměrný případ, kdy představuje dolní Fréchetovu mez (1). Její tvar je

$$C_{max}(u_1, u_2) = \max(u_1 + u_2 - 1, 0).$$

Kontramonotónní kopula představuje *perfektní negativní závislost*, a to ve smyslu, že např. náhodná veličina  $X_2$  vzniká klesající transformací náhodné veličiny  $X_1$ .

Důvod, proč se kontramonotónní kopula zavádí pouze pro dvourozměrný případ je ten, že pro dimenzi  $d > 2$  nemusí být splněn požadavek, že kopula je jako distribuční funkce  $d$ -rostoucí. Tuto vlastnost musí splňovat každá kopula, a proto tedy pro tuto dimenzi není dolní Fréchetova mez obecně kopulou.



Obrázek 5: Dvourozměrné elementární kopuly - Dolní Fréchetova mez, Nezávislostní kopula, Horní Fréchetova mez (obrázek převzatý z [12])

## Eliptické kopuly

Tyto kopuly jsou také nazývány jako implicitní kopuly, jelikož je nelze vyjádřit explicitně. Mezi eliptické kopuly patří:

### GAUSSOVA (NORMÁLNÍ) KOPULA

Ve dvourozměrném případě je Gaussova kopula definována vztahem:

$$C_\rho(u_1, u_2) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \quad u_1, u_2 \in \langle 0, 1 \rangle,$$

kde  $\Phi$  představuje distribuční funkci standardizovaného normálního rozdělení  $N(0, 1)$ , která má tvar:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in R$$

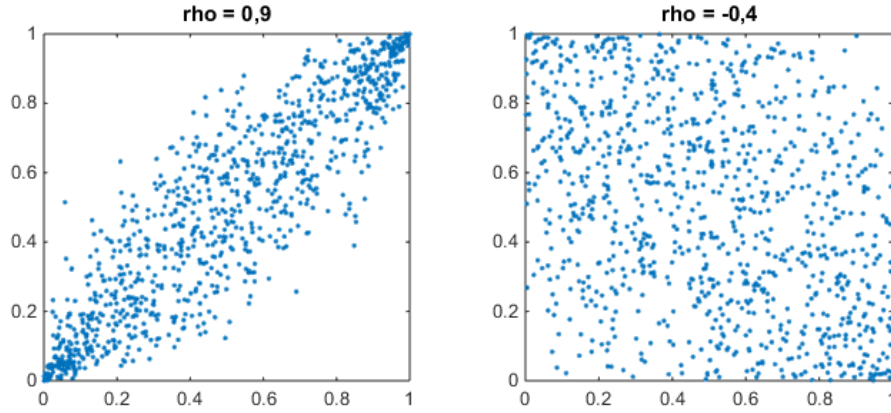
a  $\Phi_\rho$  představuje distribuční funkci dvourozměrného standardizovaného normálního rozdělení  $N(0, 1)$  s korelačním koeficientem  $\rho$ , který znázorňuje korelaci daných náhodných veličin. Odtud

$$\begin{aligned} C_\rho(F_1(x_1), F_2(x_2)) &= \Phi_\rho(\Phi^{-1}(F_1(x_1)), \Phi^{-1}(F_2(x_2))) = \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(F_1(x_1))} \int_{-\infty}^{\Phi^{-1}(F_2(x_2))} \exp\left\{-\frac{1}{2(1-\rho^2)}(z_1^2 + z_2^2 - 2\rho z_1 z_2)\right\} dz_1 dz_2. \end{aligned}$$

V tomto případě je tedy kopula závislá pouze na jediném parametru  $\rho$ , který nese veškerou informaci o závislostní struktuře. Díky integrálům ji však nelze vyjádřit v explicitním tvaru. Jelikož tato kopula generuje pro veličiny se standardizovaným normálním rozdělením sdruženou standardizovanou normální distribuční funkci, je proto označována také jako normální kopula. Jestliže

- $\rho = -1 \dots$  Gaussova kopula je kontramonotónní (maximální),
- $\rho = 0 \dots$  Gaussova kopula je nezávislostní,
- $\rho = 1 \dots$  Gaussova kopula je komonotónní (minimální).

Dvourozměrnou Gaussovou kopulu můžeme tedy chápat jako závislostní strukturu, která představuje interpolaci mezi perfektní negativní závislostí a perfektní pozitivní závislostí. Síla této závislosti je poté dána korelačním koeficientem  $\rho$ . Vliv korelačního koeficientu v případě Gaussovy kopuly můžeme pozorovat na obrázku 6.



Obrázek 6: 1 000 náhodných výběrů z Gaussovy kopuly s parametry  $\rho = 0,9$ ,  $\rho = -0,4$

#### STUDENTOVA $t$ -KOPULA

Ve dvourozměrném případě vypadá Studentova  $t$ -kopula následovně:

$$C_{\nu,\rho}(u_1, u_2) = t_{\nu,\rho}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)), \quad u_1, u_2 \in \langle 0, 1 \rangle,$$

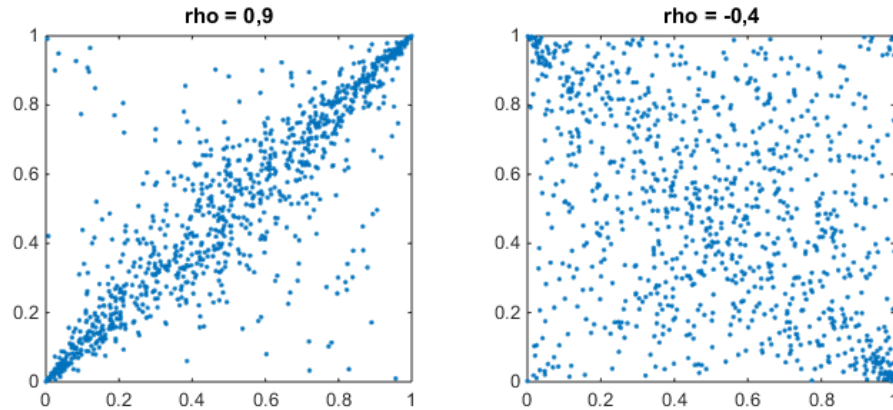
kde  $t_\nu$  představuje distribuční funkci Studentova  $t$ -rozdělení, která má tvar:

$$t_\nu(x) = \int_{-\infty}^x \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\nu/2)} \left(1 + \frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}} ds$$

a  $t_{\nu,\rho}$  představuje distribuční funkci dvourozměrného Studentova  $t$ -rozdělení s korelačním koeficientem  $\rho$ . Odtud

$$\begin{aligned} C_{\nu,\rho}(F_1(x_1), F_2(x_2)) &= t_{\nu,\rho}(t_\nu^{-1}(F_1(x_1)), t_\nu^{-1}(F_2(x_2))) = \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_\nu^{-1}(F_1(x_1))} \int_{-\infty}^{t_\nu^{-1}(F_2(x_2))} \left[1 + \frac{z_1^2 + z_2^2 - 2\rho z_1 z_2}{\nu(1-\rho^2)}\right]^{-\frac{\nu+1}{2}} dz_1 dz_2. \end{aligned}$$

Nyní vidíme, že je kopula závislá na dvou parametrech, a to na parametru  $\rho$ , který odpovídá korelačnímu koeficientu, a také na parametru  $\nu$ , který představuje počet stupňů volnosti Studentova  $t$ -rozdělení. Vliv korelačního koeficientu v případě Studentovy  $t$ -kopuly můžeme pozorovat na obrázku 7.



Obrázek 7: 1 000 náhodných výběrů ze Studentovy  $t$ -kopuly s parametry  $\rho = 0,9$ ,  $\rho = -0,4$  a  $\nu = 2$

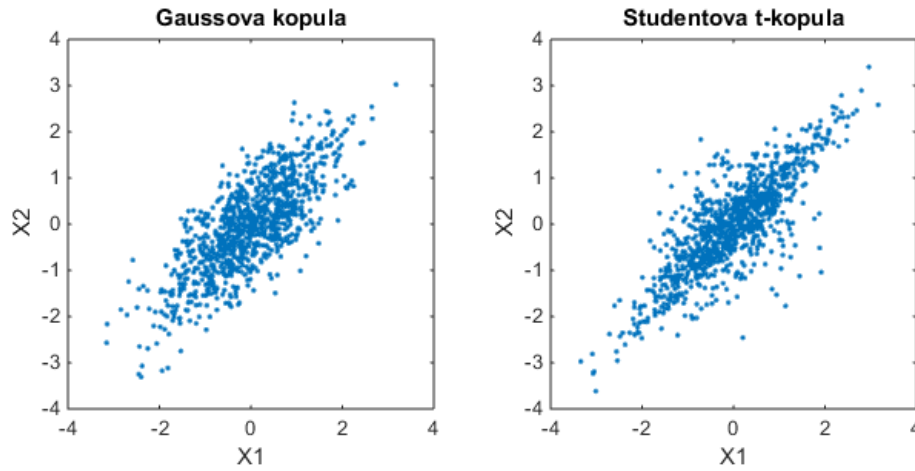
Pro Studentovu  $t$ -kopulu platí také následující vlastnosti:

- $\lim_{\rho \rightarrow -1} C_{\nu, \rho} = C_{max}$ ,
- $\lim_{\rho \rightarrow 1} C_{\nu, \rho} = C_{min}$ ,
- pro žádné  $\nu < \infty$  se nerovná kopula  $C_{\nu, 0}$  nezávislostní kopuly  $C_I$ .

Při srovnání simulací pro Gaussovu kopulu a Studentovu  $t$ -kopulu můžeme na obrázku 8 vidět, že pro vysoké hodnoty náhodných veličin  $X_1$  a  $X_2$  nám Gaussova kopula vykazuje *nezávislost v obou koncích*. Naopak v případě Studentovy  $t$ -kopuly vidíme *závislost v obou koncích*, kdy se nasimulovaná data přibližují k ose grafu.

Nechť  $X_1, X_2$  jsou rizikové faktory, které mají distribuční funkce  $F_1, F_2$ , poté postup při simulaci Monte Carlo v softwaru Matlab probíhá následovně. Nejdříve pomocí příkazu `copularnd` vygenerujeme hodnoty z Gaussovy kopuly či Studentovy  $t$ -kopuly, které představují spárované dvojice  $(U_{11}, U_{12}), \dots, (U_{n1}, U_{n2})$ .





Obrázek 8: Srovnání simulací pro Gaussovu kopulu a Studentovu  $t$ -kopulu s parametrem  $\rho = 0,8$

Následně si spočítáme hodnoty  $(F_1^{-1}(U_{11}), F_2^{-1}(U_{12})), \dots, (F_1^{-1}(U_{n1}), F_2^{-1}(U_{n2}))$ , které již vstupují do jednotlivých scénářů.

### Archimédovské kopuly

Archimédovské kopuly představují velmi důležitou třídu kopulí. Aplikace těchto kopulí je v praxi hodně častá, a to díky své relativně snadné konstrukci. Jedná se totiž o kopuly, které můžeme vyjádřit pomocí speciální jednorozměrné funkce, tzv. generátoru.

**Definice 2.3** *Nechť  $\varphi : \langle 0, 1 \rangle \rightarrow \langle 0, \infty \rangle$  je spojitá klesající konvexní funkce taková, že  $\varphi(1) = 0$ . Nechť*

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

*Funkce*

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)), \quad u_1, u_2 \in \langle 0, 1 \rangle$$

*se nazývá Archimédovská kopula a funkce  $\varphi$  (aditivní) generátor kopuly  $C$ .*

**Poznámka 2.2** *Symbol  $\varphi^{[-1]}$  značí pseudoinverzi  $\varphi$ .*

Pokud  $\varphi(0) = \infty$ , pak mluvíme o tzv. *striktním* generátoru kopuly. V tomto případě poté platí, že  $\varphi^{[-1]} = \varphi^{-1}$  a následně dostáváme *striktní Archimédovskou kopulu*, která má tvar

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2)), \quad u_1, u_2 \in \langle 0, 1 \rangle.$$

### Vlastnosti Archimédovských kopulí

Mezi základní vlastnosti Archimédovské kopuly  $C$  patří:

- symetrie, tj.  $C(u_1, u_2) = C(u_2, u_1)$ ,  $u_1, u_2 \in \langle 0, 1 \rangle$ ,
- asociativita, tj.  $C(C(u_1, u_2), u_3) = C(u_1, C(u_2, u_3))$ ,  $u_1, u_2, u_3 \in \langle 0, 1 \rangle$ ,
- je-li  $\varphi$  generátor  $C$  a  $k > 0$  libovolná konstanta, potom  $k\varphi$  je také generátor  $C$ .

Na základě volby generátoru rozlišujeme různé třídy Archimédovských kopulí. V tabulce 3 jsou uvedeny příklady nejznámějších jednoparametrických Archimédovských tříd spolu s jejich generátory  $\varphi_\alpha(t)$ .

Třída	$\varphi_\alpha(t)$	$\alpha \in$	$C_\alpha(u_1, u_2)$
<b>Clayton</b>	$\frac{1}{\alpha}(t^{-\alpha} - 1)$	$\langle -1, \infty \rangle \setminus \{0\}$	$(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}}$
<b>Gumbel</b>	$(-\ln t)^\alpha$	$\langle 1, \infty \rangle$	$\exp \left\{ - [(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{\frac{1}{\alpha}} \right\}$
<b>Frank</b>	$-\ln \frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}$	$(-\infty, \infty) \setminus \{0\}$	$-\frac{1}{\alpha} \ln \left( 1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1} \right)$

Tabulka 3: Příklady Archimédovských kopulí

Nyní se podíváme na vztah mezi Kendallovým koeficientem a Archimédovskými kopulemi. V obecném případě bylo potřeba k vyjádření Kendallova koeficientu určit hodnotu dvojného integrálu (viz vztah (2)). V případě Archimédovských kopulí je situace jednodušší, jelikož Kendallův koeficient můžeme vyjádřit přímo z generátoru kopuly  $\varphi$  jako

$$\tau = 1 + 4 \int_0^1 \frac{\varphi_\alpha(t)}{\varphi'_\alpha(t)} dt.$$

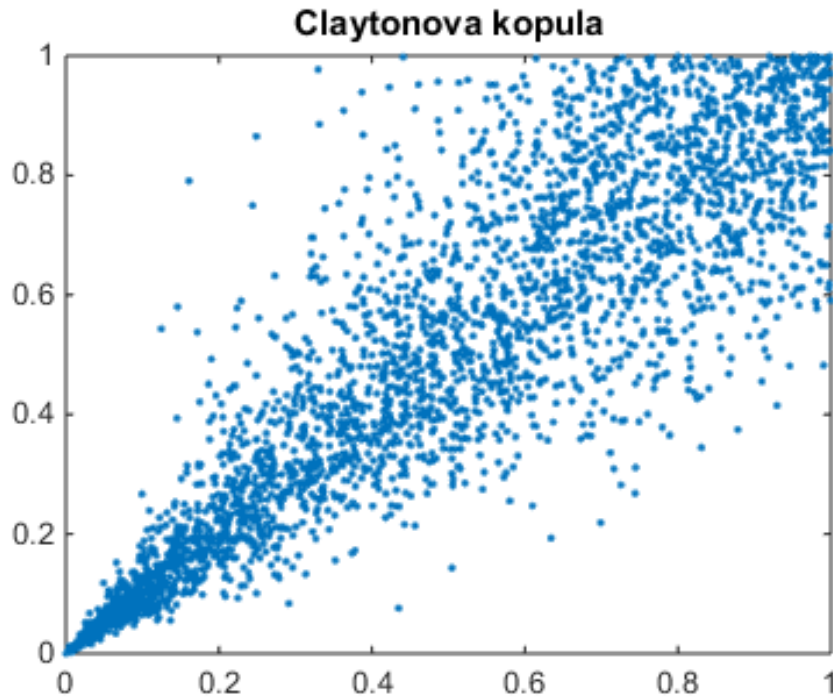
## CLAYTONOVA KOPULA

Claytonova kopula je asymetrickou Archimédovskou kopulí projevující větší závislost v *dolních záporných* koncích než v *horních kladných* koncích (obrázek 9), kde koeficient dolní koncové závislosti je  $\lambda_L = 2^{-\frac{1}{\alpha}}$ . Ve dvourozměrném případě je definovaná jako

$$C_\alpha(u_1, u_2) = \left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{-\frac{1}{\alpha}}$$

a její generátor je

$$\varphi_\alpha(t) = \frac{1}{\alpha}(t^{-\alpha} - 1).$$



Obrázek 9: 3 000 náhodných výběrů z Claytonovy kopuly s parametrem  $\alpha = 5$

Pro  $\alpha \rightarrow 0$  přechází Claytonova kopula na nezávislostní kopulu a pro  $\alpha \rightarrow \infty$  přechází na komonotónní kopulu (minimální). Můžeme ji tedy považovat za závislostní strukturu, která představuje interpolaci mezi nezávislostí a perfektní pozitivní závislostí. Sílu této závislosti pak měří parametr  $\alpha$  (viz tabulka 3).

Poté speciálně pro  $\alpha = -1$  se jedná o kontramonotónní kopulu (maximální).

Naopak v případě  $\alpha < 0$  Claytonova kopula neprojevuje závislost ani v dolních záporných koncích.

Vztah mezi Kendallovým koeficientem a parametrem  $\alpha$  Claytonovy kopuly je dán vzorcem

$$\tau = \frac{\alpha}{\alpha + 2}.$$

#### GUMBELOVA KOPULA

Gumbelova kopula je asymetrickou Archimédovskou kopulí projevující větší závislost *v horních kladných koncích než v dolních záporných koncích* (obrázek 10), kde koeficient horní koncové závislosti je  $\lambda_U = 2 - 2^{\frac{1}{\alpha}}$ . Ve dvourozměrném případě je definovaná jako

$$C_\alpha(u_1, u_2) = \exp \left\{ - [(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{\frac{1}{\alpha}} \right\}$$

a její generátor je

$$\varphi_\alpha(t) = (-\ln t)^\alpha.$$

Pro  $\alpha = 1$  přechází Gumbelova kopula na nezávislostní kopulu a pro  $\alpha \rightarrow \infty$  přechází na komonotónní kopulu (minimální). Můžeme ji tedy považovat za závislostní strukturu, která představuje interpolaci mezi nezávislostí a perfektní pozitivní závislostí. Sílu této závislosti opět měří parametr  $\alpha$  (viz tabulka 3).

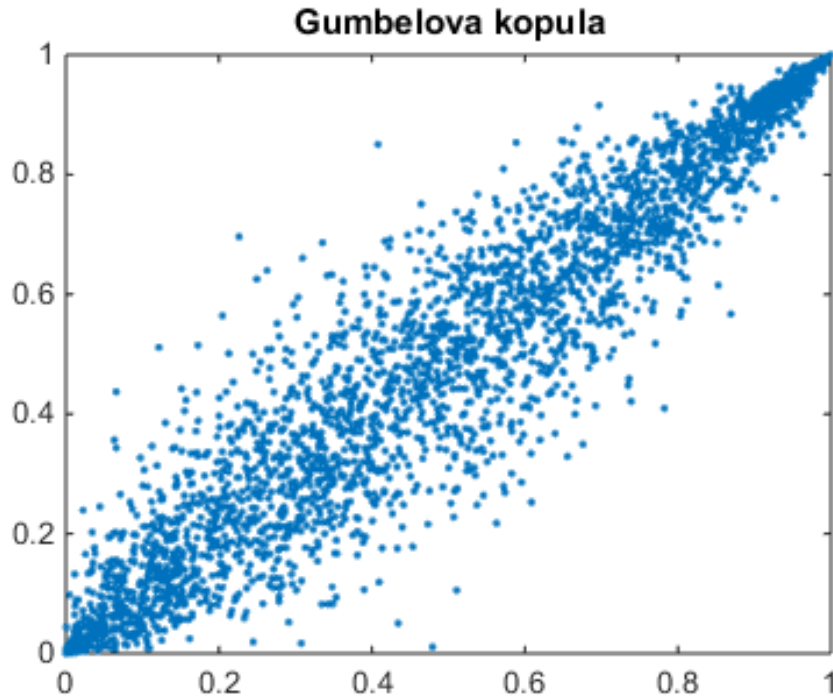
Vztah mezi Kendallovým koeficientem a parametrem  $\alpha$  Gumbelovy kopuly je dán vzorcem

$$\tau = 1 - \frac{1}{\alpha}.$$

#### FRANKOVA KOPULA

Frankova kopula na rozdíl od předchozích dvou typů kopulí představuje symetrickou Archimédovskou kopulu (obrázek 11). Ve dvourozměrném případě je definovaná jako

$$C_\alpha(u_1, u_2) = -\frac{1}{\alpha} \ln \left( 1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1} \right)$$



Obrázek 10: 3 000 náhodných výběrů z Gumbelovy kopuly s parametrem  $\alpha = 5$

a její generátor je

$$\varphi_{\alpha}(t) = -\ln \frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}.$$

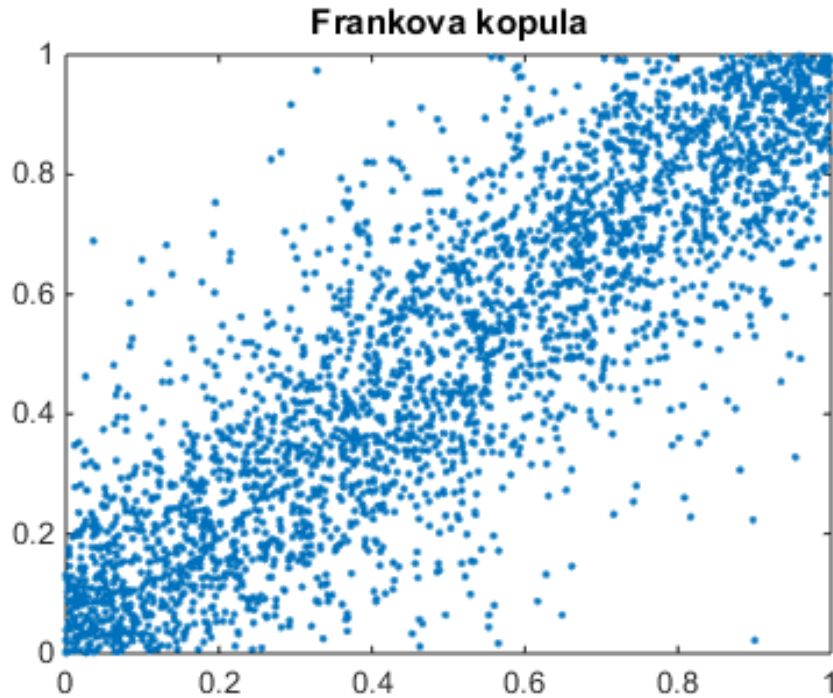
Pro  $\alpha \rightarrow \infty$  přechází Frankova kopula na komonotónní kopulu (minimální) a pro  $\alpha \rightarrow -\infty$  přechází na kontramonotónní kopulu (maximální). Tato kopula neprojevuje závislost ani v horních ani v dolních koncích.

Vztah mezi Kendallovým koeficientem a parametrem  $\alpha$  Frankovy kopuly je dán vzorcem

$$\tau = 1 + \frac{4(D_1(\alpha) - 1)}{\alpha},$$

kde  $D_1(\alpha)$  je tzv. Debye funkce, která je definovaná

$$D_k(\alpha) = \frac{k}{\alpha^k} \int_0^{\alpha} \frac{t^k}{e^t - 1} dt.$$



Obrázek 11: 3 000 náhodných výběrů z Frankovy kopuly s parametrem  $\alpha = 10$

Nechť  $X_1, X_2$  jsou rizikové faktory, které mají distribuční funkce  $F_1, F_2$ , poté postup při simulaci Monte Carlo s využitím Archimédovských kopulí podle [7] probíhá následovně:

- Nejdříve se vygenerují hodnoty  $U_1, U_2$ , které představují nezávislá náhodná čísla z rovnoměrného rozdělení pravděpodobnosti na intervalu  $(0, 1)$ .
- V dalším kroku se položí  $X_1 = F_1^{-1}(U_1)$ .
- Nakonec se  $X_2$  spočítá jako řešení

$$U_2 = F_2(X_2|X_1) = \frac{\varphi^{-1(1)}(\varphi(F_1(X_1)) + \varphi(F_2(X_2)))}{\varphi^{-1(1)}(\varphi(F_1(X_1)))}.$$

Symbol  $\varphi^{-1(1)}$  značí první derivaci inverzní funkce generátoru  $\varphi$ .

Pokud máme software, který je schopen generovat hodnoty kopulí, jako např. Matlab, tak se nám postup zjednoduší. Nejdříve zadefinujeme příslušná roz-

dělení pravděpodobnosti rizikových faktorů  $X_1, X_2$ . Dále pomocí příkazu *copularnd* vygenerujeme hodnoty z dané třídy Archimédovských kopulí, které představují spárované dvojice  $(U_{11}, U_{12}), \dots, (U_{n1}, U_{n2})$ . Nakonec si spočítáme hodnoty  $(F_1^{-1}(U_{11}), F_2^{-1}(U_{12})), \dots, (F_1^{-1}(U_{n1}), F_2^{-1}(U_{n2}))$ , které již vstupují do jednotlivých scénářů.

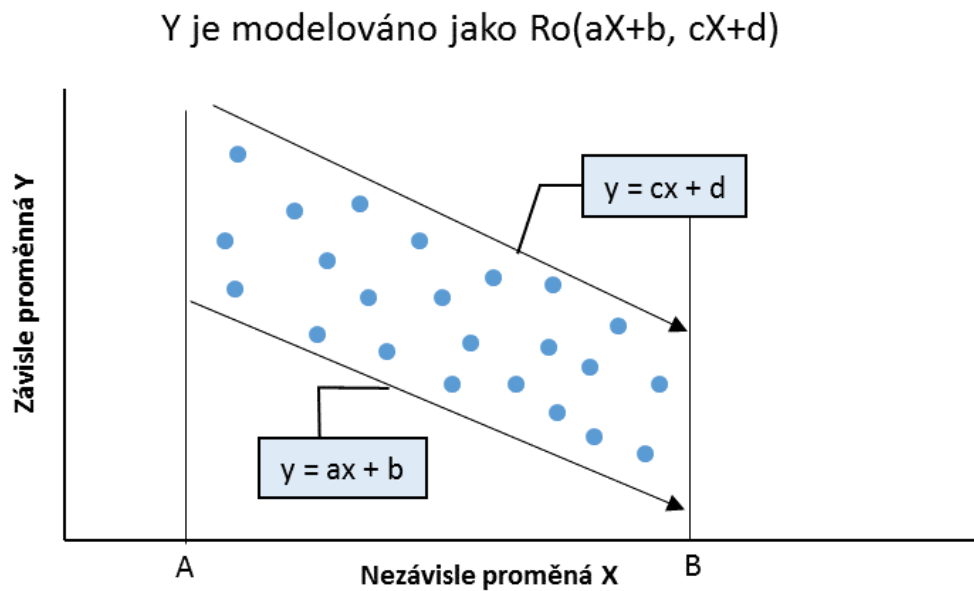
## 2.2. Obálková metoda

Obálková metoda, jejíž název pochází z anglického *Envelope method*, nabízí celkem značnou flexibilitu při modelování závislosti. Ve srovnání s pořadovou korelací nám jasně definuje kauzální směr. Princip této metody spočívá v tom, že nezávislá náhodná veličina (rizikový faktor)  $X$  statisticky určuje závislou náhodnou veličinu (rizikový faktor)  $Y$ , přičemž závislost mezi veličinami  $X$  a  $Y$  může nabývat různých podob. Její nevýhodou je dle [10], že vyžaduje mnohem větší úsilí než pořadové korelace, a proto se používá pouze tam, kde má závislostní vztah významný vliv na konečný výsledek modelu. Pro zpracování této části byla využita literatura [4], [5] a [10].

Pokud máme k dispozici data, tak v případě obálkové metody je nejjednodušší nejdříve vykreslení všech těchto dostupných dat do bodového grafu. Díky tomu získáme představu o formě závislosti v těchto datech. V bodovém grafu se nezávisle proměnná vykresluje na osu  $x$ , zatímco závisle proměnná se vykresluje na osu  $y$ . Posléze má teprve smysl zabývat se otázkou, jak danou závislost co nejlépe namodelovat. Velmi důležité je uvědomit si, že se nejedná o funkční závislost, jelikož data neleží přímo na přímce, ale jedná se o stochastickou (pravděpodobnostní) závislost.

### 2.2.1. Odvození parametrů rozdělení pravděpodobnosti rizikového faktoru $Y$ na základě realizace rizikového faktoru $X$

Nejdříve si určíme rozdělení pravděpodobnosti rizikového faktoru  $X$  a zároveň určíme, jak z hodnoty  $x$  rizikového faktoru  $X$  vypočítáme parametry pravděpodobnostního rozdělení rizikového faktoru  $Y$ . Postup si ukážeme na dvou přípa-



Obrázek 12: Použití rovnoměrného rozdělení při modelování závislostí mezi rizikovými faktory  $X$  a  $Y$

dech.

Nejjednodušší situací může být případ, kdy hodnoty rizikového faktoru  $X$  pocházející z rovnoměrného rozdělení s parametry  $A, B$  statisticky definují parametry  $ax + b, cx + d$  rovnoměrného rozdělení rizikového faktoru  $Y$ .

Na obrázku 12 máme znázorněnou situaci modelování závislostí mezi rizikovými faktory  $X$  a  $Y$ , a to pomocí rovnoměrného rozdělení.

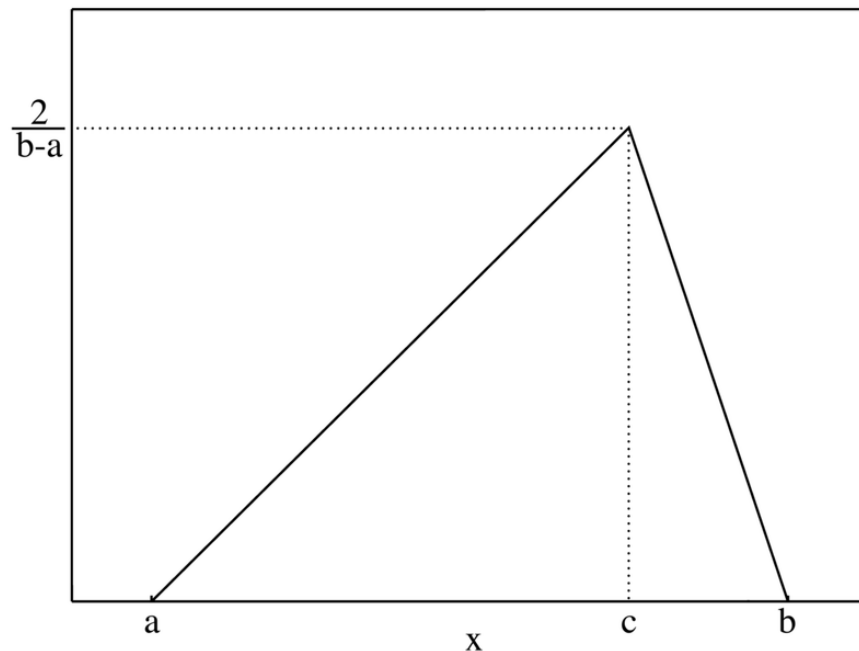
### Postup při simulaci je následující:

- Nejdříve vykreslíme data do bodového grafu, abychom viděli, jak jsou rozložena.
- Poté se budeme snažit tato data „obalit“ pomocí přímky shora a zdola, tzn. určíme odhadem minimální přímku popsanou rovnicí  $y = ax + b$  a maximální přímku popsanou rovnicí  $y = cx + d$  (viz obrázek 12).
- Nyní si vygenerujeme hodnoty rizikového faktoru  $X$  z rovnoměrného rozdělení pravděpodobnosti s parametry  $A$  a  $B$ .



- Nakonec tyto vygenerované hodnoty dosadíme do rovnic přímek, čímž získáme parametry pro modelování závislého rizikového faktoru  $Y$ , který budeme modelovat pomocí rovnoměrného rozdělení pravděpodobnosti, avšak nyní s parametry  $ax + b$  a  $cx + d$ .

Další možností pro modelování pravděpodobnostního rozdělení rizikových faktorů je využití trojúhelníkového rozdělení. Trojúhelníkové rozdělení má v grafickém zobrazení hustoty pravděpodobnosti tvar trojúhelníka se třemi parametry reprezentující nejpravděpodobnější hodnotou  $c$ , která představuje vrchol trojúhelníka a dále dolní mezí  $a$  a horní mezí  $b$ , kde pro hodnoty  $a, b, c$  platí, že  $a \leq c \leq b$ . Nemůže však nastat případ, kdy jsou si všechny tři parametry rovny.

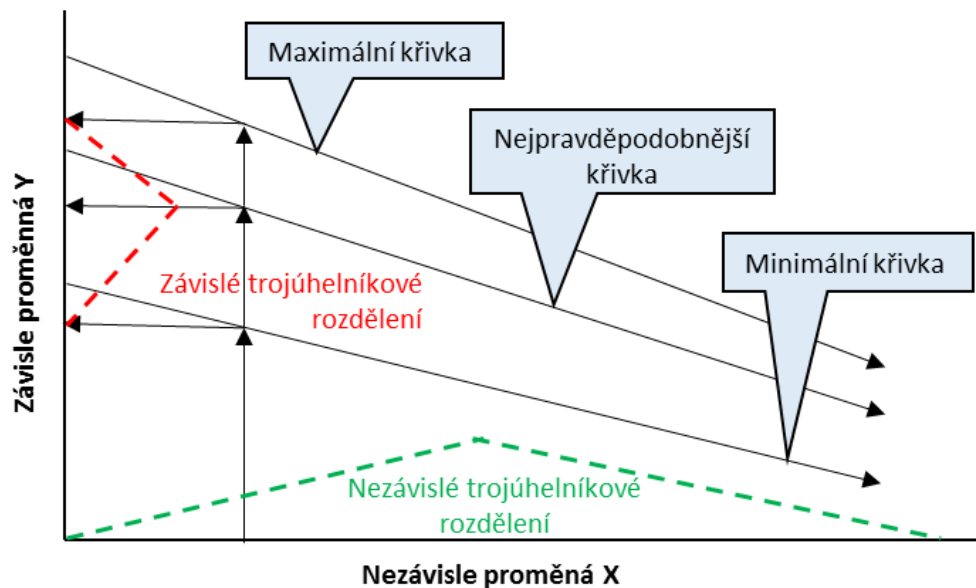


Obrázek 13: Hustota trojúhelníkového rozdělení (obrázek je převzatý z [13])

Funkce hustoty je ve tvaru (viz [10]):

$$f(x) = \begin{cases} 0 & x < a, \\ \frac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq c, \\ \frac{2(b-x)}{(b-a)(b-c)} & c < x \leq b, \\ 0 & b < x. \end{cases}$$

Nyní již potřebujeme kromě minimální a maximální křivky obalující data také křivku, která odpovídá vždy nejpravděpodobnější hodnotě závisle proměnné pro každou hodnotu nezávisle proměnné (obrázek 14). Je rozumné odhadnout tuto křivku tak, aby procházela místem největší hustoty dat. Alternativně můžeme také pro určení této křivky použít regresní přímku, kterou proložíme daty. Poslední možností je pak zvolit nepravděpodobnější křivku na půli cesty mezi minimální a maximální křivkou.



Obrázek 14: Použití trojúhelníkového rozdělení při modelování závislostí mezi rizikovými faktory  $X$  a  $Y$

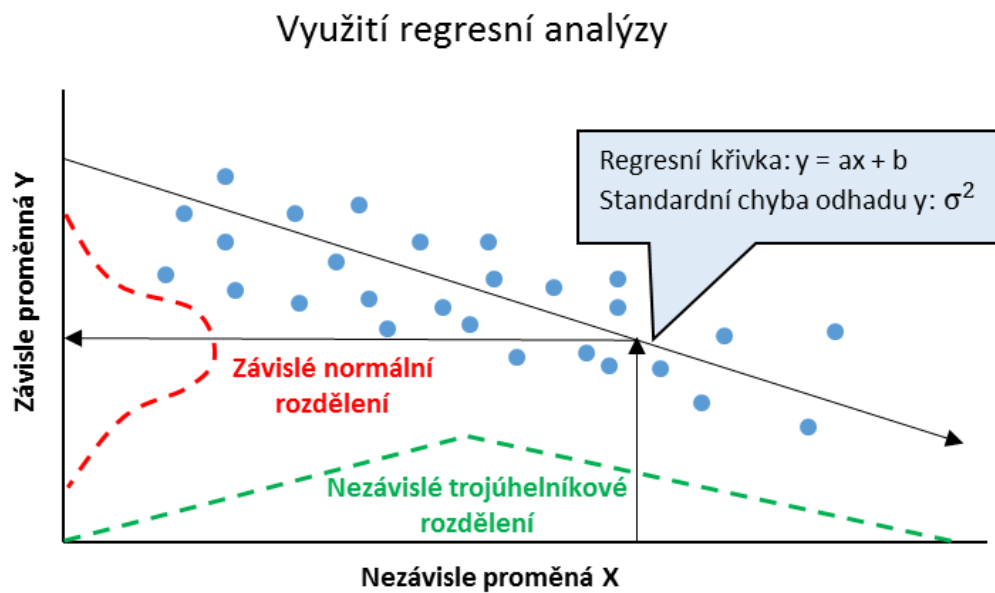
Postup při simulaci je stejný jako v případě rovnoměrného rozdělení, avšak s tou výjimkou, že nyní máme k dispozici 3 rovnice přímek a nezávislý a závislý rizikový faktor modelujeme pomocí trojúhelníkového rozdělení.

### 2.2.2. Modelování závislostí mezi rizikovými faktory $X$ a $Y$ pomocí regresní analýzy

Nejjednodušším případem je situace, kdy můžeme závislost mezi rizikovými faktory vyjádřit pomocí regresní přímky. Pro odhad neznámých parametrů této funkce se využívá metoda nejmenších čtverců, která minimalizuje součet čtverců

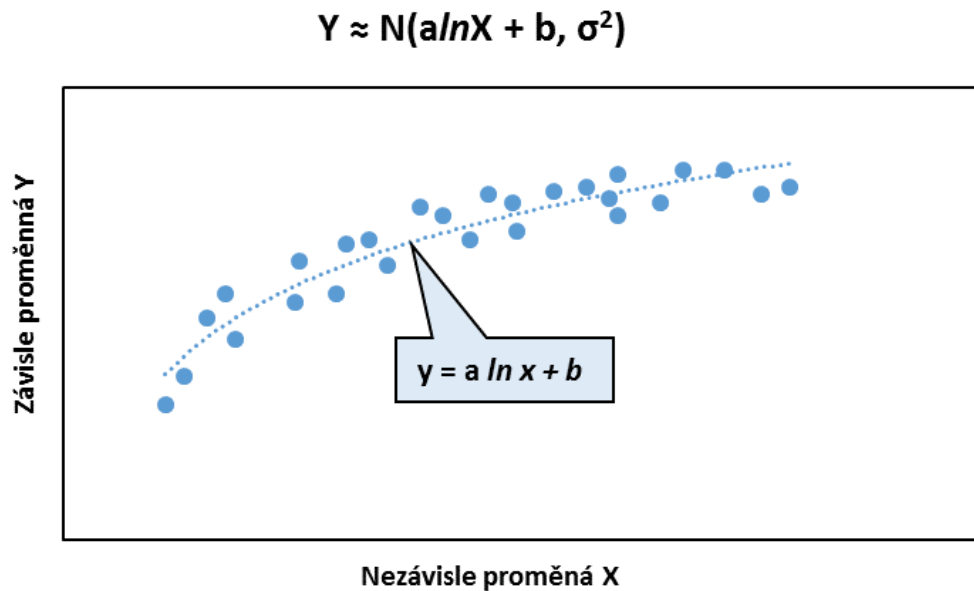
odchylek mezi pozorovanými hodnotami a odhadnutými hodnotami získanými dosazením do regresní přímky. Pokud tedy máme k dispozici historická data, můžeme jimi proložit regresní přímku, pro níž zjistíme rovnici, která je ve tvaru  $y = ax + b$  a standardní chybu odhadu této rovnice  $\sigma^2$  (obrázek 15). Parametr  $\sigma^2$  představuje standardní odchylku každého bodu od regresní přímky. Metoda nejmenších čtverců přitom předpokládá, že tato chyba pochází z normálního rozdělení. Důsledkem toho můžeme závisle proměnnou  $Y$  modelovat jako

$$Y \approx N(aX + b, \sigma^2).$$



Obrázek 15: Modelování závislosti mezi rizikovými faktory  $X$  a  $Y$  v lineárním případě

Jak jsme si již zmínili, existuje velké množství možností, jak můžeme pomocí rizikového faktoru  $X$  definovat parametry rizikového faktoru  $Y$ . Další situací může být, kdy závislost mezi rizikovými faktory  $X$  a  $Y$  není obecně lineární. Nyní je potřeba najít nejlepší nelineární křivku, která odpovídá datům (např. logaritmická viz obrázek 16).



Obrázek 16: Modelování závislosti mezi rizikovými faktory  $X$  a  $Y$  v nelineárním případě

**Obecný postup při simulaci v případě využití regresní analýzy pro lineární i nelineární případ:**

- Nejdříve si historická data proložíme vhodnou regresní křivkou (data zpravidla nikdy neleží přímo na křivce, nejedná se tedy o funkční vztah, ale jsou kolem křivky různě rozprostřena).
- Poté definujeme, podle jakého rozdělení pravděpodobnosti budeme modelovat nezávisle proměnnou  $X$  (např. pomocí trojúhelníkového rozdělení či jinak).
- Následně vygenerované hodnoty  $X$  můžeme dosadit do regresní křivky (obrázek 15) a díky tomu získáme hodnoty  $Y$ .
- K těmto odhadům je však nutné ještě přičíst chybu regrese, která ve většině případů nabývá hodnot z normálního rozdělení pravděpodobnosti s nulovou střední hodnotou a směrodatnou odchylkou rovnou směrodatné chybě z regresního odhadu. Vzhledem k tomu bude proces modelování reálnější.

## 2.3. Alternativní přístupy k modelování závislostí

Nyní si ještě stručně představíme dva alternativní přístupy popsané v [4], pomocí kterých můžeme modelovat závislost mezi rizikovými faktory. Mezi tyto přístupy patří *závislost definovaná pomocí vyhledávacích tabulek* a *závislost definovaná pomocí logických podmínek*.

### 2.3.1. Závislost definovaná pomocí vyhledávacích tabulek

Název *vyhledávací tabulky* pochází z angl. look-up tables a je přímo odvozen z funkce v MS Excel, která se při tomto postupu využívá. Tento způsob modelování je vhodné aplikovat zejména v situacích, kdy nějaký další vnější faktor rizika má významný simultánní vliv na původní faktory rizika daného modelu. Tento vnější rizikový faktor je generován nezávisle a z jeho hodnoty se následně odvodí parametry rozdělení pravděpodobností na něm závislých rizikových faktorů.

Příkladem tohoto problému může být situace, kdy se snažíme simulovat dobu, která je potřebná k vystavení určitého objektu. Výstavba daného objektu představuje proces složený z několika fází. Na celkový odhad doby projektu má však podstatný vliv faktor počasí, který musíme vzít v potaz. Faktor počasí totiž ovlivňuje odhady trvání jednotlivých fází a na jeho základě je nutné původní odhady jednotlivých rizikových faktorů ještě upravit (více viz [4]).

### 2.3.2. Závislost definovaná pomocí logických podmínek

Závislost lze definovat také pomocí různých (logických) podmínek jako např. „a“, „nebo“, „Jestliže, . . . ,pak“ atp. Ty potom mohou vést ke změně vlastností původních pravděpodobnostních rozdělení faktorů rizika.

Jako příklad této problematiky můžeme uvést situaci, kdy určitá událost může mít za následek výrazné snížení výnosu jinak nezávislých projektů. Tento příklad je podrobněji zpracován v [4].

### 3. Ilustrační příklady

V této části diplomové práce si nejprve představíme ilustrační příklad týkající se modelování expertně stanovené korelace a následně pak příklad na modelování korelace odhadnuté z dat. Pro modelování těchto korelací využijeme přístupu pomocí kopulí popsaného v podkapitole 2.1.2. Poté si ukážeme příklady, kde pro modelování závislostí využijeme obálkovou metodu. Vždy si uvedeme stručné zadání příkladu, na jehož základě si předvedeme konstrukci modelu a modelování závislostí mezi rizikovými faktory. Ke zpracování jednotlivých příkladů byly využity softwary MS Excel a Matlab.

#### 3.1. Příklad využívající kopulí pro modelování expertně stanovené korelace

Zadání tohoto příkladu bylo převzato z literatury [4].

Naším úkolem je analyzovat riziko investičního projektu pomocí metody simulace Monte Carlo. Cílem investičního projektu je výroba nového produktu, jehož roční zisk před zdaněním (výsledek hospodaření) je ovlivněn celkem šesti faktory rizika. Každý z rizikových faktorů je zatížen určitou nejistotou, vzhledem k tomu byly pro každý tento rizikový faktor zpracovány tři scénáře. Scénář nejpravděpodobnější vycházející z předpokládaných hodnot jednotlivých rizikových faktorů, scénář optimistický (v případě značně příznivého vývoje rizikového faktoru) a nakonec scénář pesimistický (v případě značně nepříznivého vývoje rizikového faktoru). Odhadované hodnoty těchto rizikových faktorů shrnuje následující tabulka.

Faktor rizika	Jednotka	Scénář		
		pesimistický	nejpravděpod.	optimistický
1. Prodej (P)	tis. ks/rok	75	100	120
2. Prodejní cena (c)	euro/ks	135	150	160
3. Měnový kurz (m)	Kč/euro	24	27	29
4. Měrná spotřeba materiálu (s)	kg/ks	62	60	58
5. Kупní cena materiálu (k)	Kč/kg	46	40	36
6. Fixní náklady (FN)	mil. Kč/rok	85	75	70

Tabulka 4: Hodnoty rizikových faktorů

Veškeré výpočty k tomuto příkladu jsou uvedeny v příložených souborech *Příklad1\_korelace.xlsx*, *příklad1\_neg\_korelace.m* a *histogram\_zisky.m*.

Jako výstupní neboli kritériální veličina v tomto modelu bude vystupovat roční zisk před zdaněním. Vstupní veličiny, tedy rizikové faktory, které ovlivňují naši výstupní proměnnou, jsou prodej, prodejní cena, měnový kurz, měrná spotřeba materiálu, kupní cena materiálu a fixní náklady.

Jelikož je naší zvolenou výstupní veličinou roční zisk před zdaněním, bude náš matematický model dán jednoduchým vztahem jako rozdíl tržeb a nákladů. Vzorec je tedy ve tvaru:

$$\begin{aligned} Z &= T - N = \\ &= P \cdot c \cdot m - (P \cdot s \cdot k + FN). \end{aligned}$$

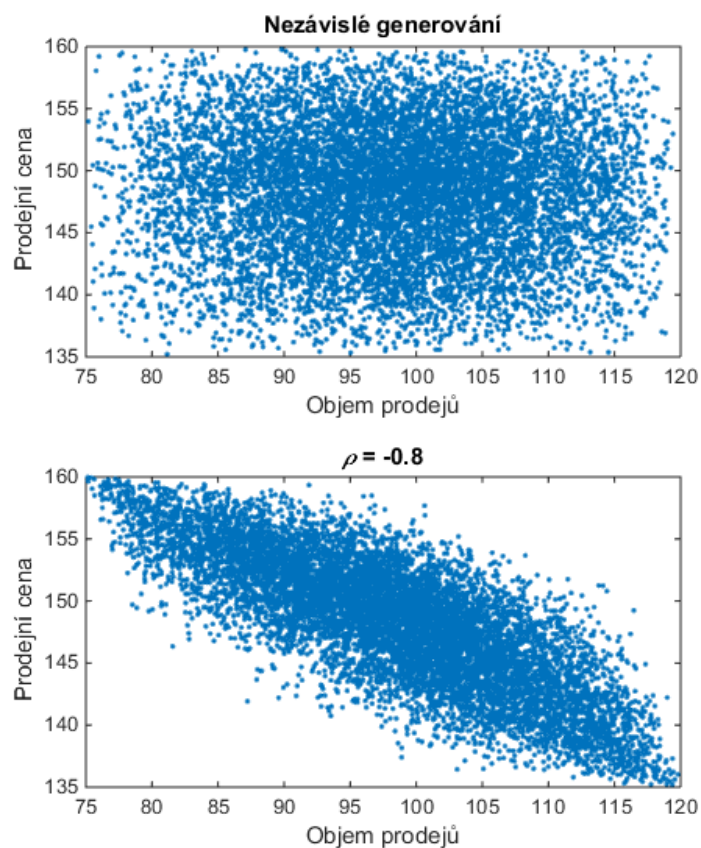
Nyní jsme schopni na základě tohoto vztahu snadno zpracovat příklad v softwaru MS Excel, který nám vypočítá v závislosti na šesti ovlivňujících rizikových faktorech požadovaný roční zisk před zdaněním z výroby nového produktu.

Pro modelování pravděpodobnostního rozdělení jednotlivých rizikových faktorů jsme použili trojúhelníkové rozdělení, které představuje jedno z nejpoužívanějších rozdělení pravděpodobnosti pro modelování expertních odhadů. Trojúhelníkové rozdělení jsme si již představili v části 2.2.1 a v našem případě tedy dolní mez (parametr  $a$ ) reprezentuje pesimistický odhad hodnoty rizikového faktoru, horní mez (parametr  $b$ ) reprezentuje optimistický odhad hodnoty rizikového faktoru a nejpravděpodobnější hodnotu (parametr  $c$ ) tvoří nejpravděpodobnější odhad rizikového faktoru.

Následně se dostáváme ke kroku simulace, kde je nutné stanovit statistickou závislost jednotlivých rizikových faktorů. Pro srovnání budeme uvažovat nejdříve simulační model, ve kterém budeme ignorovat závislost mezi jednotlivými rizikovými faktory. Poté do modelu zahrneme závislost mezi objemem prodeje a prodejní cenou. Jelikož při zvýšení prodejní ceny poklesne objem prodeje a naopak při snížení prodejní ceny vzroste objem prodeje, tak byla mezi těmito rizikovými faktory expertně stanovena silná negativní korelace ( $\rho = -0,8$ ). Postup v softwaru Matlab je následující:

- nejdříve si pomocí příkazu `makedist('Triangular',a,c,b)` zadefinujeme trojúhelníkové rozdělení pro faktory objem prodeje a prodejní cena,
- poté pomocí příkazu `copularnd('Gaussian',rho,n)` vygenerujeme zkorelované hodnoty z Gaussovy kopuly s negativní korelací  $\rho = -0,8$ ,
- nakonec pomocí příkazu `icdf` vypočítáme odpovídající dvojice hodnot z daných dvou rozdělení pravděpodobnosti.

Pro názornost si následně můžeme hodnoty vykreslit do grafu, který vidíme na obrázku 17.



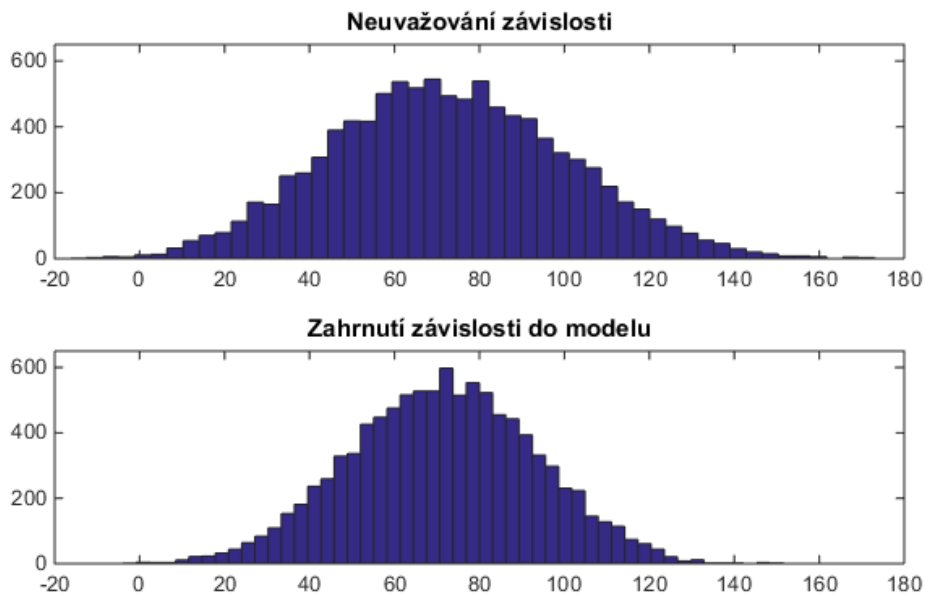
Obrázek 17: Vygenerované páry objemu prodeje a prodejní ceny z Gaussovy kopuly

V našem modelu jsme si počet iterací  $n$  zvolili 10 000. Výstupem simulace jsou



nasimulovaná data  $x_1, \dots, x_{10\,000}$ , na jejichž základě usuzujeme o rozdělení pravděpodobnosti zisku před zdaněním a také z nich konstruujeme odhady číselných charakteristik, které nám pomáhají analyzovat a srovnávat výsledky simulace.

Výsledky našeho modelování zachycují histogramy zisků na obrázku 18, kde dochází ke srovnání histogramů, a to v případě neuvažování závislosti a v případě zahrnutí závislosti do modelu.



Obrázek 18: Porovnání histogramů zisků při neuvažování závislosti a zahrnutí závislosti do modelu ( $\rho = -0,8$ ).

Nyní vidíme, že korelace v tomto případě působí jako zajišťovací prvek (při nižší ceně můžeme očekávat vyšší prodeje a naopak). Při zahrnutí nenulové korelace do modelu se nám tedy snižuje variabilita. To je patrné z obrázku 18, kde můžeme vidět, že rozdělení pravděpodobnosti zisku před zdaněním s uvažovanou korelací klesá k nule výrazně rychleji než rozdělení pravděpodobnosti zisku před zdaněním bez zahrnutí korelace (nulová korelace). Zajímavý je i dopad negativní korelace na možnost ztráty. Na obrázku 18 můžeme pozorovat, že v případě uvažování negativní korelace mezi rizikovými faktory klesne pravděpodobnost ztráty na nulu. Pro srovnání jsou v tabulce 5 uvedeny číselné charakteristiky polohy

a variability zisku před zdaněním při neuvažování a uvažování korelace. Je zde patrné, že pokud zahrneme korelaci do modelu, tak se hodnoty charakteristik variability (např. rozptyl, směrodatná odchylka) sníží.

<i>Číselné charakteristiky</i>	<i>Neuvažování závislosti</i>	<i>Uvažování závislosti</i>
Průměr	72,766	71,414
Medián	71,802	71,640
0,05-kvantil	27,920	35,442
0,95-kvantil	119,733	107,379
Rozptyl	774,000	474,297
Směrodatná odchylka	27,821	21,778
Průměrná absolutní odchylka	22,415	17,518
Variační koeficient	0,382	0,305

Tabulka 5: Srovnání číselných charakteristik zisků

### 3.2. Příklad využívající kopulí pro modelování korelace odhadnuté z dat

Nyní budeme ilustrovat závislost v podobě pozitivní korelace na příkladu modelování hodnot měnových kurzů. Situace oproti předchozímu příkladu bude v tomto případě jiná. V této chvíli totiž budeme mít k dispozici historická data, ze kterých budeme vycházet. Zadání tohoto příkladu bylo převzato z literatury [4].

Úkolem je analyzování rizikové pozice, která plyne ze dvou cizoměnových pohledávek denominovaných v amerických dolarech (USD) a britských librách (GBP). V této situaci budeme mít pohledávky 200 000 USD a 100 000 GBP splatné za jeden měsíc. Úhradu pohledávek však budeme konvertovat na české koruny (CZK). Na základě toho nám vzniká problém, jak odhadnout rozdělení pravděpodobnosti hodnot pohledávek v CZK.

V příložených souborech *Příklad2\_korelace.xlsx*, *zavislost\_kopule.m*, *normalita.m*, *extreme\_value.m* a *histogram\_pohledavky.m* jsou k dispozici veškeré výpočty k tomuto příkladu.

### 3.2.1. Vytvoření matematického modelu

Nejprve začneme s odhadem možných hodnot devizových kurzů CZK/USD a CZK/GBP, které představují *rizikové faktory*. Pro získání těchto hodnot jsme využili databázi České národní banky, ze které jsme získali hodnoty devizových kurzů vždy na konci měsíce. V důsledku toho, že se jedná o modelování pouze v rámci intervalu jednoho měsíce, tak se zaměříme jen na možné změny kurzů od aktuálních hodnot. Nejdříve tedy vypočítáme měsíční změny podle vzorce (viz [4]):

$$r_i = \frac{k_i - k_{i-1}}{k_{i-1}},$$

kde  $r_i$  jsou změny devizového kurzu a  $k_i$  jsou hodnoty kurzů na konci jednotlivých měsíců.<sup>2</sup> Tyto měsíční změny lze najít v souboru *Příklad2\_korelace.xlsx*.

Pořadí	Procentuální změna	
	CZK/USD	CZK/GBP
1	-1,14 %	0,06 %
2	2,31 %	1,73 %
3	-2,75 %	-1,16 %
...	...	...
30	2,15 %	2,34 %

Tabulka 6: Historické měsíční změny devizových kurzů CZK/USD a CZK/GBP

Na základě takto spočítaných historických změn můžeme sestavit matematický model, který využijeme pro odhad hodnot kurzů v měsíci, kdy jsou pohledávky splatné. Při výpočtu těchto odhadů budeme vycházet z aktuálních hodnot devizových kurzů  $k_{0,USD}$  a  $k_{0,GBP}$ , které budeme upravovat o možnou změnu kurzu. Výpočet *hodnot kurzů*, které budou *platné za jeden měsíc* (v době splatnosti pohledávky), jsou pak následující:

$$k_{1,USD} = k_{0,USD} \cdot (1 + r), \quad (4)$$

$$k_{1,GBP} = k_{0,GBP} \cdot (1 + s), \quad (5)$$

<sup>2</sup> Původně jsme chtěli uvažovat 36 hodnot, avšak posledních 6 hodnot (viz soubor *Příklad2\_korelace.xlsx*) bylo ovlivněno referendem o brexitu, proto jsme vzali v potaz pouze 30 hodnot.

kde  $r, s$  jsou *změny kurzu* vygenerované z příslušného rozdělení pravděpodobnosti.

Následně pak můžeme vypočítat *hodnotu pohledávek* v době splatnosti již v českých korunách:

$$p_{1,CZK} = 200\,000 \cdot k_{1,USD}, \quad (6)$$

$$p_{2,CZK} = 100\,000 \cdot k_{1,GBP}. \quad (7)$$

Celkovou pohledávku v českých korunách v době splatnosti již jednoduše vypočítáme jako:

$$p_{1,CZK} + p_{2,CZK}.$$

### 3.2.2. Určení marginálních rozdělení rizikových faktorů

Pro modelování rozdělení pravděpodobnosti rizikových faktorů v případě změny devizového kurzu CZK/USD nám na základě analýzy dat vyšlo, že můžeme použít normální rozdělení. K ověření normality lze využít hned několik nástrojů. Prvním z nich mohou být grafické nástroje, které nám umožňují vizuální porovnání dat s pravděpodobnostním rozdělením. V našem případě jsme využili porovnání distribuční funkce teoretického rozdělení s empirickou distribuční funkcí. V softwaru Matlab můžeme pro toto ověření využít příkaz *dfittool*, který nám otevře speciální prostředí pro fitování. Na obrázku 19 je zřejmé, že data odpovídají normálnímu rozdělení.

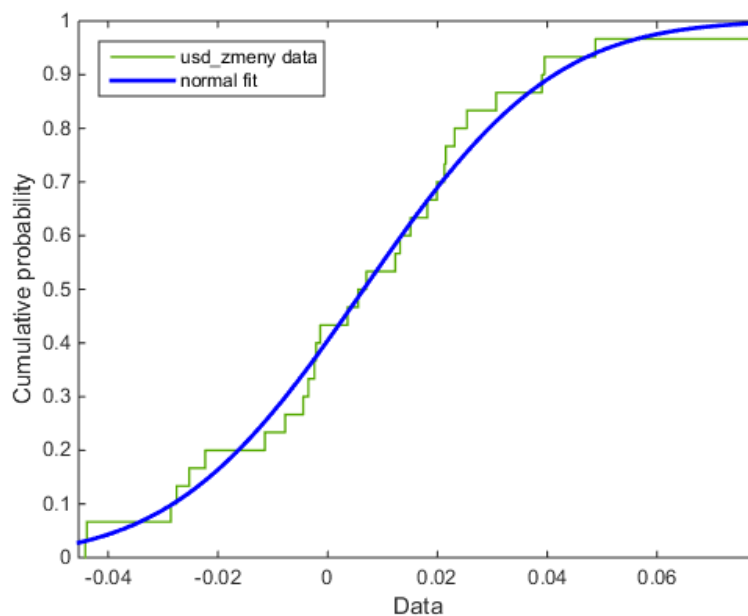
Naopak v případě změny devizového kurzu CZK/GBP jsme již normální rozdělení uvažovat nemohli, jelikož na obrázku 20 vidíme, že data neodpovídají normálnímu rozdělení. Na základě toho jsme proto zvolili rozdělení extreme value, které datům již dobře odpovídá (obrázek 21).

Rozdělení extreme value je podle [8] definováno funkcí hustoty

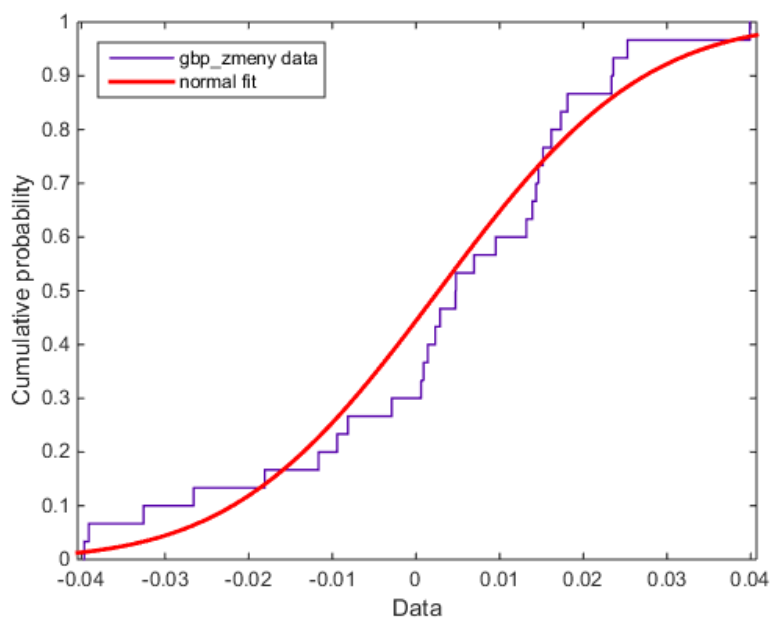
$$f(x) = \frac{1}{\beta} \exp\left(-\frac{x-\mu}{\beta} - \exp\left(-\frac{x-\mu}{\beta}\right)\right), \quad \beta > 0.$$

Další možností, jak ověřit normalitu, jsou testy dobré shody. Mezi ně můžeme například zařadit Kolmogorov-Smirnovův test, Anderson-Darlingův test či Pearsonův  $\chi^2$  test. Pro podrobnější informace odkazují na literaturu [1], [5], [10].

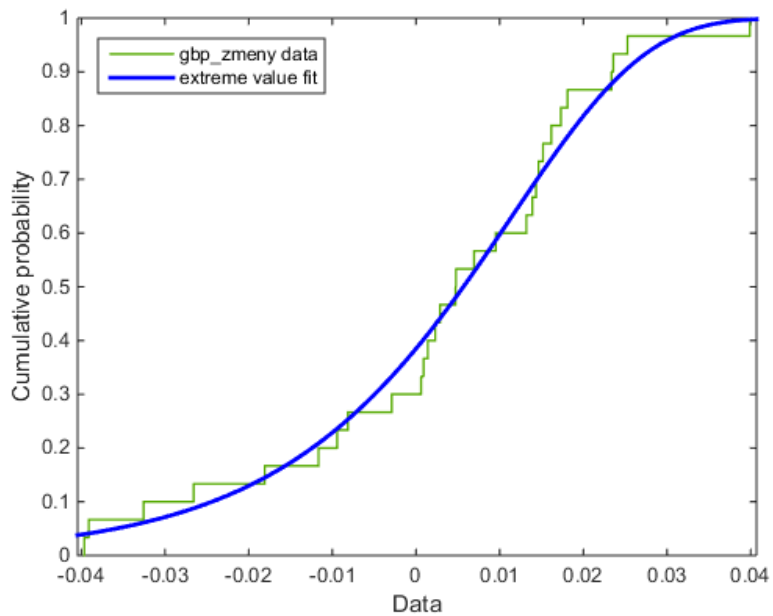
Při testování normality změny devizového kurzu CZK/USD v softwaru Matlab pomocí těchto tří testů na hladině významnosti  $\alpha = 0,05$  nám vyšlo, že pouze Kolmogorov-Smirnovův test hypotézu zamítá. I přes zamítnutí hypotézy u tohoto testu jsme se rozhodli využít pro modelování normální rozdělení, jelikož na základě grafického porovnání empirické distribuční funkce s distribuční funkcí normálního rozdělení data hezky odpovídala. Při testování normality změny devizového kurzu CZK/GBP nám opět pouze Kolmogorov-Smirnovův test hypotézu zamítl. Nyní však ještě musíme vzít u dalších testů v úvahu  $p$ -hodnotu, která představuje nejmenší hladinu, při které bychom ještě hypotézu zamítl. U Anderson-Darlingova testu totiž vyšla 0,0608 a při porovnání s hladinou testu  $\alpha = 0,05$  je tato hodnota hraniční. Proto jsme zvážili jiná rozdělení, která by lépe odpovídala těmto datům. Na základě grafického porovnání jsme tedy zvolili pro modelování změny devizového kurzu CZK/GBP rozdělení extreme value, které jsme testovali opět pomocí Anderson-Darlingova testu. Tento test hypotézu nezamítl a  $p$ -hodnota vyšla 0,6272.



Obrázek 19: Graf distribuční funkce normálního rozdělení s empirickou distribuční funkcí pro změnu kurzu CZK/USD



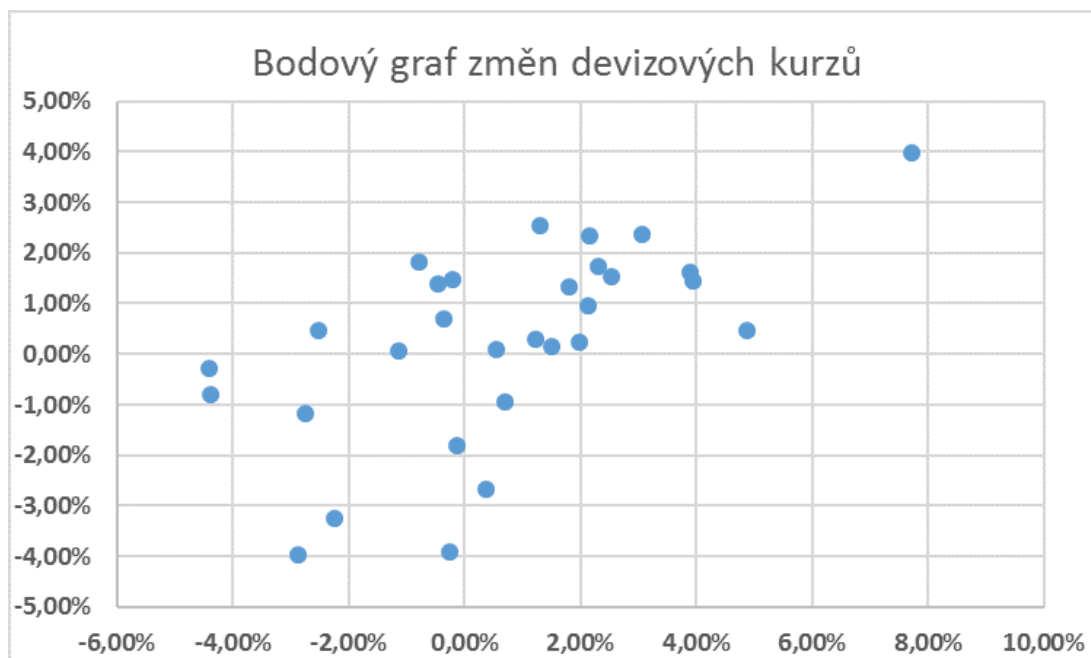
Obrázek 20: Graf distribuční funkce normálního rozdělení s empirickou distribuční funkcí pro změnu kurzu CZK/GBP



Obrázek 21: Graf distribuční funkce rozdělení extreme value s empirickou distribuční funkcí pro změnu kurzu CZK/GBP

### 3.2.3. Stanovení statistické závislosti

Nyní přichází na řadu krok simulace, ve kterém je nutné stanovit statistickou závislost mezi rizikovými faktory. Opět budeme pro srovnání uvažovat nejdříve simulační model, ve kterém budeme závislost mezi rizikovými faktory ignorovat, a následně vytvoříme model, do kterého zahrneme závislost mezi změnami devizových kurzů. V této situaci, jelikož máme k dispozici historická data, je nutné spočítat korelační koeficient mezi jednotlivými změnami devizových kurzů. Korelační koeficient nám bude sloužit pro modelování závislosti a vypočteme ho pomocí funkce *CORREL* v softwaru MS Excel. Hodnota tohoto koeficientu nám vyšla 0,602, což svědčí o pozitivní korelaci. K tomu, aby si člověk udělal představu o závislosti mezi historickými daty změn devizových kurzů, jsme vytvořili bodový graf 22.



Obrázek 22: Změny devizových kurzů CZK/USD a CZK/GBP

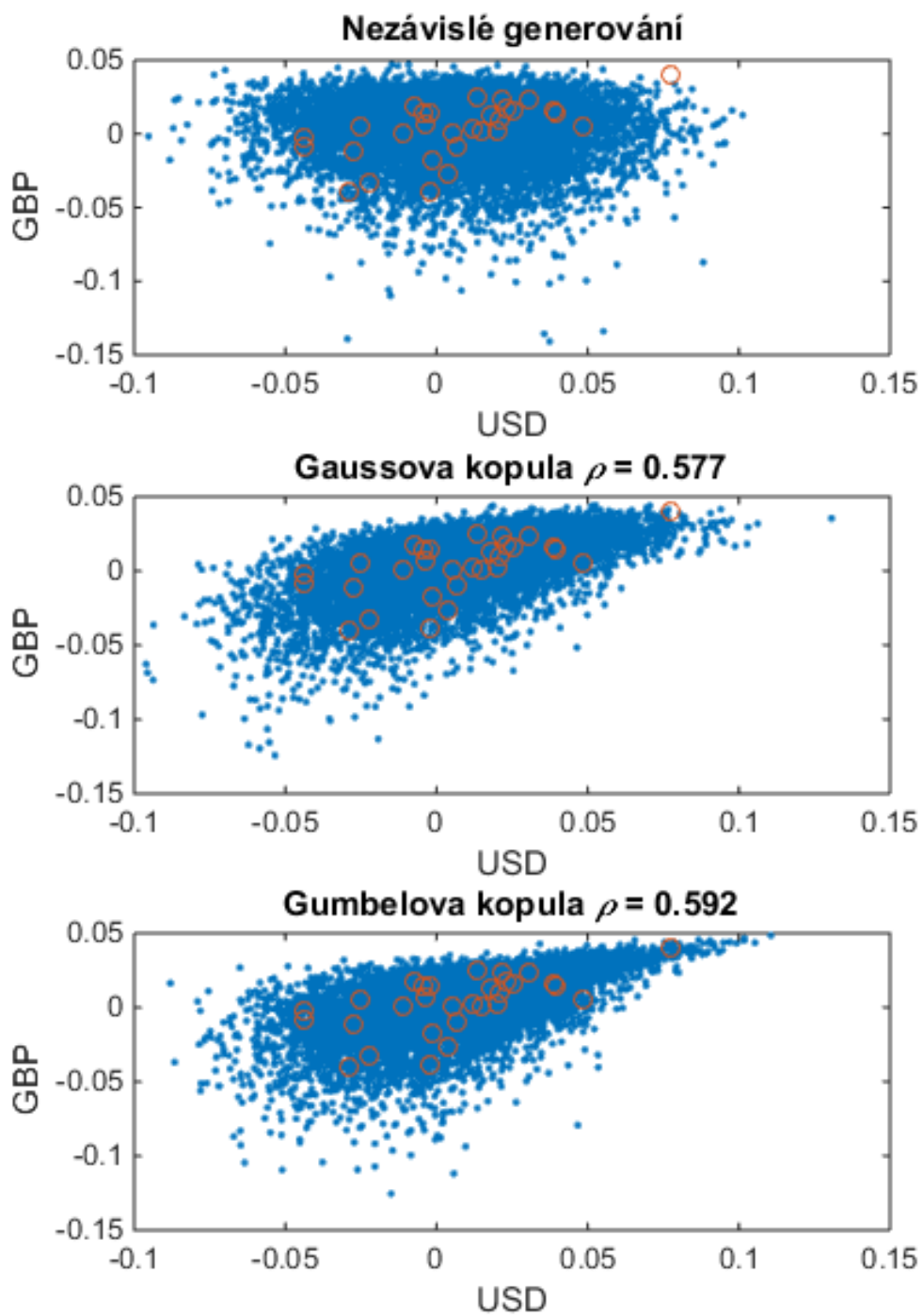
Pro vygenerování dat s pozitivní korelací pomocí Gaussovy a Gumbelovy kopuly v softwaru Matlab postupujeme takto:

- nejdříve si pomocí příkazu `makedist('Normal',mu,sigma)` zadefinujeme normální rozdělení pro změnu devizového kurzu CZK/USD, kde se parametry  $\mu$  a  $\sigma$  určí s využitím příkazu `normfit` a následně si pomocí příkazu `makedist('Extreme value',mu,sigma)` zadefinujeme rozdělení extreme value pro změnu devizového kurzu CZK/GBP, kde se parametry  $\mu$  a  $\sigma$  určí s využitím příkazu `evfit`,
- poté pomocí příkazu `copularnd('Gaussian',[1 0.602; 0.602 1],n)` vygenerujeme zkorelované hodnoty z Gaussovy kopuly,
- také pro srovnání použijeme Gumbelovu kopulu, kdy je potřeba nejdříve vypočítat z dat Kendallův korelační koeficient jako `corr(x,y,'Type','Kendall')`, který slouží pro výpočet parametru  $\alpha$ , tento parametr se vypočítá pomocí příkazu `copulaparam('Gumbel',rho)`, kde  $\rho$  představuje vypočtený Kendallův korelační koeficient, poté můžeme vygenerovat zkorelované hodnoty z této kopuly jako `copularnd('Gumbel',alpha,n)`,
- nakonec pomocí příkazu `icdf` vypočítáme odpovídající dvojice hodnot z daných dvou rozdělení pravděpodobnosti.

Opět si pro lepší představu můžeme hodnoty vykreslit do grafů společně s historickými změnami devizových kurzů, které vidíme na obrázku 23. Můžeme zde vidět srovnání modelů, kdy nejprve neuvažujeme závislost mezi změnami devizových kurzů a poté zahrneme závislost mezi těmito rizikovými faktory ve formě pozitivní korelace.

Stejně jako v předchozím příkladu jsme si počet iterací  $n$  zvolili 10 000.

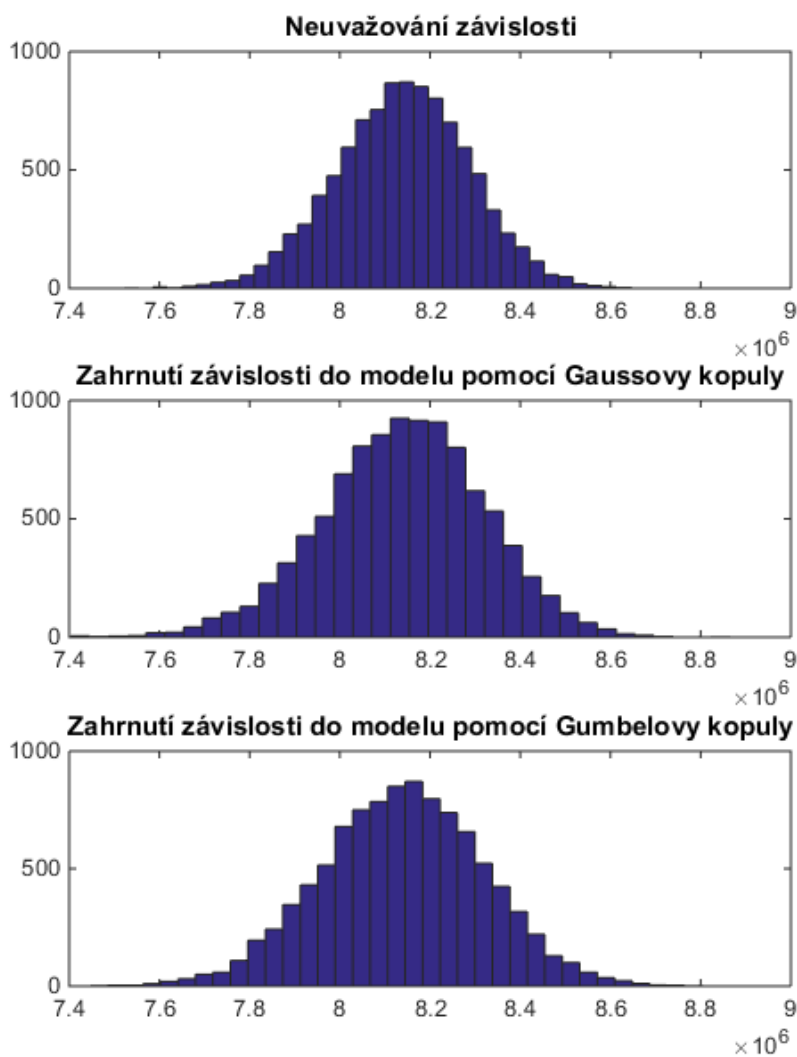




Obrázek 23: Vygenerované páry změn devizových kurzů dle jednotlivých přístupů

### 3.2.4. Výsledky modelu

Histogramy celkových pohledávek jsou znázorněny na obrázku 24, kde můžeme vidět srovnání histogramů při neuvažování závislosti a při zahrnutí závislosti do modelu s využitím Gaussovy a Gumbelovy kopuly. Z tohoto obrázku je patrné, že při zahrnutí závislosti do modelu bude rozptyl pohledávek větší než v situaci, kdybychom závislost neuvažovali.



Obrázek 24: Porovnání histogramů celkové pohledávky při neuvažování závislosti a zahrnutí závislosti do modelu

Tento fakt můžeme pozorovat i v tabulce 7, kde máme uveden souhrn číselných charakteristik polohy a variability pohledávky při neuvažování a uvažování pozitivní korelace pomocí Gaussovy a Gumbelovy kopuly. Co se týče charakteristik polohy, jako např. průměru či mediánu, vidíme, že se jejich hodnota příliš nezměnila. Naopak u charakteristik variability, jako jsou rozptyl a směrodatná odchylka, můžeme pozorovat poměrně velký nárůst.

Číselné charakteristiky	Neuvažování závislosti	Gauss	Gumbel
Průměr	8 139 823, 85	8 140 743, 71	8 138 484, 08
Medián	8 143 617, 88	8 146 590, 79	8 140 674, 45
0,05-kvantil	7 891 892, 15	7 837 079, 58	7 836 479, 82
0,95-kvantil	8 380 100, 04	8 428 211, 34	8 431 125, 14
Rozptyl	21 773 895 018	32 668 081 720	32 630 306 452
Směrodatná odchylka	147 559, 80	180 743, 14	180 638, 61
Průměrná absolutní odchylka	117 432, 75	143 300, 81	144 108, 32
Variační koeficient	0,0181	0,0222	0,0222

Tabulka 7: Srovnání číselných charakteristik celkové pohledávky

V literatuře [4] navrhují autoři J. Hnilica a J. Fotr pro modelování v tomto příkladu postup zvaný *Bootstrap*. Tento postup patří mezi neparametrické metody odhadu rozdělení pravděpodobnosti, kdy nemáme k dispozici žádné předpoklady o tvaru pravděpodobnostního rozdělení.

V rámci Bootstrapu se musí změny kurzu  $r$  v případě devizového kurzu CZK/GBP a  $s$  v případě devizového kurzu CZK/USD vybrat *náhodně* z daných 30 historických změn. Nejdříve je však nutné při simulaci vygenerovat číslo výběru tohoto historického pozorování, k němuž lze využít diskrétní rovnoměrné rozdělení v rámci hodnot 1 až 30. Nyní stačí jen postupně dosadit do vzorců (4), (5), (6), (7) a na základě takového generování se získá pravděpodobnostní rozdělení jednotlivých pohledávek. Tento model se vyznačuje důležitou vlastností, kterou je zachování historické závislosti změn mezi jednotlivými kurzy, jelikož se bere vždy celý měnový pár.

### 3.3. Příklad využívající obálkové metody pro modelování závislosti mezi rizikovými faktory na základě expertních odhadů

Zadání tohoto příkladu bylo převzato z literatury [4]. Veškeré výpočty jsou k dispozici v příložených souborech *obalkova\_metoda\_urokove\_miry.xlsx* a *urokove\_miry.m*.

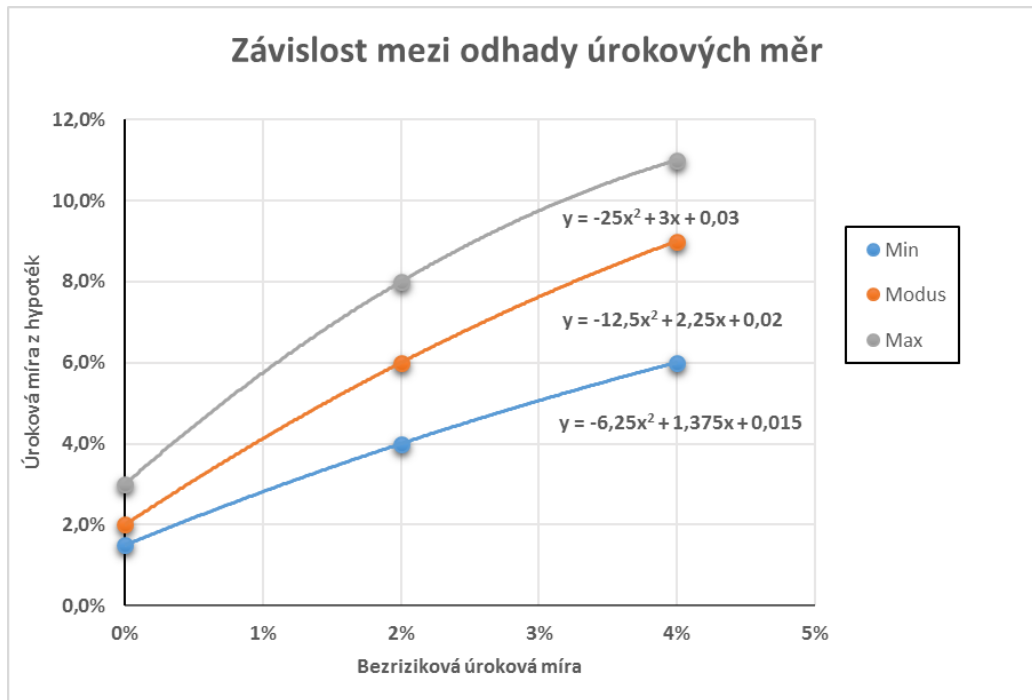
Naším úkolem je s využitím obálkové metody modelovat závislost mezi bezrizikovými úrokovými mírami a úrokovými mírami z hypoték. Nejprve potřebujeme získat expertní odhad těchto rizikových faktorů, řekněme např. na příští rok. Na základě porady s experty získáváme tabulku hodnot 8.

Bezriziková úroková míra	Úrokové míry z hypoték		
	minimum	modus	maximum
0 %	1,5 %	2 %	3 %
2 %	4 %	6 %	8 %
4 %	6 %	9 %	11 %

Tabulka 8: Závislost mezi odhady úrokových měr

Pro bezrizikovou úrokovou míru (nezávislý rizikový faktor) máme dán v tabulce expertní odhad určující interval možných hodnot, kterých tato úroková míra může nabývat. Za nejpravděpodobnější hodnotu se považuje 2 %. V případě, kdy bezriziková úroková míra bude 0 %, tak odhad úrokové míry pro hypotéky (závislý rizikový faktor) bude náležet do intervalu 1,5 % až 3 %, přičemž nejpravděpodobnější hodnota je 2 %. Obdobně pak i pro další hodnoty bezrizikové úrokové míry. V tabulce máme tedy dané scénáře možných hodnot úrokových měr z hypoték v závislosti na hodnotách bezrizikových úrokových měr, a to přes optimistické, nejpravděpodobnější a pesimistické hodnoty. Na základě toho můžeme využít pro modelování závislostí trojúhelníkové rozdělení. Nyní je potřeba zjistit odhady parametrů rozdělení úrokových měr z hypoték pro všechny možné hodnoty bezrizikové úrokové míry pocházející z rozdělení pravděpodobnosti, které je definované parametry  $\min = 0 \%$ ,  $\text{mod} = 2 \%$  a  $\text{max} = 4 \%$ . To lze jednoduše provést zakreslením jednotlivých scénářů z tabulky 8 do soustavy souřadnic a ná-

sledným proložením těchto křivek funkcí polynomu druhého stupně, což můžeme vidět na obrázku 25.



Obrázek 25: Proložení scénářů polynomickou funkcí

Následný postup v softwaru Matlab je již jednoduchý. Zdefinujeme si pomocí příkazu `makedist('Triangular',0,2,4)` trojúhelníkové rozdělení s parametry 0, 2 a 4. Poté si v rámci jednoho scénáře z tohoto rozdělení vygenerujeme s využitím příkazu `random` hodnotu  $x$  bezrizikové úrokové míry  $X$ , kterou dosadíme do každé z rovnic:

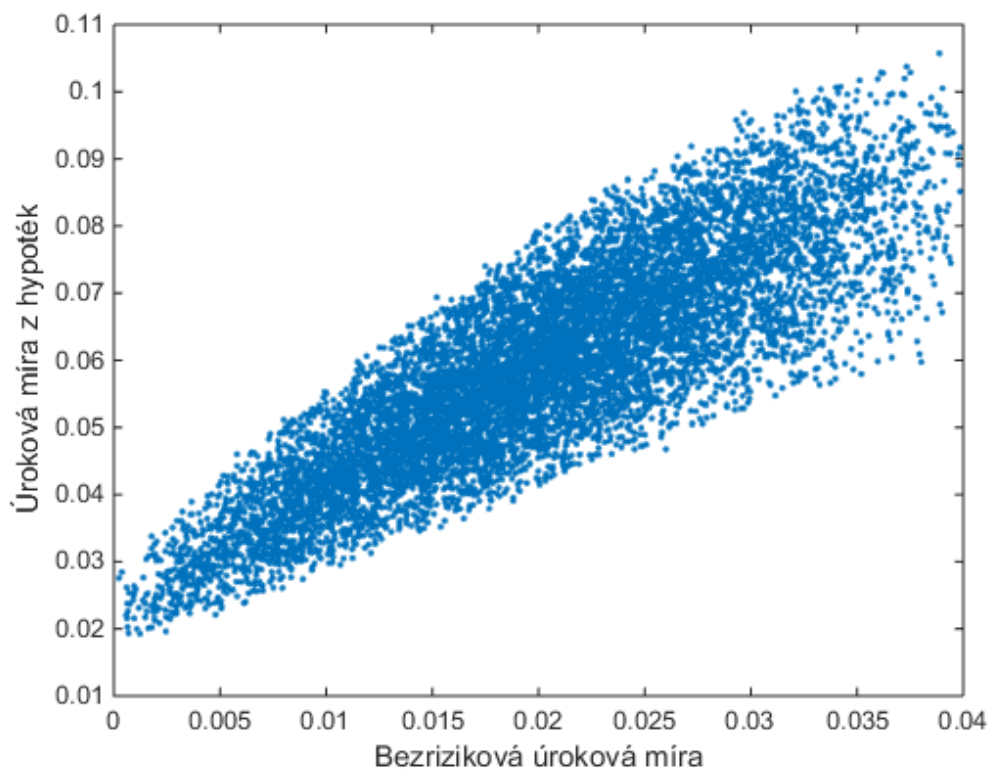
$$MIN = a_1x^2 + a_2x + a_3$$

$$MODUS = c_1x^2 + c_2x + c_3$$

$$MAX = b_1x^2 + b_2x + b_3$$

Tímto získáme parametry trojúhelníkového rozdělení pro modelování odhadu úrokových měř z hypoték, který si vygenerujeme opět pomocí příkazu `random`. V našem modelu jsme si zvolili počet iterací (scénářů) 10 000.

Výsledek modelování můžeme vidět na obrázku 26.



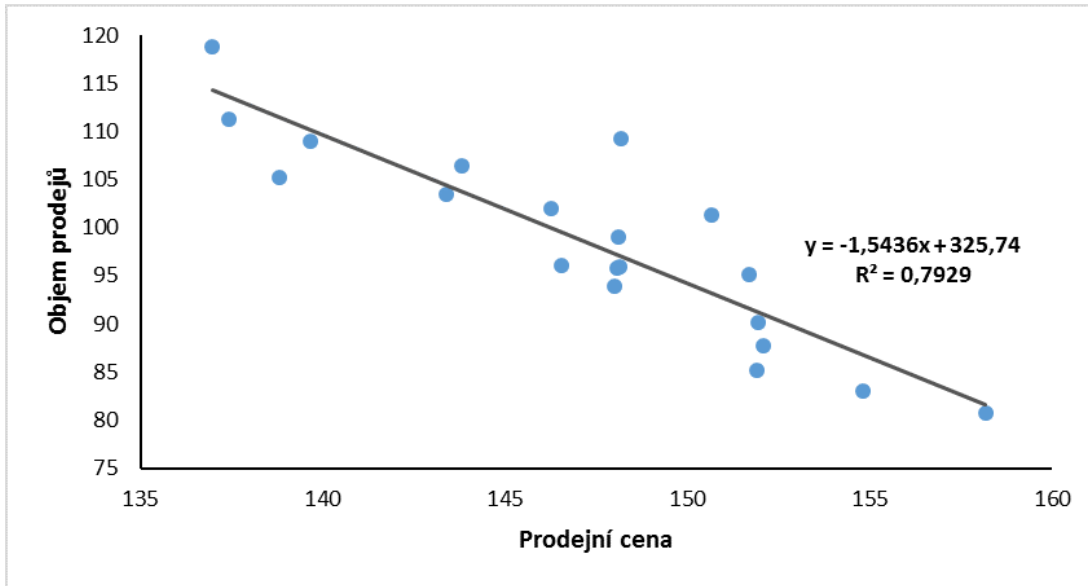
Obrázek 26: Simulovaná data

### 3.4. Příklad využívající obálkové metody pro modelování závislosti mezi rizikovými faktory na základě historických dat

V tomto příkladě budeme uvažovat stejné zadání jako v podkapitole 3.1, tedy nás bude opět zajímat roční zisk před zdaněním z výroby nového produktu. Nyní se však budeme snažit závislost mezi objemem prodeje a prodejní cenou modelovat pomocí obálkové metody s využitím regresní analýzy. Jelikož však nemáme k dispozici historická data, tak si pro jednoduchost v softwaru Matlab vygenerujeme pomocí trojúhelníkového rozdělení 20 dat, o jejichž analýzu se můžeme opřít. Veškeré výpočty lze najít v příložených souborech *obalkova\_metoda\_zisk.xlsx*, *historicka\_data.m*, *zisk.m* a *histogram\_obalkova\_metoda.m*.

Nejdříve si vygenerovaná data vykreslíme do bodového grafu (obrázek 27), ze kterého vidíme, že vztah mezi prodejní cenou a objemem prodeje je přibližně li-

neární. Dále si v softwaru MS Excel vypočítáme hodnoty prvního a druhého regresoru a také hodnotu směrodatné chyby odhadu pomocí funkce *LINREGRESE*, které vidíme v tabulce 9.

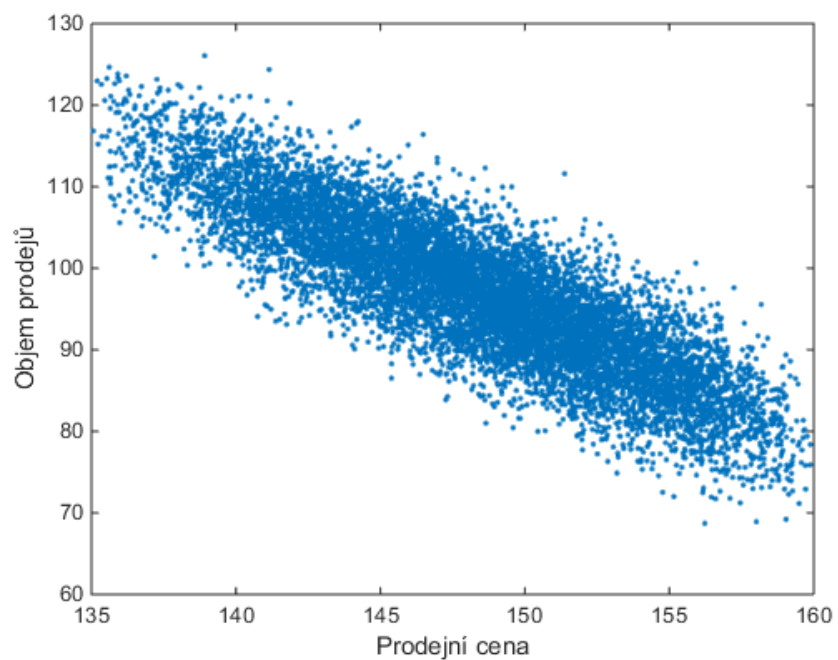


Obrázek 27: Prodejní cena a objem prodeje

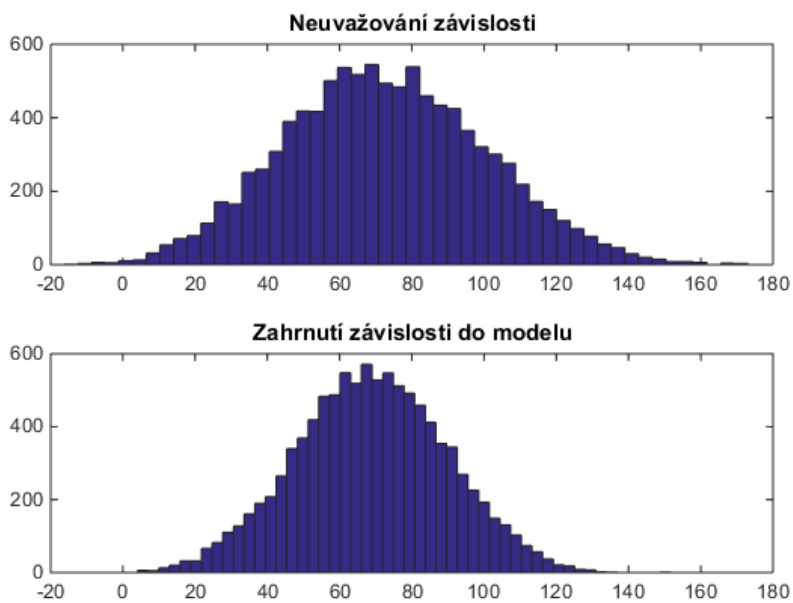
a	-1,544
b	325,737
chyba odhadu	4,697

Tabulka 9: Výsledky regresní analýzy

Nyní si pomocí příkazu *makedist('Triangular',135,150,160)* zadefinujeme trojúhelníkové rozdělení s parametry 135, 150 a 160, pomocí kterého budeme modelovat prodejní cenu představující nezávislý rizikový faktor. Pro získání odhadu objemu prodeje (závislý rizikový faktor) již stačí vygenerované hodnoty prodejní ceny dosadit do regresní rovnice a přičíst chybu odhadu vygenerovanou pomocí normálního rozdělení s nulovou střední hodnotou a směrodatnou odchylkou rovnou směrodatné chybě odhadu, tedy  $N(0; 4, 697)$ . Na obrázku 28 máme vykreslené závislé hodnoty prodejní ceny a objemu prodeje a na obrázku 29 můžeme vidět výsledné histogramy zisků.



Obrázek 28: Vygenerované páry prodejní ceny a prodeje pomocí obávkové metody



Obrázek 29: Porovnání histogramů zisků při neuvažování závislosti a zahrnutí závislosti do modelu



## Závěr

V úvodu diplomové práce jsme se seznámili s problematikou metody Monte Carlo. Nejdříve jsme si tuto metodu představili a uvedli si něco z její historie, poté jsme si ukázali postup simulace a nakonec jsme si předvedli přednosti a nedostatky této metody.

Dále jsme se v práci věnovali jednotlivým metodám, pomocí kterých můžeme modelovat závislost mezi rizikovými faktory v simulaci Monte Carlo. První metodou, kterou jsme si popsali, bylo modelování závislostí pomocí korelací. Ukázali jsme si, že se v simulačních modelech v případě, kdy existuje logický vztah mezi dvěma rizikovými faktory, využívají dva přístupy. Prvním z nich bylo využití Spearmanova koeficientu pořadové korelace, se kterým se často můžeme setkat v softwarových produktech. Druhým přístupem bylo modelování pomocí kopulí, které představuje flexibilnější metodu. Dále jsme si ukázali, že pro modelování závislostí můžeme také využít obálkovou metodu. Jedním z přístupů obálkové metody bylo odvození parametrů rozdělení pravděpodobnosti rizikového faktoru  $Y$  na základě realizace rizikového faktoru  $X$ . Dalším představeným přístupem bylo modelování pomocí regresní analýzy. Nakonec jsme si stručně představili alternativní přístupy k modelování závislostí.

Třetí část práce byla zaměřena především na demonstrování popsaných metod na konkrétních příkladech. Nejdříve jsme si ukázali příklad na modelování expertně stanovené korelace pomocí kopulí, kde jsme analyzovali riziko investičního projektu. Také jsme si představili příklad, který využíval kopulí pro modelování korelace odhadnuté z dat. Nakonec jsme poslední dva příklady zaměřili na ilustrování obálkové metody.

Při zpracování diplomové práce jsem získala mnoho nových zkušeností a znalostí, a to nejen v problematice modelování závislostí mezi rizikovými faktory. Díky jejímu vypracování jsem prohloubila své znalosti matematických softwarů Matlab a Microsoft Excel a také typografického systému  $\text{T}_{\text{E}}\text{X}$ , v němž je tato práce napsána.

## Literatura

- [1] Anděl, J.: *Základy matematické statistiky*. Matfyzpress, Praha 2005.
- [2] Cipra, T.: *Riziko ve financích a pojišťovnictví: Basel III a Solvency II*. Eko-press, Praha 2015
- [3] Fabian, F., Kluiber, Z.: *Metoda Monte Carlo a možnosti jejího uplatnění*. Prospektrum, Praha 1998.
- [4] Hnilica, J., Fotr, J.: *Aplikovaná analýza rizika*. Grada Publishing, Praha 2009.
- [5] Hron, K., Kunderová, P.: *Základy počtu pravděpodobnosti a metod matematické statistiky (1. vydání)*. Univerzita Palackého v Olomouci, 2013.
- [6] Mandl, P., Mazurová, L., Justová, I.: *Matematika a řízení rizik 2009/10*. Matfyzpress, Praha 2010.
- [7] Obdržálková, L.: *Modelování výše škod v čase pomocí kopul*. Diplomová práce. UK Praha. 2009.
- [8] Rafter, John A., Abell, Martha L., Braselton, James P.: *Statistics with Maple (1st edition)*. Academic Press, Amsterdam 2003.
- [9] Sklar, A. *Functions de Repartition an Dimension Set Leursmarges*. Publications de L'Institut de Statistique de L'Universite de Paris, 1959.
- [10] Vose, D.: *Risk Analysis: a quantitative guide (3rd edition)*. John Wiley & Sons, West Sussex 2008.
- [11] Korelace [online]. [cit. 12.11.2016]. Dostupné z:  
<https://cs.wikipedia.org/wiki/Korelace>
- [12] Copula (probability theory) [online]. [cit. 18.2.2017]. Dostupné z:  
[https://en.wikipedia.org/wiki/Copula\\_\(probability\\_theory\)](https://en.wikipedia.org/wiki/Copula_(probability_theory))
- [13] Trojúhelníkové rozdělení [online]. [cit. 21.2.2017]. Dostupné z:  
[https://cs.wikipedia.org/wiki/Trojúhelníkové\\_rozdělení](https://cs.wikipedia.org/wiki/Trojúhelníkové_rozdělení)