

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

DIPLOMOVÁ PRÁCE

Modelování nejistoty parametrů
pravděpodobnostních rozdělení v simulacích
Monte Carlo



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

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Adriana Crhonková

Název práce: Modelování nejistoty parametrů pravděpodobnostních rozdělení v simulacích 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: V rámci simulace Monte Carlo se setkáváme s důležitým krokem, a tím je určení pravděpodobnostního rozdělení rizikových faktorů. Pokud disponujeme historickými daty, odhadneme toto rozdělení z nich. Pro následnou simulaci rizikového faktoru se pak nabízí spočítat bodové odhady parametrů zvoleného rozdělení pravděpodobnosti a z nich modelovat uvažovaný rizikový faktor. Jaké nejistotě ovšem čelíme, použijeme-li k simulaci místo skutečných parametrů pouze bodové odhady? Hlavní náplní práce je zkoumat modelování této nejistoty pomocí klasické statistiky, metody Bootstrap a Bayesovy statistiky.

Klíčová slova: Simulace Monte Carlo, analýza rizika, rizikové faktory, rozdělení pravděpodobnosti, modelování nejistoty parametrů, klasická statistika, metoda Bootstrap, Bayesova statistika

Počet stran: 57

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Adriana Crhonková

Title: Modelling of uncertainty of parameters of probability distributions 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: In a Monte Carlo simulation, there is an important step - how to determine probability distribution of risk factors. If we possess historical data, we can estimate this distribution from these. For the following simulation of a risk factor we can count point estimations of parameters of chosen probability distribution and then simulate the risk factor from these estimators. However, what is the uncertainty we face, when we use these point estimators instead of real parameters? The main content of this theses is to examine modelling of this uncertainty by utilizing the classical statistics approach, the Bootstrap, and the Bayesian inference.

Key words: Monte Carlo simulation, risk analysis, risk factors, probability distribution, modelling of uncertainty of parametres, classical statistics, the Bootstrap, Bayesian inference

Number of pages: 57

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením pana RNDr. Ondřeje Pavlačky, Ph.D. s použitím uvedené literatury.

V Olomouci dne
.....
podpis

Obsah

Úvod	7
1 Metoda Monte Carlo	9
1.1 Historie	9
1.2 Postup simulace Monte Carlo	9
1.3 Konstrukce ilustračního příkladu	11
2 Statistická analýza historických dat	15
2.1 Úvod	15
2.2 Metody odhadu pravděpodobnostních rozdělení	15
2.2.1 Neparametrické metody	16
2.2.2 Parametrické metody	17
3 Modelování nejistoty parametrů pravděpodobnostních rozdělení v simulacích Monte Carlo	18
3.1 Úvod do problematiky	18
3.2 Klasická statistika	19
3.2.1 Ilustrační příklad - odvození důležitých vztahů	19
3.2.2 Simulace rizikového faktoru <i>Prodej</i> z 12-ti původních dat	22
3.2.3 Simulace rizikového faktoru <i>Prodej</i> z 30-ti původních dat	25
3.2.4 Simulace rizikového faktoru <i>Prodej</i> ze 100 původních dat	28
3.2.5 Souhrnný příklad k metodě klasická statistika	33
3.3 Bootstrap	35
3.3.1 Parametrický Bootstrap	37
3.3.2 Simulace rizikového faktoru <i>Prodej</i> z 12-ti původních dat	37
3.3.3 Simulace rizikového faktoru <i>Prodej</i> z 30-ti původních dat	41
3.3.4 Simulace rizikového faktoru <i>Prodej</i> ze 100 původních dat	44
3.3.5 Souhrnný příklad k metodě parametrický Bootstrap	48
3.4 Bayesova statistika	51
Závěr	55
Literatura	57

Poděkování

Ráda bych poděkovala vedoucímu diplomové práce za spolupráci i za čas, který mi věnoval při konzultacích. Poděkování patří také mým rodičům, kteří mi byli psychickou i finanční oporou během mého studia.

Úvod

Tato práce se bude věnovat problematice modelování nejistoty parametrů pravděpodobnostních rozdělení v simulacích Monte Carlo. Simulace Monte Carlo je stochastická metoda, jejíž podstatou je generování velkého počtu hodnot zvažované výstupní veličiny, za respektování jejího pravděpodobnostního rozdělení. Každá hodnota vlastně představuje jeden možný scénář, tj. jednu potenciální situaci, která může v budoucnu nastat. Dnes je uplatnění této metody stále více četnější, a to např. v oblastech řešení fyzikálních úloh, řízení dopravy, hromadné obsluhy a dále také v ekonomické oblasti nebo ve finanční sféře ve spojení s analýzou rizika.

V diplomové práci částečně navážeme na mou práci bakalářskou, která se zabývala porovnáváním rizik založené na výsledcích simulace Monte Carlo, a ve které byla celá tato simulace podrobněji představená. V rámci této práce se budeme důkladněji zabývat pouze jedním z kroků simulace Monte Carlo, a to konkrétně stanovením pravděpodobnostních rozdělení klíčových rizikových faktorů. Jednou z možností, jak toto stanovení provést, je analýza historických dat rizikového faktoru, jsou-li tato data k dispozici. Nejjednodušším způsobem, jak následně nasimulovat rizikový faktor, je spočítat z těchto dat bodové odhady parametrů již určeného rozdělení pravděpodobnosti. Jaké nejistotě ale čelíme, použijeme-li tyto odhady namísto jejich skutečných hodnot? Cílem práce je tedy studovat, jak lze modelovat nejistotu parametrů těchto pravděpodobnostních rozdělení za pomoci tří různých metod.

V první kapitole se seznámíme se simulací Monte Carlo a jejím postupem konstrukce, který je ilustrován na příkladu, jež nás bude provázet celou prací. Následující kapitola se bude zabývat statistickou analýzou historických dat, kterou můžeme provést pomoci parametrických nebo neparametrických metod, podle toho, zda známe či neznáme rozdělení pravděpodobnosti, ze kterého historická data pochází. Poslední kapitola bude zaměřena na metody modelování nejistoty pravděpodobnostních rozdělení v simulacích Monte Carlo, přičemž budeme uvažovat tři přístupy, kterými jsou metoda Bootstrap, klasická statistika a Bayesova

statistika. První dvě metody budou podrobně popsány a názorně ilustrovány na stejném příkladu, který je představen v první kapitole. Pro obě metody pak budeme uvažovat tři varianty, lišící se pouze počtem historických dat, kterými disponujeme k uvažovanému rizikovému faktoru. Následně budeme sledovat vliv počtu těchto dat na míru nejistoty spojenou s odhadem parametrů rozdělení pravděpodobnosti tohoto faktoru. Celá simulace Monte Carlo pak bude uskutečněna v softwaru Matlab. Metoda Bayesovy statistiky pak bude popsána teoreticky a ilustrována na jednoduchém příkladu.

1. Metoda Monte Carlo

Pro začátek se seznámíme s pojmem simulace Monte Carlo za pomoci literatury [1] a [3]. Simulace Monte Carlo představuje stochastickou metodu, která je založená na používání náhodných čísel. Metoda spočívá ve vytvoření velkého počtu budoucích scénářů pomocí generování náhodných hodnot jednotlivých zvolených rizikových faktorů a jejich následného propočtu na výstupní kritérium. Výsledkem pak může být stanovení rozdělení pravděpodobnosti tohoto výstupního kritéria nebo určení jeho číselných charakteristik.

1.1. Historie

Metoda Monte Carlo je světu známa už více jak šedesát let. Poprvé byla zformulována a použita v USA v době druhé světové války, a to významnými vědci jako je John von Neumann a Stanislav Ulam.

Metoda Monte Carlo byla dříve používána hlavně k řešení složitých fyzikálních úloh. Později byla uplatňována k řešení problémů technických, ekonomických, z oblasti činnosti telefonních centrál, řízení dopravy, hromadné obsluhy, řízení zásob a mnoha dalších.

V dnešní době je simulace Monte Carlo hojně užívána jako jedna z metod, jak lze v podnikatelských projektech modelovat riziko. V takovém případě mluvíme o jedné z činností risk-managementu, která se nazývá analýza rizika. Analýza rizika představuje začleňování rizika a nejistoty do přípravy nějaké, převážně podnikatelské, aktivity. Tato práce se však samotnou analýzou rizika zabývat nebude, a tak čtenáře odkazují pro více informací na např. [1], [2].

1.2. Postup simulace Monte Carlo

Simulaci Monte Carlo můžeme rozdělit do několika kroků:

1. Sestavení **matematického modelu** reprezentujícího reálný problém, který je potřeba řešit. To v překladu znamená, že dojde k vytvoření matematického vzorce. Důležité je stanovení *kritériálních veličin* neboli výstupů,

a následně identifikace všech *faktorů rizika*, tedy vstupů tohoto modelu. Příkladem takových výstupů může být zisk firmy, cash flow, čistá současná hodnota, hospodářský výsledek a další.

2. Stanovení **klíčových faktorů rizika**. Jsou vybrány takové vstupní veličiny, které významně ovlivňují nejistotu výstupů matematického modelu. Jinými slovy, výstupy modelu budou na změny těchto klíčových faktorů rizika značně citlivé. Tyto klíčové faktory reprezentují náhodné veličiny, na rozdíl od ostatních neklíčových faktorů rizika, které budou považovány za konstanty ve formě bodových odhadů.
3. Určení **rozdělení pravděpodobnosti klíčových faktorů rizika**, které se odvozuje podle povahy rizikových faktorů. Tyto faktory mohou být:

- *Diskrétní*, kdy lze rozdělení pravděpodobnosti stanovit přímo (vypíše se hodnoty a jejich pravděpodobnosti) nebo pomocí teoretických rozdělení (binomické, Poissonovo atd.).
- *Spojité*, kdy se využijí některá z teoretických spojitých rozdělení (normální, exponenciální atd.).

Rozhodnutí o rozdělení pravděpodobnosti rizikových faktorů se děje na základě výběru jedné ze dvou možných variant:

- *Názory expertů*.
Některé uvažované modely mohou popisovat reálný projekt, který je značně specifický a v minulosti buďto nikdy nenastal nebo nebyl zaznamenán. V takovém případě je značně obtížné odhadnout vývoj a chování jeho rizikových faktorů. Východiskem je využít názory expertů, kteří pomohou stanovit vhodná expertní rozdělení pravděpodobnosti.
- *Statistická analýza historických dat*.
Existují-li historické záznamy některého rizikového faktoru, pak lze uplatnit aproximaci nějakým teoretickým rozdělením pravděpodobnosti.

4. Popsání **statistické závislosti mezi rizikovými faktory**. V praxi není výjimečnou situací, když hodnoty jednoho rizikového faktoru reagují na změny hodnot faktoru druhého (typickým příkladem je závislost poptávky po produktu na jeho prodejní ceně). V rámci simulace je zapotřebí tuto závislost respektovat, jinak se může stát, že dojde ke zkreslení výsledků simulace.
5. Provedení **samotné simulace**, která je realizována s pomocí nějakého počítačového **softwaru**. Příkladem takového softwaru je CrystalBall, @RISK, MC Excel nebo Matlab, který byl v této práci použit.

Tato simulace, která proběhne v námi zvoleném softwaru, je složená ještě z několika kroků. Nejprve se vygeneruje dostatečně velký počet možných scénářů, kde každý scénář reprezentuje jednu kombinaci hodnot vstupních veličin. Z těchto scénářů se následně propočítá shodné množství možných výsledků výstupního kritéria.

Výstupem celé simulace je *statistický soubor dat*, který se může dále analyzovat. Typickým příkladem této analýzy je vizualizace neboli grafické vyjádření, kterým je například empirická distribuční funkce nebo histogram. Dále připadají v úvahu různé číselné charakteristiky této výstupní veličiny jako jsou výběrový průměr, výběrová směrodatná odchylka, výběrový rozptyl, pravděpodobnost dosažení či nedosažení určitých výsledků atd.

1.3. Konstrukce ilustračního příkladu

Výše popsanou simulaci si nyní představíme na ilustračním příkladu, jehož zadání je převzato z [1]. Tento příklad nás bude provázet celým textem, neboť na něm budeme ilustrovat problematiku, kterou se tato práce zabývá.

Uvažujme investiční projekt, jehož cílem je výroba nového produktu firmy. Naším úkolem je provést analýzu rizika pomocí simulace Monte Carlo. Postup simulace Monte Carlo je následující:

Sestavení finančního a matematického modelu

V tomto případě je výsledná kritériální veličina zisk z nového výrobku před zdáněním, který můžeme vypočítat podle vztahu:

$$Z = P \cdot (c \cdot m - s \cdot k) - FN \quad (1)$$

Jak vidíme, zisk je ovlivněn šesti rizikovými faktory znázorněnými v tabulce 1. Pět z nich je uvedeno včetně jejich možného vývoje, popsaného pomocí pesimistického, optimistického a nejpravděpodobnějšího scénáře.

Rizikový faktor	Jednotka	Scénáře		
		pesimis.	nejpravdě.	optimis.
1. Počet prodaných kusů - P	tis. ks/rok	-	-	-
2. Prodejní cena - c	euro/ks	135	150	160
3. Měnový kurz - m	CZK/EUR	24	27	29
4. Měrná spotřeba materiálu - s	kg/ks	62	60	58
5. Kupní cena materiálu - k	Kč/kg	46	40	36
6. Fixní náklady - FN	mil. Kč/rok	85	75	70

Tabulka 1: Scénáře hodnot rizikových faktorů

Stanovení rozdělení pravděpodobnosti rizikových faktorů

Pro většinu rizikových faktorů využijeme expertní hodnocení a na jejich základě zvolíme aproximaci trojúhelníkovým rozdělením, přičemž za parametry tohoto rozdělení budou sloužit odhadnuté scénáře, které se nachází v tabulce 1.

To ovšem nebude platit pro jeden z rizikových faktorů, konkrétně pro *Počet prodaných kusů*, který pro jednoduchost budeme v dalším textu označovat jako *Prodej*. Pro ten budeme předpokládat, že k němu jsou k dispozici historická data a využijeme tedy aproximaci nějakým teoretickým rozdělením pravděpodobnosti - v našem případě normálním rozdělením pravděpodobnosti.

Analýza statistické závislosti

Pro jednoduchost nebudeme uvažovat žádnou statistickou závislost mezi jednotlivými rizikovými faktory.

Realizace simulace Monte Carlo

Nyní se můžeme pustit do samotné simulace, kterou provedeme v matematickém softwaru **Matlab** a výsledky zobrazíme v softwaru **MS Excel** v souboru s názvem *KlasickaStatistika.xlsx*, konkrétně v záložce *1D simulace*.

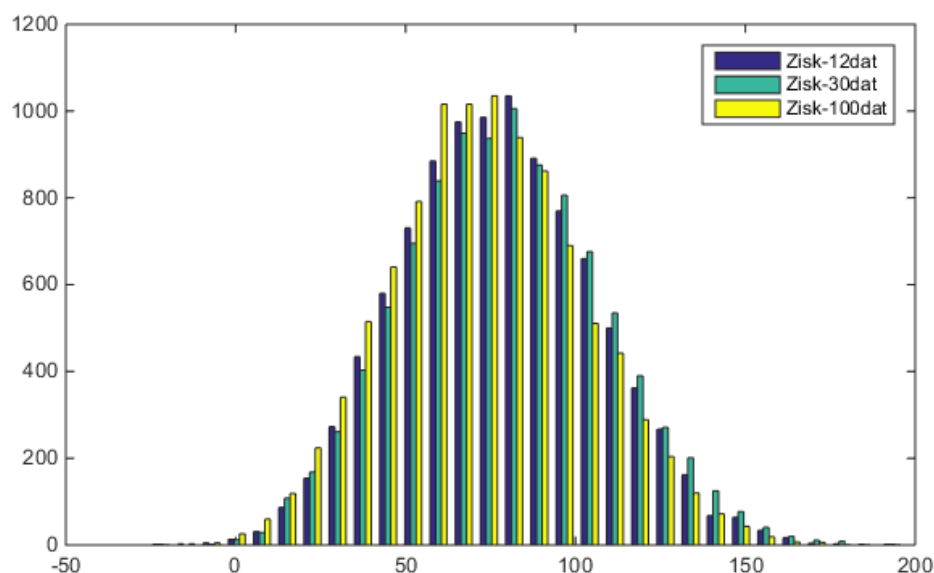
Nejprve si v softwaru Matlab pro každý z rizikových faktorů *Prodejní cena*, *Měnový kurz*, *Měrná spotřeba materiálu*, *Kupní cena materiálu* a *Fixní náklady* náhodně vygenerujeme 10 000 scénářů z trojúhelníkového rozdělení. Za parametry tohoto rozdělení zvolíme možný vývoj uvedených rizikových faktorů, znázorněný v tabulce 1. Výsledné scénáře pak použijeme k propočtu výstupního kritéria *Zisk*. Příslušný skript pro generování hodnot z trojúhelníkového rozdělení je přiložen k práci pod názvem *trojuhel.m*.

Největší pozornost budeme ovšem věnovat simulaci rizikového faktoru *Prodej*, který budeme generovat z obdržených historických dat. Zde je důležité poznamenat, že budeme uvažovat tři případy: nejprve budeme mít k dispozici pouze dvanáct historických dat, posléze třicet a nakonec sto. Tato data jsme si náhodně vygenerovali ve zmíněném softwaru Matlab z normálního rozdělení $N(100, 9)$ s použitím příkazu **normrnd**. Budeme tedy pro tuto práci předpokládat, že známe rozdělení pravděpodobnosti, ze kterého historická data pochází, což je situace, která by v praxi nenastala. Vygenerovaná data jsou zachycena v přiloženém souboru *KlasickaStatistika.xlsx* v záložce *Data*, včetně jejich číselných charakteristik.

Samotná simulace rizikového faktoru *Prodej* pak vypadá tak, že z původního souboru dvanácti (resp. třiceti, sto) dat vypočítáme výběrový průměr \bar{x} a výběrovou směrodatnou odchylku s , přičemž obě tyto veličiny představují bodové odhady parametrů μ a σ z původního rozdělení $N(100, 9)$. Spočítanou dvojici bodových odhadů pak použijeme pro simulaci 10 000 scénářů v softwaru Matlab z normálního rozdělení s těmito parametry, tj. z rozdělení $N(\bar{x}, s)$. Tento postup provedeme tři krát - pro dvanáct, třicet a sto historických dat a obdržíme tak tři soubory nasimulovaných scénářů pro rizikový faktor *Prodej*. Následně provedeme propočet na výstupní kritérium pro všechny tři případy, a to s použitím již vygenerovaných ostatních rizikových faktorů. Rozdíly mezi třemi výstupy můžeme

vidět na obrázku 1, kde jsou zachyceny tři histogramy kritériální veličiny *Zisk*. V tabulce 2 se pak nachází vybrané číselné charakteristiky k těmto třem výstupům. Nabízí se otázka, proč se jednotlivé varianty od sebe odlišují. Důvodem je skutečnost, že jsme pro modelování rizikového faktoru *Prodej* použili pouze bodové odhady parametrů normálního rozdělení namísto jejich skutečných hodnot. Setkáváme se tedy s jakousi *nejistotou* ohledně těchto parametrů, která je tím větší, čím méně máme původních historických dat, ze kterých jsme simulaci provedli. Právě těmito úvahami se pak budeme zabývat v dalším textu práce.

Celá tato simulace je zachycena ve zmíněném souboru *KlasickaStatistika.xlsx*, tentokrát v záložce *Zisk 1D*.



Obrázek 1: Histogramy výstupu *Zisk* ve třech variantách

	Výběr. průměr	Výběr. směr.odch.	5% - kvantil	95% - kvantil
Zisk - 12 dat	77,6601	28,5624	32,2091	125,7442
Zisk - 30 dat	79,2328	29,5980	31,6056	129,7928
Zisk - 100 dat	73,4509	28,4429	27,7523	121,9760

Tabulka 2: Číselné charakteristiky výstupu *Zisk* v mil. Kč ve třech variantách

2. Statistická analýza historických dat

Tato kapitola se zabývá vhodnou volbou pravděpodobnostního rozdělení rizikového faktoru na základě druhé z výše zmíněných možností, a to analýzy historických dat. Představíme si zde postupy, jak tuto analýzu provést. Níže uvedený text byl zpracován s pomocí literatury [1],[3].

2.1. Úvod

Vraťme se nejprve k předchozí kapitole, a to konkrétně ke třetímu kroku postupu simulace Monte Carlo, jež se týká volby vhodného rozdělení pravděpodobnosti rizikového faktoru. Samotná volba náležitého rozdělení je velmi náročná a jak víme, má dva různé přístupy. Uvažujme nyní druhý přístup, tedy že se volba děje na základě analýzy historických dat. Jinými slovy, disponujeme nějakými vybranými daty, která můžeme statisticky testovat a následně rozhodnout, které rozdělení pravděpodobnosti na daný rizikový faktor uplatnit. Typickým příkladem je situace, kdy máme nějaká pozorování a snažili bychom se zjistit, zda pochází z normálního rozdělení. Data, která máme k dispozici, mohou pocházet z různých zdrojů jako např. databáze historických časových řad, vědecké experimenty, dotazníky či počítačové simulace. V každém případě je ovšem důležité si ověřit kvalitu těchto dat.

2.2. Metody odhadu pravděpodobnostních rozdělení

Metody pro odhad vhodného rozdělení pravděpodobnosti rizikového faktoru se dělí do dvou skupin podle toho, zda předpokládáme, že analyzovaná data pochází z nějakého předem známého rozdělení či nikoli. V případě, kdy máme představu o původu těchto data, tak mluvíme o *parametrických metodách*. Naopak, nemáme-li žádnou představu o tvaru pravděpodobnostního rozdělení, pak hovoříme o *neparametrických metodách*.

2.2.1. Neparametrické metody

Neparametrický Bootstrap

Jedná se o velmi praktický a nejspíše nejjednodušší způsob, jak přímo z historických dat generovat možné hodnoty sledovaného rizikového faktoru a tím obdržet jeho rozdělení pravděpodobnosti. Metoda v překladu znamená „pomoz si sám“ a je založená na náhodném výběru s vracením, jehož postup je následující:

1. Disponujeme historickými daty, která použijeme pro predikci možných hodnot rizikového faktoru.
2. Předpokládáme, že každá z těchto historických hodnot může nastat se stejnou pravděpodobností.
3. Provedeme náhodný výběr s vracením, kdy z historických dat vybereme právě jednu hodnotu a tu po vybrání „vrátíme“.
4. Takto obdržíme libovolný počet možných scénářů vývoje rizikového faktoru a získáme jeho diskretní pravděpodobnostní rozdělení.

Jelikož se jedná o velmi jednoduchou metodu, představíme si postup jejího řešení na následující úloze. Uvažujme znovu příklad z kapitoly 1, kde se konkrétně zaměříme na rizikový faktor *Měnový kurz CZK/EUR*, který chceme modelovat. K tomuto rizikového faktoru disponujeme dvaadvaceti historickými daty, získanými jako měsíční průměry od ledna 2012 do října 2013, tj. před intervencemi prováděnými ČNB. Uvažované měnové kurzy jsou zobrazeny v tabulce 3 a pochází z [5]. Předpokládáme, že každý z těchto měnových kurzů může nastat se stejnou pravděpodobností, tj. jedná se o rovnoměrné diskretní rozdělení s parametrem $p = \frac{1}{22}$. V dalším kroku bychom pak provedli náhodný výběr s vracením, kdy bychom obdrželi např. 10 000 možných scénářů tohoto rizikového faktoru, které by následně vstoupily do propočtu výsledné kriteriální veličiny *Zisk*.

Leden 2012	Únor	Březen	Duben	Květen	Červen
25,532	25,041	24,676	24,799	25,322	25,641
Červenec	Srpen	Září	Říjen	Listopad	Prosinec 2012
25,434	25,020	24,731	24,938	25,365	25,216
Leden 2013	Únor	Březen	Duben	Květen	Červen
25,563	25,476	25,476	25,840	25,840	25,840
Červenec	Srpen	Září	Říjen 2013		
25,840	25,840	25,840	25,840		

Tabulka 3: Měsíční vývoj EURA za období leden 2012 - říjen 2013

2.2.2. Parametrické metody

U těchto metod předpokládáme, že hodnoty rizikového faktoru jsou generovány na základě předem známého rozdělení pravděpodobnosti. Právě na analytikovi pak záleží, aby zvolil vhodné rozdělení a odhadl jeho parametry z obdržených historických dat. Vhodnost vybraného rozdělení pravděpodobnosti můžeme testovat mnoha způsoby. Speciálně k tomuto účelu slouží tzv. *testy dobré shody*, z nichž zmíníme například Kolmogorův-Smirnovův test, Anderson-Darlingův test nebo χ^2 test. Pro více informací o těchto testech čtenáře odkazují např. na [4]. Tyto zmíněné testy jsou poměrně složité na ruční provedení, a tak se nabízí další možnost a tou je nechat statistické softwaru¹, aby tyto testy provedly za nás. Ještě větším ulehčením je použít simulační softwaru², které nám dokonce sami vyberou nejvhodnější pravděpodobnostní rozdělení. Je ovšem potřeba tento výběr dobře zvážit, neboť napozorovaná data mohou naznačovat, že pochází z rozdělení, které teoreticky nemusí dávat smysl.

¹Příkladem takových softwarů je Matlab, R.

²Příkladem je @RISK nebo CrystalBall.

3. Modelování nejistoty parametrů pravděpodobnostních rozdělení v simulacích Monte Carlo

V této kapitole se budeme věnovat hlavní náplni této práce. Nejprve čtenáře uvedeme do problematiky modelování nejistoty parametrů pravděpodobnostních rozdělení, což provedeme za pomoci literatury [1]. Následně představíme metody navržené na řešení této situace a to včetně ilustračních příkladů. První dvě metody, klasická statistika a metoda Bootstrap, byly sepsány s použitím publikací [1] a [3]. Poslední metoda Bayesovy statistiky byla zformulována díky literatuře [1], [3] a [6].

3.1. Úvod do problematiky

Pro začátek předpokládejme, že jsme již zvolili vhodné rozdělení pravděpodobnosti na základě analýzy historických dat z předchozí kapitoly. Nyní se tímto problémem zabývejme hlouběji a položme si otázku: Jaké nejistotě čelíme při odhadu parametrů tohoto zvoleného rozdělení pravděpodobnosti? V odborné literatuře je s touto otázkou spojen termín *Second order distribution*, což můžeme přeložit jako *volba pravděpodobnostního rozdělení druhého řádu*. Druhým řádem pak myslíme druhý stupeň analýzy, kdy se již nezabýváme analýzou samotného rozdělení pravděpodobnosti, ale analyzujeme příslušné parametry.

Pro simulační účely se pak objevuje termín *2D simulace* neboli dvourozměrné simulace. Jejich význam je pak takový, že první rozměr reprezentuje náhodný proces generující rizikový faktor a druhý rozměr představuje nejistotu parametrů tohoto procesu. Tyto dvě situace je třeba od sebe pojmově separovat, neboť se jedná o dva různé problémy:

1. Náhodný proces popisuje vlastní nahodilost systému.
2. Parametry tohoto procesu neznáme s jistotou, můžeme je pouze odhadovat.

Tyto parametry reprezentují naši znalostní nejistotu, neboť je odhadujeme na základě původních pozorování. Tato skutečnost je v mnoha modelech velmi často opomíjená a parametry se definují nějakým pevně daným číslem.

Postupy pro modelování nejistoty spojené s odhadem parametrů můžeme rozdělit do tří skupin:

- klasická statistika,
- metoda Bootstrap,
- Bayesova statistika.

Zde je potřeba podotknout, že v mnoha případech se statističtí experti shodují pouze s jedním z možných přístupů a někdy je také zjevná vzájemná nevraživost mezi různě zaměřenými „tábory“ odborníků.

3.2. Klasická statistika

Klasická statistika je založena na jistých počátečních předpokladech, při jejichž dodržení dochází k exaktním závěrům. Právě díky tomu je tento přístup často pokládán za jedinou nesubjektivní statistickou metodu. To je pravda, ale pouze v případě, kdy výchozí předpoklady jsou v praxi splněny.

Nejběžnějším počátečním předpokladem pro odhad parametrů rozdělení je podmínka normálního nebo binomického rozdělení pravděpodobnosti, ze kterého lze poté odvodit celou řadu závěrů. Takovým závěrem může být např. hodnota regresních parametrů, interval spolehlivosti atd.

Výhodou klasické statistiky je uplatnění tzv. centrální limitní věty, jejímž důsledkem je skutečnost, že se mnoho náhodných pozorovaných veličin chová, jako by byly generovány normálním rozdělením. Bohužel jsou i situace, které se s tímto předpokladem neztotožňují.

Postupy této metody jsou mnohdy značně složité a rozsáhlé, proto se v této práci zaměříme jen na řešení nejjednodušších problémů. Celý přístup pak bude ilustrován na příkladu, který byl představen v kapitole 1.3.

3.2.1. Ilustrační příklad - odvození důležitých vztahů

Pro lepší pochopení této metody se podíváme na klasický problém odhadu parametrů normálního rozdělení. Jak víme např. z [4], problém odhadu střední hod-

noty μ , resp. směrodatné odchyly σ závisí na situaci, zda známe nebo neznáme ten druhý parametr. V této práci si popíšeme případ, který je ve finanční praxi častější, kdy odhadujeme střední hodnotu a směrodatnou odchylku za neznalosti ani jednoho z uvedených parametrů. Nejprve provedme následující označení:

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

jako výběrový průměr z n pozorování. Zároveň se jedná o nestranný bodový odhad střední hodnoty μ normálního rozdělení. Bohužel je to pouze odhad, a tak nemůžeme mezi těmito dvěma veličinami položit rovnítko. Naopak, jelikož výběrový průměr \bar{x} je závislý na pozorováních, ze kterých je počítán, jedná se vlastně o náhodnou veličinu a s tou je spojena nejistota. Zde se nabízí otázka, jaké nejistotě čelíme, pokud za parametr normálního rozdělení μ vezmeme jeho bodový odhad \bar{x} . Využijme nyní centrální limitní věty, ze které víme:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \approx \frac{1}{n} \sum_{i=1}^n N(\mu, \sigma^2) = \frac{1}{n} N(n\mu, n\sigma^2) = N(0, 1) \frac{\sigma}{\sqrt{n}} + \mu.$$

Díky symetrii normálního rozdělení kolem nuly pak můžeme napsat:

$$\mu \approx N(0, 1) \frac{\sigma}{\sqrt{n}} + \bar{x}. \quad (3)$$

Za předpokladu znalosti parametru σ bychom tak dostali vztah pro výpočet μ . My ovšem směrodatnou odchylku neznáme, a tak se v našich úvahách pustíme ještě o krok dál. Označme:

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

jako výběrovou směrodatnou odchylku v souboru o n pozorováních při neznalosti střední hodnoty μ . Opět se jedná o bodový odhad σ . Potom můžeme napsat:

$$s \approx \sqrt{\frac{\sum_{i=1}^n [N(\mu, \sigma) - \bar{x}]^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^{n-1} [N(\mu, \sigma) - \mu]^2}{n - 1}} = \sigma \sqrt{\frac{\sum_{i=1}^{n-1} [N(0, 1)]^2}{n - 1}}. \quad (5)$$

Dále nechť rozdělení χ^2 je definováno jako:

$$\chi = U_1^2 + U_2^2 + \dots + U_n^2,$$

kde U_i představují náhodnou veličinu s rozdělením $N(0, 1)$. Díky čemuž můžeme vztah (5) upravit následovně:

$$s \approx \sigma \sqrt{\frac{\chi^2(n-1)}{n-1}}$$

odkud si vyjádříme σ a dostaneme tak konečný vzorec pro výpočet směrodatné odchylky normálního rozdělení při neznalosti střední hodnoty μ :

$$\sigma \approx \sqrt{\frac{n-1}{\chi^2(n-1)}}. \quad (6)$$

Nyní tento vztah dosadíme do vzorce (3) za neznámou σ :

$$\mu \approx N(0, 1) \sqrt{\frac{n-1}{\chi^2(n-1)}} \frac{s}{\sqrt{n}} + \bar{x}. \quad (7)$$

Víme, že vztah:

$$N(0, 1) \sqrt{\frac{n-1}{\chi^2(n-1)}}$$

představuje standardizované Studentovo rozdělení, tudíž můžeme vzorec (7) přepsat do konečné podoby:

$$\mu \approx t(n-1) \frac{s}{\sqrt{n}} + \bar{x}. \quad (8)$$

Jelikož Studentovo rozdělení konverguje pro $n > 20$ k normálnímu rozdělení, můžeme tento vztah pro větší n psát jako:

$$\mu \approx N(0, 1) \frac{s}{\sqrt{n}} + \bar{x}. \quad (9)$$

Díky výše popsanému odvození jsme nyní dostali výsledné vztahy (8), popř. (9) pro výpočet odhadu střední hodnoty a (6) pro odhad směrodatné odchylky, které nyní použijeme v rámci simulace pro stanovení nejistoty, které čelíme, použijeme-li odhady těchto parametrů namísto jejich skutečných hodnot.

3.2.2. Simulace rizikového faktoru *Prodej* z 12-ti původních dat

Vraťme se opět k našemu úvodnímu příkladu, který byl představen v kapitole 1.3 a nejprve předpokládejme, že disponujeme 12-ti historickými daty k rizikovém faktoru *Prodej*. Jak víme, tato data pochází z normálního rozdělení $N(100, 9)$ a jsou zachycena v tabulce 4. Nyní je využijeme pro simulaci zmíněného faktoru tak, že si z nich spočítáme výběrový průměr, pomocí vzorce (2), který je $\bar{x} = 101,826$ a výběrovou směrodatnou odchylku, díky (4), která činí $s = 9,095$. Tyto charakteristiky následně použijeme místo původních parametrů μ a σ k simulaci 10 000 možných hodnot rizikového faktoru *Prodej* z normálního rozdělení, což je identický proces, jaký jsme využili v rámci úvodního příkladu. Co jsme tam ovšem nezmnili je, že tímto postupem jsme provedli výše zmíněnou 1D simulaci, generující hodnoty náhodné veličiny *Prodej*. Nagenované hodnoty pro 1D simulaci můžeme opět najít v souboru *KlasickaStatistika.xlsx*, do kterého budou exportovány i všechny následující výsledky.

92,837	93,023	94,818	96,050	97,199	97,531
99,878	101,779	105,282	106,535	110,782	126,197

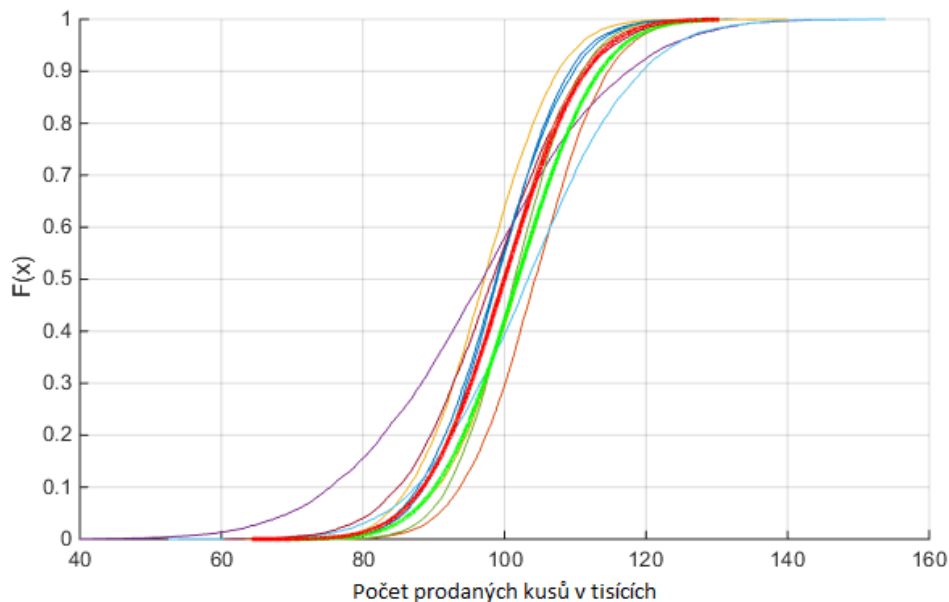
Tabulka 4: Vzestupně uspořádaná historická data pro *Prodej* v tis. ks

V dalším kroku již do našich úvah zahrneme i nejistotu parametrů normálního rozdělení uvažovaného rizikového faktoru a provedeme tedy 2D simulaci. Postupujeme tak, že si nejprve vygenerujeme deset dvojic parametrů μ a σ . Proměnné μ vypočítáme pomocí výše odvozeného vzorce (8), kde $t(n-1)$ jsou náhodně vygenerované hodnoty studentova rozdělení o $(n-1)$ stupních volnosti. Proměnné σ dostaneme díky vzorci (6), kde $\chi^2(n-1)$ jsou náhodně vygenerované hodnoty χ^2 rozdělení o $(n-1)$ stupních volnosti. V obou vzorcích jsme přitom za \bar{x} a s

dosadili charakteristiky vypočtené v rámci 1D simulace. Pro každou dvojici parametrů následně dojde k vygenerování 10 000 hodnot uvažovaného rizikového faktoru z normálního rozdělení. Provedeme tedy simulaci z rozdělení $N(\mu, \sigma)$, pro každou dvojici proměnných, a výsledkem je deset souborů nasimulovaných hodnot faktoru *Prodej*, které nyní však zahrnují naši nejistotu ohledně skutečných hodnot parametrů. Celý postup byl uskutečněn v softwaru Matlab pomocí přiloženého skriptu *KlasStatistika.m* a výsledky jsou opět vloženy do zmíněného excelovského souboru v odpovídající záložce *Prodej12*.

Porovnání rizikového faktoru *Prodej* pro 1D a 2D simulace

Nyní se podívejme na rozdíl mezi simulacemi, zachyceném na obrázku 2, který ztvárňuje empirické distribuční funkce rizikového faktoru *Prodej* pro simulaci 1D, která je světle zeleně, a deset simulací 2D. Červeně je znázorněna skutečná distribuční funkce normálního rozdělení $N(100, 9)$, ze kterého pochází naše historická data. Vidíme, že některé křivky se značně odlišují od ostatních a celkově jsou všechny distribuční funkce dost roztáhlé okolo červené křivky skutečného rozdělení pravděpodobnosti.



Obrázek 2: Porovnání 1D a 2D simulace pro 12 původních dat

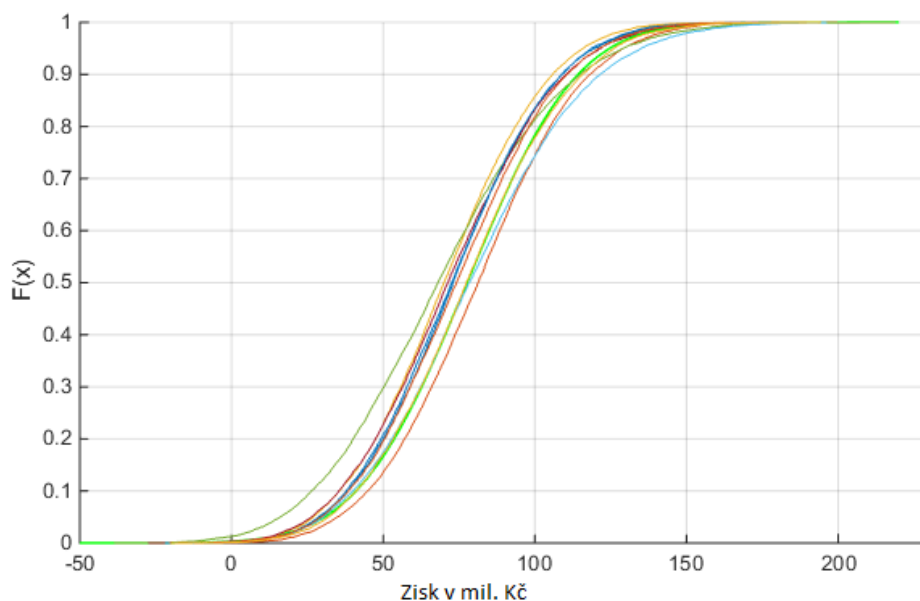
Z tabulky 5, kde se nachází vybrané číselné charakteristiky těchto simulací, můžeme vyčíst, že výběrové průměry i výběrové směrodatné odchylky dost kolísají. Tuto rozdílnost způsobuje skutečnost, že rizikový faktor *Prodej*, byl simulován na základě pouhých 12-ti výchozích dat. Nejistota ohledně parametrů normálního rozdělení, ze kterého je simulace uskutečněna, je v tomto případě velická, což právě naznačuje obrázek, resp. tabulka. Je také vidět, že světle zelená empirická distribuční funkce je rozdílná od té skutečné, nicméně skutečná distribuční funkce leží v pásu empirických distribučních funkcí z jednotlivých simulací.

	Výběrový průměr	Výběrová směr. odchylka
Prodej 1D	101,7911	9,0898
Prodej 2Da	98,7975	8,5145
Prodej 2Db	104,2856	8,1625
Prodej 2Dc	97,0944	8,1988
Prodej 2Dd	96,5497	16,3008
Prodej 2De	101,4410	7,4598
Prodej 2Df	103,3130	12,5351
Prodej 2Dg	98,2855	10,4519
Prodej 2Dh	98,9880	8,0571
Prodej 2Di	99,6922	8,8315
Prodej 2Dj	102,0057	9,0269

Tabulka 5: Vybrané číselné charakteristiky v tis. ks

Porovnání výstupní veličiny *Zisk* při zahrnutí a nezahrnutí nejistoty parametrů do rizikového faktoru *Prodej*

Podívejme se na celkovou simulaci výstupní veličiny *Zisk*, kterou jsme dostali příslušnými matematickými operacemi dle výše uvedeného vzorce (1). Propoččet byl proveden jak s 1D rizikovým faktorem *Prodej*, tak s jeho deseti 2D obměnami. Situace je znázorněná na obrázku 3 pomocí empirických distribučních funkcí, kde jsme zachytili, jak by situace dopadla, pokud bychom v rizikovém faktoru *Prodej* nezahrnuli (světle zelená křivka) a zahrnuli nejistotu spojenou s odhady parametrů normálního rozdělení. Celý propoččet na výsledné kritérium *Zisk* byl proveden v příloženém souboru v záložce *Zisk12*.



Obrázek 3: Empirické distribuční funkce pro *Zisk* z 12-ti historických dat

Jak můžeme ze zmíněného obrázku vidět, jednotlivé křivky se od sebe vcelku odlišují. Tuto situaci si můžeme ověřit také z tabulky 6, obsahující vybrané číselné charakteristiky, ze které zmíníme např. velké rozdíly u očekávaného průměrného zisku. Zatímco z 1D simulace bychom mohli očekávat zisk 77,66 mil. Kč, ze čtvrté 2D simulace je tento očekávaný zisk pouze 69,71 mil. Kč, a naopak ze druhé 2D simulace je 81,45 mil. Kč. Tyto velké rozdíly mezi 1D a 2D simulacemi jsou způsobeny právě nezahrnutím nejistoty parametrů u 1D simulací.

3.2.3. Simulace rizikového faktoru *Prodej* z 30-ti původních dat

Uvažujeme situaci, kdy se počet historických dat zvětšil a my nyní disponujeme 30-ti daty, na jejichž základě budeme opět simulovat rizikový faktor *Prodej* pro 1D a 2D simulace. To uskutečníme pomocí stejného skriptu v softwaru Matlab, který byl použit již v minulé části. Výsledky simulací jsou zobrazeny v záložce *Prodej30*. Historická data pak můžeme vidět v tabulce 7. Číselné charakteristiky, nutné k výpočtu 1D a 2D simulace jsou $\bar{x} = 102,860$ a $s = 10,361$.

	Výběr. průměr	Výběr. směr.odch.	5% - kvantil	95% - kvantil	Medián	Prav. ztráty
Zisk 1D	77,6601	28,5624	32,2091	125,7442	77,1198	0,0014
Zisk 2Da	73,1476	27,6391	28,5285	119,6940	72,6058	0,0017
Zisk 2Db	81,4459	28,4117	35,0018	129,1721	81,2796	0,0007
Zisk 2Dc	70,5645	27,0233	26,9406	115,9937	70,0149	0,0026
Zisk 2Dd	69,7114	34,4657	16,0538	129,0696	68,0415	0,0125
Zisk 2De	77,1179	27,1937	32,8374	121,9050	76,9004	0,0012
Zisk 2Df	79,9740	31,8414	30,2365	134,9725	78,3806	0,0022
Zisk 2Dg	72,3544	28,9253	26,3375	121,6368	71,3029	0,0031
Zisk 2Dh	73,4161	27,2338	29,0955	118,6664	73,0605	0,0017
Zisk 2Di	74,4982	27,9517	29,4667	121,1557	74,1917	0,0022
Zisk 2Dj	78,0319	28,7584	32,0300	126,0493	77,4434	0,0014

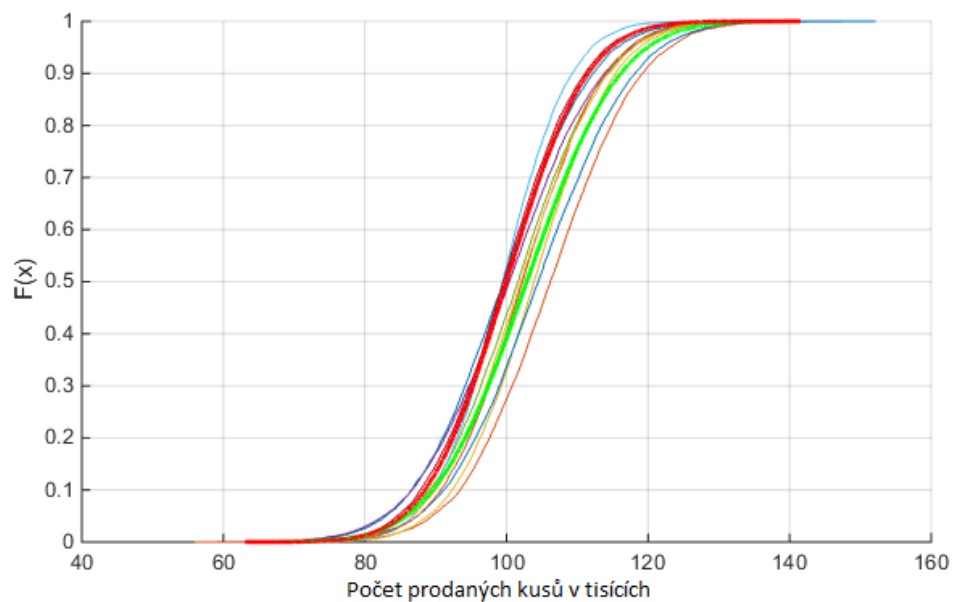
Tabulka 6: Vybrané číselné charakteristiky v mil. Kč

86,744	87,025	87,946	91,946	91,980	92,837
93,023	93,408	94,818	96,050	97,198	97,531
99,878	101,779	102,532	103,163	103,424	104,664
105,282	106,143	106,535	107,401	110,782	112,157
113,944	115,922	116,955	117,831	120,670	126,197

Tabulka 7: Vzestupně uspořádaná historická data pro *Prodej* v tis. ks

Porovnání rizikového faktoru *Prodej* pro 1D a 2D simulace

Znovu se zaměříme na rozdíl mezi 1D simulací a deseti 2D simulacemi pro rizikový faktor *Prodej*. Všechny varianty jsou znázorněny na obrázku 4, přičemž světle zelená je křivka pro 1D simulaci a červená je křivka původního rozdělení. Setkáváme se zde s větší vzájemnou podobností všech empirických distribučních funkcí. Na rozdíl od předchozí situace, žádná z těchto funkcí významně nevyčnívá z hlavního „prutu“ všech křivek, který je seskupen okolo skutečné tj. červené distribuční funkce. Tato zesílená podobnost je následkem zvýšeného počtu původních dat, ze kterého je rizikový faktor *Prodej* počítán, neboť čím více je původních dat, tím menší je naše nejistota ohledně odhadu parametrů. Tuto situaci potvrzují i údaje z tabulky 8, ve které jsou zachyceny vybrané číselné charakteristiky. Jak vidíme, tyto veličiny se od sebe již méně odlišují, a to hlavně pro výběrové směrodatné odchylky.



Obrázek 4: Porovnání 1D a 2D simulace pro 30 původních dat

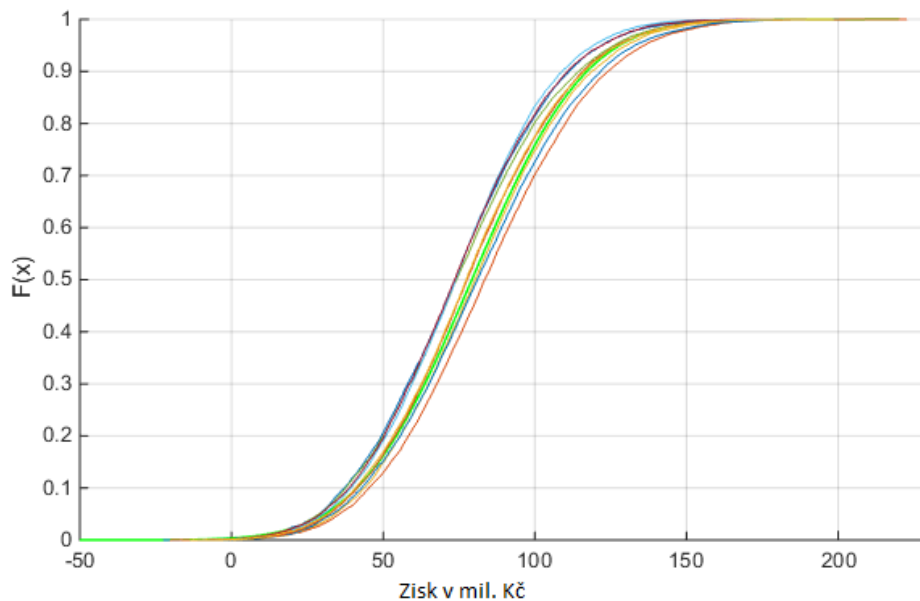
	Výběrový průměr	Výběrová směr. odchylka
Prodej 1D	102,8346	10,2467
Prodej 2Da	99,4513	10,0029
Prodej 2Db	106,2120	10,1813
Prodej 2Dc	103,9010	9,0568
Prodej 2Dd	100,2471	10,6214
Prodej 2De	101,6074	9,8037
Prodej 2Df	99,4331	7,6790
Prodej 2Dg	99,5281	9,0141
Prodej 2Dh	104,5496	10,5834
Prodej 2Di	102,3222	9,1376
Prodej 2Dj	102,1059	9,5813

Tabulka 8: Vybrané číselné charakteristiky v tis. ks

Porovnání výstupní veličiny *Zisk* při zahrnutí a nezahrnutí nejistoty parametrů do rizikového faktoru *Prodej*

Podívejme se na celkovou simulaci výstupu *Zisk* v jedenácti variantách, při zahrnutí a nezahrnutí (světle zelená křivka) nejistoty parametrů při výpočtu faktoru *Prodej*. Výsledná veličina pak byla propočítána dle potřebného vzorce jako v předchozím případě, a to ve stejném souboru v záložce *Zisk30*. Situace je znázorněná na obrázku 5 pomocí empirických distribučních funkcí a pro každou variantu jsou v tabulce 9 vypsány vybrané číselné charakteristiky.

Z obrázku, a hlavně z tabulky je patrné, že rozdíly mezi výsledným výstupem pro 1D a 2D simulace se snížily. Zatímco dle 1D simulace můžeme očekávat průměrný zisk 79,23 mil. Kč, pak dle 2D simulací můžeme nejvýše očekávat 84,88 mil. Kč a nejméně 74,07 mil. Kč.



Obrázek 5: Empirické distribuční funkce pro *Zisk* z 30-ti historických dat

3.2.4. Simulace rizikového faktoru *Prodej* ze 100 původních dat

Do třetice uvažujeme situaci, kdy disponujeme 100 původními historickými daty, které znovu uplatníme pro 1D a 2D simulace rizikového faktoru *Prodej*,

	Výběr. průměr	Výběr. směr.odch.	5% - kvantil	95% - kvantil	Medián	Prav. ztráty
Zisk 1D	79,2328	29,5980	31,6058	129,7928	78,5415	0,0008
Zisk 2Da	74,0966	28,6389	28,7586	121,9337	73,6053	0,0023
Zisk 2Db	84,3968	30,4745	35,5376	136,0562	83,4376	0,0010
Zisk 2Dc	80,8803	29,0921	34,1225	130,2962	80,2963	0,0006
Zisk 2Dd	75,3423	29,5935	28,8519	126,1117	74,2469	0,0021
Zisk 2De	77,4177	29,2428	30,8232	126,6538	76,9568	0,0021
Zisk 2Df	74,0710	26,9507	29,9800	119,2954	73,8985	0,0019
Zisk 2Dg	74,2289	28,0244	29,2727	122,1469	73,5775	0,0017
Zisk 2Dh	81,8854	30,4805	33,2625	133,0213	81,0144	0,0010
Zisk 2Di	78,4873	28,7834	32,1646	127,1873	77,6678	0,0018
Zisk 2Dj	78,1587	29,0801	31,6952	127,6208	77,4420	0,0013

Tabulka 9: Vybrané číselné charakteristiky v mil. Kč

a to totožným postupem jako tomu bylo v předchozích dvou částech. Využijeme k tomu i shodných skriptů v softwaru Matlab a výsledky opět exportujeme do souboru *KlasickaStatistika.xlsx* tentokrát v záložce *Prodej100*. Uvažovaná historická data jsou zobrazena v tabulce 10, přičemž číselné charakteristiky, nezbytné pro provedení simulací, jsou: $\bar{x} = 98,975$ a $s = 9,542$.

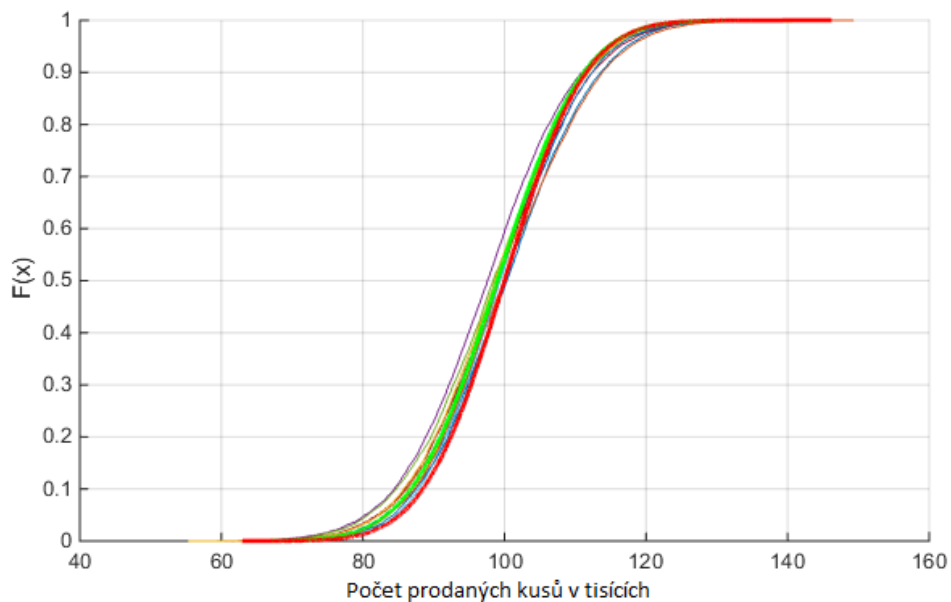
Porovnání rizikového faktoru *Prodej* pro 1D a 2D simulace

Naposledy se podíváme na rozdíl u faktoru *Prodej* v případě, že zahrneme a nezahrneme nejistotu parametrů. Situace je znázorněna na obrázku 6, přičemž světle zelená je opět křivka pro 1D simulaci a červená je křivka skutečného rozdělení pravděpodobnosti historických dat. Pro lepší srovnání jsou v tabulce 11 uvedeny vybrané číselné charakteristiky.

Z obrázku je na první pohled zřejmé, že došlo k výraznému snížení variability mezi jednotlivými simulacemi rizikového faktoru *Prodej*. Tuto skutečnost podtrhuje i zmíněná tabulka, kde vidíme, že rozdíly mezi číselnými charakteristikami ještě více poklesly oproti předchozí situaci. Tento závěr je ovšem logický, neboť celá simulace je již založena na vcelku velkém počtu historických dat, a tak naše nejistota ohledně odhadů z těchto dat je poměrně nízká. Simulace 2D tak pomalu začínají ztrácet na významnosti.

74,183	79,688	81,897	82,747	82,872
84,843	86,111	86,126	86,683	86,744
87,000	87,025	87,160	87,445	87,946
88,609	89,552	90,114	90,168	91,233
91,583	91,615	91,946	91,980	92,482
92,837	93,023	93,408	93,690	93,762
94,070	94,818	95,050	95,226	95,315
95,658	95,713	95,788	95,948	96,017
96,027	96,050	96,246	96,502	96,622
97,199	97,236	97,497	97,531	97,743
97,746	98,234	99,027	99,093	99,101
99,497	99,679	99,811	99,878	100,644
100,819	101,779	102,309	102,532	102,835
102,887	103,163	103,424	104,607	104,664
104,897	105,282	105,720	105,784	105,861
106,143	106,176	106,203	106,535	106,669
107,081	107,342	107,401	107,618	107,768
109,009	109,787	110,076	110,398	110,782
112,157	112,769	112,928	113,128	113,944
115,928	116,955	117,831	120,670	126,197

Tabulka 10: Vzestupně uspořádaná historická data pro *Prodej* v tis. ks



Obrázek 6: Porovnání 1D a 2D simulace pro 100 původních dat

	Výběrový průměr	Výběrová. směr. odchylka
Prodej 1D	98,9908	9,4602
Prodej 2Da	100,2546	10,3475
Prodej 2Db	99,9991	10,8422
Prodej 2Dc	99,4034	9,6658
Prodej 2Dd	97,5625	10,4292
Prodej 2De	98,4051	10,6122
Prodej 2Df	99,4353	9,0973
Prodej 2Dg	99,1262	10,4609
Prodej 2Dh	100,0532	9,6265
Prodej 2Di	99,1961	9,3570
Prodej 2Dj	98,6710	10,0922

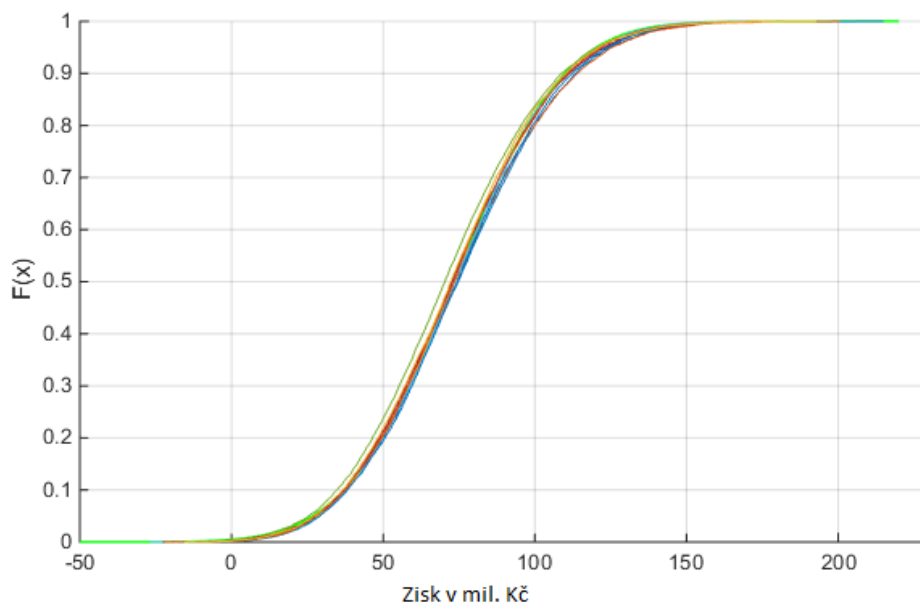
Tabulka 11: Vybrané číselné charakteristiky v tis. ks

Porovnání výstupní veličiny *Zisk* při zahrnutí a nezahrnutí nejistoty parametrů do rizikového faktoru *Prodej*

Nakonec se zaměříme na výslednou kriteriální veličinu *Zisk*, kterou jsme zobrazili na obrázku 7 pro jednotlivé varianty. Vybrané číselné charakteristiky jsou potom zachyceny v tabulce 12.

Jak můžeme z obrázku vidět, jednotlivé empirické distribuční funkce kriteriální veličiny *Zisk* jsou si již velmi podobné. Také z tabulky dostáváme jasné informace, že rozdíly mezi výsledným výstupem pro 1D a 2D simulace se ještě více snížily a to ve všech uvedených číselných charakteristikách. Očekávaný průměrný zisk dle 1D simulace je 73,45 mil. Kč, přičemž podle 2D simulací je největší očekávaný zisk pouze 75,36 mil. Kč a naopak nejmenší 71,28 mil. Kč. Simulace 2D nám poskytují výstupy velmi podobné výstupům z 1D simulací a jejich užitečnost tak upadá.

Tato poslední simulace se pak stává nejvíce věrohodnou, neboť parametry pravděpodobnostního rozdělení rizikového faktoru *Prodej* byly odhadnuty z největšího vzorku historických dat. Nejistota, které čelíme při použití odhadů parametrů pravděpodobnostního rozdělení namísto jejich skutečných hodnot, je v tomto případě nejmenší.



Obrázek 7: Empirické distribuční funkce pro *Zisk* ze 100 historických dat

	Výběř. průměr	Výběř. směr.odch.	5% - kvantil	95% - kvantil	Medián	Prav. ztráty
Zisk 1D	73,4509	28,4429	27,7523	121,9760	72,5822	0,0017
Zisk 2Da	75,3593	29,3644	28,9556	124,4983	74,7691	0,0023
Zisk 2Db	74,9477	29,6960	28,2774	125,4058	73,5949	0,0017
Zisk 2Dc	74,0182	28,3149	28,5869	121,5151	73,2961	0,0017
Zisk 2Dd	71,2795	28,9283	25,3505	120,3143	70,3108	0,0037
Zisk 2De	72,5573	29,2657	25,3698	122,2470	71,5563	0,0026
Zisk 2Df	74,0945	28,0242	28,9576	120,9874	73,4697	0,0022
Zisk 2Dg	73,6108	29,0196	27,4729	122,8289	72,7401	0,0025
Zisk 2Dh	75,0444	28,6797	29,1059	123,9624	74,4437	0,0016
Zisk 2Di	73,7540	28,3790	27,7957	121,9655	73,3050	0,0021
Zisk 2Dj	72,9087	28,5928	26,9091	121,0366	72,3021	0,0021

Tabulka 12: Vybrané číselné charakteristiky v mil. Kč

3.2.5. Souhrnný příklad k metodě klasická statistika

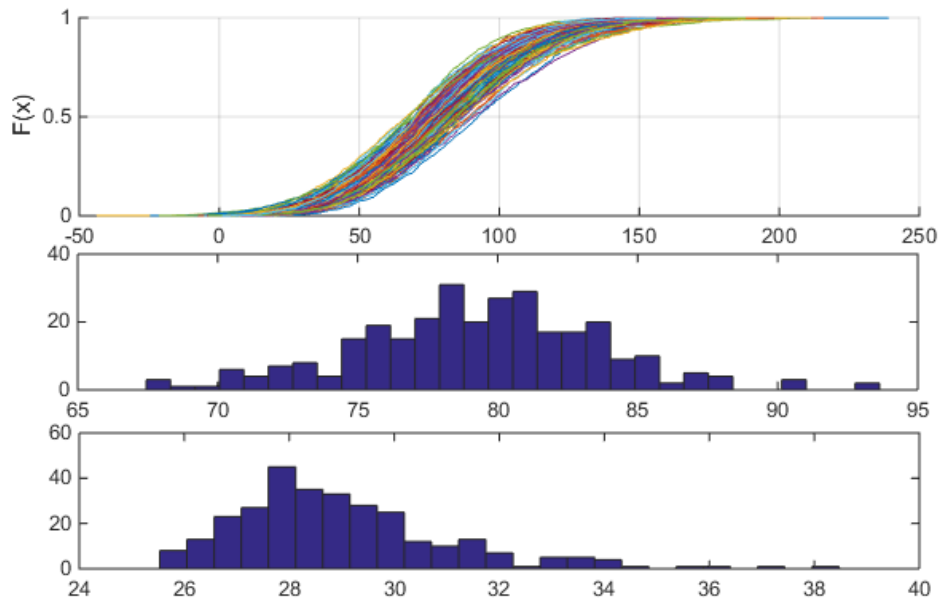
Princip metody byl pro jednoduchost ilustrována na velmi malém vzorku výsledných kritériálních veličin *Zisk*, jejichž rizikový faktor *Prodej* byl zkonstruován z normálního rozdělení z pouhých deseti dvojic parametrů μ a σ , tj. jen z deseti 2D scénářů. V této části se podíváme, jak se tato metoda bude chovat, použijeme-li místo deseti dvojic parametrů 300 těchto dvojic.

Metodu znovu uplatníme pro původních 12, 30 a následně 100 historických dat. Celý postup bude totožný s tím, který byl popsán v odpovídajících úsecích k metodě klasická statistika. Jediným rozdílem bude, jak již bylo zmíněno, že vygenerujeme 300 dvojic parametrů a následně namodelujeme z normálního rozdělení 1 000 hodnot rizikového faktoru *Prodej*, pro každou z dvojic parametrů. Toto snížení počtu hodnot rizikového faktoru z 10 000 na 1 000 je zde nutné z důvodu velké náročnosti na čas. Výstupem bude tři sta namodelovaných souborů o tisíci datech k rizikovému faktoru *Prodej*, které použijeme pro přepočítání na výstupní kritérium *Zisk*. Ostatní rizikové faktory zůstanou neměnné, jenom místo původních 10 000 hodnot vezmeme pouze prvních 1 000. Celý proces byl proveden v Matlabu, v příloženém skriptu *KlasStatistikaSouhrn.m*, odkud byly rovnou získány všechny grafické výstupy, které jsou znázorněny na následujících obrázcích 8, 9, 10, pro původních 12, 30 a 100 historických dat.

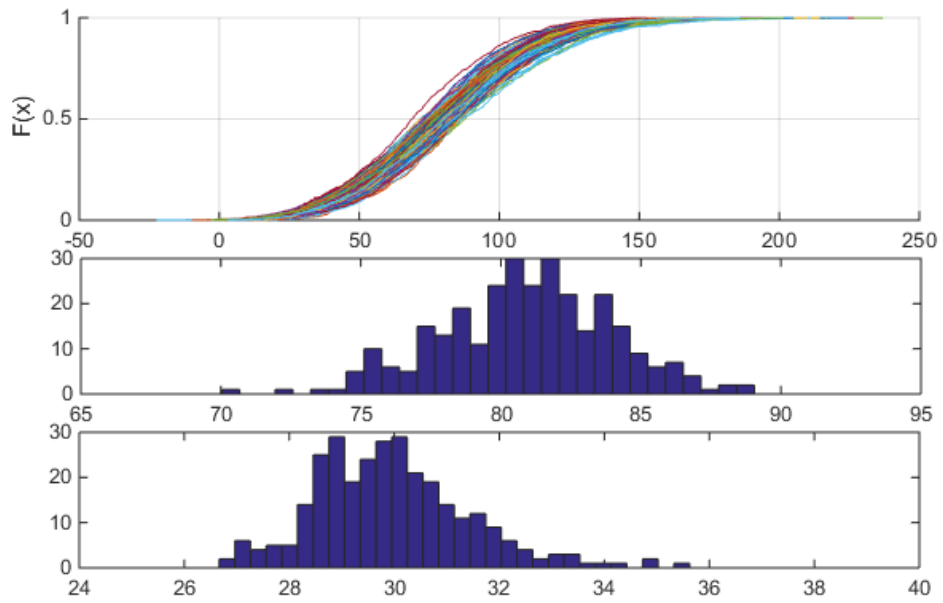
Každá z ilustrací je rozdělena na tři části:

- první, ve které vidíme empirické distribuční funkce všech tří set výstupních veličin *Zisk*,
- druhá, kde je zobrazen histogram výběrových průměrů, spočítaných z jednotlivých výstupů,
- třetí, v níž se nachází histogram výběrových směrodatných odchylek, vypočítaných z jednotlivých výstupů.

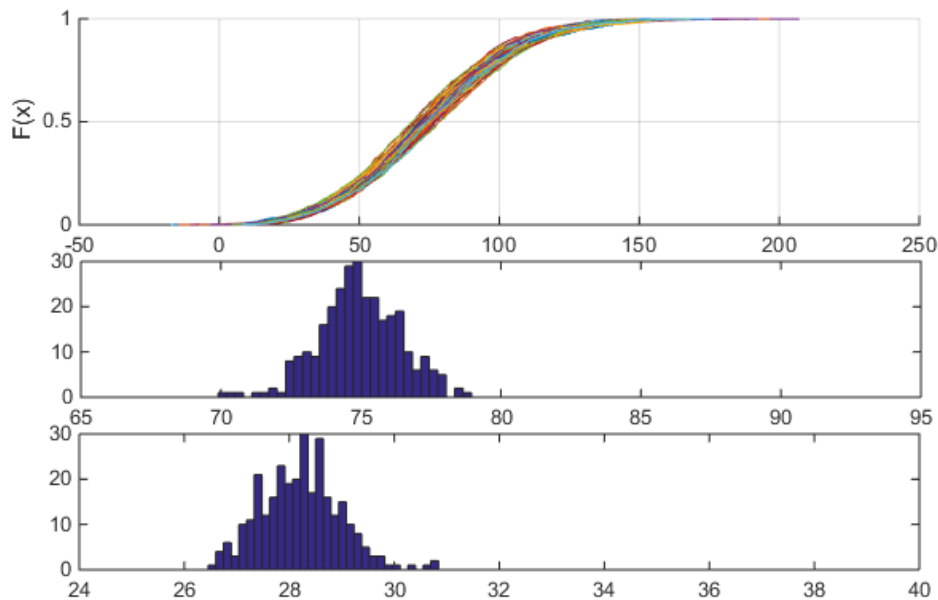
Z obrázků názorně vyplývá, že trend, který byl zachycen pro pouhých deset 2D scénářů, je v tomto případě ještě jasněji viditelný. Můžeme pozorovat, jak



Obrázek 8: Výsledné grafické výstupy pro 12 historických dat



Obrázek 9: Výsledné grafické výstupy pro 30 historických dat



Obrázek 10: Výsledné grafické výstupy pro 100 historických dat

s přibývajícím počtem původních dat dochází k významnému zužování celého „prutu“ empirických distribučních křivek. Dalším významným atributem je rozhodně skutečnost, že s přibývajícím počtem dat, dochází ke snížení variability očekávaného průměrného zisku i odchylky od tohoto průměru. Závěrem můžeme říct, že 2D simulace nám umožňují matematicky modelovat nejistotu ohledně výsledného rozdělení pravděpodobnosti nebo ohledně hodnot číselných charakteristik kritériální veličiny, způsobenou počtem historických dat, které máme k dispozici k uvažovanému rizikovému faktoru.

3.3. Bootstrap

Metoda Bootstrap byla veřejnosti představena poprvé v roce 1979 americkým statistikem Bradley Efronem v odborné publikaci *Bootstrap Methods: Another Look at the Jackknife*. Od té doby byla předmětem mnoha dalších vědeckých prací.

Původně byla tato metoda vyvinutá z techniky světu známé již mnohem dříve, která je pojmenována *The jackknife*. The jackknife se používal ke stanovení přes-

nosti statistik počítaných z datového souboru. Postup byl takový, že k danému bodovému odhadu θ se spočítala *the jackknife value* $\hat{\theta}_i$ a to stejným způsobem jako původní odhad, ale s tím rozdílem, že to bylo z modifikovaného souboru původních dat, kde byla vyjmuta i -tá hodnota. V souboru o n datech jsme tak ke statistice θ dostali n hodnot $\hat{\theta}_i$, které představují rozdělení pravděpodobnosti týkající se nejistoty tohoto odhadu. Tato metoda se ovšem ukázala jako nepřesná a byla později nahrazena právě metodou Bootstrap.

Na první pohled bychom mohli Bootstrap zařadit do metod, které nejsou příliš věrohodné. Zdání ovšem klame a tato technika si vysloužila své právoplatné místo jako velmi účinný nástroj a to hned ze dvou důvodů:

1. velmi dobře koresponduje s klasickými metodami,³ a to zejména při rozsáhlejších statistických souborech,
2. nabízí možnost modelovat nejistotu parametrů rozdělení pravděpodobnosti v případech, kde užití klasických metod není možné.

Jedná se tedy o metodu, která při správném a opatrném použití, poskytne velmi rychlým a jednoduchým způsobem správné výsledky.

Zajímavý název této metody, který jak již víme, znamená v překladu „pomož si sám“, koresponduje s frází „to pull yourself up by your bootstraps“, která pramení z příběhu v knize *Adventures of Baron Munchausen* od autora Rudolpha Ericha. V jednom z příběhů se hlavní hrdina Baron Munchausen ocitá v potížích na dně hlubokého jezera, dokud ho nenapadne se z té svízelné situace dostat *svépomocí*.

V rámci této techniky se setkáváme se dvěma přístupy, odvozenými od toho, zda známe nebo neznáme rozdělení pravděpodobnosti, ze kterého uvažovaná data pochází. Známe-li toto rozdělení, mluvíme o parametrickém Bootstrapu, v opačném případě o neparametrickém, který by popsán v kapitole 2.2.

K odhadu nejistoty parametrů využijeme pouze parametrický Bootstrap.

³např. klasická statistika

3.3.1. Parametrický Bootstrap

Pokud disponujeme znalostí o pravděpodobnostním rozdělení, ze kterého pochází uvažovaná data, potom parametrický Bootstrap uskutečníme v následujících čtyřech krocích:

1. Máme k dispozici soubor n hodnot, o kterém předpokládáme znalost jeho pravděpodobnostního rozdělení.
2. Určíme parametry tohoto rozdělení a spočítáme si ze souboru n dat jejich příslušné odhady.
3. Na základě vypočítaných odhadů parametrů si z příslušného rozdělení pravděpodobnosti náhodně vygenerujeme n nových dat. To provedeme N krát.
4. Pro každý z N souborů dat znovu spočítáme příslušné odhady parametrů, přičemž takto vzniklá rozdělení pravděpodobnosti budou představovat naši nejistotu ohledně skutečné hodnoty parametrů.

Tuto metodu budeme nyní ilustrovat na příkladu, s jehož zadáním jsme se setkali v kapitole 1.3. Vycházet budeme z **totožných historických dat**, která budou stejným způsobem rozdělena do tří skupin. Tato data, jak víme, pochází z normálního rozdělení s parametry $\mu = 100$ a $\sigma = 9$. Naším cílem pak bude v rámci každé skupiny provést simulaci rizikového faktoru *Prodej* pomocí metody parametrického Bootstrapu a následně provést propočet na výslednou kritériální veličinu *Zisk*.

3.3.2. Simulace rizikového faktoru *Prodej* z 12-ti původních dat

Znovu se nacházíme v situaci, kdy disponujeme dvanácti historickými daty k rizikovému faktoru *Prodej*, která chceme využít k jeho simulaci. Nejprve potřebujeme vypočítat příslušné bodové odhady parametrů normálního rozdělení. Ty ovšem známe z předchozí kapitoly a jsou to hodnoty $\bar{x} = 101,826$ a $s = 9,095$. Nyní si vygenerujeme deset dvanáctic z normálního rozdělení s vypočítanými odhady parametrů, tj. z rozdělení $N(\bar{x}, s)$. V rámci každé z deseti dvanáctic dat si

znovu spočítáme příslušné bodové odhady, které pro lepší přehled označíme jako $\hat{\mu}_i$, a $\hat{\sigma}_i$, kde $i = 1, \dots, 10$. Tyto dvojice odhadů jsou pak zobrazeny v tabulce 13. Nakonec uskutečníme deset simulací rizikového faktoru, a to tak, že vygenerujeme 10 000 scénářů z rozdělení $N(\hat{\mu}_i, \hat{\sigma}_i)$ pro $i = 1, \dots, 10$. Tímto postupem jsme znovu provedli 2D simulaci, tentokrát pomocí metody parametrického Bootstrapu. Jednotlivé kroky byly opět provedeny v matematickém softwaru Matlab, pomocí přiloženého skriptu *ParametBootstrap.m* a výsledky byly pro lepší přehled exportovány do souboru *Bootstrap.xlsx* do příslušné záložky *Prodej12*.

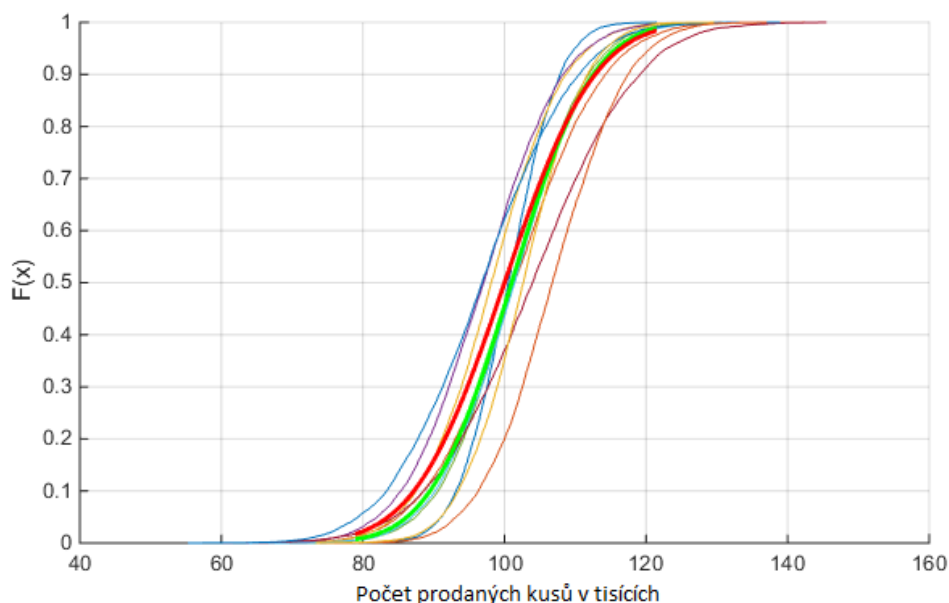
	$\hat{\mu}_i$	$\hat{\sigma}_i$
i=1	100,5079	5,7507
i=2	101,3679	10,0148
i=3	98,0803	8,4141
i=4	96,8592	9,0664
i=5	101,0432	8,3266
i=6	101,2937	8,7225
i=7	104,0090	11,7298
i=8	96,9660	10,6325
i=9	106,8022	8,3271
i=10	102,6436	7,0938

Tabulka 13: Hodnoty odhadů parametrů pro 12 původních dat

Porovnání rizikového faktoru *Prodej* pro 1D a 2D simulace

Pojďme se podívat na rozdíly mezi 1D simulací, při nezahrnutí nejistoty ohledně parametrů, a 2D simulacemi, při zahrnutí této nejistoty. Pro srovnání využijeme 1D simulaci, která byla zkonstruována v rámci předchozí metody klasická statistika pro tentýž případ, kdy disponujeme dvanácti historickými daty. Odpovídající empirická distribuční funkce bude tedy stejného tvaru a opět bude světle zelené barvy. Do úvah také zahrneme i červenou distribuční funkci normálního rozdělení, ze kterého data pochází. Všechny distribuční funkce můžeme vidět na obrázku 11.

Můžeme zpozorovat, že empirické distribuční funkce jsou dost roztažené okolo červené křivky původního rozdělení pravděpodobnosti. Tento závěr podtrhují i číselné charakteristiky, zobrazené v tabulce 14, kde si můžeme všimnout celkem



Obrázek 11: Porovnání 1D a 2D simulací pro 12 původních dat

velké variability v těchto charakteristikách. Například pro první 2D křivku je vy počítaná výběrová směrodatná odchylka pouze 5,74 tis. kusů, což je velký rozdíl oproti skutečné hodnotě 9 tis. kusů.

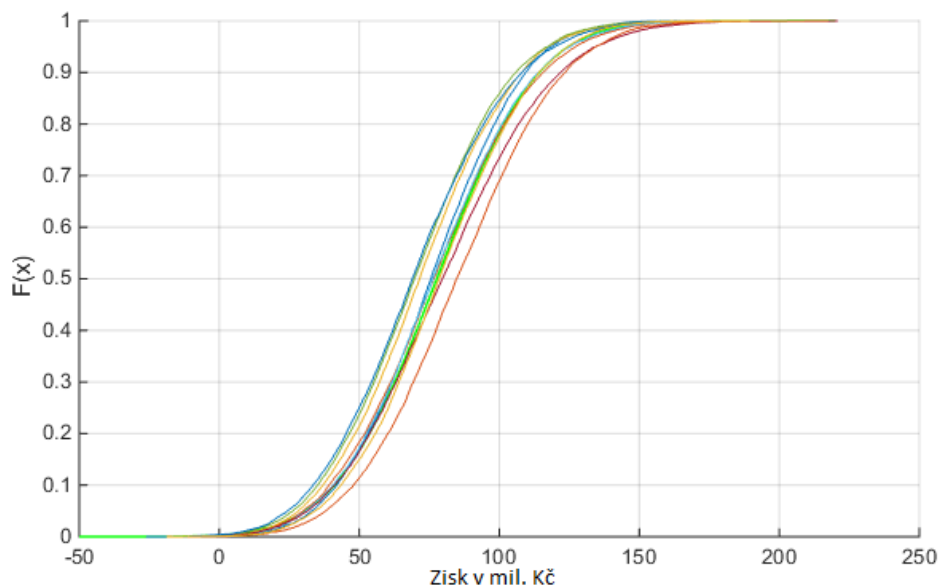
Porovnání výstupní veličiny *Zisk* při zahrnutí a nezahrnutí nejistoty parametrů do rizikového faktoru *Prodej*

Zaměříme se nyní na celkovou simulaci výstupní veličiny *Zisk*, kterou jsme opět dostali propočtem rizikových faktorů pomocí uvedeného matematického vzorce (1). Situace je znázorněná na obrázku 12, kde je zachycena světle zelená empirická distribuční funkce kritériální veličiny *Zisk* pro 1D simulaci a deset empirických distribučních funkcí pro 2D simulaci, které byly spočteny z deseti obměn rizikového faktoru *Prodej*. Celý propočet byl proveden v příslušném souboru v záložce *Zisk12*.

Jak lze z obrázku vidět, jednotlivé křivky se od sebe odlišují, což je zřejmé i z tabulky vybraných číselných charakteristik 15. Příkladem může být např. očekávaný průměrný zisk, který je u 1D simulace 77,66 mil. Kč. U čtvrté 2D simulace lze očekávat zisk pouze 70,26 mil. Kč a naopak u deváté 2D simulace

	Výběrový průměr	Výběrová směr. odchylka
Prodej 1D	101,7911	9,0898
Prodej 2Da	100,5250	5,7377
Prodej 2Db	101,4558	9,9845
Prodej 2Dc	98,1257	8,3343
Prodej 2Dd	96,9045	9,0247
Prodej 2De	101,1369	8,3553
Prodej 2Df	101,3664	8,7913
Prodej 2Dg	103,8488	11,8271
Prodej 2Dh	96,8092	10,5868
Prodej 2Di	106,8681	8,2665
Prodej 2Dj	102,6210	7,1222

Tabulka 14: Vybrané číselné charakteristiky v tis. ks



Obrázek 12: Empirické distribuční funkce pro *Zisk* z 12-ti historických dat

85,41 mil. Kč. Tyto velké rozdíly mezi 1D a 2D simulacemi jsou opět způsobeny nezahrnutím nejistoty parametrů u výpočtu 1D simulací.

	Výběr. průměr	Výběr. směr.odch.	5% - kvantil	95% - kvantil	Medián	Prav. ztráty
Zisk 1D	77,6601	28,5624	32,2091	125,7442	77,1198	0,0014
Zisk 2Da	75,7368	26,0937	33,1140	118,6389	75,6112	0,0011
Zisk 2Db	77,1997	29,3926	29,9076	127,2230	76,6582	0,0015
Zisk 2Dc	72,1187	27,3122	28,2103	117,3072	71,4906	0,0018
Zisk 2Dd	70,2619	27,5178	26,2210	117,3701	69,6756	0,0028
Zisk 2De	76,6789	27,9277	31,7389	123,6457	76,2926	0,0012
Zisk 2Df	77,0682	28,5167	30,8023	125,3616	76,3212	0,0012
Zisk 2Dg	80,8100	31,4290	31,4333	134,3176	79,6710	0,0013
Zisk 2Dh	70,1271	28,8246	24,2874	119,9295	68,9488	0,0040
Zisk 2Di	85,4125	29,3025	37,9195	134,5426	84,7512	0,0006
Zisk 2Dj	78,9189	27,3377	34,2035	124,5239	78,5273	0,0011

Tabulka 15: Vybrané číselné charakteristiky v mil. Kč

3.3.3. Simulace rizikového faktoru *Prodej* z 30-ti původních dat

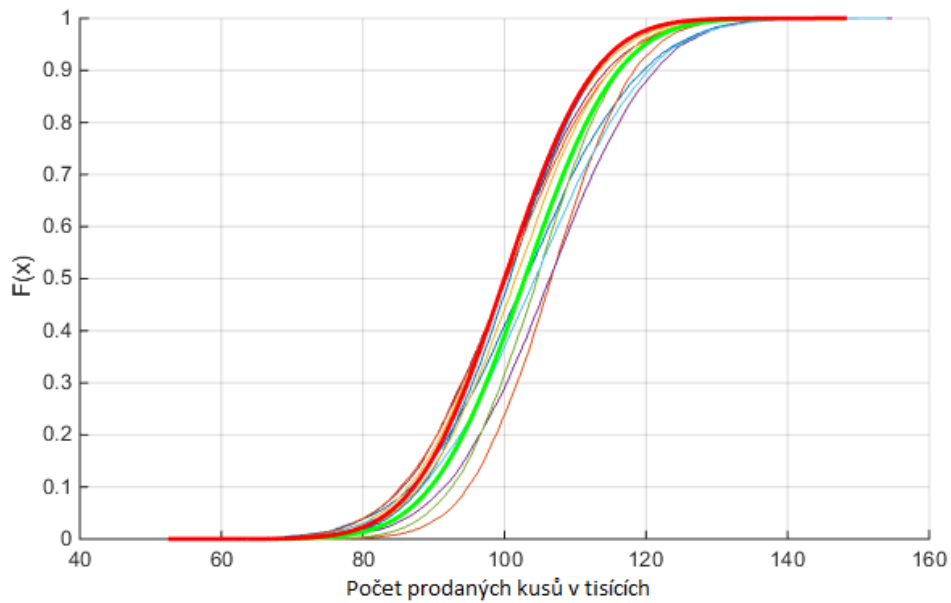
Nyní disponujeme 30-ti historickými daty, které využijeme pro simulaci rizikového faktoru *Prodej* a to s pomocí stejného skriptu v softwaru Matlab. Celý postup konstrukce 2D simulace je totožný s předchozím případem, akorát s tím rozdílem, že nyní vygenerujeme deset třicetic dat, namísto předchozích deseti dvanáctic dat. Následně v každé desetici spočítáme příslušné dvojice bodových odhadů $\hat{\mu}_i$, a $\hat{\sigma}_i$, které najdeme v tabulce 16, a ze kterých uskutečníme simulace rizikového faktoru. Potřebné číselné charakteristiky jsou $\bar{x} = 102, 860$ a $s = 10, 361$. Výsledky simulace jsou zobrazeny v záložce *Prodej30*.

Porovnání rizikového faktoru *Prodej* pro 1D a 2D simulace

Podíváme se na rozdíly mezi 1D a 2D simulacemi rizikového faktoru *Prodej*, přičemž pro toto srovnání využijeme 1D simulaci, která byla zkonstruována v rámci minulé metody pro původních třicet dat. Výsledek vidíme na obrázku 13, kde 1D simulace je vyobrazena opět světle zelenou empirickou distribuční funkcí a červenou distribuční funkcí je znázorněno rozdělení, ze kterého data pochází.

	$\hat{\mu}_i$	$\hat{\sigma}_i$
i=1	102,8475	12,8812
i=2	106,6868	9,1897
i=3	99,9470	10,5100
i=4	106,2408	11,4743
i=5	104,3367	9,3055
i=6	104,3052	12,4833
i=7	100,1378	11,2090
i=8	100,5719	9,7797
i=9	100,2635	11,5178
i=10	101,5763	10,5039

Tabulka 16: Hodnoty odhadů parametrů pro 30 původních dat



Obrázek 13: Porovnání 1D a 2D simulací pro 30 původních dat

Jak je patrné z obrázku, došlo ke zúžení pásu empirických funkcí. Jednotlivé křivky si jsou více podobné, což je následkem zvýšeného počtu historických dat, ze kterého jsou příslušné simulace provedeny. Naše nejistota ohledně parametrů uvažovaného rozdělení totiž se zvýšeným počtem vstupních dat klesá. Menší rozdílnost mezi empirickými distribučními funkcemi zobrazuje také tabulka vybraných číselných charakteristik 17.

	Výběrový průměr	Výběrová směr. odchylka
Prodej 1D	102,8346	10,2467
Prodej 2Da	102,9885	12,8640
Prodej 2Db	106,6059	9,1793
Prodej 2Dc	99,9420	10,4899
Prodej 2Dd	106,3486	11,5613
Prodej 2De	104,5038	9,3088
Prodej 2Df	104,2558	12,6021
Prodej 2Dg	99,9906	11,3185
Prodej 2Dh	100,6136	9,6937
Prodej 2Di	100,1947	11,4836
Prodej 2Dj	101,5116	10,5647

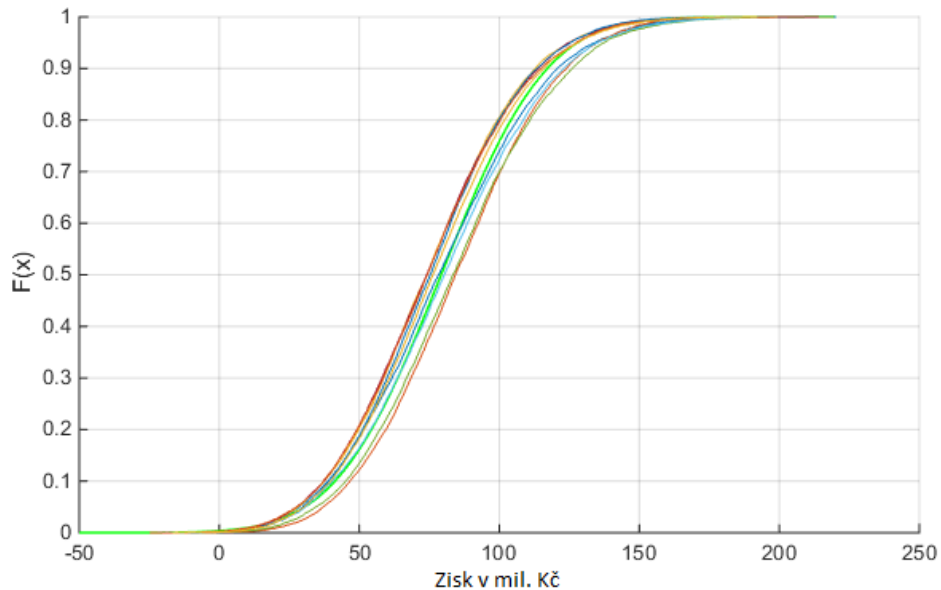
Tabulka 17: Vybrané číselné charakteristiky v tis. ks

Porovnání výstupní veličiny *Zisk* při zahrnutí a nezahrnutí nejistoty parametrů do rizikového faktoru *Prodej*

Zaměříme se na celkovou simulaci výstupní veličiny *Zisk*, zobrazenou v jednácti variantách. Situace je znázorněná na obrázku 14, kde světle zelená empirická distribuční funkce představuje 1D simulaci a zbývajících deset funkcí pak znázorňuje varianty 2D simulace. Celý propočítání na výslednou výstupní veličinu byl proveden v příslušném souboru v záložce *Zisk30*.

Znovu se setkáváme se stejným trendem jako v předchozí situaci, kdy z obrázku vidíme, že došlo ke zúžení celého pásu empirických distribučních funkcí. Jednotlivé křivky se od sebe již tolik neodlišují. Tento závěr je zřejmý i z tabulky vybraných číselných charakteristik 18. Příkladem může být očekávaný průměrný zisk, který je u 1D simulace 79,23 mil. Kč, u 2D simulací je pak nejméně 74,86 mil.

Kč a naopak nejvíce 84,96 mil. Kč. Oproti předchozí situaci tak došlo ke značné změně v rozdílu mezi očekávanými průměrnými zisky u 1D a 2D simulací.



Obrázek 14: Empirické distribuční funkce pro *Zisk* z 30-ti historických dat

3.3.4. Simulace rizikového faktoru *Prodej* ze 100 původních dat

Naposledy se podíváme na simulaci rizikového faktoru *Prodej*, tentokrát za předpokladu, že disponujeme sto historickými daty. Jediný rozdíl bude opět spočívat v deseti vygenerovaných Nticích dat, neboť nyní se bude jednat o stotici. Příslušné dvojice bodových odhadů pro každou z deseti skupin dat jsou zobrazeny v tabulce 19. Následně znovu provedeme 2D simulaci. Využijeme k tomu stejného souboru v softwaru Matlab jako v předchozí situaci a výsledky exportujeme do příslušné záložky v souboru *Bootstrap.xlsx*. Potřebná dvojice číselných charakteristik pak je $\bar{x} = 89,975$ a $s = 9,542$.

Porovnání rizikového faktoru *Prodej* pro 1D a 2D simulace

Potřetí se zaměříme na rozdíly u faktoru *Prodej*, pro případy, kdy nezahrneme a zahrneme nejistotu parametrů. Pro srovnání použijeme 1D simulaci, která byla

	Výběr. průměr	Výběr. směr.odch.	5% - kvantil	95% - kvantil	Medián	Prav. ztráty
Zisk 1D	79,2328	29,5980	31,6058	129,7928	78,5415	0,0008
Zisk 2Da	79,5116	32,0874	28,5158	134,1804	78,3141	0,0024
Zisk 2Db	84,9606	29,6061	37,0441	134,9163	84,2776	0,0010
Zisk 2Dc	74,8584	29,3205	28,0289	125,1156	74,1666	0,0012
Zisk 2Dd	84,5586	31,4794	34,4414	137,2791	83,4705	0,0012
Zisk 2De	81,7745	29,2557	34,6476	130,4380	81,3584	0,0009
Zisk 2Df	81,4050	31,9788	30,8223	134,8721	80,3179	0,0018
Zisk 2Dg	74,8907	29,8157	27,5049	125,0322	74,0825	0,002
Zisk 2Dh	75,8867	28,8423	29,7317	125,3826	75,3212	0,0018
Zisk 2Di	75,2336	30,1699	27,5128	127,3665	74,3085	0,0026
Zisk 2Dj	77,2357	29,6712	30,3069	128,0245	76,2857	0,0015

Tabulka 18: Vybrané číselné charakteristiky v mil. Kč

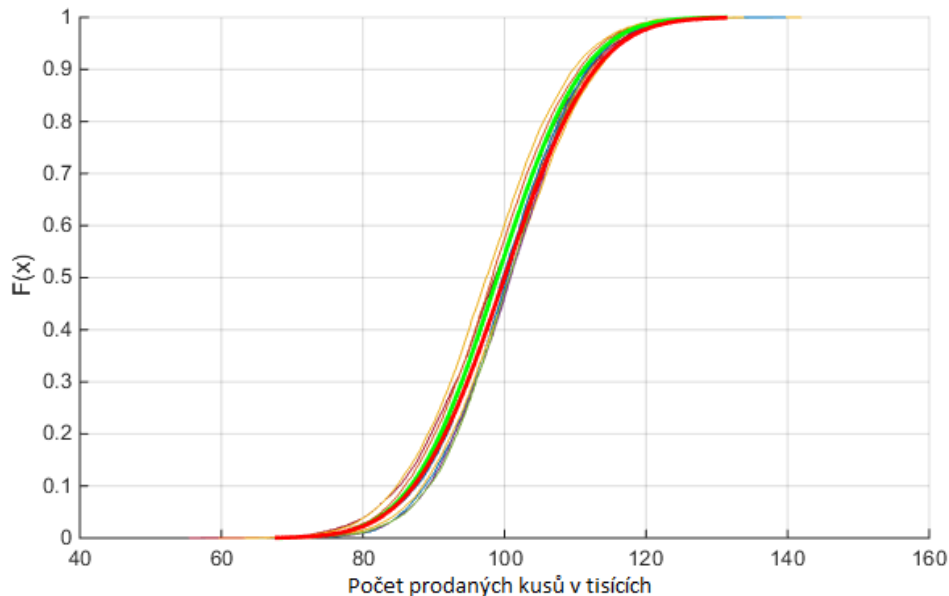
	$\hat{\mu}_i$	$\hat{\sigma}_i$
i=1	100,4968	8,9027
i=2	99,5720	9,7842
i=3	100,8708	9,7641
i=4	100,6535	9,1682
i=5	100,6735	8,9387
i=6	98,9443	9,4791
i=7	98,6391	10,4140
i=8	99,5021	9,43457
i=9	98,3557	9,5823
i=10	97,4205	9,8106

Tabulka 19: Hodnoty odhadů parametrů pro 100 původních dat

zkonstruovaná v rámci minulé metody pro původních sto dat. Výsledné empirické distribuční funkce můžeme vidět na obrázku 15, kde světle zelená funkce značí 1D simulaci a červená funkce je pro normální rozdělení, ze kterého původní data pochází. V tabulce 20 pak nalezneme vybrané číselné charakteristiky pro jednotlivé funkce.

Z obrázku si na první pohled všimneme, že došlo k výraznému zúžení pásu empirických distribučních funkcí. Odlišnost mezi jednotlivými křivkami značně poklesla, což je patrné i ze zmíněné tabulky. Tento závěr je pak jedině logický, neboť simulace rizikového faktoru je provedena na základě již celkem velkého

počtu původních dat a naše nejistota ohledně odhadu parametrů je malá.



Obrázek 15: Porovnání 1D a 2D simulací pro 100 původních dat

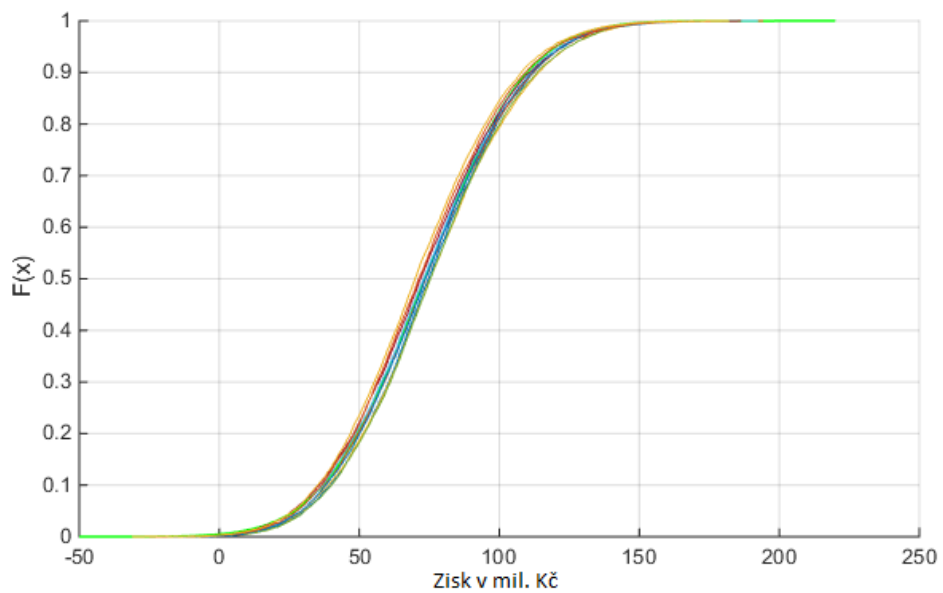
Porovnání výstupní veličiny *Zisk* při zahrnutí a nezahrnutí nejistoty parametrů do rizikového faktoru *Prodej*

Nakonec posoudíme výsledné simulace kritériální veličiny *Zisk*, zobrazené na obrázku 16. Světle zelená empirická distribuční funkce představuje 1D simulaci a zbývajících deset empirických distribučních funkcí pak znázorňuje 2D simulace. Celý propočtení na výslednou výstupní veličinu byl proveden v příslušném souboru v záložce *Zisk100*.

Z obrázku je na první pohled vidět, že všechny výsledné distribuční funkce jsou si velmi podobné. O tom vypovídá také tabulka číselných charakteristik 21, kde se rozdíly mezi jednotlivými variantami kritériální veličiny *Zisk* podstatně zmenšily. Pro ilustraci se zaměříme na očekávaný průměrný zisk, který je pro 1D simulaci 73,45 mil. Kč, přičemž pro 2D simulaci je nejméně 71,00 mil. Kč a nejvíce 76,22 mil. Kč. Simulace 2D se tak pomalu přestávají odlišovat od 1D simulací, neboť pracujeme s poměrně velkým počtem historických dat a naše nejistota

	Výběrový průměr	Výběrová směr. odchylka
Prodej 1D	98,9908	9,4602
Prodej 2Da	100,3577	8,8789
Prodej 2Db	99,7390	9,7022
Prodej 2Dc	100,7900	9,6895
Prodej 2Dd	100,8451	9,1530
Prodej 2De	100,6538	8,8670
Prodej 2Df	99,0544	9,5649
Prodej 2Dg	98,5595	10,4838
Prodej 2Dh	99,7051	9,5083
Prodej 2Di	98,2300	9,5274
Prodej 2Dj	97,3865	9,7135

Tabulka 20: Vybrané číselné charakteristiky v tis. ks



Obrázek 16: Empirické distribuční funkce pro *Zisk* ze 100 historických dat

ohledně parametrů pravděpodobnostního rozdělení uvažovaného kritéria se stává zanedbatelnou.

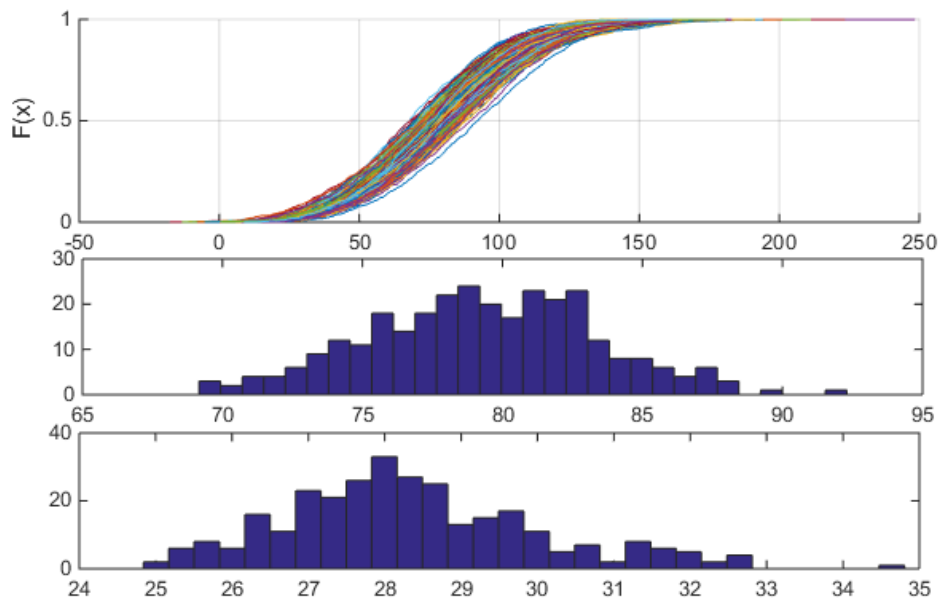
	Výběr. průměr	Výběr. směr.odch.	5% - kvantil	95% - kvantil	Medián	Prav. ztráty
Zisk 1D	73,4509	28,4429	27,7523	121,9760	72,5822	0,0017
Zisk 2Da	75,4779	28,0218	30,3845	123,1321	74,6752	0,0014
Zisk 2Db	74,5686	28,7121	29,3424	122,7671	73,6948	0,0020
Zisk 2Dc	76,1451	28,7880	29,7311	124,5969	75,3351	0,0016
Zisk 2Dd	76,2211	28,3939	30,5147	124,7284	75,3187	0,0019
Zisk 2De	75,9708	28,3022	29,7023	123,6329	75,1399	0,0010
Zisk 2Df	73,5275	28,4024	27,8372	121,6759	72,8810	0,0021
Zisk 2Dg	72,7949	29,1109	25,8664	122,3796	71,9218	0,0028
Zisk 2Dh	74,4858	28,3675	29,4335	122,8440	73,5723	0,0016
Zisk 2Di	72,2903	28,3609	27,6487	120,0226	71,1823	0,0026
Zisk 2Dj	70,9970	28,2130	26,2150	118,9896	69,9561	0,0031

Tabulka 21: Vybrané číselné charakteristiky v mil. Kč

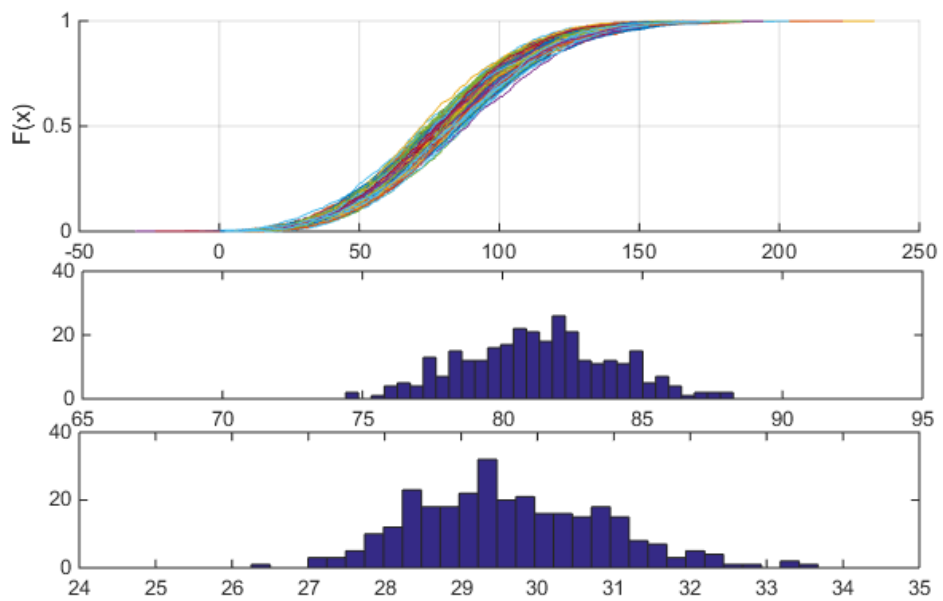
3.3.5. Souhrnný příklad k metodě parametrický Bootstrap

Metoda byla opět pro jednoduchost ilustrována na vzorku pouhých deseti výsledných kritériálních veličin *Zisk*, kde rizikový faktor *Prodej* byl zkonstruován z deseti dvojic parametrů μ a σ , tj. jen z deseti 2D scénářů. Podívejme se, co se stane, použijeme-li místo deseti dvojic parametrů, 300 těchto dvojic.

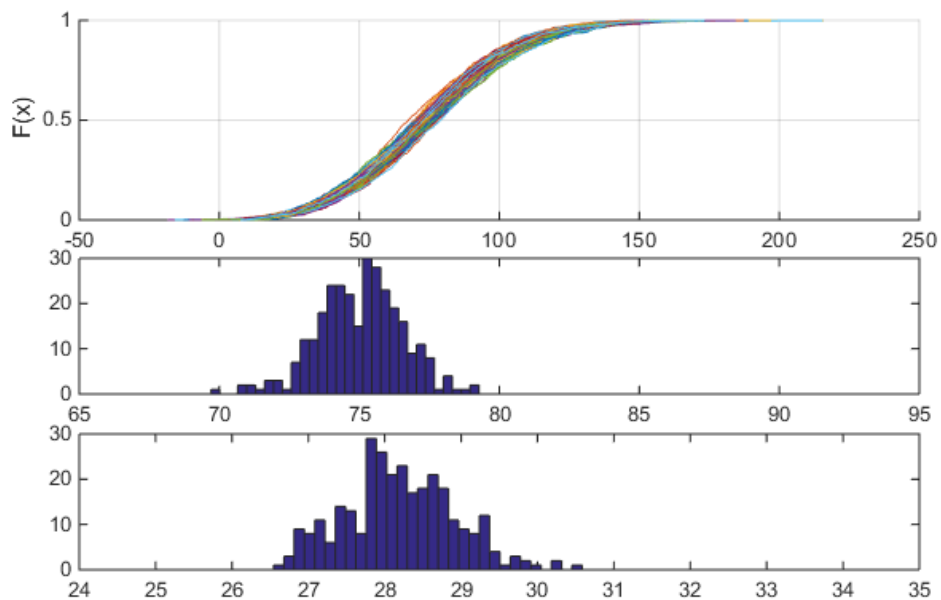
Znovu budeme uvažovat tři situace, kdy disponujeme 12-ti, 30-ti a následně 100 historickými daty. Postup konstrukce bude shodný s tím, který byl k této metodě popsán výše. Rozdíl spočívá v tom, že budeme pracovat s 300 dvojicemi parametrů, ze kterých následně vygenerujeme z normálního rozdělení 1 000 hodnot rizikového faktoru *Prodej*. Těchto tři sta namodelovaných souborů o tisíci datech použijeme pro přepočítání na výstupní kritérium *Zisk*, přičemž k ostatním rizikovým faktorům uvažujeme pouze jejich prvních 1 000 hodnot. Uvedený postup byl proveden v Matlabu, v příloženém souboru *ParametBootstrapSouhrn.m*, ze kterého jsme ihned obdrželi následující grafické výstupy 17, 18, 19, pro jednotlivé situace.



Obrázek 17: Výsledné grafické výstupy pro 12 historických dat



Obrázek 18: Výsledné grafické výstupy pro 30 historických dat



Obrázek 19: Výsledné grafické výstupy pro 100 historických dat

Každý z obrázků je rozdělen na tři části jako v předchozím případě, nejprve vidíme empirické distribuční funkce výstupní veličiny $Zisk$, posléze histogram výběrových průměrů a nakonec histogram výběrových směrodatných odchylek.

Z těchto ilustrací vyplývá, že tendence, které byly zachyceny při pouhých deseti 2D scénářích, jsou zcela opodstatněné a nejednalo se o nahodilou situaci. Můžeme jasně pozorovat, že čím větší je počet původní dat, tím více jsou si jednotlivé empirické distribuční funkce podobnější. Také s přibývajícím počtem historických dat dochází ke snížení variability očekávaného průměrného zisku i směrodatné odchylky. Závěrem tedy je, že čím více máme původních dat, tím stabilnější výstupy dostáváme.

Nakonec si porovnejme tyto výstupy s těmi, které jsme dostali v rámci souhrnného příkladu u klasické statistiky. Jak vidíme, výsledky se zvláště neodlišují. Z obou metod vplynuly vcelku shodné závěry a záleží na nás, kterou z metod v praxi uplatníme.

3.4. Bayesova statistika

Bayesova statistika je podle [3] velmi schopná metoda. Dle [1] se používá většinou pro pokročilou analýzu rizika a bohužel se zatím uplatňuje hlavně v ne-finanční oblasti, např. při analýze rizika šíření epidemií, testování působení léku atd.

Metoda je založená na Bayesově větě, která plyne ze vztahů podmíněné pravděpodobnosti. Pro jevy A a B platí:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (10)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (11)$$

Značí-li jev A například, že „osoba je kuřák“ a jev B , že „osoba má nadváhu“, potom symbol $(A \cap B)$ znamená, že se jedná o člověka, který je kuřák a má nadváhu. Důležitá je v tomto textu tzv. podmíněná pravděpodobnost $P(A|B)$, značící pravděpodobnost, s jakou když bude uvažovaná osoba kuřák, bude mít nadváhu, resp. $P(B|A)$, představující pravděpodobnost, že osoba mající nadváhu je kuřák.

Nyní si ze vztahu (11) vyjádříme:

$$P(A \cap B) = P(B|A)P(A) \quad (12)$$

a dosadíme do vztahu (10). Dostaneme tak slavnou Bayesovu větu:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (13)$$

kteřou můžeme rozšířit pro více jevů A_1, \dots, A_n , za podmínky B , na následující vztah:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}, i = 1, \dots, n. \quad (14)$$

Pokud uvažujeme jev X , jehož rozdělení pravděpodobnosti závisí na parametru $\theta_0 \in \Theta$, pak rozdělení pravděpodobnosti tohoto parametru můžeme vyjádřit

pomocí následujícího vztahu:

$$f(\theta_0|X) = \frac{\pi(\theta_0)l(X|\theta_0)}{\int_{\Theta} \pi(\theta)l(X|\theta)d\theta}, \quad (15)$$

kde:

- $\pi(\theta_0)$ je apriorní rozdělení pravděpodobnosti. Vyjadřuje naši předem danou představu o možných hodnotách parametru θ_0 , ještě předtím, než obdržíme hodnoty jevu X . Jinými slovy, $\pi(\theta_0)$ nevyjadřuje klasické rozdělení pravděpodobnosti, ale spíše *rozdělení nejistoty*, kterému jsme vystaveni při odhadu hodnot parametru θ_0 .
- $l(X|\theta_0)$ je věrohodnostní funkce, která představuje pravděpodobnost, že budeme pozorovat jevy X při stanovených hodnotách θ_0 .
- $f(\theta_0|X)$ je aposteriorní rozdělení, které popisuje naši znalost o hodnotách θ_0 poté, co jsme obdrželi data X . Toto rozdělení je jmenovatelem normalizováno tak, aby jeho integrál (pro diskrétní případ - jeho součet) byl roven jedné. Tato normalizace je zde nutná z toho důvodu, aby hodnoty této funkce představovaly pravděpodobnosti.

Metoda Bayesovy statistiky pro odhad nejistoty parametrů pravděpodobnostního rozdělení představuje jakýsi *proces učení*. V prvním kroku zformujeme naši subjektivní představu o potenciálních hodnotách parametru θ_0 , tedy dojde ke stanovení jeho apriorního rozdělení. Posléze pomocí empirických pozorování upřesňujeme toto apriorní rozdělení, a to přes věrohodnostní funkci. Tvar této funkce se mění v závislosti na množství informací, které z napozorovaných dat získáváme. Bude-li informace v napozorovaných datech malá, pak věrohodnostní funkce bude pokrývat široký interval hodnot a naopak. Zároveň platí, že čím více si budou tvary věrohodnostní funkce a apriorního rozdělení pravděpodobnosti podobnější, tím méně dodatečných informací z pozorovaných dat získáváme.

Pojďme si tento proces učení ilustrovat na jednoduchém příkladu. Uvažujme, že máme v peněžence pět mincí, přičemž čtyři z nich jsou pravé a jedna je falešná.

Falešná mince je těžší na jedné straně a tak se liší pravděpodobností, se kterou padne panna. Zatímco u pravé mince je tato pravděpodobnost 50 %, u falešné je 80 %. Nás bude zajímat, jaká je pravděpodobnost, že námi vybraná mince je mince falešná. Nejprve určíme apriorní pravděpodobnosti:

$$\pi(\text{pravá mince}) = 0,80,$$

$$\pi(\text{falešná mince}) = 0,20.$$

Následně budeme sestavovat věrohodnostní funkci. Hodíme si náhodně vybranou mincí a předpokládejme, že padne panna. Obdržíme tak dodatečnou informaci a můžeme se znovu ptát, jaká je pravděpodobnost, že je tato mince falešná? Věrohodnostní funkce je pak následující:

$$l(X|\theta_1): l(\text{padne panna}|\text{pravá mince}) = 0,50,$$

$$l(X|\theta_2): l(\text{padne panna}|\text{falešná mince}) = 0,80.$$

Nyní můžeme spočítat aposteriorní pravděpodobnosti pomocí Bayesovy věty:

$$f(\theta_1|X): f(\text{pravá mince}|\text{padne panna}) = 0,80 \cdot 0,50 / (0,80 \cdot 0,50 + 0,20 \cdot 0,80) = 0,714,$$

$$f(\theta_2|X): f(\text{falešná mince}|\text{padne panna}) = 0,20 \cdot 0,80 / (0,80 \cdot 0,50 + 0,20 \cdot 0,80) = 0,286.$$

Na základě provedeného pozorování se upravilo pravděpodobnostní rozdělení toho, zda je mince falešná či nikoli. Pravděpodobnost, že se jedná o falešnou minci, se zvýšila z 20 % na 28,6 %.

V dalším kroku znovu hodíme mincí a to celkově tři krát. Uvažujme, že ve všech případech nám padne panna - znovu přehodnotíme míru naší neznalosti pomocí Bayesovy věty:

$$\pi(\text{pravá mince}) = 0,714,$$

$$\pi(\text{falešná mince}) = 0,286,$$

$$l(X|\theta_1): l(\text{padne 3x panna}|\text{pravá mince}) = 0,50 \cdot 0,50 \cdot 0,50 = 0,125,$$

$$l(X|\theta_2): l(\text{padne 3x panna}|\text{falešná mince}) = 0,80 \cdot 0,80 \cdot 0,80 = 0,512,$$

$$f(\theta_1|X): f(\text{pravá mince}|\text{padne 3x panna}) = 0,714 \cdot 0,125 / (0,714 \cdot 0,125 + 0,286 \cdot 0,512) = 0,379,$$

$$f(\theta_2|X): f(\text{falešná mince}|\text{padne 3x panna}) = 0,286 \cdot 0,512 / (0,714 \cdot 0,125 + 0,286 \cdot 0,512) = 0,621.$$

Znovu pozorujeme, že s dodatečnou informací, že padla tři krát po sobě panna, se významně upravila pravděpodobnost ohledně odhadu pravosti mince. Nyní už existuje 62,1% šance, že mince je falešná.

Pro složitější ilustrační příklady, může čtenář nahlédnout např. do [1] nebo [3].

Závěr

V diplomové práci jsme si nejprve představili pojem simulace Monte Carlo, dozvěděli se něco o historii a nakonec popsali postup konstrukce této metody. Simulaci jsme následně uplatnili na ilustračním příkladu, na kterém jsme naznačili problematiku, kterou se tato práce zabývá.

Ve druhé kapitole jsme se více věnovali jednomu z kroků simulace Monte Carlo - stanovení rozdělení pravděpodobnosti klíčových rizikových faktorů. Jak jsme zjistili z předchozí kapitoly, tak toto rozdělení můžeme stanovit dvěma způsoby, buď využít názory expertů nebo provést analýzu historických dat, máme-li taková data k dispozici. V této části práce jsme se zabývali právě druhou z uvedených možností, a to analýzou historických dat. Dozvěděli jsme se, že tato analýza může být provedena dvěma způsoby, lišícími se skutečností, zda máme či nemáme informaci o rozdělení pravděpodobnosti, ze kterého data pocházejí. Pokud neznáme pravděpodobnostní rozdělení historických dat, pak můžeme uplatnit metodu neparametrického Bootstrapu, která byla následně popsána a ilustrována na příkladu. V opačném případě můžeme provést testy o vhodnosti předpokládaného rozdělení pravděpodobnosti nebo využít některých softwarů, které tyto testy uskuteční za nás.

Třetí a také poslední kapitola byla kapitolou stěžejní. Zabývali jsme se zde problémem, že pokud na základě analýzy historických dat určíme rozdělení pravděpodobnosti, které následně využijeme k tomu, abychom nasimulovali rizikový faktor, pak se dostáváme do potíží ohledně parametrů tohoto rozdělení. K modelování rizikového faktoru totiž použijeme namísto skutečných hodnot parametrů pouze jejich bodové odhady a s těmi je spojena nejistota. S touto záležitostí koresponduje 2D simulace, která je v této části charakterizována. Následně jsme si popsali tři přístupy zabývající se touto problematikou. V první řadě jsme se seznámili s metodou klasické statistiky a ilustrovali jsme ji na příkladu z první kapitoly, který byl sestaven ve třech variantách. Uvažovali jsme totiž, že k jednomu z rizikových faktorů máme k dispozici historická data, která získáváme postupně, nejprve máme 12, následně 30 a nakonec 100 těchto dat. Za použití 2D simulace

jsme ověřili, že nejistota ohledně parametrů je ovlivněna právě počtem těchto historických dat. Čím více máme těchto dat, tím menší je naše nejistota ohledně parametrů pravděpodobnostního rozdělení. Další metodu parametrického Bootstrapu jsme detailně popsali a opět ilustrovali na stejném příkladu jako metodu předchozí. Závěr byl obdobný. Se zvětšujícím se počtem vstupních historických dat se zmenšovala naše nejistota ohledně parametrů. Obě metody pak vykazovaly velmi podobné výsledky. Poslední technika Bayesovy statistiky byla teoreticky zpracována a použita na jednoduchém příkladu, kde jsme si ukázali, jak metoda funguje.

Tato práce pro mě byla velkým přínosem, neboť jsem prohloubila své znalosti v oblasti simulace Monte Carlo a seznámila se s jedním z možných úskalí v rámci této metody. Dozvěděla jsem se, jak pracují různé metody, zabývající se touto problematikou, které ovšem mohou uplatnit i v jiných oblastech. Nakonec jsem rozšířila své schopnosti pro práci v softwaru Matlab a MS Excel.

Literatura

- [1] Fotr, J., Hnilica, J.: Aplikovaná analýza rizika: ve finančním managementu a investičním rozhodování (2. vydání). Grada, Praha, 2014.
- [2] Crhonková, A.: Bakalářská práce: Porovnávání rizik na základě simulace Monte Carlo, Univerzita Palackého v Olomouci, 2015.
- [3] Vose, D.: Risk analysis: a quantitative guide (3rd edition). John Wiley Sons, West Sussex 2008.
- [4] Hron, K., Kunderová, P.: Základy počtu pravděpodobnosti a metod matematické statistiky(1.vydání). Olomouc, Univerzita Palackého v Olomouci, 2013
- [5] https://www.cnb.cz/cs/financni_trhy/devizovy_trh/kurzy_devizoveho_trhu/prumerne_mena.jsp?mena=EUR [online 17.03.2017] (Kurzy devizového trhu)
- [6] http://gchd.cz/fygyz/2012_2013/david_cerny-bayesovska_statistika.pdf [online 08.04.2017] (Bayesova statistika)