

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

BAKALÁŘSKÁ PRÁCE

Stanovení číselných charakteristik rizika pomocí
simulace Monte Carlo



Vedoucí bakalářské práce: **RNDr. Ondřej Pavlačka, Ph.D.**
Vypracovala: **Barbora Sedláčková**
Studijní program: B1103 Aplikovaná matematika
Studijní obor Matematika–ekonomie se zaměřením na bankovníctví
Forma studia: prezenční
Rok odevzdání: 2015

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Barbora Sedláčková

Název práce: Stanovení číselných charakteristik rizika pomocí simulace Monte Carlo

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

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

Rok obhajoby práce: 2015

Abstrakt: První část bakalářské práce je zaměřena na problematiku simulace Monte Carlo. Je zde uveden popis a konstrukce této simulace. Další část je věnována číselným charakteristikám a vysvětlení jejich konstrukce z výsledků simulace. Poslední část práce se zabývá analýzou vlivu počtu iterací z hlediska přesnosti odhadu číselných charakteristik. Jsou zde popsány a na ilustračním příkladu uvedeny postupy výpočtu iterací pro stanovení dostatečné přesnosti odhadu střední hodnoty a parametru p u empirického kvantilu.

Klíčová slova: simulace Monte Carlo, analýza rizika, rizikové faktory, číselné charakteristiky, průměr, směrodatná odchylka, rozdělení pravděpodobnosti, iterace

Počet stran: 44

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Barbora Sedláčková

Title: Determining the quantitative characteristics of a risk by Monte Carlo simulation

Type of thesis: Bachelor's

Department:

Department of Mathematical Analysis and Application of Mathematics

Supervisor: RNDr. Ondřej Pavlačka, Ph.D.

The year of presentation: 2015

Abstract: The first part of the thesis is focused on Monte Carlo simulation. Description and construction of the simulation are presented here. The next part is dedicated to quantitative characteristics and to explanation of their construction from generated data. The last part of the thesis concerns analysis of the influence of the number of iterations regarding the accuracy of estimation of quantitative characteristics. It is described here how to calculate iterations for determining sufficient accuracy of estimates of mean and parameter p of empirical percentile. The problem is illustrated on the demonstrative example.

Key words: Monte Carlo simulation, risk analysis, risk factors, quantitative characteristics, mean, standard deviation, probability distribution, iteration

Number of pages: 44

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem tuto bakalářskou 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

Úvod	7
1 Simulace Monte Carlo	9
1.1 Historie simulace Monte Carlo	9
1.2 Princip simulace Monte Carlo	10
1.3 Popis simulace Monte Carlo	11
1.4 Přednosti a nedostatky simulace Monte Carlo	14
1.5 Konstrukce modelu řešeného problému	14
2 Číselné charakteristiky	18
2.1 Charakteristiky polohy	18
2.1.1 Aritmetický průměr	18
2.1.2 Empirické kvantily	19
2.1.3 Podmíněný průměr	19
2.2 Charakteristiky variability	20
2.2.1 Výběrový rozptyl	20
2.2.2 Výběrová směrodatná odchylka	20
2.2.3 Průměrná absolutní odchylka	21
2.2.4 Semi-variance a semi-směrodatná odchylka	21
2.2.5 Variační rozpětí	22
2.2.6 Kvantilové rozpětí	23
2.2.7 Relativní charakteristiky variability	24
2.3 Charakteristiky tvaru	24
2.3.1 Šikmost	24
2.3.2 Špičatost	26
2.4 Výsledky číselných charakteristik vypočtené z nasimulovaných dat ukázkového příkladu	27
3 Analýza vlivu počtu iterací na přesnost odhadu číselných charakteristik	29
3.1 Posuzování vlivu počtu iterací	29
3.2 Postupy výpočtu iterací pro dostatečnou přesnost odhadu střední hodnoty a parametru p u empirického p -kvantilu	31
3.2.1 Kolik iterací je potřeba pro dostatečnou přesnost odhadu střední hodnoty?	32
3.2.2 Kolik iterací je potřeba pro dostatečnou přesnost odhadu parametru p u empirického p -kvantilu?	40
Závěr	43
Literatura	44

Poděkování

Na tomto místě bych chtěla poděkovat vedoucímu mé bakalářské práce panu RNDr. Ondřeji Pavlačkovi Ph.D. za obětavou spolupráci, trpělivost a za čas, který mi věnoval na konzultacích. Ráda bych také poděkovala mé rodině za velkou podporu při studiu.

Úvod

Simulace Monte Carlo je statistická metoda, která má široké použití v mnoha vědních oborech, mimo jiné i v oblasti analýzy rizika, které se bude týkat má bakalářské práce.

Riziko a také nejistota jsou významnými prvky lidské činnosti, a to především činnosti podnikatelské. Mnohdy firmy stojí před důležitými rozhodnutími, ať už se jedná o vývoj nových produktů, zavádění nových technologií, vstupy na nové trhy či investiční projekty, které mohou významně ovlivnit chod firmy. Budoucí výsledky těchto rozhodnutí jsou však nejisté a mohou se odklánět od výsledků, které byly předpokládány nebo plánované. Proto je důležité analyzovat toto riziko a zhodnotit negativní dopady těchto rizik a nejistot na budoucí výsledky projektů.

Dle [2] můžeme riziko definovat jako možnost (pravděpodobnost) vzniku ztráty, dále také jako možnost výskytu událostí, které by mohly zabránit či ohrozit dosažení cílů jednotlivce či organizace. Riziko můžeme také chápat jako nebezpečí negativních odchylek od stanovených úrovní cílů jednotlivce či organizace.

Problematikou rizika se zabývá analýza rizika, jejíž proces rozčleňujeme do dvou fází. První fáze představuje identifikaci rizika jakožto jevů, událostí, faktorů, které by mohly mít negativní či pozitivní dopady na výsledky činnosti firmy. Cílem identifikace je dosáhnout souboru rizikových faktorů, které mohou jakkoliv (negativně i pozitivně) ovlivňovat hospodářské výsledky firmy. Druhá fáze představuje stanovení velikosti rizika, a to z hlediska dopadů výskytu rizik na výsledky firemních činností, projektů aj. a jejich pravděpodobností. Jedná se o měření rizika, jehož základem je stanovení číselných charakteristik většinou v podobě charakteristik variability. Jednou z hlavních metod aplikovaných v této fázi je simulace Monte Carlo, jíž se tato práce věnuje. [2]

Cílem této práce je seznámit čtenáře s problematikou číselných charakteristik, které budeme konstruovat na základě dat získaných ze simulace Monte Carlo. Především se však budeme zabývat vlivem počtu iterací na přesnost odhadu těchto číselných charakteristik.

V první kapitole se budeme věnovat simulaci Monte Carlo. Popíšeme si její

charakter, ukážeme si, jak postupovat při simulaci a uvedeme si některé její přednosti a nedostatky. V druhé kapitole se zaměříme především na popis číselných charakteristik a vysvětlení jejich konstrukce z výsledků simulace. Třetí kapitola bude zaměřena na analýzu vlivu počtu iterací týkající se přesnosti odhadu číselných charakteristik. Ukážeme si zde postupy výpočtu iterací pro stanovení dostatečné přesnosti odhadu střední hodnoty a parametru p u empirických kvantilů.

1. Simulace Monte Carlo

V této kapitole se seznámíme především s metodou simulace Monte Carlo. Nejdříve si však představíme něco z historie této metody. Poté si popíšeme, v čem tato metoda spočívá a ukážeme si, jak postupovat při konstrukci modelů, založených na této simulaci. Také si představíme ukázkový příklad, na kterém si vysvětlíme postup při simulaci. Tato kapitola byla napsána s využitím literatury [1], [2] a [4].

1.1. Historie simulace Monte Carlo

Simulace, které jsou podobné simulacím Monte Carlo, jsou již známy z počátků počtu pravděpodobnosti. Avšak s prvním opravdovým využitím simulace Monte Carlo, jakožto výpočetního postupu, se můžeme setkat až ve čtyřicátých letech 20. století. Jejimi zakladateli byli vědečtí pracovníci Stanislav Ulam a John von Neumann, kteří v té době pracovali v Los Alamos National Laboratory ve Spojených státech amerických na vývoji jaderné zbraně v rámci projektu Manhattan. Zkoumali zde chování neutronů, a zvláště je zajímalo, jaké množství neutronů projde různými materiály. Tento problém nemohli vyřešit teoreticky ani prakticky, přestože disponovali s důležitými údaji, jako např. s pravděpodobností srážky neutronu s atomem vodíku nebo kyslíku, s množstvím energie, kterou ztrácí neutron během srážky atd. I přes znalost těchto informací nemohli najít metodu pro předpověď dalšího chování neutronu (předpověď „historie jeho života“). K výsledku jim dopomohla až simulace Monte Carlo, kdy k modelování předpovědi „historie života neutronu“ autoři využili techniku kola rulety. Metoda Monte Carlo jim pak pomohla předpovědět trajektorii každého neutronu.

Bylo jim známo, že pohlcení neutronu atomem vodíku při jejich srážce nastane průměrně v jednom případě ze sta možných případů. Každé roztočení kola rulety pak simuluje pohyb neutronu. Jestliže se kolo rulety zastaví na dílku, který znázorňuje pohlcení neutronu atomem vodíku, znamená to „konec života“ neutronu. Simulace „historie života“ neutronu se opakuje vždy tak dlouho, dokud neutron není pohlcen nebo dokud se mu nepodaří úspěšně projít celou cestu.

Simulace Monte Carlo dostala název podle slavných kasin v Monte Carlu, poněvadž princip simulací Monte Carlo zahrnuje prvky nahodilosti a opakování obdobně jako hry v kasinu.

1.2. Princip simulace Monte Carlo

Máme-li více významných faktorů rizika, které ovlivňují výsledky analýzy rizika objektu (firmy, investičních projektů aj.), mnohdy nelze vzhledem k možnému počtu kombinací potenciálních stavů využít běžné nástroje rizika jako např. kvantitativní scénáře, pravděpodobnosti apod. Řešením této situace je využití simulace Monte Carlo. Simulace Monte Carlo spočívá v generování velkého počtu scénářů, řádově stovek až desetitisíců, a v následném propočtu hodnot kvantitativních kritérií (např. finančních) pro každý scénář. Výstup simulace je v podobě grafického zobrazení rozdělení pravděpodobnosti těchto kvantitativních kritérií a jejich číselných charakteristik.

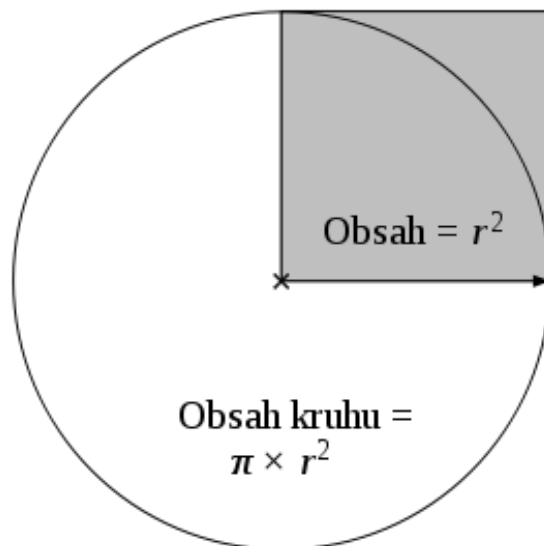
Simulace Monte Carlo jsou dnes využívány především v situacích, kdy nemůžeme předem požadovaný výsledek přímo spočítat (např. ve formě vzorce), a proto je potřebné s pomocí výpočetní techniky postup simulovat.

Typickým příkladem, který nám ilustruje postup při simulaci Monte Carlo, (aniž bychom nezbytně potřebovali výpočetní techniku), je odhad Ludolfova čísla π . Základem tohoto příkladu je nejprve nakreslení čtverce o rozměru 1x1 metr, do kterého vepíšeme kružnici o poloměru 1 metr (obrázek 1).

Následně se budeme do čtverce náhodně střílet např. oblázky, zrnky rýže apod. Jestliže bude rovnoměrně zasahována celá oblast čtverce, potom poměr počtu objektů, které zasahují část čtverce pod kružnicí (označíme si ji písmenem q) k celkovému počtu hodů n do čtverce je roven:

$$\lim_{n \rightarrow \infty} \frac{q}{n} = \frac{\pi \cdot r^2}{4} = \frac{\pi}{4}$$

Poté stačí poměr počtu zásahů $\frac{q}{n}$ vynásobit číslem čtyři a získáme tak číslo π .



Obrázek 1: Geometrický odhad čísla π (obrázek je převzat z [5])

Tento postup by však zabral mnoho času, a jenom za předpokladu velice vysokého počtu „hodů“ bychom dostávali přibližnou hodnotu čísla π . Tento příklad však názorně demonstruje obecný princip simulací Monte Carlo:

1. Nejprve se vymezí soubor možných výstupů, ze kterého se náhodně vybírá, tj. počet objektů zasahujících čtverec (n);
2. Pak dochází k analýze, tzn., zjistíme, jestli se náhodně vybraný objekt nachází v části čtverce, který jsme vymezili kružnicí (q);
3. Nakonec dojde k vyhodnocení všech získaných výsledků, tj. k dosažení čtyřnásobku poměru počtu objektů v části q čtverce pod kružnicí a celkového počtu n objektů ve čtverci, čímž získáme odhad čísla π .

1.3. Popis simulace Monte Carlo

Nyní si popíšeme jednotlivé kroky, do kterých lze simulaci Monte Carlo rozdělit.

1. Určení celkového kontextu finančního modelu, vymezení cílů, tedy stanovení *kritériálních veličin* (výstupů) a dále identifikace *faktorů rizika* (vstupů).

Je potřebné si uvědomit, čeho má být dosaženo, co pomocí modelu chceme získat. Tedy je nutné stanovit si kritériální veličiny, kterými mohou být např. zisk, čistá současná hodnota apod. Dále je také potřebné si uvědomit, v jakém prostředí se budou aktivity pro dosažení cílů realizovat. Poté je teprve možné mluvit o možných náhodných vlivech, které mohou požadovaný či předpokládaný stav ohrozit nebo naopak, které mohou postup k požadovanému cíli usnadnit. Teprve pak je možné provést analýzu rizika.

2. Tvorba matematického modelu projektu (objektů analýzy rizika) a zpracování jeho programu v Microsoft Excelu či jiném matematickém softwaru.

Tvorba matematického modelu představuje v podstatě zkonstruování určitých matematických vzorců, kdy z jednotlivých rizikových faktorů (vstupních veličin) vypočteme hodnotu kvantitativního kritéria (výstupní veličiny). Na základě těchto vzorců poté zpracujeme model např. v MS Excelu.

Tento model má, při analýze rizika firmy nebo určitého finančního projektu, nejčastěji podobu výkazu zisku a ztráty nebo rozvahy peněžních toků. Také může být ve formě vztahů pro výpočet jednotlivých finančních kritérií či ukazatelů pro hodnocení firmy nebo investičního projektu (např. ukazatel rentability, doba návratnosti, čistá současná hodnota apod.).

3. Vytyčení klíčových faktorů rizika.

Klíčovými faktory rizika rozumíme ty vstupní veličiny modelu, které zásadně ovlivňují nejistotu výstupů simulace ve formě finančních ukazatelů či kritérií. Z toho vyplývá, že budeme nejistotu těchto faktorů rizika respektovat, a současně ostatní vstupní veličiny modelu budeme uvažovat jako konstanty ve formě jejich bodových odhadů. Mezi klíčové faktory rizika budou především patřit ta rizika, jejichž změny na výstupy simulace budou velmi citlivé. Nástrojem pro výběr těchto klíčových faktorů může proto sloužit *analýza citlivosti* (viz např. [2]).

4. Stanovení rozdělení pravděpodobnosti klíčových faktorů rizika.

Pokud máme k dispozici historická data určitého faktoru (např. prodeje, měnového kurzu aj.), lze tato data aproximovat některým teoretickým rozdělením. Naopak, jestliže nemáme historická data k dispozici, musíme vycházet ze znalostí a zkušeností expertů z oblastí, kterých se dané faktory rizika týkají.

5. Stanovení statistické závislosti rizikových faktorů.

Hodnoty určitých rizikových faktorů, které modelujeme, mohou být na sobě různě závislé (např. závislost mezi poptávkou po určitém produktu a jeho prodejní cenou, jelikož růst prodejní ceny ovlivní pokles poptávky a naopak). Při simulaci pak nelze generovat nezávisle na sobě tyto hodnoty faktorů, ale je třeba jejich větší či menší závislost respektovat, jinak bychom dostali nepřesné výsledky. Rozlišuje dva typy statistické závislosti faktorů rizika.

- párová závislost = vzájemná závislost dvou faktorů ve stejném období (např. závislost poptávky na prodejní ceně);
- časová závislost = závislost téhož faktoru rizika avšak ve dvou časových obdobích (např. prodej určitého nového produktu v roce jeho nastoupení na trh ovlivňuje prodej tohoto výrobku v letech dalších).

6. Vlastní proces simulace s použitím počítačového programu.

Vlastní proces simulace probíhá tak, že pomocí speciálního softwaru (např. Crystal Ball, MS Excel, Matlab) vygenerujeme možné kombinace hodnot rizikových faktorů z jejich rozdělení pravděpodobnosti, tzn., že vytvoříme určitý scénář a pro každý tento scénář propočteme výstupní hodnotu kritéria modelu objektu analýzy rizika. Po dostatečně velkém počtu iterací získáváme výsledky, které mohou být v podobě:

- grafické - rozdělení pravděpodobnosti výstupní veličiny (finančního kritéria),

- číselné - číselné charakteristiky rizika (průměr, rozptyl, směrodatná odchylka, variační koeficient, šikmost, špičatost nebo také pravděpodobnost překročení či nedosažení požadované hodnoty kritéria aj.).

1.4. Přednosti a nedostatky simulace Monte Carlo

Simulace Monte Carlo je užitečným nástrojem analýzy rizika, která zvyšuje kvalitu významných rozhodnutí za rizika a nejistoty v různých oblastech managementu. Předností simulace Monte Carlo je především to, že dokáže přinutit manažery či další subjekty, které analyzují a hodnotí určité objekty (firmy, složky jejich aktiv, investiční projekty aj.), aby více přemýšleli a analyzovali tyto objekty z hlediska jednotlivých rizikových faktorů, jejich nejistot a vzájemných závislostí. Tato schopnost vede tedy manažery k hlubšímu poznání rizikové stránky objektů a také především k lépe podloženému rozhodování, které se týká např. přijetí či zamítnutí určitých investičních projektů.

Nedostatkem simulace Monte Carlo je obtížnost, která se především týká stanovení rozdělení pravděpodobnosti rizikových faktorů a respektování jejich závislosti. Dále je nevýhodou poskytnutí jen přibližného výsledku, což budeme v práci dále zkoumat a snažit se tento nedostatek eliminovat. Dalším nedostatkem je, že nejvýznamnější rizikové faktory (tedy nejvíce ovlivňující výsledky analýzy rizika), jsou často na základně hodnocení současnosti a minulosti nepředvídatelné. Jelikož se vychází pouze ze známých, v minulosti a přítomnosti vystupujících, faktorů rizika a nedochází k opakované identifikaci rizikových faktorů, může to vést při uplatnění simulace k tzv. *tunelovému efektu*. To vše pak může vést ke kvantifikaci nesprávných rizik, což představuje hlavní riziko simulace.

1.5. Konstrukce modelu řešeného problému

Nejdříve si uvedeme zadání ilustrativního příkladu, na jehož základě si předvedeme konstrukci modelu a na němž budeme objasňovat problémy popsané v následujících částech práce. Tento příklad byl převzat z literatury [2].

Příklad 1 Úkolem je vypracovat analýzu rizika investičního projektu s vy-

užitím simulace Monte Carlo. Cílem projektu je výroba nového produktu, jehož výsledek hospodaření před zdaněním, tedy roční zisk před zdaněním, ovlivňuje celkem šest rizikových faktorů. Vzhledem k tomu, že každý z nich je zatížen určitou nejistotou, byly pro každý tento rizikový faktor zpracovány tři scénáře. Scénář nejpravděpodobnější (výchozí, resp. realistický), který vychází z předpokládaných hodnot jednotlivých faktorů rizika, dále scénář optimistický (v případě značně příznivého vývoje faktoru rizika) a scénář pesimistický (v případě značně nepříznivého vývoje faktoru rizika). Jejich odhadované hodnoty shrnuje následující tabulka.

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

Tabulka 1: Hodnoty rizikových faktorů

Roční výsledek hospodaření (zisk před zdaněním) bude vystupovat v našem příkladu jako kritériální veličina neboli výstupní veličina. Rizikové faktory (vstupní veličiny), které ovlivňují naši výstupní proměnnou, jsou prodej, prodejní cena, měnový kurz, měrná spotřeba materiálu, kupní cena materiálu a fixní náklady. Pokud máme tyto prvky ujasněny, můžeme začít vytvářet matematický model.

Součástí našeho finančního modelu budou následující matematické vztahy. Prvním vztahem bude výpočet ročních tržeb, dále bude následovat vztah pro výpočet nákladů, a to variabilních a celkových, a nakonec vztah pro výpočet zisku před zdaněním.

Roční tržby (T) z prodeje nového produktu se vypočítají jako součin objemu prodeje (P), který je v naturálním vyjádření, dále prodejní ceny (c) a měnového kurzu Kč/euro (m). Dostaneme tedy vztah:

$$T = P \cdot c \cdot m$$

Roční náklady vypočteme jako součet variabilních nákladů (VN) a fixních nákladů (FN):

$$N = VN + FN$$

Fixní náklady nebudeme rozkládat do jednotlivých složek a budeme s nimi pracovat jako s jedinou souhrnnou položkou. Variabilní náklady, které jsou tvořeny náklady na materiál, vypočteme jako součin ročních prodejů, měrné spotřeby materiálu (s) a jeho kupní ceny (k). Dostaneme:

$$VN = P \cdot s \cdot k$$

Nyní můžeme vypočítat roční zisk (Z) z výroby nového produktu jako rozdíl tržeb a nákladů. Vzorec je ve tvaru:

$$Z = T - N$$

Podle těchto čtyř vztahů můžeme snadno zpracovat program v MS Excelu, který nám vypočítá v závislosti na šesti ovlivňujících veličinách (faktorech rizika) tržby, variabilní náklady, celkové náklady a nakonec zisk před zdaněním.

Dále budeme předpokládat, že všechny vstupní veličiny jsou rizikové faktory modelu, tj. prodej, prodejní cena, měnový kurz, norma spotřeby materiálu a fixní náklady.

Pro modelování rozdělení pravděpodobnosti jednotlivých faktorů rizika jsme použili trojúhelníkové rozdělení. Trojúhelníkové rozdělení představuje jedno z nej-používanějších rozdělení pro modelování expertních odhadů. Toto rozdělení má v grafickém zobrazení hustoty pravděpodobnosti tvar trojúhelníka se třemi parametry. Tyto parametry jsou tvořeny nejpravděpodobnější hodnotou, která představuje vrchol trojúhelníka a dále jsou tvořeny dolní a horní mezí. Dolní mez prezentuje pesimistický odhad hodnoty faktoru rizika, horní mez prezentuje optimistický odhad hodnoty faktoru rizika a nejpravděpodobnější hodnotu tvoří nejpravděpodobnější (realistický) odhad.

V případě stanovení statistické závislosti jednotlivých rizikových faktorů jsme v našem simulačním modelu ignorovali závislost mezi těmito rizikovými faktory. Implicitně jsme tedy předpokládali vzájemnou nezávislost rizikových faktorů.

Důležitým krokem v simulaci je zvolit si určitý počet iterací, tedy počet vygenerovaných scénářů. V našem případě jsme si počet iterací zvolili 10 000. Následně tato vygenerovaná data (výstupní data) budeme dále analyzovat a určovat jejich číselné charakteristiky.

Veškeré výpočty jsou uvedeny v příloženém souboru *Simulace + číselné charakteristiky.xlsx* na listě *Simulace*.

2. Číselné charakteristiky

Tato kapitola byla zpracována na základě literatury [2], [3], [4], [6], [7], [8] a [9].

Základním výstupem simulace Monte Carlo jsou nasimulovaná data x_1, \dots, x_n , na jejichž základě usuzujeme o rozdělení pravděpodobnosti kritériální veličiny, kterou analyzujeme, a také z nich konstruujeme odhady číselných charakteristik. Číselné charakteristiky nám pomáhají analyzovat a srovnávat výsledky simulace a pomocí nich můžeme určit rizikovost projektu. V této kapitole se zaměříme na popis jednotlivých číselných charakteristik, které si rozdělíme do tří skupin, a to na charakteristiky polohy, variability a tvaru. Také si ukážeme, jak lze tyto charakteristiky konstruovat z nasimulovaných dat a popíšeme si některé jejich vlastnosti. V příloze *Simulace + číselné charakteristiky.xlsx* na listě *Číselné charakteristiky* jsou uvedeny vypočtené odhady následně popsanych číselných charakteristik.

2.1. Charakteristiky polohy

Charakteristiky polohy nám popisují, jak ze statistického souboru dat (např. z výsledků pozorování) určit hodnotu, kolem které se data soustřeďují. Některé nám pomáhají stanovit jakýsi „střed“ těchto dat. Patří mezi ně především průměr a medián.

2.1.1. Aritmetický průměr

Aritmetický průměr je jednou z nejznámějších charakteristik polohy. Použijeme ho pro výpočet odhadu *střední hodnoty* a značí se \bar{x} . Aritmetický průměr vypočteme jako:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Je nejpoužívanější statistikou v analýze rizika a také je někdy označován jako *očekávaná hodnota*. Pro aritmetický průměr platí následující vlastnosti:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0,$$

$$y_i = x_i + c, i = 1, \dots, n, c \in R \Rightarrow \bar{y} = \bar{x} + c,$$

$$z_i = kx_i, i = 1, \dots, n, k \in R \Rightarrow \bar{z} = k\bar{x}.$$

2.1.2. Empirické kvantily

V případě, že máme k dispozici nasimulovaná data a chceme na jejich základě odhadnout daný kvantil, použijeme empirický kvantil. Symbolem (\tilde{x}_p) neboli empirickým p -kvantilem, kde p je z intervalu $(0, 1)$, nazveme hodnotu, pro kterou platí, že nejméně $100p\%$ čísel v souboru hodnot jsou menší nebo rovna \tilde{x}_p a také nejméně $100(1 - p)\%$ čísel v souboru hodnot jsou větší nebo rovna \tilde{x}_p . Kvantil $\tilde{x}_{0,75}$ se nazývá horní kvartil a $\tilde{x}_{0,25}$ se nazývá dolní kvartil. Často konstruovaným kvantilem je také medián, pro který platí, že hodnota $p = 0,5$ a značí se $\tilde{x}_{0,5}$. Horní a dolní kvartil společně s mediánem rozdělují statistický soubor dat na čtyři stejně početné části.

Pro určení empirického p -kvantilu je nejprve nutné soubor dat setřídít od nejmenší hodnoty po největší:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Indexy $(1), (2), \dots, (n)$ udávají pořadí jednotlivých zjištěných hodnot. Poté pro určení empirického p -kvantilu platí vztah:

$$\tilde{x}_p = \begin{cases} x_{([np]+1)}, & np \neq [np], \\ \frac{x_{(np)} + x_{(np+1)}}{2}, & np = [np]. \end{cases} \quad (1)$$

Symbolem $[\cdot]$ značíme funkci „celá část“. Vzorec byl převzat z literatury [3].

2.1.3. Podmíněný průměr

Podmíněný průměr používáme, když se zajímáme jen o část výstupů, které splňují určitou požadovanou podmínku. Příkladem může být očekávaná ztráta, která vznikne v případě, kdy podnik nevydělává. Podmíněný průměr vypočteme jako průměr, z těch hodnot, které splňují požadovanou podmínku. Například očekávanou ztrátu můžeme tedy vypočítat jako průměr všech negativních zisků.

Označme si A jako množinu dat, které splňují požadovanou podmínku. Potom pro podmíněný průměr použijeme tento vzorec:

$$\bar{x}_A = \frac{1}{\text{card}A} \sum_{i:x_i \in A} x_i,$$

kde $\text{card}A$ značí počet prvků množiny A .

2.2. Charakteristiky variability

Charakteristiky variability vyjadřují koncentraci (variabilitu) hodnot kolem některé míry polohy.

2.2.1. Výběrový rozptyl

Výběrový rozptyl se využívá pro výpočet odhadu *rozptylu*. Představuje jednu z nejpoužívanějších charakteristik variability a je dán vzorcem:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

což vyjadřuje průměr čtverců odchylek všech generovaných hodnot od aritmetického průměru. Čím větší je hodnota výběrového rozptylu, tím větší je variabilita hodnot sledované proměnné. Výběrový rozptyl je vyjádřen ve čtverečných měrných jednotkách. Pokud tedy budeme mít výstup v korunách českých, tak výběrový rozptyl bude měřen v korunách českých na druhou (bez ohledu na to, že takové jednotky nemají žádný význam). Výběrový rozptyl je také velmi citlivý na hodnoty, které se nacházejí na koncích rozdělení, tedy na odlehlá pozorování.

2.2.2. Výběrová směrodatná odchylka

Odhad *směrodatné odchylky* vypočteme pomocí výběrové směrodatné odchylky, a to jako druhou odmocninu z výběrového rozptylu:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}.$$

Výběrová směrodatná odchylka má na rozdíl od výběrového rozptylu tu výhodu, že je ve stejných měrných jednotkách jako výstup. Z definice dále vyplývá, že musí nabývat pouze kladných hodnot a stejně jako výběrový rozptyl je citlivá na odlehlá pozorování.

2.2.3. Průměrná absolutní odchylka

Průměrná absolutní odchylka je definována vztahem:

$$MD = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n - 1}$$

a vyjadřuje průměr absolutních odchylek jednotlivých hodnot od aritmetického průměru. Průměrná absolutní odchylka je vyjádřena ve stejných jednotkách jako výstupní hodnoty.

2.2.4. Semi-variance a semi-směrodatná odchylka

Rozptyl a směrodatnou odchylku používáme jako charakteristiky rizika ve finančním sektoru, jelikož představují nejistotu. Nicméně, v rozdělení pravděpodobnosti toku peněz, velký pozitivní konec rozdělení neboli šance velkého příjmu, nepředstavuje ve skutečnosti riziko.

Semi-směrodatná odchylka a semi-variance kompenzují tento problém tak, že berou v úvahu pouze ty hodnoty pod nebo nad (podle potřeby) tzv. „prahem“. Práh nám vymezuje určitý scénář, který představuje riziko a hodnoty, které toto riziko nepředstavují, jsou vyloučeny. Semi-variance a semi-směrodatná odchylka jsou dány tímto vztahem:

$$V_s = \frac{\sum_{i=1}^k (x_i - x_0)^2}{k},$$

$$s_s = \sqrt{V_s},$$

kde hodnota x_0 je zadaný práh a x_1, \dots, x_k jsou hodnoty ležící pod (nad) danou prahovou hodnotou x_0 .

2.2.5. Variační rozpětí

Variační rozpětí vyjadřuje rozdíl mezi největší a nejmenší hodnotou výstupu:

$$R = x_{max} - x_{min}.$$

Variační rozpětí však není příliš přesnou charakteristikou variability a nedoporučuje se konstruovat z nasimulovaných dat, jelikož velikost krajních (extrémních) hodnot je značně ovlivněna náhodou (hodnoty dosti kolísají). Tento problém si budeme ilustrovat na následujícím příkladu.

Podle zadání příkladu z předchozí kapitoly provedeme 10 simulací při 10 000 iteracích a pro každou simulaci vypočítáme variační rozpětí. Následující tabulka shrnuje hodnoty výsledných variačních rozpětí:

	<i>Variační rozpětí</i>
1	192,982
2	186,465
3	204,766
4	197,968
5	195,245
6	182,856
7	189,971
8	193,363
9	208,313
10	186,941

Tabulka 2: Srovnání hodnot variačního rozpětí v 10 simulacích při 10 000 iteracích v mil. Kč

Z tabulky 2 můžeme vidět, že hodnoty variačního rozpětí opravdu dosti kolísají. Pohybují se již od hodnoty 182,856 mil. Kč až po hodnotu 208,313 mil. Kč. Proto tedy variační rozpětí nepředstavuje příliš přesnou charakteristiku variability.

Výsledky výpočtu jsou uvedeny v příloženém souboru s názvem *Simulace + číselné charakteristiky* na listě *Rozpětí*.

2.2.6. Kvantilové rozpětí

Kvantilové rozpětí vypočítáme jako rozdíl mezi dvěma kvantily:

$$R_D = \tilde{x}_p - \tilde{x}_{1-p},$$

kde pro parametr p platí $0,5 < p < 1$. Kvantilové rozpětí je stabilní charakteristikou variability, což znamená, že ji získáme rychle z poměrně nízkého počtu iterací.

Příkladem kvantilového rozpětí může být:

$$R_D = \tilde{x}_{0,95} - \tilde{x}_{0,05}.$$

Toto rozpětí nám určuje šířku intervalu, ve kterém leží 90 % prostředních hodnot.

Nyní si ukážeme srovnání hodnot tohoto kvantilového rozpětí na příkladu. Použijeme 10 simulací při 10 000 iteracích z ukázkového příkladu z předchozí kapitoly a pro každou simulaci vypočítáme toto kvantilové rozpětí. Následující tabulka shrnuje výsledné hodnoty:

	<i>Kvantilové rozpětí</i>
1	92,046
2	91,440
3	91,340
4	90,848
5	91,039
6	91,206
7	91,709
8	90,944
9	93,081
10	90,933

Tabulka 3: Srovnání hodnot kvantilového rozpětí v 10 simulacích při 10 000 iteracích v mil. Kč

Můžeme vidět, že na rozdíl od variačního rozpětí jsou hodnoty kvantilového rozpětí, uvedené v tabulce 3, stabilnější.

Výsledky výpočtu jsou opět uvedeny v příloženém souboru s názvem *Simulace + číselné charakteristiky* na listě *Rozpětí*.

2.2.7. Relativní charakteristiky variability

Relativní charakteristiky variability se vypočítají jako podíl charakteristik variability a charakteristik polohy. Nejznámější je *variační koeficient*, který vypočítáme jako podíl směrodatné odchylky a aritmetického průměru. Platí tedy pro něj vztah:

$$s_x = \frac{s}{\bar{x}} \cdot 100.$$

Variační koeficient v tomto tvaru nám vyjadřuje, z kolika procent se podílí směrodatná odchylka na aritmetickém průměru. Tato statistika dovoluje srovnávat variabilitu více proměnných, které jsou vyjádřené v různých jednotkách. Dále nám dovoluje srovnávat variabilitu více proměnných s výrazně rozdílnými aritmetickými průměry.

2.3. Charakteristiky tvaru

2.3.1. Šikmost

Šikmost nám udává do jaké míry je rozdělení hodnot kolem zvoleného středu „nerovnoměrné“. Negativní šikmost vyjadřuje, že levý konec rozdělení je delší než pravý, tzn., že data jsou vychýlená k záporným hodnotám. Naopak pozitivní šikmost vyjadřuje, že pravý konec rozdělení je delší než levý, tzn., že data jsou vychýlená ke kladným hodnotám. Nulová šikmost potom znamená, že rozdělení je symetrické kolem aritmetického průměru. Čím více jsou hodnoty charakteristik šikmosti různé od nuly, tím větší asymetrii rozdělení hodnot představují. Jednotlivé situace jsou znázorněny na obrázku 2.

Koeficient šikmosti se vypočítá jako:

$$S = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}.$$

Další charakteristikou šikmosti, která je však využívána jen zřídka, je *kvanti-*

lová šikmost S_p pro niž platí vztah:

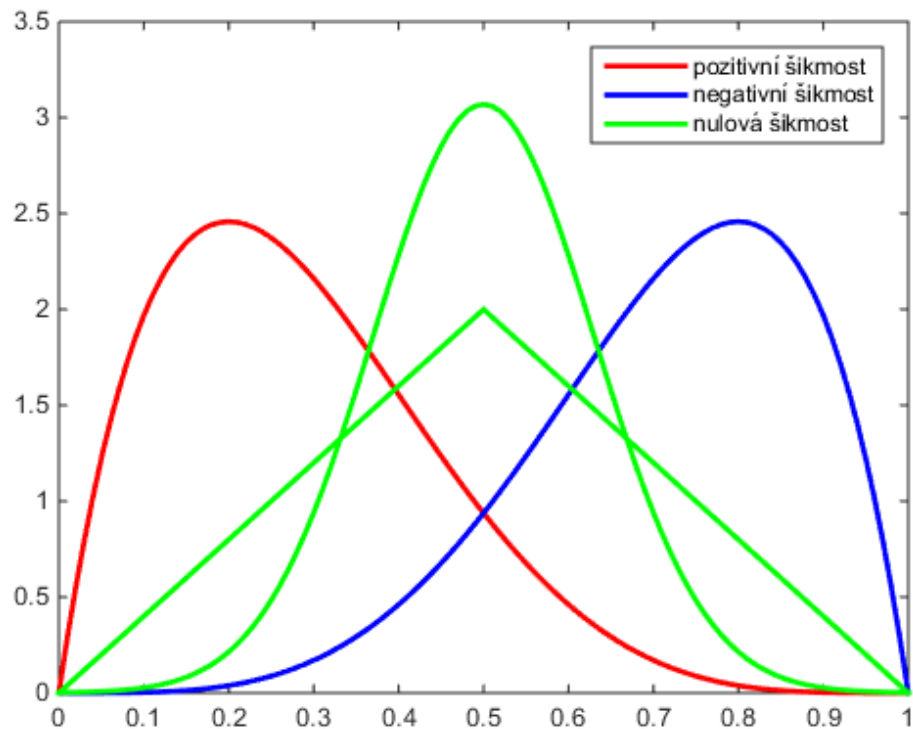
$$S_p = \frac{(\tilde{x}_{0,9} - \tilde{x}_{0,5})}{(\tilde{x}_{0,5} - \tilde{x}_{0,1})}.$$

Tato šikmost má oproti standardní šikmosti tu výhodu, že je dost stabilní, protože není ovlivněna extrémními hodnotami. Kvantilová šikmost nabývá jen kladných hodnot s následující interpretací:

$0 < S_p < 1 \dots$ rozdělení je negativně zešikmené,

$S_p \doteq 1 \dots$ rozdělení je zhruba symetrické,

$S_p > 1 \dots$ rozdělení je pozitivně zešikmené.



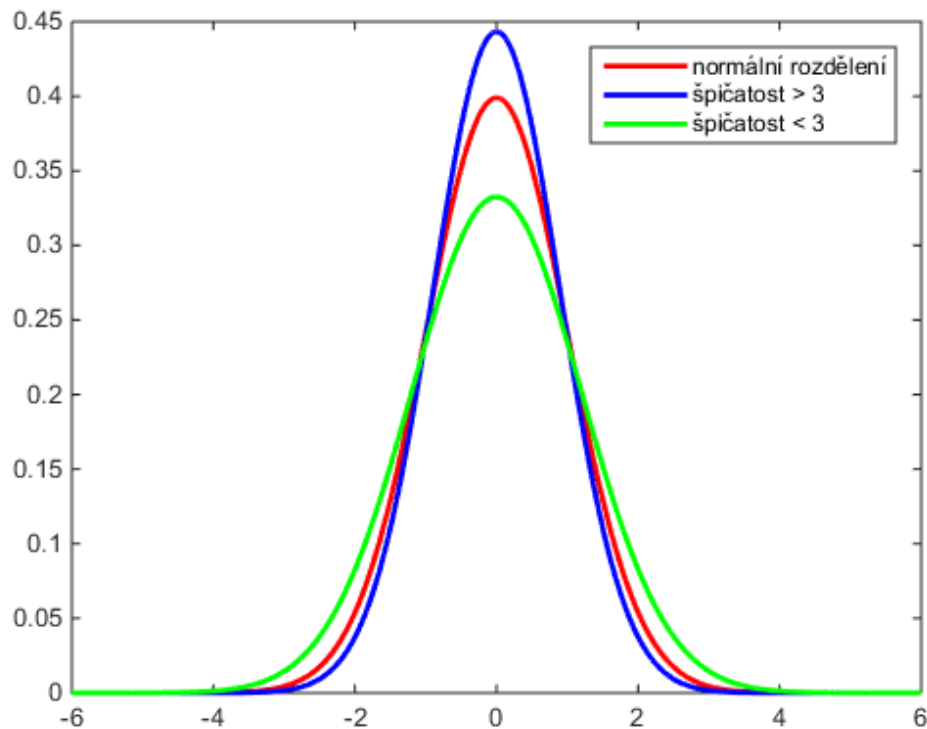
Obrázek 2: Příklady hustot rozdělení pravděpodobností s pozitivní, negativní a nulovou šikmostí

2.3.2. Špičatost

Koeficient špičatosti nám udává míru koncentrace dat kolem středu. Zaměřuje se především na strmost těchto dat a je dán vzorcem:

$$K = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4}.$$

Čím větší hodnotu tohoto koeficientu dostaneme, tím špičatější bude zkoumané rozdělení a tím budou data více koncentrována kolem středu. Naopak nízké hodnoty tohoto koeficientu budou značit větší vzdálenost hodnot od středu. Proto, abychom dostali stabilní hodnoty špičatosti, je potřeba mít větší počet iterací, než pro ostatní charakteristiky (viz [4]).



Obrázek 3: Příklady hustot rozdělení pravděpodobností se špičatostí menší nebo větší než špičatost normálního rozdělení

Špičatost se také někdy používá společně se šikmostí k určení, zda výstupní

hodnoty mají normální rozdělení. Normální rozdělení má špičatost rovnu 3, takže jakékoliv výstupní hodnoty, které vypadají symetricky, ve tvaru zvonu a které mají nulovou šikmost a špičatost rovnu 3, můžeme pravděpodobně považovat za normální.

Jelikož je podle předchozího vzorce špičatost normálního rozdělení rovna 3, používá se pro výpočet koeficientu špičatosti také následující vzorec, pomocí něhož dostaneme špičatost normálního rozdělení rovnu 0:

$$K = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3. \quad (2)$$

Tento vzorec se používá například v MS Excelu.

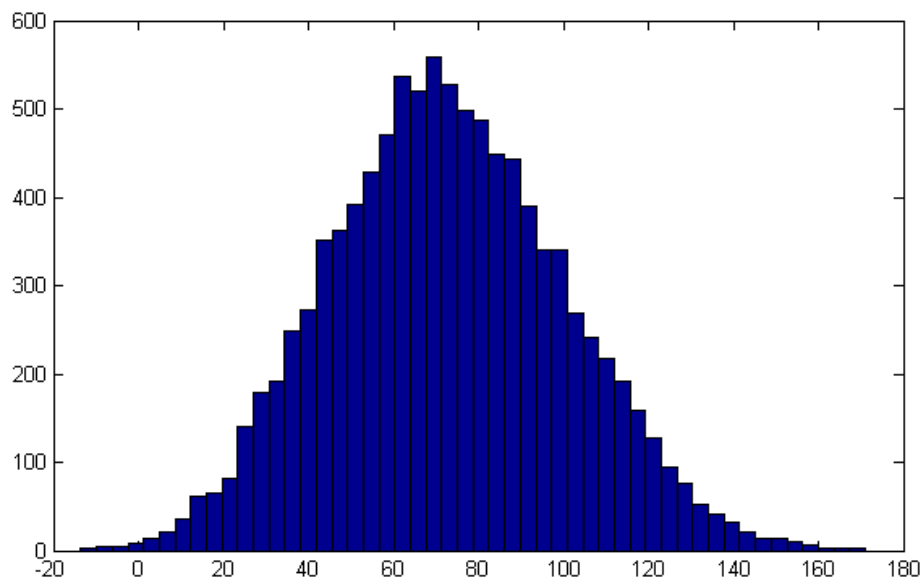
2.4. Výsledky číselných charakteristik vypočtené z nasimulovaných dat ukázkového příkladu

V podkapitole 1.5 jsme si uvedli a popsali postup simulace na ukázkovém příkladu, kde zkoumanou kritériální veličinou byl zisk před zdaněním. Nyní si ukážeme výsledky simulace tohoto příkladu a také výsledky vypočtených odhadů číselných charakteristik.

Základním výstupem (výsledkem) jsou nasimulovaná data získaná ze simulace při 10 000 iteracích, která zachycuje histogram na obrázku 4.

Na základě těchto nasimulovaných dat jsme vypočítali odhady číselných charakteristik a jednotlivé výsledky shrnuje tabulka 4.

Z tabulky 4 vidíme, že například průměrně dosahovaný zisk byl přibližně 73 mil. Kč. Pro výpočet podmíněného průměru jsme uvažovali pouze ztrátu. Tedy můžeme říci, že průměrná ztráta se pohybovala kolem 5 mil. Kč. Maximální zisk byl v tomto případě roven 171,05 mil. Kč a naopak minimální zisk byl -13,5 mil. Kč. Šikmost nám vyšla 0,14, tzn., že hodnoty zisku jsou mírně vychýlené ke kladným hodnotám. Špičatost se rovná hodnotě -0,17. Pro výpočet špičatosti jsme použili vzorec (2), podle něhož má normální rozdělení špičatost rovnu 0. V našem případě je tedy rozdělení zisku méně špičaté ve srovnání s normálním rozdělením.



Obrázek 4: Histogram

Průměr	72,667	Semi-variance	40,594
0,95-kvantil	119,230	Semi-směrodatná odchylka	6,371
0,90-kvantil	109,326	Maximum	171,052
Medián	71,655	Minimum	-13,525
0,10-kvantil	37,326	Variační rozpětí	184,577
0,05-kvantil	28,368	Kvantilové rozpětí	90,862
0,01-kvantil	12,684	Variační koeficient	0,379
Podmíněný průměr	-5,062	Šikmost	0,141
Rozptyl	759,556	Kvantilová šikmost	1,097
Směrodatná odchylka	27,560	Špičatost	-0,167
Průměrná absolutní odchylka	22,116		

Tabulka 4: Číselné charakteristiky

3. Analýza vlivu počtu iterací na přesnost odhadu číselných charakteristik

V této kapitole se zaměříme na problematiku přesnosti odhadů střední hodnoty a parametru p u empirického p -kvantilu v závislosti na počtu iterací, které jsme získali ze simulace našeho ukázkového příkladu popsáno v první kapitole. Budeme tedy analyzovat charakteristiky zisku před zdaněním, který představuje naši kritériální veličinu. Nejdříve si však ukážeme vliv počtu iterací na odhady číselných charakteristik popsanych v předchozí kapitole.

3.1. Posuzování vlivu počtu iterací

Při posuzování vlivu počtu iterací jsme postupovali následovně. Nejdříve jsme si zvolili počet iterací roven 1 000. Provedli jsme 30 simulací a z těchto nasimuovaných dat jsme si vypočítali 30 odhadů jednotlivých číselných charakteristik. Poté jsme pro každou číselnou charakteristiku určili její směrodatnou odchylku. Následně jsme tento postup opakovali i pro 3000 a 10000 iterací. Veškeré výpočty jsou uvedeny v příloženém souboru s názvem *Srovnání číselných charakteristik*.

Z tabulky 5 vidíme, že u charakteristik polohy, tedy u průměru a kvantilů, se s rostoucím počtem iterací hodnoty směrodatné odchylky snižují. Znamená to tedy, že čím větší je počet iterací, tím jsou odhady číselných charakteristik méně rozkolísané. Dále také můžeme vidět, že ze všech kvantilů, co se týká variability, je medián nejvíce stabilní. Jedinou výjimkou je směrodatná odchylka při 3 000 iteracích, kdy nejmenší hodnotu směrodatné odchylky má 0,10-kvantil, avšak v tomto případě jde jen o náhodu.

V případě charakteristik variability vidíme, že se opět s rostoucím počtem iterací hodnoty směrodatné odchylky snižují. Jedinou výjimku tvoří charakteristiky maximum a minimum a tedy i variační rozpětí. Hodnoty těchto charakteristik většinou dosti kolísají. Tento problém je způsoben tím, že maximum a minimum jsou značně závislé na náhodných vlivech, jak bylo uvedeno v podkapitole 2.2.5. Například u charakteristiky maximum je zřejmé, že hodnota směrodatné odchylky

při 3 000 iteracích nejdříve klesla a poté při 10 000 iteracích narostla.

Co se týče charakteristik tvaru, tak z tabulky 5 můžeme vidět, že pro šikmost, kvantilovou šikmost a špičatost platí, že se s rostoucím počtem iterací hodnoty směrodatných odchylek snižují. Opět tedy platí, že čím větší je počet iterací, tím jsou odhady číselných charakteristik méně rozkolísané.

<i>Číselné charakteristiky</i>	<i>n = 1 000</i>	<i>n = 3 000</i>	<i>n = 10 000</i>
Průměr	0,669	0,495	0,237
0,95-kvantil	1,553	0,939	0,518
0,90-kvantil	1,124	0,795	0,418
Medián	0,964	0,711	0,308
0,10-kvantil	1,293	0,666	0,375
0,05-kvantil	1,687	0,863	0,443
0,01-kvantil	3,064	1,507	0,790
Rozptyl	28,918	17,479	9,599
Směrodatná odchylka	0,517	0,315	0,172
Průměrná absolutní odchylka	0,470	0,285	0,157
Maximum	8,176	7,891	8,525
Minimum	6,457	6,321	5,195
Variační rozpětí	9,495	9,238	8,597
Kvantilové rozpětí	2,039	1,117	0,663
Variační koeficient	0,009	0,005	0,002
Šikmost	0,069	0,042	0,018
Kvantilová šikmost	0,060	0,037	0,021
Špičatost	0,152	0,072	0,047

Tabulka 5: Hodnoty směrodatných odchylek číselných charakteristik

V tabulce jsme však neuvažovali podmíněný průměr, semi-varianci a semi-směrodatnou odchylku. Tyto číselné charakteristiky jsme nebrali v úvahu z toho důvodu, že jejich chování záleží na zvolené podmínce.

3.2. Postupy výpočtu iterací pro dostatečnou přesnost odhadu střední hodnoty a parametru p u empirického p -kvantilu

Ke zpracování této kapitoly byla využita literatura [4].

Přesnost odhadů střední hodnoty a parametru p u empirického p -kvantilu nás bude zajímat ze dvou pohledů. Za prvé budeme řešit úlohu, kdy máme danou požadovanou přesnost odhadu dané charakteristiky a budeme chtít zjistit, kolik je potřeba provést iterací, aby jí bylo dosaženo. Druhou možností je pak řešit úlohu, kdy naopak máme daný počet iterací, a úkolem bude zjistit, jaká je „nepřesnost“ výsledného odhadu. Přesnost určité statistiky může být popsána následovně: „Potřebuji odhadnout statistiku Z tak, aby mi vyšla s určitou přesností $\pm \delta$ na hladině pravděpodobnosti α “.

S číselnými charakteristikami se pojí často frekventovaná otázka, kolik je potřeba iterací pro dostatečně přesný odhad těchto charakteristik. Problém je v tom, že v jednom modelu pro dostatečně přesné odhadnutí střední hodnoty může být 500 postačující počet iterací, avšak pro dostatečně přesné určení parametru $p = 0,99$ empirického p -kvantilu v témže modelu je potřeba provést již 10 000 iterací. Jak můžeme vidět, odpověď na tuto otázku není jednoduchá. Proto, abychom si mohli učinit alespoň nějakou představu o rozdělení pravděpodobnosti, je nutné mít minimálně 300 iterací (dle [4]). Při rozhodování o počtu iterací máme obvykle dvě protichůdné tendence:

- Pokud máme příliš málo iterací, tak dostaneme nepřesné výstupní hodnoty.
- Pokud máme příliš mnoho iterací, tak nám simulace modelu může zabrat hodně času a také může být časově náročné vytvoření grafu a analýza výstupních hodnot.

My se tedy v této kapitole budeme snažit určit počet iterací tak, aby byla zajištěna určitá úroveň přesnosti týkající se statistiky, kterou budeme chtít určit z výstupů modelu.

3.2.1. Kolik iterací je potřeba pro dostatečnou přesnost odhadu střední hodnoty?

Následující postup byl převzat z literatury [4].

Odhad skutečné střední hodnoty pomocí simulace Monte Carlo je dán součtem všech generovaných hodnot x_i , který je vydělený počtem iterací n :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Platí zde, že generované hodnoty x_i jsou stejně rozdělené nezávislé náhodné veličiny. S touto charakteristikou jsme se již seznámili ve třetí kapitole a zde jde jen o její připomenutí. *Centrální limitní věta*¹ nám poté říká, že rozdělení pravděpodobnosti odhadu skutečné střední hodnoty je dáno:

$$\bar{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right), \quad (3)$$

kde μ představuje skutečnou střední hodnotu a σ je skutečná směrodatná odchylka.

Vztah (3) můžeme napsat také jako:

$$\bar{x} \approx \mu + N\left(0, \frac{\sigma}{\sqrt{n}}\right).$$

Poté si z tohoto vztahu vyjádříme skutečnou střední hodnotu μ . Jelikož je rozdělení $N\left(0, \frac{\sigma}{\sqrt{n}}\right)$ symetrické kolem 0, můžeme psát:

$$\mu \approx \bar{x} - N\left(0, \frac{\sigma}{\sqrt{n}}\right) = \bar{x} + N\left(0, \frac{\sigma}{\sqrt{n}}\right).$$

¹ *Nechť $\{X_n\}_1^\infty$ je posloupnost nezávislých, stejně rozdělených náhodných veličin s konečnou střední hodnotou μ a s konečným nenulovým rozptylem σ^2 . Potom posloupnost distribučních funkcí normovaných součtů*

$$Z_n = \frac{\sum_1^n X_k - E(\sum_1^n X_k)}{\sqrt{\text{var}(\sum_1^n X_k)}} = \frac{\sum_1^n (X_k - \mu)}{\sigma\sqrt{n}}$$

konverguje na R k distribuční funkci normovaného normálního rozdělení $N(0,1)$, tj.

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_1^n (X_k - \mu)}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x), \forall x \in R.$$

Věta převzata z [3].

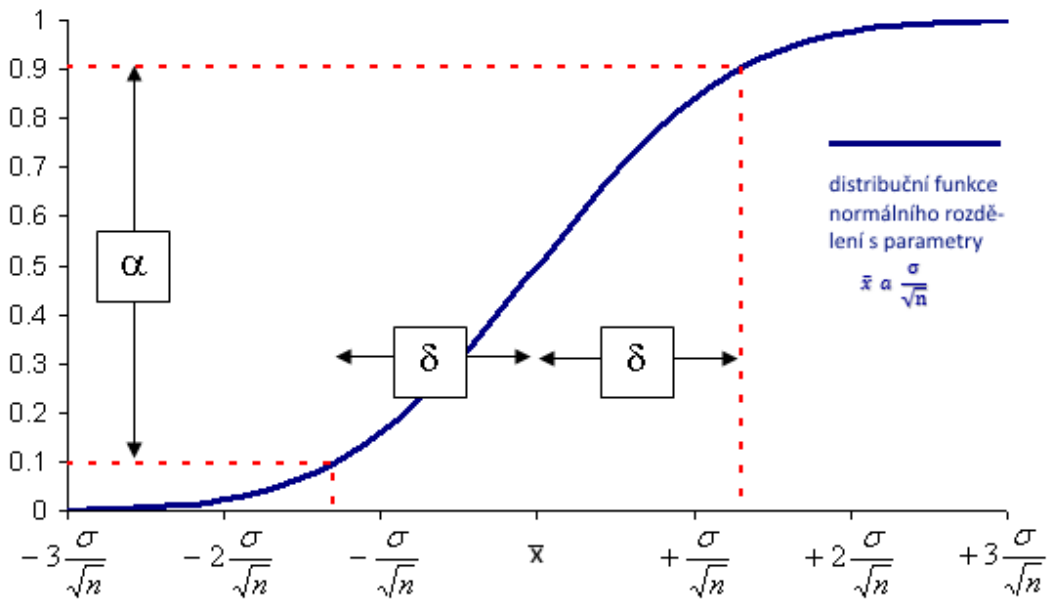
Opět lze tento vztah přepsat do tohoto tvaru:

$$\mu \approx N\left(\bar{x}, \frac{\sigma}{\sqrt{n}}\right). \quad (4)$$

Obrázek 5 znázorňuje distribuční funkci normálního rozdělení s parametry \bar{x} a $\frac{\delta}{\sqrt{n}}$ pro vztah (4). Také je na tomto obrázku znázorněno stanovení určité úrovně přesnosti pro odhad střední hodnoty na určité hladině pravděpodobnosti α , které se promítá do vztahu mezi δ , σ a n . Tento vztah můžeme zapsat jako:

$$\delta = \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{1+\alpha}{2}\right), \quad (5)$$

kde Φ^{-1} je inverzní funkce k distribuční funkci normovaného normálního rozdělení $N(0, 1)$ a α představuje uvažovanou hladinu pravděpodobnosti.



Obrázek 5: Graf ilustrující vztah mezi δ , σ , α a n pro normální rozdělení

Nyní z rovnice (5) zjistíme minimální počet iterací, které potřebujeme pro určitou úroveň přesnosti odhadu střední hodnoty:

$$n = \left\lceil \left(\frac{\sigma}{\delta} \Phi^{-1}\left(\frac{1+\alpha}{2}\right) \right)^2 \right\rceil + 1, \quad (6)$$

kde symbolem $[\cdot]$ značíme funkci „celá část“.

Jediným problémem je, že neznáme skutečnou hodnotu směrodatné odchyly. Avšak dle literatury [4] lze směrodatnou odchytku velmi dobře odhadnout z prvních 50 iterací.

Nyní se podíváme, jak odhadnout minimální počet iterací n potřebný pro dostatečnou přesnost odhadu střední hodnoty v praxi, pokud neznáme skutečnou hodnotu směrodatné odchyly σ . V literatuře [4], jak jsme uvedli výše, je doporučeno provést odhad směrodatné odchyly z 50 iterací. Na tomto základě jsme v případě našeho ukázkového příkladu provedli 10 simulací o 50 iteracích, z kterých jsme vypočítali 10 odhadů směrodatné odchyly σ . Výsledky těchto odhadů shrnuje následující tabulka.

	<i>Směrodatná odchytky</i>
1	25,220
2	27,312
3	29,920
4	25,485
5	30,563
6	28,915
7	29,788
8	27,100
9	25,413
10	28,291

Tabulka 6: Hodnoty směrodatné odchyly při 50 iteracích

Z tabulky 6 vidíme, že rozpětí hodnot směrodatné odchyly je poměrně velké, z toho tedy vyplývá, že odhad směrodatné odchyly je málo přesný. To se následně promítne i do odlišnosti počtu iterací. Situaci budeme ilustrovat na dvou hladinách pravděpodobnosti, například $\alpha = 0,95$ a $\alpha = 0,975$, pro δ si zvolíme hodnotu 1 mil. Kč. Pro výpočet hodnoty Φ^{-1} jsme v MS Excelu použili funkci *NORM.S.INV*, která nám vrátí inverzní funkci k distribuční funkci normovaného normálního rozdělení. Nyní známe všechny potřebné hodnoty a můžeme dosadit do rovnice (6) pro zjištění minimálního počtu iterací.

V tabulce 7 máme hodnoty minimálního počtu iterací, které jsou nezbytné

pro určitou úroveň správnosti odhadu střední hodnoty. Vidíme, že hodnoty pro $\alpha = 0,95$ se pohybují od 2 444 iterací až po hodnotu 3 589, tedy rozpětí je velké. Pro $\alpha = 0,975$ jsou potřebné již vyšší počty iterací a hodnoty se pohybují nyní od 3 196 iterací až po hodnotu 4 693, opět je zde vidět velké rozpětí.

1.	$\alpha = 0,95, \delta = 1 \text{ mil.}$	2.	$\alpha = 0,975, \delta = 1 \text{ mil.}$
1	2444	1	3196
2	2866	2	3748
3	3439	3	4498
4	2496	4	3264
5	3589	5	4693
6	3212	6	4201
7	3409	7	4458
8	2822	8	3690
9	2481	9	3245
10	3075	10	4021

Tabulka 7: Hodnoty minimálního počtu iterací při použití směrodatné odchytky vypočtené z 50 iterací

Jelikož odhady směrodatné odchytky z 50 iterací se ukázaly jako velmi rozptýlené, uvažovali jsme tedy dále, že pro dobrý odhad směrodatné odchytky bude stačit 100 iterací. Opět jsme provedli 10 simulací, tentokrát o 100 iteracích, a vypočítali 10 odhadů směrodatné odchytky. Výsledky jsme shrnuli do tabulky 8.

	<i>Směrodatná odchytká</i>
1	26,401
2	27,672
3	29,645
4	28,527
5	27,252
6	27,430
7	27,880
8	26,701
9	29,883
10	27,792

Tabulka 8: Hodnoty směrodatné odchytky při 100 iteracích

Vidíme, že oproti výpočtu odhadu směrodatné odchylky z 50 iterací jsou nyní hodnoty méně rozptýlené. Zmenšení rozpětí těchto hodnot se projeví i v počtu iterací, jelikož tyto hodnoty již nebudou tak rozkolísané, což ukazuje tabulka 9. V této tabulce můžeme pozorovat, že hodnoty počtu iterací pro $\alpha = 0,95$ se pohybují od 2 678 až po hodnotu 3 431, z toho vyplývá, že se nám rozkolísanost jen mírně zmenšila. Pro $\alpha = 0,975$ jsou opět potřebné vyšší hodnoty počtu iterací a pohybují se od 3 502 iterací až po hodnotu 4 487.

1.	$\alpha = 0,95, \delta = 1 \text{ mil.}$	2.	$\alpha = 0,975, \delta = 1 \text{ mil.}$
1	2678	1	3502
2	2942	2	3847
3	3376	3	4416
4	3127	4	4089
5	2853	5	3731
6	2891	6	3780
7	2986	7	3905
8	2739	8	3582
9	3431	9	4487
10	2968	10	3881

Tabulka 9: Hodnoty minimálního počtu iterací při použití směrodatné odchylky vypočtené ze 100 iterací

Jelikož jsou opět hodnoty směrodatné odchylky i hodnoty jednotlivých počtů iterací rozkolísané, uvažovali jsme ještě 300 iterací. Postup výpočtu byl totožný jako v předchozích dvou případech. Výsledky odhadu směrodatné odchylky ukazuje tabulka 10.

Z této tabulky je zřejmé, že hodnoty směrodatné odchylky jsou opět méně rozptýlené, což se projeví i menším rozpětím v počtu iterací.

V tabulce 11 vidíme, že hodnoty počtu iterací pro $\alpha = 0,95$ se pohybují od 2 554 až po hodnotu 3 024 a mají menší rozpětí než v předchozích dvou případech. Pro $\alpha = 0,975$ jsou opět potřebné vyšší hodnoty počtu iterací a pohybují se od 3 341 iterací až po hodnotu 4 030.

	<i>Směrodatná odchylka</i>
1	27,889
2	27,731
3	28,166
4	28,056
5	27,188
6	28,116
7	28,322
8	25,784
9	26,125
10	26,425

Tabulka 10: Hodnoty směrodatné odchylky při 300 iteracích

1.	$\alpha = 0,95, \delta = 1 \text{ mil.}$	2.	$\alpha = 0,975, \delta = 1 \text{ mil.}$
1	2988	1	3908
2	2955	2	3864
3	3048	3	3986
4	3024	4	3955
5	2840	5	3714
6	3037	6	3972
7	3082	7	4030
8	2554	8	3341
9	2622	9	3429
10	2683	10	3509

Tabulka 11: Hodnoty minimálního počtu iterací při použití směrodatné odchylky vypočtené z 300 iterací

Je zřejmé, že i při odhadu směrodatné odchylky z 300 iterací jsou hodnoty rozkolísané, proto při nastavení počtu iterací, které jsou potřebné pro dostatečnou přesnost odhadu střední hodnoty, doporučujeme postupovat následovně. Opět budeme předpokládat, že pro dobrý odhad směrodatné odchylky bude stačit 50 iterací. Provedeme 10 simulací, na jejichž základě vypočteme 10 odhadů směrodatné odchylky. Nyní však z těchto 10 odhadů vypočítáme průměr směrodatné odchylky. Tabulka 12 porovnává průměry odhadů směrodatných odchylek vypočtených jak z 50 iterací, tak i ze 100 a 300 iterací.

	50	100	300
průměrná směrodatná odchylka	27,801	27,918	27,380

Tabulka 12: Průměry odhadů směrodatných odchylek

Hodnoty průměrů směrodatných odchylek jsou si velmi podobné, což značí o tom, že stačí průměrovat hodnoty směrodatných odchylek pouze z 50 iterací.

Nyní zvolíme opět hladiny pravděpodobnosti $\alpha = 0,95$ a $\alpha = 0,975$, pro $\delta = 1$ mil. a můžeme dosadit do rovnice (6). Výsledky počtu iterací potřebných pro dostatečnou přesnost odhadu střední hodnoty shrnuje tabulka 13, v níž jsou porovnávány opět možnosti výpočtu odhadu směrodatné odchylky z 50, 100 a 300 iterací.

	50	100	300
$\alpha = 0,95; \delta = 1$ mil. Kč	2969	2995	2880
$\alpha = 0,975; \delta = 1$ mil. Kč	3883	3916	3767

Tabulka 13: Minimální počty iterací potřebné pro dostatečnou přesnost odhadu střední hodnoty

Z tabulky vidíme, že minimální počty iterací jsou již málo rozkolísané. Proto tedy při zjišťování počtu iterací potřebných pro dostatečnou přesnost odhadu střední hodnoty nám bude stačit odhadnout směrodatnou odchylku jako průměr 10 směrodatných odchylek vypočítaných z 10 simulací o 50 iteracích.

Veškeré výpočty nalezneme v příloženém souboru s názvem *Analýza vlivu počtu iterací.xlsx*.

Nyní budeme řešit naopak situaci, kdy známe počet iterací n a naším úkolem bude zjistit, jaká je „nepřesnost“ výsledného odhadu. Pro tuto situaci použijeme rovnici (5):

$$\delta = \frac{\sigma}{\sqrt{n}} \Phi^{-1} \left(\frac{1 + \alpha}{2} \right).$$

V příloženém souboru s názvem *Analýza vlivu počtu iterací.xlsx* na listu *delta pro stř. hodnotu I.* je uveden postup určení δ pro počty iterací, které jsme stanovili v předchozím případě.

Tento postup si však představíme na příkladu, kde jsme si zvolili pro simulaci počet iterací nejprve 10 000 a poté 3 000 a 1 000. Známe tedy počty iterací, nyní si vypočteme z těchto nasimulovaných dat odhady směrodatných odchylek a také průměrné zisky:

	10 000	3 000	1 000
průměrný zisk	72,667	71,687	71,435
směrodatná odchylka	27,560	27,429	28,034

Tabulka 14: Hodnoty průměrných zisků a směrodatných odchylek

Zvolíme si hladiny pravděpodobnosti $\alpha = 0,95$, $\alpha = 0,975$, $\alpha = 0,99$ a pro výpočet hodnoty Φ^{-1} použijeme funkci *NORM.S.INV*, která nám vrátí inverzní funkci k distribuční funkci normovaného normálního rozdělení. Nyní již můžeme dosadit do rovnice (5) a dostaneme hodnotu *delta* pro uvažované hladiny pravděpodobnosti. Přičtením a odečtením hodnoty *delta* od průměrného zisku získáme *interval* odhadů středních hodnot, pro jejichž dostatečně dobré odhady stačí 10 000, 3 000 a 1 000 iterací. Výsledky shrnuje tabulka 15.

n = 10 000		
	<i>delta</i>	<i>interval</i>
$\alpha = 0,95$	0,540	72,127; 73,207
$\alpha = 0,975$	0,618	72,049; 73,285
$\alpha = 0,99$	0,710	71,957; 73,377
n = 3 000		
	<i>delta</i>	<i>interval</i>
$\alpha = 0,95$	0,982	70,705; 72,668
$\alpha = 0,975$	1,122	70,564; 72,809
$\alpha = 0,99$	1,290	70,397; 72,977
n = 1 000		
	<i>delta</i>	<i>interval</i>
$\alpha = 0,95$	1,738	69,698; 73,173
$\alpha = 0,975$	1,987	69,448; 73,422
$\alpha = 0,99$	2,284	69,152; 73,719

Tabulka 15: Intervaly odhadů středních hodnot

3.2.2. Kolik iterací je potřeba pro dostatečnou přesnost odhadu parametru p u empirického p -kvantilu?

Následující postup byl převzat z literatury [4].

Empirické p -kvantily, kde p se blíží 0,5, budou dosahovat stabilní hodnoty daleko rychleji než empirické p -kvantily, kde p se blíží extrémním hodnotám 0 a 1. Na druhé straně nás právě hodnoty blížíící se konci rozdělení výstupů často zajímají více, jelikož zde leží rizika a příležitosti. Následující postup nám ukáže, jak můžeme zaručit, že dostaneme požadovanou úroveň přesnosti pro parametr p u empirického p -kvantilu.

Odhad parametru p u empirického p -kvantilu spojeného s určitou hodnotou x z rozdělení výstupů se provede tak, že se určí, jaký podíl iterací simulace Monte Carlo spadá pod nebo nad hodnotu x . Představme si, že nyní hodnota x představuje skutečnou hodnotu 80% kvantilu. Poté pro simulaci Monte Carlo platí, že vygenerované hodnoty (získané nezávisle na simulaci kteréhokoli modelu) s 80% pravděpodobností spadají pod hodnotu x . Tedy, pokud máme n iterací a s hodnot, které spadají pod nebo nad hodnotu x , rozdělení pravděpodobnosti $Beta(s + 1, n - s + 1)$ popisuje nejistotu související se skutečným parametrem p .

Jestliže budeme odhadovat parametr p , který je blízko hodnotě 0,5 (mediánu) nebo budeme-li provádět velký počet iterací, tedy jestliže hodnoty s a n budou velké, můžeme použít aproximaci normálním rozdělením k beta rozdělení:

$$Beta(s + 1, n - s + 1) \approx N \left(\hat{P}, \sqrt{\frac{\hat{P}(1 - \hat{P})}{n}} \right), \quad (7)$$

kde hodnota $\hat{P} = \frac{s}{n}$ představuje nejlepší odhad pro parametr p . Nyní můžeme odvodit vztah, který je podobný vztahu v rovnici (5) z předchozí podkapitoly:

$$\delta = \sqrt{\frac{\hat{P}(1 - \hat{P})}{n}} \Phi^{-1} \left(\frac{1 + \alpha}{2} \right). \quad (8)$$

Následně z rovnice (8) zjistíme n , které nám vyjadřuje minimální počet iterací

potřebný pro požadovanou přesnost odhadu parametru p :

$$n = \left[\hat{P}(1 - \hat{P}) \left(\frac{\Phi^{-1} \left(\frac{1+\alpha}{2} \right)}{\delta} \right)^2 \right] + 1, \quad (9)$$

kde symbolem $[\cdot]$ značíme funkci „celá část“.

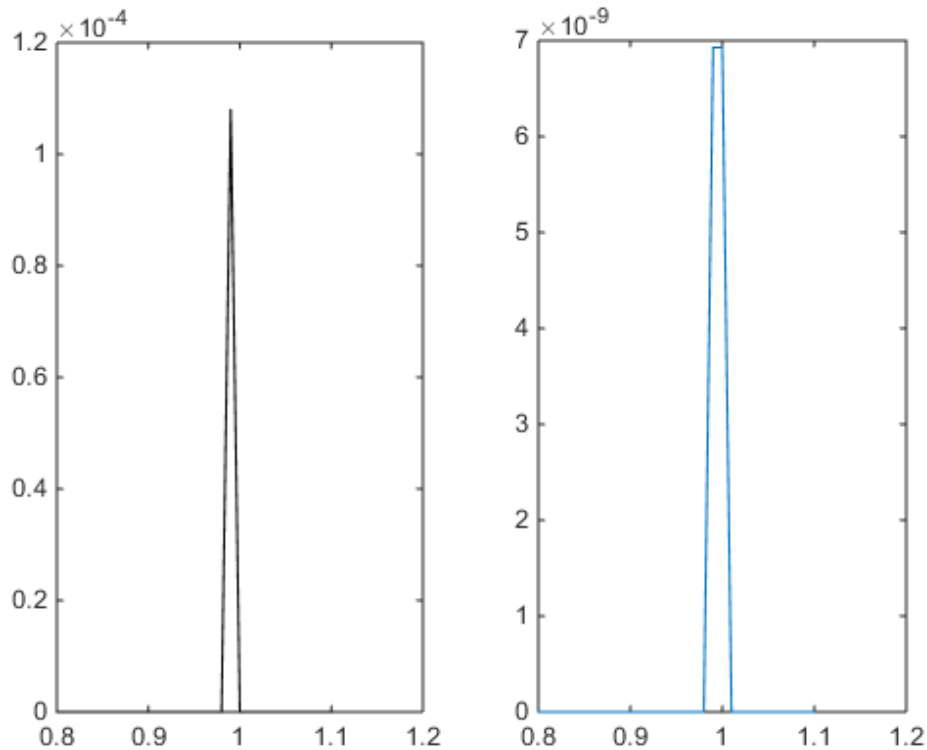
V příloženém souboru s názvem *Analýza vlivu počtu iterací* máme na listu *n_min pro kvantil* soubor umožňující výpočet minimálního počtu iterací potřebných pro dostatečnou přesnost odhadu parametru $p = 0,995; 0,5$ a $0,095$ u empirického p -kvantilu. Nejprve jsme si zvolili $\delta = 0,005$, tedy budeme se snažit najít minimální počet iterací pro to, aby na určité hladině pravděpodobnosti představoval vypočtený empirický kvantil skutečný p -kvantil, kde p je z intervalu $\langle 0,99; 1 \rangle$ pro odhad $p = 0,995$, $\langle 0,495; 0,505 \rangle$ pro odhad $p = 0,5$ a $\langle 0,09; 0,1 \rangle$ pro odhad $p = 0,095$. Hladinu pravděpodobnosti jsme si zvolili $\alpha = 0,99$. Nyní můžeme dosadit do rovnice (9) a vyjdou nám minimální počty iterací:

<i>Parametr</i>	<i>Počet iterací</i>
p=0,995	1 321
p=0,5	66 349
p=0,095	22 818

Tabulka 16: Minimální počty iterací pro přesnost odhadu parametru p

Z tabulky je zřejmé, že pro odhad parametru $p = 0,995$ je potřeba jen velmi malý počet iterací. Tento vypočtený údaj však není přesný. Na začátku této kapitoly jsme uvedli, že nejistotu spojenou s odhadem skutečného parametru p popisuje beta rozdělení. Na obrázku 6 vlevo můžeme vidět hustotu beta rozdělení s parametry $(s + 1, n - s + 1)$ odpovídající $p = 0,995$ při 10 000 iteracích. Proto, abychom mohli odvodit vztah pro výpočet odhadu parametru p , jsme beta rozdělení aproximovali pomocí normálního rozdělení $N \left(\hat{P}, \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right)$. Avšak pro $p = 0,995$ nemůžeme tuto aproximaci provést, jelikož dle obrázku 6 vpravo vidíme, že graf hustoty odpovídajícího normálního rozdělení se výrazně liší. Jelikož nelze aproximaci normálním rozdělením provést, nemůžeme tedy obecně pro extrémní parametry použít vzorec (9) pro výpočet minimálního počtu iterací.

Pro odhad $p = 0,5$ jsme vypočítali, že je potřeba 66 349 iterací. Tento výsledek plyne ze vztahu (7), kde pro $\hat{P} = 0,5$ má normální rozdělení (a tedy i příslušné beta rozdělení) největší rozptyl. Pro dostatečně přesný odhad tohoto parametru je tedy potřeba velký počet iterací.



Obrázek 6: Hustota beta rozdělení s parametry (9 951; 51) a hustota normálního rozdělení s parametry (0, 995; 0, 000705)

V tomto souboru na listu *delta pro kvantil* nalezneme také opačný postup. Nyní známe počet iterací, například $n = 100\,000$, dále $\alpha = 0,99$ a naším úkolem je zjistit interval parametrů kolem $p = 0,5$, pro jejichž dostatečně dobré odhady stačí 100 000 iterací. Dosadíme vše do rovnice (8) a vyjde nám hodnota $\delta = 0,00407$. Hledaný interval je $(0,49593; 0,50407)$, tzn., že na hladině $\alpha = 0,99$ při 100 000 iterací představuje vypočtený empirický kvantil skutečný p -kvantil, kde p je hodnota z tohoto intervalu.

Závěr

V rámci bakalářské práce jsme se seznámili s problematikou simulace Monte Carlo. Nejdříve jsme si uvedli něco z historie této metody, poté jsme se zaměřili na její princip a také na to, jak bychom měli postupovat při její konstrukci. Popis jednotlivých kroků simulace jsme ilustrovali na ukázkovém příkladu, jehož výsledkem byla nasimulovaná data, s kterými jsme následně pracovali.

Dále jsme se v práci věnovali číselným charakteristikám. Na základě nasimulovaných dat jsme si popsali konstrukci jednotlivých číselných charakteristik, které jsme si rozdělili do tří skupin, a to na charakteristiky polohy, variability a tvaru. Poté jsme si představili výsledky ukázkového příkladu pomocí histogramu a také výsledky příslušných číselných charakteristik.

Stěžejní část práce jsme zaměřili na analýzu vlivu počtu iterací. Nejprve jsme posuzovali vliv počtu iterací u charakteristik polohy, variability a tvaru a ověřili jsme, že s rostoucím počtem iterací jsou odhady číselných charakteristik méně rozkolísané. Dále jsme se zaměřili na zjištění počtu iterací, který je potřebný pro dostatečnou přesnost odhadu střední hodnoty. Zanalyzovali jsme postup popsany v literatuře [4] a došli jsme k závěru, že není úplně věrohodný. Proto byl navržen postup, který nám dává již stabilnější hodnoty směrodatných odchylek. Nakonec jsme se zabývali určením počtu iterací, který je potřebný pro dostatečnou přesnost odhadu parametru p u empirického kvantilu. Zjistili jsme, že pro dostatečně přesný odhad $p = 0,5$ je potřeba velký počet iterací. Naopak pro odhad $p = 0,995$ a celkově pro extrémní hodnoty parametru jsme zjistili, že nelze použít aproximaci normálním rozdělením, jelikož hustota normálního rozdělení odpovídající tomuto parametru se výrazně liší od hustoty beta rozdělení pro parametr p tohoto kvantilu.

Zpracování bakalářská práce mi přineslo mnoho nových zkušeností a znalostí, a to nejen v problematice číselných charakteristik. Díky jejímu vypracování jsem prohloubila své znalosti matematických softwarů Matlab a Microsoft Excel. Také jsem se lépe seznámila s typografickým systémem \TeX , ve kterém je tato práce napsána.

Literatura

- [1] Fabian, F., Klumber, Z.: *Metoda Monte Carlo a možnosti jejího uplatnění*. Prospektrum, Praha 1998.
- [2] Hnilica, J., Fotr, J.: *Aplikovaná analýza rizika*. Grada Publishing, Praha 2009.
- [3] Hron, K., Kunderová, P.: *Základy počtu pravděpodobnosti a metod matematické statistiky (1. vydání)*. Univerzita Palackého v Olomouci, 2013.
- [4] Vose, D.: *Risk Analysis: a quantitative guide (3rd edition)*. John Wiley & Sons, West Sussex 2008.
- [5] Pí (číslo) [online]. [cit. 1.11.2014]. Dostupné z:
[http://cs.wikipedia.org/wiki/Pí_\(číslo\)](http://cs.wikipedia.org/wiki/Pí_(číslo))
- [6] Charakteristiky variability (proměnlivosti souboru) [online]. [cit. 1.2.2015]. Dostupné z:
<http://cit.vfu.cz/statpotr/POTR/Teorie/Predn1/variabil.htm>
- [7] Číselné charakteristiky a jejich výpočet [online]. [cit. 1.2.2015]. Dostupné z:
http://k101.unob.cz/neubauer/pdf/ciselne_charakteristiky.pdf
- [8] Základní statistické zpracování dat [online]. [cit. 7.2.2015]. Dostupné z:
<http://home.zcu.cz/sediva/stav/kap01.pdf>
- [9] Popisné statistiky [online]. [cit. 11.3.2015]. Dostupné z:
http://amper.ped.muni.cz/jenik/nejistoty/html_tree/node13.html