Univerzita Pardubice

Fakulta ekonomicko–správní

Bayesovské odhady v neživotním pojištění

Bc. Bronislava Sehnalová

Diplomová práce

2011

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2010/2011

ZADÁNÍ DIPLOMOVÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Bc. Bronislava SEHNALOVÁ**
Osobní číslo:	**E080073**
Studijní program:	**N6209 Systémové inženýrství a informatika**
Studijní obor:	**Pojistné inženýrství**
Název tématu:	**Bayesovské odhady v neživotním pojištění**
Zadávající katedra:	**Ústav matematiky**

Z á s a d y p r o v y p r a c o v á n í :

Základní pojmy a princip bayesovské teorie odhadů.
Využití bayesovských odhadů v aktuárských vědách.
Bayesovký odhad parametrů binomického, Poissonovho a normálního rozdělení a jejich aplikace v neživotním pojištění.
Význam a využití empirické bayesovské teorie odhadu v pojistné praxi.
Ukázka aplikace karedibilních bayesovských odhadů.

Rozsah grafických prací:	–
Rozsah pracovní zprávy:	**cca 50 stran**
Forma zpracování diplomové práce:	**tištěná/elektronická**

Seznam odborné literatury:

BOLAND, P. J.: Statistical and Probabilistic Methods in Actuarial Science. London: Chapman&Hall/CRC, 2007. ISBN 1-58488-695-1

BOOTH, P., CHADBURN, R., COOPER, D., HABERMAN, S., JAMES, D.: Modern Actuarial Theory and Practice. Chapmann&Hall/CRC, 1999. ISBN 1-4020-2952-7

CIPRA, T.: Pojistná matematika - teorie a praxe. Praha: EKOPRESS, 2006. ISBN 80-86929-11-6

KOTLEBOVÁ, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách. Bratislava: Vydavateľstvo EKONÓM, 2009. 79 s. ISBN 978-80-225-2788-0.

PACÁKOVÁ, V.: Aplikovaná poistná štatistika - 3. prepracované a doplnené vydanie. Bratislava: IURA EDITION, 2004. ISBN 80-8078-004-8

ŠOLTÉS, E.: Modely kredibility na výpočet poistného. Bratislava: Vydavateľstvo EKONÓM, 2009. ISBN 978-80-225-2798-9

TSE Y.-K.: Nonlife Actuarial Models. Cambridge: University Press, 2009. ISBN 978-0-521-76465-0

Vedoucí diplomové práce:	**prof. RNDr. Viera Pacáková, Ph.D.** Ústav matematiky
Datum zadání diplomové práce:	**30. června 2010**
Termín odevzdání diplomové práce:	**6. května 2011**

doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.

doc. Ing. Jiří Křupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 4. srpna 2010

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích dne 20. 4. 2011 Bronislava Sehnalová

Poděkování

Velice ráda bych touto cestou poděkovala vedoucí diplomové práce paní prof. Viere Pacákové za její čas, připomínky a trpělivost při vedení diplomové práce.

Abstrakt

Tato diplomová práce, zvaná Bayesovské metody v neživotním pojištění, se zabývá využitím bayesovských odhadů v oblasti neživotního pojištění. Důraz je kladen na model binomické/beta, model Poissonovo/gama a model normální/normální. Pro každý model je uveden příklad aplikace v pojistné praxi. Práce je dále zaměřena na Bühlmannův model a Bühlmann-Straubův model.

Klíčová slova

bayesovská statistika, bayesovské odhady, teorie kredibility, model binomické/beta, model Poissonovo/gama, model normální/normální, Bühlmannův model kredibility, Bühlmann-Straubův model kredibility

Abstract

This thesis called the Bayesian estimates in non-life insurance is focused on using bayesian estimates in non-life insurance. Emphasis is placed on model binomial/beta, model Poisson/gama and model normal/normal. This thesis contains the exampls of applications for each model in the insurance practice. The thesis is also focused on Bühlmann model and Bühlmann-Straub model.

Keywords

Bayesian statistics, bayesian estimates, credibility theory, model binomial/beta, model Poisson/gama, model normal/normal, Bühlmann model, Bühlmann-Straub model

OBSAH

SEZNAM GRAFŮ

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

SEZNAM ZKRATEK

ISBA	Society for bayesian analysis
ETH	Swiss federal institute of technology
EBCT 1	Bühlmannův model kredibility
EBCT 2	Bühlmann-Straubův model kredibility

Úvod

Pojišťovnictví v průběhu svého vývoje získávalo na své důležitosti a v nynější době je neoddělitelnou součástí národního hospodářství zvláště z toho důvodu, že dnešní doba představuje velké množství rizik, plynoucích z nastání náhodných událostí. Pojišťovnictví tedy plní významnou funkci „ochránce“ ekonomických subjektů, protože kryje ztráty při nastání takovéto události a stabilizuje ekonomickou úroveň.

Vzhledem k tomu, že na našem území dlouhou dobu panoval monopol v oblasti pojišťovnictví, nedocházelo k rozvoji v oblasti teorie a aplikace pojistných věd. Konkurenční prostředí se začalo vytvářet opět od roku 1991, kdy se začaly hledat nové metody výpočtu pojistného, odhadů pojistného plnění atd. Jednou z moderních metod je právě využití teorie kredibility, pro úpravu pojistných sazeb, vycházející z bayesovské statistiky, a výpočet netto pojistného převážně v neživotním pojištění. Přestože v České republice (i na Slovensku) nemají bayesovské metody zatím širší uplatnění, začínají získávat na své důležitosti a stávají se stále častěji užívanými postupy nejen v pojišťovnictví, ale i v jiných odvětvích lidské činnosti. Český pojistný trh se stále více začíná zaměřovat na analýzu vlastních dat z minulosti, jejich zpracování a praktické užití, což je dobrý základ pro častější aplikaci bayesovských metod aktuárských věd.

Diplomová práce je rozdělena do několika dílčích částí. V první části se práce bude zabývat popsáním teoretického základu bayesovských metod. Budou tedy popsány důležité pojmy z teorie pravděpodobnosti vzhledem k bayesovským metodám, jako např. střední hodnota, rozptyl, a Bayesova věta.

Ve druhé části práce bude vysvětlen pojem bayesovská statistika, základní pojmy a princip bayesovské teorie odhadů. Důraz bude kladem především na rozdíly mezi bayesovským a klasickým přístupem. Vzhledem k tomu, že v bayesovská statistice hraje důležitou roli subjektivní stránka, jsou tyto metody hodně kritizovány. V práci budou uvedeny jak důvody pro kritiku, tak i důvody tuto kritiku vyvracející.

Následující část bude zaměřena na vysvětlení pojmu kredibilita a kredibilní bayesovské odhady. Bude popsán princip bayesovkých odhadů parametrů binomického, Poissonovo a normálního rozdělení, přičemž budou uvedeny příklady aplikace v neživotném pojištění.

Důležitou část práce tvoří popis empirické bayesovské teorie a její význam v pojistné praxi. Důraz bude kladen na dva modely, Bühlmannův model a Bühlmann-Straubův model. První z nich není moc vhodný pro praxi, ale tvoří teoretický základ pro model druhý, který se začíná stávat jeden z nejpoužívanějších ve světě. V České republice tomuto modelu není zatím věnována pozornost, jakou by si zasloužil. Je naděje, že v budoucnosti se i u nás stane jedním z často používaných modelů pro odhad kredibilního pojistného.

Cílem diplomové práce je na jedné straně teoreticky popsat, vysvětlit a na praktické ukázce předvést princip využití bayesovských metod v oblasti pojišťovnictví a na straně druhé poskytnout ucelený materiál zabývající se touto problematikou.

1. Teorie pravděpodobnosti v pojištění

Teorie pravděpodobnosti tvoří základ pojistných věd i pojistné praxe, proto je vhodné vysvětlit si její základní pojmy. Teorie pravděpodobnosti prochází neustále rozvojem a tak i definice pravděpodobnosti prošla určitým vývojem. Během vývoje společnosti se začaly objevovat jevy, u kterých se nedal přesně určit výsledek (popř. jevy, které mohou a nemusí nastat). Jako příklad mohou sloužit hazardní hry, které se hrály již ve středověku. Samozřejmě nejen kvůli hazardu začala vznikat teorie pravděpodobnosti. Hlavní důvod byl ten, že bylo potřeba takové případy matematicky interpretovat. Hlavní rozvoj proběhl v 17. století, kdy si dva matematikové Pascal a Fermat vyměňovali dopisy. Řešili otázku, jak spravedlivě rozdělit bank mezi hráče, jestliže série hazardních her musela být předčasně ukončena. Tak tedy vznikal Laplaceova (klasická) definice pravděpodobnosti, která „*odvozuje pravděpodobnost z objektivních reálních vlastností zkoumaného náhodného jevu. Oproti statistické definici umožňuje určit pravděpodobnost náhodné události i bez realizace náhodného pokusu. Předpoklad symetrie možných výsledků je však příliš silný a neumožňuje využití této definice v mnohých reálních situací*“.[1]

Pravděpodobnost by se dala definovat i takto: „*Míru možnosti vzniku náhodné události A, chápanou jako její objektivní vlastnost, bez ohledu na to, zda ji známe nebo ne, nazýváme pravděpodobnost náhodné události A a značíme P(A).*“[2] Existují i další definice pravděpodobnosti – statistická, geometrická, axiomatická. Statistická definice: „*Poskytuje názornou představu pojmu, nevysvětluje však jeho podstatu. Vede k přesvědčení, že bez opakování náhodného pokusu není možné stanovit pravděpodobnost náhodného jevu.*“[3] Geometrická definice je určitým zevšeobecněním klasické definice a je založena na porovnání objemů, délek, ploch geometrických útvarů. Tato definice také vychází z předpokladu, že všechny výsledky náhodného pokusu jsou stejně pravděpodobné. Kolmogorova axiomatická definice pravděpodobnosti byla publikována v roce 1933. Tato definice nedává návod na výpočet pravděpodobnosti. Tyto definice jsou založené na objektivních informacích a nedokážou modelovat informace, které mají subjektivní charakter. Z toho důvodu nemohou být použité při bayesovských metodách.

[1] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 16
[2] Pacáková, V.: Aplikovaná poistná statistika [5], s. 15
[3] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 16

Jak píše V. Pacáková ve své knize Aplikovaná poistná štatistika: „*Teorie pravděpodobnosti je teoretickým základem moderních pojistných věd.*“[4] Je důležité si tedy vyjasnit některé základní pojmy. Prvním z takových pojmů je náhodný pokus. Náhodný pokus je činnost prováděna za stejných podmínek, přičemž nevede k jednoznačnému výsledku. Např. v pojišťovnictví je pojistná smlouva náhodný pokus, protože může a nemusí dojít k pojistné události. S náhodným pokusem souvisí náhodná událost. Náhodná událost může a nemusí nastat a je podmnožinou množiny všech možných výsledků. Kdybychom to opět měli vztáhnout na pojišťovnictví, tak každá pojistná událost je náhodná událost. Další důležitý pojem je náhodná proměnná, která je číselným výsledkem náhodných pokusů. Jako příklad můžeme uvést výšku pojistného plnění.

V práci se bude pracovat s pojmem aritmetický průměr, který se vypočítá podle vztahu $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ [5]. Hodnota x_i je konkrétní realizace náhodného výběru.

Důležité pojmy pro pochopení výpočtů v diplomové práci jsou střední hodnota a rozptyl. Střední hodnota náhodné proměnné je charakteristika úrovní, je zevšeobecněním pojmu aritmetická průměr. V případě diskrétní proměnné je vztah pro střední hodnotu $E(x) = \sum_i p_i x_i$ [6], kde p_i je pravděpodobnost x_i. Rozptyl je základní charakteristikou variability náhodné proměnné X. Vypočítá se podle vzorce $D(X) = EX^2 - (EX)^2$[7].

Další dva důležité pojmy jsou distribuční funkce a hustota pravděpodobnosti. Distribuční funkce je pravidlo, které každému reálnému *x* přiřazuje určitou pravděpodobnost. Je definovaná $F(x) = P(X \leq x)$[8]. Ze vztahu vyplývá, že distribuční funkce vyjadřuje pravděpodobnost, že náhodná veličina X má hodnoty menší nebo rovny hodnotě reálného čísla *x*. Vztah pro výpočet distribuční funkce je $F(x) = P(X \leq x) =$ $= \sum_{x_i \leq x} P(x_i)$ [9]. Vlastnosti distribuční funkce jsou:[10]

- $0 \leq F(x) \leq 1$ pro každé reálné *x*,
- je neklesající, tedy pro $x_1 < x_2$ platí $F(x_1) \leq F(x_2)$,
- je spojitá zprava,

[4] Pacáková, V.: Aplikovaná poistná statistika [5], s. 15
[5] Kubanová, J.: Statistické metody pro ekonomickou a technickou praxi [3], s. 34
[6] Kubanová, J., Linda, B.: Sbírka příkladů z pravděpodobnosti [4], s. 78
[7] Kubanová, J., Linda, B.: Sbírka příkladů z pravděpodobnosti [4], s. 72
[8] Pacáková, V.: Aplikovaná poistná statistika [5], s. 19
[9] Pacáková, V.: Aplikovaná poistná statistika [5], s. 19
[10] Pacáková, V.: Aplikovaná poistná statistika [5], s. 19

- v bodě - ∞ konverguje k nule, v + ∞ k jedné.

Hustota pravděpodobnosti je nezáporná funkce *f(x)* taková, že pro každé reálné číslo x_0 je možné vyjádřit distribuční funkci $F(x_0)$ vztahem $F(x_0) = \int_{-\infty}^{x_0} f(x)dx$ [11], pro $-\infty < x_0 < +\infty$[12]. Hustota pravděpodobnosti *f*(x) má tyto vlastnosti:[13]

- $f(\mathrm{x}) \geq 0$,
- $\int_{-\infty}^{+\infty} f(x)dx = 1$,
- $P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$, pro a < b.

[11]

[12] Pacáková, V.: Aplikovaná poistná statistika [5], s. 20

[13] Pacáková, V.: Aplikovaná poistná statistika [5], s. 20

2. Základní principy bayesovské statistiky

Bayesovská statistika jako vědní disciplína je docela mladá, začala se rozvíjet až na konci 20. století. První publikace začaly vznikat v 80. a 90. letech. Další významná publikace navazuje však až v roce 2004. Tuto knihu napsal W. M. Bolstat, který podrobně popisuje problematiku bayesovské statistiky. Samozřejmě bayesovská statistika se objevuji ve velké části statistické literatury, jsou jí však věnovány jen některé kapitoly. V České republice nebyla publikována samostatná knížka věnovaná pouze bayesovským metodám do roku 1985, kdy prof. M. Hušková z Karlovy univerzity publikovala první skripta[14] na dané téma.

Příznivci bayesovské statistiky jsou sdruženi v celosvětové organizaci International Society for Bayesian Analysis (ISBA). ISBA bylo založeno v roce 1992 pro podporu aplikace Bayesovské statistiky při řešení teoretických a praktických problémů v různých oblastech, např. věda, průmysl, dále také i na úrovni vládních organizací. ISBA má široký záběr své působnosti, pravidelně pořádají konference, vydávají svůj časopis Bayesian Analysis, jehož čísla jsou k dispozici online na internetových stránkách http://ba.stat.cmu.edu/. Pro členství se stačí zaregistrovat na internetových stránkách organizace http://www.bayesian.org/.

Jak již bylo zmíněno, bayesovská statistika se uplatňuje v celé řadě odvětví. Její důležitost lze vidět v ekonometrii, marketingu, důležitá oblast je marketing, společenských vědách a důležitou roli hraje v medicíně, především ve výzkumu. Tomu nasvědčuje i množství literatury na tohle téma z oblasti medicíny.

Bayesovská statistika nám nabízí alternativní přístup k řešení problémů statistické indukce. Je založena na myšlence, že naše informace (zkušenost) o hodnotě neznámého parametru může být vyjádřena pomocí pravděpodobnostního rozdělení, tj. neznámý parametr můžeme považovat za náhodnou proměnnou s určitým rozdělením pravděpodobnosti a ne za neznámou konstantu, jak je tomu v klasické statistice. Parametr tohoto rozdělení a i samotný typ tohoto rozdělení se může měnit vlivem údajů.[15]

---

[14] http://www.karlin.mff.cuni.cz/~huskova/backup_files/huskova_bayes.pdf, 1. 2. 2011[14]

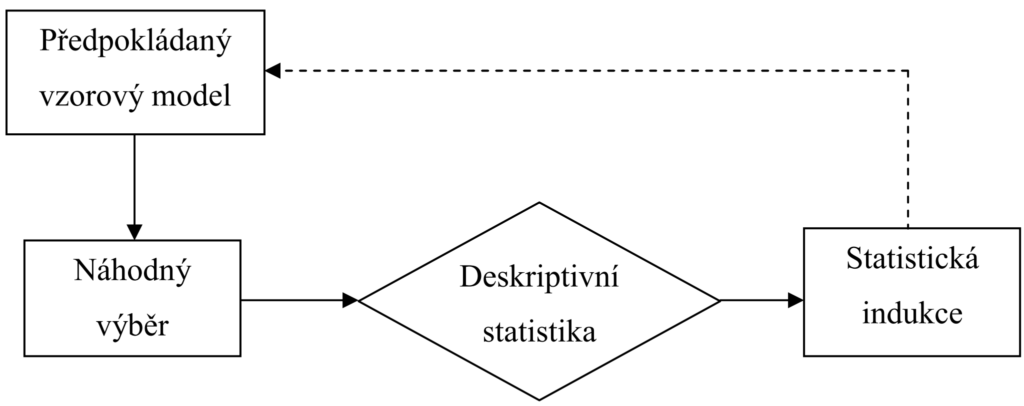
[15] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2],

2.1. Základní rozdíly mezi klasickým a byesovským přístupem

Na obrázcích níže můžeme vidět rozdíl mezi bayesovským a klasickým přístupem.

Na obr. 1 je znárodněný proces, kterým se řídí klasická statistika. Tento proces začíná u modelu, který chceme odvodit, popř. ověřit jeho správnost (např. odhad počtu pojistných plnění). Díky testovacím datům (např. data z minulých let) se dostaneme pomocí různých testů k závěru, který si poté zpětně ověříme.

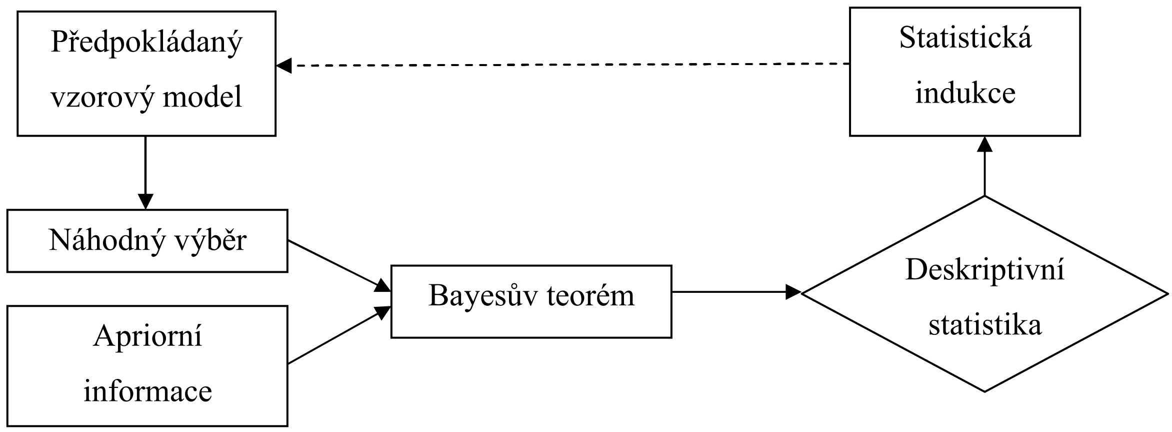
Obrázek 1: Dedukce založená na klasické teorii



[zdroj: Waller, R. A., Martz, H. F. Bayesia realiability analysis [8], s. 166]

Na obr. 2 můžeme vidět rozdíl oproti předchozímu případu. Jako u klasického přístupu začneme u modelu, který chceme odhadnout, popř. ověřit jeho správnost. K tomu však využíváme nejen data z předchozích let, ale i apriorní informace. Bayesův teorém tedy bere v potaz oba zdroje dat. V tomto lze spatřit hlavní rozdíl. Dále je postup stejný jako u klasického přístupu, tedy otestujeme data a zpětně ověříme.

Obrázek 2: Dedukce založená na bayesovské teorii



[zdroj: Waller, R. A., Martz, H. F. Bayesia realiability analysis [8], s. 166]

Jsou zde dva zřetelné rozdíly mezi klasickou teorií a byesovským přístupem. Statistická odhady založené na klasické teorii jsou většinou více omezené než je tomu u odhadů založených na byesovské teorii díky výhradnímu používání dat z náhodného výběru. U byesovského přístupu využíváme informace z předchozích zkušeností, které nazýváme apriorní informace.

Druhý hlavní rozdíl můžeme vidět v tom, že na rozdíl od klasické teorie, bayesovská teorie většinou vyžaduje méně dat z náhodného výběru k dosažení stejné kvality odhadu. Ve většině případů je tohle důvod pro užití bayesovského přístupu a využití apriorních informací, převážně v těch oblastech aplikace, kde data mohou být buď drahá, nebo velice složitá na získání. V tab. 1 je přehledné porovnání vybraných charakteristik klasické teorie odhadu a bayesovské teorie odhadu.

Tabulka 1: Rozdíl v základních charakteristikách u klasické a bayesovské teorie

Chareakteristika	Klasická teorie	Bayesovská teorie
Parametr	Neznámá konstanta	Náhodná proměnná
Apriorní rozdělení	Neexistuje	Existuje, jasně definované
Odhad	Předpokládaný	Předpokládaný
Aposteriorní rozdělení	Neexistuje	Přesně odvozený
Metody uvažování	Induktivní	Deduktivní
Interval odhadů	Jistý interval	Pravděpodobný interval
Význam předcházejících zkušeností	Nevyužívané	Využívané
Účel	Využití výběrových dat pro vyvození závěrů	Potvrdit nebo popřít očekávané výsledky vzhledem k předchozím zkušenostem z minulých období
Kvalita odhadu	Limitovanější kvůli výhradně objektivním informacím	Závisí na schopnosti kvantitativně popsat zkušenosti pomocí objektivních dat
Množství objektivních dat	Čím více, tím lépe.	Bayesovský přístup většinou požaduje méně těchto dat, protože využívá data z minulosti

[zdroj: Waller, R. A., Martz, H. F., Bayesian realiability analysis [8], s. 169]

2.2. Bayesova věta

Je potřeba vysvětlit základní princip bayesovské statistiky a to je Bayesova věta. Bayesova věta o podmíněné pravděpodobnosti souvisí se subjektivní pravděpodobností a je teoretickým základem bayesovské statistiky. Filozofie související s Bayesovou větou byla příčinou vytvoření dvou škol statistické indukce, klasické a bayesovské. Tyto dvě školy se již od začátku vyvíjely odděleně a soupeřily mezi sebou. Hlavní rozdíl a důvod k rozporu je interpretace náhodných jevů.

Bayesovu větu formuloval Thomas Bayes. Datum narození Thomase Bayese se odhaduje na rok 1702. Tento reverend se celý život zabýval studováním statistiky. Paradoxem je, že se proslavil až po své smrti roku 1761. Ve své posmrtně vydané práci „*An essey towards solving a problem in the doctrine of chances*“ popsal, že pravděpodobnost náhodného jevu je ovlivněna nejen podmínkami náhodného pokusu, ale také informacemi, které má konkrétní osoba k dispozici a jejím postojem k těmto informacím.[16] Tato práce byla vydaná v Londýně roku 1763 a publikovaná v „*Philosophical Transactions of the Royal Society*".[17] V diskrétní verzi má Bayesova věta tvar:[18]

$$P(A_i|B) = \frac{P(B|A_i).P(A_i)}{\sum_{j=i}^{n} P(B|A_j).P(A_j)} \qquad (2.1)$$

Jak můžeme vidět, že vzorec vychází z pravidel pro násobení podmíněných a nepodmíněných pravděpodobností výskytu jevů A (hypotéza) a B (předpoklad). Jinými slovy věta nám umožňuje vypočítat pravděpodobnost jednotlivých událostí A_j (j = 1, 2, …n), když víme, že nastala událost B. Pravděpodobnosti P (A_j) pro j = 1, 2, …n, resp. P (A_i), jsou pravděpodobnosti jednotlivých událostí, které vychází z informací před provedením pokusu B – tzv. apriorní pravděpodobnosti. Ty se však zohledněním informace o nastání události B změní na tzv. aposteriorní pravděpodobnosti P $(A_i|B)$, přičemž Bayesova věta kvantifikuje vzájemné vztahy těchto pravděpodobností. Tzn. apriorní pravděpodobnosti se po provedení jednoho nebo více náhodných pokusů mění na aposteriorní pravděpodobnosti. Na její využití je potřebné znát i podmíněné pravděpodobnosti P $(B|A_j)$.[19] Aposteriorní a apriorní pravděpodobnosti budou vysvětleny podrobněji dále.

- *Spojitá verze Bayesovy věty*

V pojišťovnictví se však setkáme spíše se spojitou verzi Bayesovy věty. Využívá se např. při odhadu frekvence pojistných plnění a čistého pojistného. Spojitá verze má tvar:[20]

$$f(\theta/\boldsymbol{x}) = \frac{f(\boldsymbol{x}/\theta)\ .\ f(\theta)}{\int_\theta f(\boldsymbol{x}/\theta)\ .\ f(\theta)\, d\theta} \qquad (2.2)$$

$f(\theta)$ je hustota apriorního rozdělení neznámého parametru θ,

$f(\theta/\boldsymbol{x})$ je hustota aposteriorního rozdělení parametru θ,

16 http://www.karlin.mff.cuni.cz/~huskova/backup_files/huskova_bayes.pdf, 8. 1. 2011[14]
[17] http://www.wyomingbioinformatics.org/~achurban/docs/presentationBayesian.pdf , 1. 2. 2011[15]
[18] Kubanová, J., Linda, B.: Sbírka příkladů z pravděpodobnosti [4], s. 31
[19] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 12
[20] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 15

$f(\boldsymbol{x}/\theta)$ je podmíněná hustota náhodné proměnné X (funkce věrohodnosti),

$\int_\theta f(\boldsymbol{x}/\theta) \,.\, f(\theta)\, d\theta$ je marginální hustota $f(\boldsymbol{x})$, která nezávisí na parametru θ .

Říkáme, že $\mathrm{f}(\theta/\mathrm{x})$ je proporcionální součin u $\mathrm{f}(\mathbf{x}/\theta)\cdot \mathrm{f}(\theta)$. $\mathrm{c} = \frac{1}{\mathrm{f(x)}}$ je konstanta, o kterou se aposteriorní rozdělení $\mathrm{f}(\theta/\mathrm{x})$ liší od součinu. Podmíněná hustota $\mathrm{f}(\mathbf{x}/\theta) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{f}(\mathrm{x_i}/\theta)$ [21] se nazývá funkcí věrohodnosti. Platí zde, že čím je odhad parametru přesnější, tím funkce věrohodnosti nabývá vyšších hodnot. Slovně můžeme vztah aposteriorního rozdělení vyjádřit:[22]

Aposteriorní rozdělení $\propto$ *funkce věrohodnost . apriorní rozdělení*

Tedy při vynásobení apriorního rozdělení věrohodnostní funkcí získáme aposteriorní rozdělení. Můžeme je přesně vyjádřit vztahem: $\frac{Apriorní\ rozdělení\ x\ \ Funkce\ věrohodnosti}{Marginální\ distribuční\ funkce}$ [23]. Přesná definice aposteriorního rozdělení zní: „*Aposteriorní rozdělení $f(\theta/\boldsymbol{x})$ je rozdělení neznámého (odhadovaného) parametru θ, které využívá kromě znalostí apriorního rozdělení $f(\theta)$ i výsledek vlastního výběrového zjišťování $x = (x_1,\ x_2, \dots\ x_{n,})$.*“[24]

Apriorní rozdělení získáme z informací, které jsou o parametru známé ještě před výběrovým zjišťováním. Tyto informace mohou být jak objektivní, tak i subjektivní. Objektivní informace jsou např. informace z podobných úloh, problému z minulosti, kdežto subjektivní vyjadřují názor či zkušenost. Jako zdroj těchto informací mohou být považovány např. zkušenosti jiných pojišťoven o počtu nebo výši pojistných plnění při určitém typu pojištění. Důležité je zmínit, že hlavní je kvalita těchto informací, protože zásadním způsobem ovlivňují aposteriorní rozdělení a tím pádem i závěr, který vyvodíme (na základě apriorní informace dvou různých subjektů můžeme dojít k odlišným závěrům). Můžeme se však také dostat do situace, kdy o parametru žádné apriorní informace neznáme a známe jen množinu možných hodnot. Taková informace se nazývá vágní nebo difusní. V takovémto případě předpokládáme, že každá možná hodnota odhadovaného parametru je stejně pravděpodobná.

[21] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 15
[22] Pacáková, V.: Aplikovaná poistná statistika [5], s. 182
[23] Waller, R. A.; Martz, H. F.: Bayesian reiability analysis. [8], s. 176
[24] Pacáková, V.: Aplikovaná poistná statistika [5], s. 181

Z toho vyplývá, že „*Bayesovský postup odhadu parametru* θ *tedy nutně vyžaduje znalost apriorního rozdělení* $f(\theta)$ *před výběrovým zjišťováním – přesného nebo přibližného, v souladu s dostupnými informacemi*“[25]. Zde však narážíme na kritiku byesovských odhadů. Důvodem této kritiky je přílišná subjektivita, která se projevuje v míře přesvědčení subjektu, že určitý jev nastane. Může to tedy být pocit jistoty, skepse, úspěchu akce atd. V případě, že má apriorní informace subjektivní charakter, pak je i model založený na akceptaci subjektivní pravděpodobnosti. Samozřejmě to je v naprosté opozici k objektivisticky smýšlejícím statistikům. Subjektivní pravděpodobnosti totiž nejsou nikde definovány, říkáme jen, že určitý výsledek je nejpravděpodobnější, méně pravděpodobný či málo pravděpodobný. Dochází zde však k určitým snahám o nalezení způsobu o objektivizaci míry přesvědčení subjektu. Např. se mohou využívat stupnice logické pravděpodobnosti. Zde bychom mohli uvést citaci z publikace Texty k Bayesovské statistice od P. Hábka[26]: „*Při analýze dat je nejdůležitější (což řada statistiků opomíjí), že „mnohem lepší je přibližná odpověď na správnou otázku, která je často vágní, než přesná odpověď na špatnou otázku.*“

[25] Pacáková, V.: Aplikovaná poistná statistika [5], s. 181
[26] Hábek, P.: Texty k Bayesovské statistice [1], s. 91

3. Teorie kredibility

Pojištění je postaveno na myšlence tarifování. To znamená, že jedinci, kteří jsou vystaveni stejnému pojistnému riziku, se začleňují do tříd. Cílem tarifování je poznat tohle riziko. Existuje předpoklad, že intenzitu tohoto rizika můžeme přesně stanovit ze statistických údajů o pojistném riziku příslušné třídy. Nemůže však vytvořit homogenní třídy, protože stupeň rizika je ovlivněn i náhodnými vlivy a nepředvídatelnými událostmi. Homogenizace však můžeme dosáhnout pomocí výběru vhodných faktorů rizika. Musíme dát pozor, protože při segmentaci portfolia nemůžeme uvažovat faktory, které by mohly být pokládány za diskriminační (např. národnost), nejsou dostatečně objektivní, jsou těžce kontrolovatelné, atd. Teorie kredibility je zaměřená na co nejrealističtější rozdělení portfolia na tarifní třídy s co nejnižším stupněm heterogenity.

V aktuárských vědách slovo „kredibilita" vyjadřuje „*míru věrohodnosti, kterou aktuár prokazuje zkušenostmi s příslušnou třídou pojištění s cílem stanovit pojistné pro tuto třídu pojištění*".[27] Můžeme si to uvést na příkladu, kdy je škodový průběh pro uvažovanou třídu pojištění málo kredibilní. Pro nás to znamená, že budoucí škodový průběh může být odlišný od toho, který jsme do této doby pozorovali.

Teorie kredibility, která vznikla v Severní Americe na začátku století, poskytuje moderní přístup kalkulace netto pojistného, zvláště v neživotním pojištění. Využívá se u krátkodobých pojistných smluv při výpočtu a úpravě permanentní výše pojistného. Jak uvádí V. Pacáková ve své knize Aplikovaná poistná statistika[28], využívají se dva zdroje údajů na určení výše pojistného v následujícím účetním období:

- údaje z vlastního portfolia pojistek,
- údaje z jiných porovnatelných zdrojů.

V letech těsně před 1. světovou válkou byla industrializace Spojených státu a Kanady ovlivněna mnohými změnami. Jedna z těchto změn zahrnovala nové zaopatření rodinných jistot. Např. první druh životního pojištění pro zaměstnance byl aplikován ve Spojených státech v roce 1911. Ve stejném čase byl ustanoven zákon o odškodnění pracovníků (úrazové a nemocenské pojištění). V podstatě tím bylo vytvořeno nové zákonné pojištění, které s sebou však přineslo problém, jakým způsobem stanovit sazby pojistného. Právě

[27] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 10
[28] Pacáková, V.: Aplikovaná poistná statistika [5]

tohle úsilí najít co nejpřesnější sazby pojistného dalo základ pro teorii kredibility. Poté se teorie krediblity začala uplatňovat i u jiných typů pojištění, především u pojištění vozidel.

Na rozvoji této teorie se významně podíleli Albert H. Mowbray a Albert W. Whitney. První jmenovaný svou prací publikovanou v roce 1914 a Whitney prací publikovanou v roce 1918 formuloval „*problém stanovení pojistného P, definovaného jako očekávané pojistné náklady připadající na jedno riziko, pro individuální riziko vybrané z portfolia podobných rizik*“.[29] Odvodil vztah kredibilní pojistné P_k na základě kombinovaného použití zkušeností s individuálním rizikem a zkušenostní z portfolia tak, aby bylo lineární kombinací individuálního P_r a kolektivního P_c pojistného:[30]

$$P_k = Z\,P_r + (1 - Z).P_C \qquad (3.1)$$

Jako individuální riziko zde chápeme riziko, kterému je vystavená uvažovaná třída pojištění. Celé portfolio je tvořeno jednotlivými individuálními riziky. Můžeme vyvodit, že každé individuální riziko je součástí kolektivního rizika. Z logické úvahy vyplývá, že k ohodnocení individuálního rizika by se měli využít oba zdroje informací. Krediblní pojistní se tedy skládá z individuálního i kolektivního pojistného. Individuální pojistné je pozorovaná výše pojistných plnění připadající na jednu pojistnou smlouvu a kolektivní pojistné je střední hodnota za celé portfolio.

Kredibilní pojistné je chápáno jako kompromis mezi individuálním pojistným (vypočítaným výhradně z údajů o vlastním riziku) a kolektivním pojistným (z údajů o podobných rizicích). Je to vážený aritmetický průměr individuálního a kolektivního pojistného. Faktor krediblity Z je vyjádřením relativní váhy individuálního pojistného a (1 – Z) je vyjádřením relativní váhy kolektivního pojistného.[31]

Základní vlastnosti kredibilního pojistného:[32]

- je nelineární kombinací dvou odhadů pojistného, založených na údajích o vlastním riziku a na údajích o jiných, podobných rizikách,
- faktor kredibility Z je mírou spolehnutí se na údaje o vlastním riziku, tedy jeho hodnota je tím vyšší, čím víc se pojišťovatel spoléhá na vlastní údaje,

[29] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 10
[30] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 10
[31] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 11
[32] Pacáková, V.: Aplikovaná poistná statistika [5], s. 172

- faktor (1 – Z) určuje míru, jakou se pojišťovatel spoléhá na údaje o cizích, porovnatelných rizicích,
- faktor kredibility Z závisí od počtu údajů o vlastním riziku a s jejich rostoucím počtem roste i jeho hodnota.

4. Bayesovská teorie kredibility

Při vzniku Bayesovské teorie kredibility hráli klíčovou úlohu Arthur L. Bailey a A. Mayerson. Díky prvnímu jmenovanému se bayesovský přístup k určení kredibilního pojistného dostal do podvědomí aktuárů. A. Mayerson poté tento přístup k určení kredibilního pojistného uvedl do souvislosti s moderní bayesovskou statistikou. Arthur L. Bailey vystudoval pojistnou matematiku a statistiku na Michiganské univerzitě. Zabýval se bayesovskou statistikou ještě dříve, než se dostal k aktuárským vědám. Jeho matematice bylo velice složité porozumět a byl za to velmi kritizován. To také byl jeden z důvodů, proč nebyla jeho práce všeobecně uznána a byla dlouhý čas zapomenuta. To trvalo až do doby, kdy se o ni začal zajímat současný švýcarský aktuár Hans Bühlmann (viz. kapitola 5).

Předchůdce bayesovské teorie krediblility se označuje jako americká teorie krediblity nebo přístup ohraničeného kolísání. Jak již bylo zmíněno, v roce 1914 Mowbray vydává publikaci, ve které navrhuje postup určení počtu údajů o individuálním riziku, aby byly úplně kredibilní. Neboli jinými slovy, abychom spoléhali pouze na údaje o individuálním riziku a zanedbali údaje o kolektivním riziku. Za podmínky úplné kredibility byl faktor kredibility nahrazen hodnotou 1. S tím vznikla tzv. parciální kredibilita, protože ne vždy máme k dispozici tolik informací o vlastním riziku, aby byly úplně věrohodné při ohodnocení tohoto rizika. Tedy jinými slovy potřebujeme určit, kolik pozorování o vlastním riziku je nutné znát, abychom mohli položit Z = 1. V řeči statistiků řekneme, že *„údaje pro stanovení věrohodného pojistného se považují za úplně kredibilní, resp. úplně (k, p) kredibilní, když platí vztah:* $P(|\hat{\beta} - \beta| \leq k\beta) \geq p.$*“*[33] β označuje skutečný, ale neznámý počet pojistných událostí, aby platilo Z = 1. $\hat{\beta}$ je odhadem β. Poté tedy můžeme říci, že vlastní riziko je úplně kredibilní, pokud odchylka $\hat{\beta}$ od neznámého počtu β je 100 . k % z počtu β alespoň s pravděpodobností p.[34] V Severní Americe vznikla řada prací věnující se této problematice, v žádná z nich však nebyly zahrnuty všechny speciální případy a neposkytla jednotný princip.

Po 2. světové válce se opět odborná veřejnost začala zabývat teorií kredibility. Americká teorie nedávala návod na využití údajů o cizích rizicích při určování sazeb čistého pojistného, proto byla rozvinuta o různé modely a metody, které tento nedostatek

[33] Pacáková, V.: Aplikovaná poistná statistika [5], s. 172
[34] Pacáková, V.: Aplikovaná poistná statistika [5], s. 172

odstraňovaly. Vytvořila se tzv. bayesovská teorie, někdy nazývána i jako teorie největší přesnosti, jejíž základ tvoří bayesovská teorie odhadu.

V bayesovské teorii kredibility se setkáme s odlišným přístupem v porovnání s klasickými metodami odhadu, využívajícími převážně výběrové rozdělení a metodu maximální věrohodnosti. Jak již bylo uvedeno hlavní rozdíl, kterým se odlišuje bayesovský přístup, spočívá v tom, že neznámý parametr θ považujeme za náhodnou proměnnou a ne za neznámou konstantu, jak je tomu u klasických metod odhadu.[35] Dále můžeme vidět rozdíl i v tom, že u bayesovského odhadu je nutné znát nejen výběrové rozdělení, ale i apriorní rozdělení. O apriorním rozdělení byla řeč v kapitole 2.2. Apriorní rozdělení chápeme jako rozdělení pravděpodobnosti výše pojistného plnění připadajícího na jednu pojistnou smlouvu portfolia.[36]

Nechť $\boldsymbol{X}$ je náhodná proměnná, jejíž rozdělení můžeme vyjádřit pravděpodobnostní funkcí $f(x;\theta)$ s neznámým parametrem θ. Parametr θ je náhodná proměnná, která má apriorní rozdělení pravděpodobnosti $f(\theta)$. Jak již bylo řečeno, apriorní rozdělení poskytuje první, základní informaci o odhadovaném parametru. Jako nejlepší odhad θ před výběrovým zjišťováním je $E(\theta)$, kdežto rozptyl $D(\theta)$ je stupeň nejistoty prvního odhadu $\hat{\theta} = E(\theta)$ na základě apriorního rozdělení. θ_B nazýváme bayesovský odhad parametru θ, pokud odhad θ provádíme pomocí náhodného výběru $\boldsymbol{x}$ = $(x_1, x_2\ \ ..., x_n)$ vzhledem na aposteriorní rozdělení $f\ (\theta\ /\ x)$. Jako bayesovský odhad neznámého parametru θ se v praxi nejčastěji užívá střední hodnota, medián nebo modus aposteriorního rozdělení $f\ (\theta\ /\ \boldsymbol{x})$.[37]

4.1. Konjugované systémy

O konjugovaném rozdělení k rozdělení, ze kterého pochází náhodný výběr, mluvíme v takových případech, kdy jsou apriorní a aposteriorní rozdělení stejného typu (za předpokladu, že náhodný výběr pochází z některého ze známých pravděpodobnostních rozdělení). Konjugované rozdělení definujeme takto: „*Nechť náhodný výběr pochází z pravděpodobnostního rozdělení typu R s neznámým parametrem* θ. *Říkáme, že typ*

[35] Pacáková, V.: Aplikovaná poistná statistika [5], s. 179
[36] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 12
[37] Pacáková, V.: Aplikovaná poistná statistika [5], s. 189

rozdělení F je konjugovaný pro typ rozdělení R, když apriorní rozdělení parametru θ typu F vede k aposteriornímu rozdělení stejného typu F.“[38]

Výhody konjugovaného systému hustot jsou především ty, že je to velmi efektivní prostředek na jednoduché určení aposteriorního rozdělení. Není třeba počítat složité integrály a umožňuje určit parametr aposteriorního rozdělení dosazením parametrů apriorního rozdělení a výběrových charakteristik do odvozených vzorců. Problém může představovat splnění předpokladů o výběrovém a apriorním rozdělení. Další výhoda je v jednoznačných a dobře interpretovatelných výsledcích, avšak kvalita výsledu je odvozena od volby apriorního rozdělení.

Ke konstrukci konjugovaných systémů využíváme tzv. postačující statistiku. V této výběrové statistice je obsažená celá informace z výběru potřebná k určení aposteriorního rozdělení. Informace obsažená v aposteriorním rozdělení a vztahující se k odhadovanému parametru θ závisí od údajů z výběru jen prostřednictvím postačující statistiky. Definice zní: „*Statistika T (X_1, X_2, …, X_n) je postačující statistika pro parametr θ právě tehdy, když aposteriorní rozdělení $f(\theta / X_1, X_2, \dots, X_n)$ je stejné jako aposteriorní rozdělení $f(\theta / T)$.*“[39]

Konjugovaný systém hustot je možné vytvořit takto:[40]

1. Funkci věrohodnosti upravíme tak, že části obsahující odhadovaný parametr vyjádříme jako funkci postačující statistiky a ostatní výrazy zanedbáme.
2. Za parametry apriorního rozdělení považujeme postačující statistiky.
3. Jakmile vytvořené parametry dosadíme do funkce věrohodnosti a vzniklý výraz vynásobíme vhodnou konstantou tak, aby určitý integrál měl hodnotu rovnou jedné, dostaneme funkci hustoty apriorního rozdělení.
4. Vyjádření hustoty aposteriorního rozdělení dostaneme, jakmile dosadíme funkci hustoty vytvořeného apriorního rozdělení a funkce věrohodnosti do Bayesovy věty.

Nejčastěji používané konjugované systémy hustot jsou binomické/beta, Poissonovo/gama a normální/normální. V názvu stojí na prvním místě typ rozdělení, z kterého pochází náhodný výběr a kterého parametru odhadujeme, a na druhém typ

[38] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 17
[39] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 18
[40] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 18

apriorního rozdělení, resp. aposteriorního. Nevýhodou konjugovaných systémů je to, že pokud bychom chtěli využít některý z konjugovaných systémů, je nutné, aby měl sledovaný znak takové rozdělení, ke kterému existuje konjugovaný systém. Také dostupná apriorní informace by měla být modelovatelná příslušným konjugovaným rozdělením. V tomto právě můžeme vidět hlavní nevýhodu při širším využití bayesovských metod. Výhodou konjugovaných systémů je dobrá interpretovatelnou a názornost.

- *Model binomické/beta*

Ukážeme, že beta rozdělení je konjugovaným rozdělením pro bayesovský odhad parametru θ binomického rozdělení. Výběr pochází z binomického rozdělení $X/\theta \approx Bi\ (n;\ \theta)$ s neznámým parametrem θ. Pravděpodobnostní funkce má tvar:[41]

$$f(x/\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad (4.1)$$

Apriorní rozdělení parametru θ je rozdělení beta $Be(\alpha;\ \beta)$, tedy $\theta \sim Be(\alpha;\ \beta)$. Z výběrového zjišťování známe hodnotu $X = x$, což je počet pokusů, při kterých nastal sledovaný jev. Poté aposteriorní rozdělení neznámého parametru θ je opět rozdělení beta s parametry $\alpha + x, \beta + n - x$, tedy:[42]

$$\theta/x \sim Be\ (\alpha + x;\ \beta + n - x) \qquad (4.2)$$

kde *n* je rozsah výběrového souboru. Beta rozdělení $Be(\alpha;\ \beta)$ je vhodným typem pro modelování rozdělení pravděpodobnosti náhodné proměnné, která nabývá hodnot z intervalu $\langle 0; 1 \rangle$, z toho důvodu, že je dost flexibilní. Tento konjugovaný systém má velký význam a je dobře využitelný v praxi neživotního pojištění, hlavně v neživotním pojištění na úmrtí.

- *Model Poissonovo/gama*

Dalším typem konjugovaných systému je model Poissonovo/gama. Ukážeme, že rozdělení gama je konjugované apriorní rozdělení pro náhodnou proměnnou s Poissonovým rozdělením. Výběr pochází z Poissonova rozdělení $X/\lambda \approx Po\ (\lambda)$

[41] Pacáková, V.: Aplikovaná poistná statistika [5], s. 186
[42] Pacáková, V.: Aplikovaná poistná statistika [5], s. 186

s neznámým parametrem λ. *X* je náhodná proměnná znamenající počet náhodných událostí (např. počet pojistných plnění). Když je apriorní rozdělení typu gama s parametry α, β, tj. $\lambda \sim G(\alpha;\ \beta)$, tak aposteriorní rozdělení $f(\lambda/\boldsymbol{x})$ je též typu gama rozdělení s parametry $\alpha_1 = \alpha + \sum_{i=1}^{n} x_i$ a $\beta_1 = \beta + n$, takže platí:[43]

$$\lambda/\boldsymbol{x} \sim G(\alpha + \sum x_i\ ;\ \beta + n) \tag{4.3}$$

Gama rozdělení $G(\alpha;\ \beta)$ je velmi často používaným apriorním rozdělením. Důvod je takový, že vzhledem na svoji flexibilitu umožňuje modelování rozdělení proměnných nabývající kladné hodnoty.

- *Model normální/normální*

Normální rozdělení má významné místo ve statistické indukci. Normální rozdělení má, nebo alespoň přibližně, mnoho statistických znaků. Dokazuje to centrální limitní věta, podle které má výběrový průměr pro velké *n* přibližně normální rozdělení. To nám tedy umožňuje pracovat s tímto rozdělením i v takovém případě.

Normální rozdělení je konjugovaným rozdělením pro bayesovský odhad střední hodnoty θ normálního rozdělení. Výběr pochází z normálního rozdělení $X/\theta \approx N\ (\theta;\ \sigma_1^2)$ se střední hodnotou θ, od které závisí rozdělení náhodné proměnné *X*. Apriorní rozdělení parametru θ je také normální rozdělení se střední hodnotou μ a rozptylem σ_2^2, tedy $\theta \sim N\ (\mu;\ \sigma_2^2)$. Poté je i aposteriorní rozdělení parametru θ normální rozdělení $\theta/\boldsymbol{x} \sim N\ (\tilde{\mu};\ \widetilde{\sigma^2})$. Pro parametry platí:[44]

$$\tilde{\mu} = \frac{\mu\sigma_1^2 + n\bar{x}\sigma_2^2}{\sigma_1^2 + n\sigma_2^2} \tag{4.4}$$

$$\widetilde{\sigma^2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + n\sigma_2^2} \tag{4.5}$$

[43] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 21
[44] Pacáková, V.: Aplikovaná poistná statistika [5], s. 188

4.2. Bayesovské odhady

Bayesovský odhad neznámého parametru θ, který se označuje θ_B, vychází z aposteriorního rozdělení. Apriorní rozdělení zohledňuje výběrové údaje i apriorní informaci. Pro toto rozdělení se vybírá určitá charakteristika, tedy funkce náhodného výběru g(x), s požadovanými vlastnostmi. Často se posuzuje podle střední kvadratické chyby. Nejčastějším odhadem parametru θ je střední hodnota aposteriorního rozdělení.[45]

Nyní si vysvětlíme bayesovské odhady neznámých parametrů, jejichž aposteriorní rozdělení bylo vysvětleno v předchozí kapitole.

V případě binomické/beta je Bayesovský odhad θ_B parametru θ binomického rozdělení $Bi(\theta;\, n)$ střední hodnota aposteriorního beta rozdělení $Be(\alpha + x;\, \beta + n - x)$, která podle V. Pacákové má tvar:[46]

$$\theta_B = \frac{\alpha + x}{\alpha + \beta + n} \tag{4.6}$$

V případě modelu Possinovo/gama je bayesovským odhadem λ_B parametru λ Poissonova rozdělení $Po(\lambda)$ střední hodnota aposteriorního gama rozdělení $G(\alpha + \sum_{i=1}^{n} x_i\,;\, \beta + n)$. Podle vzorce pro střední hodnotu gama rozdělení, dostaneme:[47]

$$\lambda_B = \frac{\alpha + \sum_{i=1}^{n} x_i}{\beta + n} = \frac{\alpha + n\bar{x}}{\beta + n} \tag{4.7}$$

U modelu normální/normální je bayesovským odhadem θ_B střední hodnoty θ normálního rozdělení $N(\theta;\, \sigma_1^2)$ střední hodnota aposteriorního normálního rozdělení, kterou můžeme vyjádřit vztahem $\tilde{\mu} = \frac{\mu\sigma_1^2 + n\bar{x}\sigma_2^2}{\sigma_1^2 + n\sigma_2^2}$, tedy:[48]

$$\mathrm{\theta_B} = \frac{\mathrm{\mu\sigma_1^2 + n\bar{x}\sigma_2^2}}{\mathrm{\sigma_1^2 + n\sigma_2^2}} \tag{4.8}$$

[45] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 34
[46] Pacáková, V.: Aplikovaná poistná statistika [5], s. 191
[47] Pacáková, V.: Aplikovaná poistná statistika [5], s. 192
[48] Pacáková, V.: Aplikovaná poistná statistika [5], s. 192

Střední hodnota je považována za náhodnou proměnnou. V porovnání s klasickým bodovým odhadem $\bar{x}$ má bayesovský odhad větší vychýlení, ale menší rozptyl.

4.3. Bayesovké a maximálně věrohodné odhady

V této kapitole si ukážeme výhody bayesovských odhadů parametru v porovnání s maximálně věrohodnými odhady.

Velmi častým odhadovaným parametrem je podíl statistických jednotek v základním soboru, které jsou nositeli určité informace. Jednoduše bychom to mohli popsat jako zjištění podílu statistických jednotek z náhodně vybraného vzorku s určitou vlastností.

- *Model binomické/beta*

V případě modelu binomické/beta můžeme upravit vztah pro odhad parametru θ rozdělení $Bi\ (\theta; n)$:[49]

$$\theta_B = \frac{n}{\alpha + \beta + n} \cdot \frac{x}{n} + \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \frac{\alpha}{\alpha + \beta} \qquad (4.9)$$

Parametr *n* značí počet opakování náhodného pokusu, tzn. počet náhodně vybraných statistických jednotek ve výběrovém souboru. Vyjádření $\frac{x}{n} = \bar{x}$ znační maximálně věrohodný odhad parametru θ rozdělení $Bi\ (\theta; n)$, kdežto $\frac{\alpha}{\alpha+\beta}$ je vyjádřením střední hodnoty apriorního beta rozdělení $Be\ (\alpha; \beta)$ parametru θ.[50] Bayesovský odhad θ_B je tedy váženým aritmetickým průměrem maximálně věrohodného odhadu a střední hodnoty apriorního rozdělení. Střední hodnota aposteriorního rozdělení je váženým průměrem střední hodnoty apriorního rozdělení a výběrového průměru, přičemž čím je větší rozsah výběrového souboru, tím větší váhu má maximálně věrohodný odhad parametru θ.

- *Model Poissonovo/gama*

V tomto případě se bayesovský odhad λ_B vyjádří ve tvaru:[51]

[49] Pacáková, V.: Aplikovaná poistná statistika [5], s. 192
[50] Pacáková, V.: Aplikovaná poistná statistika [5], s. 192
[51] Pacáková, V.: Aplikovaná poistná statistika [5], s. 193

$$\lambda_B = \frac{\alpha + n\bar{x}}{\beta + n} = \frac{n}{\beta + n} \cdot \bar{x} + \frac{\beta}{\beta + n} \cdot \frac{\alpha}{\beta} \tag{4.10}$$

Bayesovský odhad λ_B je vyjádřen pomocí váženého aritmetického průměru maximálně věrohodného odhadu $\hat{\lambda} = \bar{x}$ a střední hodnota apriorního rozdělení $\lambda_A = \frac{\alpha}{\beta}$. I zde platí, že čím větší je rozsah *n* výběrového souboru, tím více závisí bayesovský odhad na výběrových údajích ve srovnání s apriorním rozdělením.

- *Model normální/normální*

Vztah pro bayesovský odhad parametru θ normálního rozdělení můžeme zapsat takto:[52]

$$\theta_B = \frac{n\sigma_2^2}{\sigma_1^2 + n\sigma_2^2}\bar{x} + \frac{\sigma_1^2}{\sigma_1^2 + n\sigma_2^2}\mu = \frac{\frac{n}{\sigma_1^2}}{\frac{1}{\hat{\sigma}^2}}\bar{x} + \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\hat{\sigma}^2}}\mu \tag{4.11}$$

Bayesovský odhad θ_B parametru θ normálního rozdělení je váženým průměrem střední hodnoty μ apriorního rozdělení $N(\mu; \sigma_2^2)$ a výběrového průměru $\hat{\theta} = \bar{x}$. Jako váhy jsou použité převrácené hodnoty rozptylů apriorního rozdělení a výběrového průměru. Váhy ovlivňují výsledek tímto způsobem:[53]

- Čím je variabilita apriorního rozdělení vyšší, tím je menší hodnota $\frac{1}{\sigma_2^2}$, tedy střední hodnota apriorního rozdělení má nižší váhu.
- Čím vyšší je rozptyl σ_1^2 rozdělení, tím má výběrový průměr menší vliv.
- Rozsah výběrového souboru zvyšuje váhu výběrového průměru.

4.4. Bayesovské modely v pojišťovnictví

Kredibilní bayesovské odhady se dají výhodně aplikovat do pojistné praxe. Je potřeba mít dostatek dat a pak již můžeme snadno odhadnout potřebné parametry.

[52] Pacáková, V.: Aplikovaná poistná statistika [5], s. 193
[53] Kotlebová, E.: Bayesovská štatistická indukcia v ekonomických aplikáciách [2], s. 50

- *Model binomické/beta*

U modelu binomické/beta chceme najít bayesovský odhad θ_B pravděpodobnosti θ. Víme, že počet pojistných událostí má binomické rozdělení $Bi(n;\ \theta)$. To značí, že při každé pojistce je konstantní pravděpodobnost θ pojistné události. Při určování odhadu známe, že při *n* pojistkách bylo právě *x* pojistných plnění, resp. pojistná událost nastala právě x-krát.

Odhad θ_B ve tvaru kredibilního bayesovského odhadu vyjadřuje vztah:[54]

$$\theta_B = Z \cdot \frac{x}{n} + (1 - Z) \cdot \mu \qquad (4.12)$$

ve kterém pro faktor kredibility Z platí:[55]

$$Z = \frac{n}{\alpha + \beta + n} \qquad (4.13)$$

Pokud by se pojišťovna měla spolehnout jen na údaje ze svého portfolia, pro Z platí, že se blíží jedné. V tomto případě dostaneme $\theta_B = \frac{x}{n}$, což je vyjádření výběrového podílu počtu pojistek, u kterých nastala škoda. Pokud $Z = 0$, tak se pojišťovna spoléhá pouze na informace od cizích pojišťoven, protože nemá žádné informace z vlastního portfolia. Tento případ se vyskytne u nově začínajících pojišťoven.

- *Model Poissonovo/gama*

Tento model můžeme využít k odhadu počtu pojistných plnění v následujícím roce. Předpokládáme, že počet pojistných plnění má v průběhu roku Poissonovo rozdělení s neznámým parametrem λ. Ten je považován za náhodnou proměnnou. Apriorní rozdělení tohoto parametru je gama rozdělení s parametry α, β. Při odhadu průměrného počtu pojistných plnění za rok, tedy parametru λ, využíváme apriorní informace o rozdělení parametru λ a informace o počtu pojistných plnění pojišťovny v předchozích letech. Počet let značíme *n*, jednotlivé údaje o počtu pojistných plnění $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

Podle kapitoly 4. 2. můžeme bayesovský odhad λ_B formulovat do tvaru:[56]

[54] Pacáková, V.: Aplikovaná poistná statistika [5], s. 194
[55] Pacáková, V.: Aplikovaná poistná statistika [5], s. 194

$$\lambda_B = E(\lambda / \mathbf{x}) = Z\,\bar{x} + (1 - Z)\mu \qquad (4.14)$$

Tento vztah nám vyjadřuje kredibilní bayesovský odhad průměrného ročního počtu pojistných plnění. Z je faktor kredibility a v tomto případě je definován jako:[57]

$$Z = \frac{n}{\beta + n} \qquad (4.15)$$

Hodnota Z je podmíněna hodnotou *n* a β, závisí tedy údaji z vlastního portfolia a na apriorním rozdělení. Pokud je tedy $n = 0$, pak i $Z = 0$. Na druhou stranu, čím větší je *n*, tím víc se hodnota Z blíží 1. Dále platí, že čím je vyšší β, tím menší je variabilita apriorního rozdělení a tím je i nižší hodnota faktoru Z.

- *Model normální/normální*

V tomto modelu odhadujeme hodnotu čistého (netto) pojistného, nebo-li střední hodnotu celkového pojistného plnění. *X* značíme celkové roční plnění pojišťovny a její hustota závisí od neznámé hodnoty parametru θ. $f(X\,/\,\theta)$ vyjadřuje podmíněné rozdělení celkových ročních plnění pojišťovny a má normální rozdělení $N(\theta;\, \sigma_1^2)$. Neznámý parametr θ je střední hodnotou rozdělení a rozptyl σ_1^2 je známá konstanta. Víme, že parametr θ je považován za náhodnou proměnnou s normálním apriorním rozdělením $N(\mu;\, \sigma_2^2)$, na druhou stranu μ, σ_1^2 a σ_2^2 považujeme za známé konstanty. Stejně jako v předešlém případě značíme počet let *n* a hodnoty ročních pojistných plnění $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

Podle informací z předchozích odstavců můžeme tedy říci, že bayesovský odhad čistého pojistného je střední hodnota aposteriorního rozdělení, které upravíme do tvaru:[58]

$$E(\theta/\boldsymbol{x}) = Z\,\bar{x} + (1 - Z)\mu \qquad (4.16)$$

Dostali jsme formulaci čistého kredibilního pojistného. Pro faktor kredibility Z v období (n +1) platí:[59]

[56] Pacáková, V.: Aplikovaná poistná statistika [5], s. 196
[57] Pacáková, V.: Aplikovaná poistná statistika [5], s. 196
[58] Pacáková, V.: Aplikovaná poistná statistika [5], s. 198

$$Z = \frac{n\sigma_2^2}{\sigma_1^2 + n\sigma_2^2} = \frac{n}{n + \frac{\sigma_1^2}{\sigma_2^2}} \quad (4.17)$$

4.5. Ukázky aplikace byesovských modelů

- *Model binomické/beta*

Uvažujeme zdravotní pojišťovnu, u níž chceme porovnat maximálně věrohodný a bayesovský odhad pravděpodobnosti nastání pojistné události. Postupně aktualizujeme neznámý parametr θ, což je pravděpodobnost pojistné události v průběhu roku. Pojistnou událostí je v tomto případě potřeba zdravotní péče a budeme ji značit *x*. Apriorní rozdělení parametru θ je rovnoměrné rozdělení na intervalu <0; 1>. Rozsah výběrového souboru *n* = 100 pojištění. Zjistili jsme, že v prvním roce potřebuje zdravotní péči 58 osob, v druhém 63, další údaje o počtu pojistných událostí jsou uvedeny v tab. 2.

Tabulka 2: Počet pojistných událostí

Roky	Počet pojistných událostí
1	58
2	63
3	49
4	60
5	68
6	57
7	72
8	64
9	59
10	70

[zdroj: vlastní zpracování]

V prvním roce jsou parametry α = 1 a β = 1, což je speciální případ beta rozdělení. Pro první rok platí podle (4.6) $\theta_B = \frac{1}{1+1} = 0{,}5$. V následujících letech budeme postupně aktualizovat parametry α a β podle $Be(\alpha + x;\ \beta + n - x)$. Pro rok 2 platí $\alpha = 1 + 58 =$ $= 59$ a $\beta = 1 + 100 - 58 = 43$. Bayesovský odhad dostaneme pomocí výpočtu $\theta_B =$

[59] Pacáková, V.: Aplikovaná poistná statistika [5], s. 198

$= \frac{59}{59+43} = 0{,}57843.$ Pro další roky postupujeme stejným způsobem. Výsledky jsou zaznamenány v tab. 3. V posledním řádku tabulky je uveden bayesovský odhad pro následující rok. Můžeme tedy předpokládat, že v dalším roce bude asi 62 pojistných událostí.

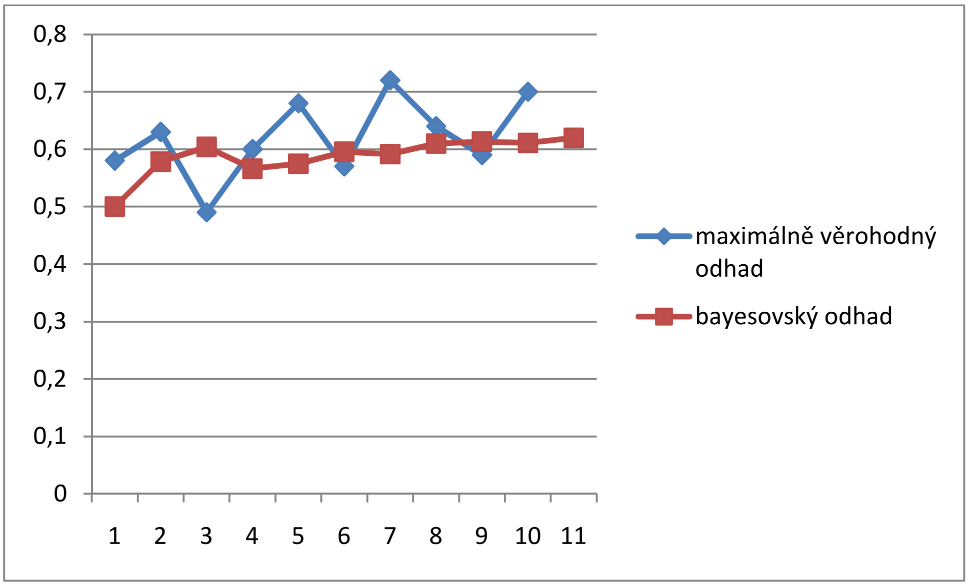
Tabulka 3: Výpočty pro model binomické/beta

Roky	**x**	**x/n**	**α**	**β**	**Bayesovský odhad**
1	58	0,58	1	1	0,5
2	63	0,63	59	43	0,57843
3	49	0,49	122	80	0,60396
4	60	0,6	171	131	0,56623
5	68	0,68	231	171	0,57463
6	57	0,57	299	203	0,59562
7	72	0,72	356	246	0,59136
8	64	0,64	428	274	0,60969
9	59	0,59	492	310	0,61347
10	70	0,7	551	351	0,61086
11			621	381	0,61976

[zdroj: vlastní zpracování]

V grafu 1 názorně vidíme porovnání maximálně věrohodného a bayesovského odhadu pravděpodobnosti, že nastane sledovaná událost, v našem případě pravděpodobnost návštěvy lékaře, resp. potřeby lékařské péče v průběhu roku.

Graf 1: Porovnání maximálně věrohodného a bayesovského odhadu



[zdroj: vlastní zpracování]

- *Model Poissonovo/gama*

Uvažujeme pojišťovnu, z jejíchž údajů známe počty pojistných plnění v letech 2005, 2006, …, 2010. Chceme odhadnout počet pojistných plnění v roce 2011, když víme, že předchozí počty mají v průběhu roku Poissonovo rozdělní s neznámým parametrem λ a jejich apriorní rozdělení je gama rozdělení $G\,(240; 1{,}2)$. Počty pojistných plnění v předchozích letech můžeme vidět v tab. 4.

Tabulka 4: Počet pojistných plnění

Roky	**Počet pojistných plnění**
2005	175
2006	184
2007	180
2008	196
2009	195
2010	202

[zdroj: V. Pacáková, cvičení Teorie rizika [22]]

V prvním kroku výpočtu vyjádříme hodnotu $\mu = \frac{240}{1{,}2} = 200$. Při kdredibilním odhad počtu pojistných plnění λ_B budeme dále vycházet ze vztahu (4.14)

$\lambda_B = E(\lambda/\mathrm{x}) = \mathrm{Z}\,\bar{\mathrm{x}} + (1-\mathrm{Z})200$. Tab. 4 upravíme a rozšíříme o sloupce, ve kterých postupně budeme vyjadřovat hodnoty průměru, faktoru kredibility a samotného odhadu λ_B. Průměrný počet pojistného plnění počítáme z dat z předchozích let. V prvním roce je nulový, protože nemáme žádné informace o počtech pojistných plnění z dřívějších let. Pro rok n = 2 dosadíme do vzorce $\bar{x}_2 = \frac{\sum_{i=1}^{n-1} x_i}{n-1} = \frac{175}{1} = 175$, pro n = 3 dostaneme $\bar{x}_3 = = \frac{175+184}{2} = 179{,}5$. Stejným postupem budeme pokračovat až po n = 6.

Apriorní gama rozdělení $G\ (240; 1{,}2)$ má parametr $\beta = 1{,}2$, hodnotu faktoru krediblity Z pro výpočet kredibiliního bayesovského odhadu λ_B v období $(n+1) = 2, \dots, 7$ pro libovolné *n* vypočítáme podle vztahu (4.15). V prvním roce bude tato hodnota opět nulová: $Z_1 = \frac{0}{1{,}2+0} = 0$. V dalších letech dostáváme postupně hodnoty: pro n = 2 je $Z_2 = \frac{1}{1{,}2+1} = = 0{,}4545455$, pro n = 3 je $Z_2 = \frac{2}{1{,}2+2} = 0{,}625$. Stejným způsobem postupujeme až k n = 7. Hodnoty jsou zapsány v tab. 5. Faktor kredibility Z odečteme od jedné a pomocí ostatních vypočítaných hodnot můžeme vyjádřit kredibilní odhad λ_B. Pro rok 2005 dostaneme $\lambda_{B(2005)} = 200$, což je střední hodnota apriorního rozdělení. Pro rok 2006 dostaneme $\lambda_B = 0{,}4545455\,.\,175 + 0{,}5454545\,.\,200 = 188{,}636$. Pro další roky postupujeme stejným způsobem, výsledky jsou zapsány v posledním sloupci tab. 5. V posledním řádku je uvedena hodnota odhadu pro rok 2011, kterou jsme získali na základě informací z předchozích let. Odhadujeme tedy, že očekávaný počet pojistných plnění bude přibližně 191.

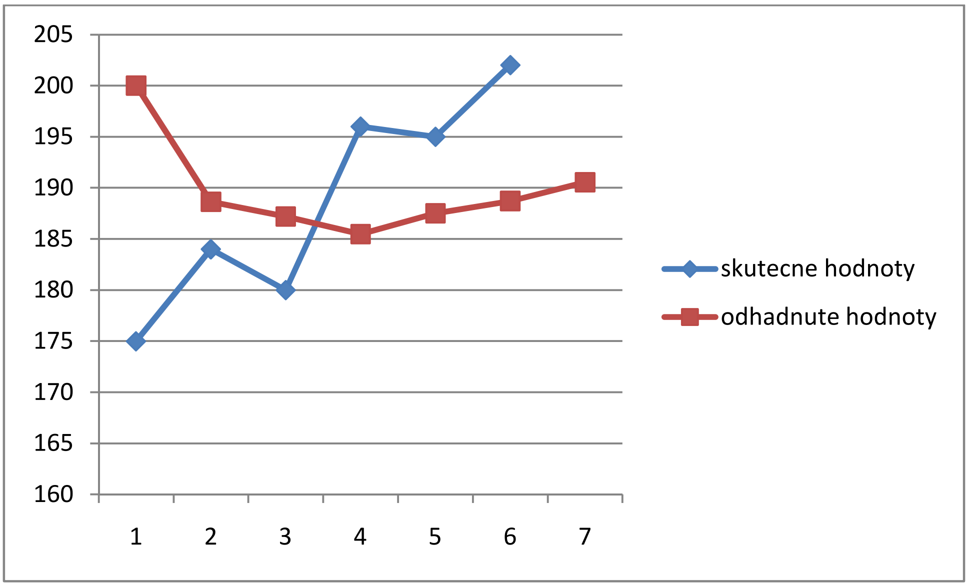
Tabulka 5: Odhad počtu pojistných plnění u modelu Poissonovo/gama

n	**x**	**$\bar{x}$**	**Z**	**1-Z**	**λ_B**
1	175	0	0	1	**200**
2	184	175,000	0,4545455	0,5454545	**188,636**
3	180	179,500	0,6250000	0,3750000	**187,188**
4	196	179,667	0,7142857	0,2857143	**185,476**
5	195	183,750	0,7692308	0,2307692	**187,500**
6	202	186,000	0,8064516	0,1935484	**188,710**
7		188,667	0,8333333	0,1666667	**190,556**

[zdroj: vlastní zpracování]

Můžeme provést porovnání skutečných a odhadnutých hodnot. V Grafu 2 vidíme, že čím více údajů z minulosti máme, tím více se odhadnutý počet přibližuje počtu skutečnému.

Graf 2: Porovnání skutečného a odhadnutého počtu pojistného plnění u modelu Poissonovo/gama



[zdroj: vlastní zpracování]

- *Model normální/normální*

Uvažujeme pojišťovnu, jejíž pojistné plnění X při určitém typu neživotního pojištění v jednotlivých letech označíme x_i. X_i má v každém roku normální rozdělení se střední hodnotu θ. Rozptyl je roven $\sigma_1^2 = 135\ 000^2$. Víme, že parametr θ má apriorní normální rozdělení s parametry $\mu = 2\ 100\ 000$ a $\sigma_2^2 = 150\ 000^2$.

Naším cílem je určit bayesovský odhad $\theta_B$výše čistého pojistného. Budeme vycházet z dat o výšce pojistných plnění ze 7 předchozích let.

Tabulka 6: Výše pojistného plnění

Roky	Výše pojistného plnění (x_i)
2004	2 112 000
2005	2 140 000
2006	1 955 000
2007	2 315 000
2008	2 280 000
2009	2 035 000
2010	2 215 000

[zdroj: Pacáková, V., cvičení Teorie rizika]

K tomu, abychom našli bayesovský odhad θ_B v roce 2011, potřebujeme pro každý rok znát průměrné pojistné plnění a faktor kredibility Z. Tab. 6 tedy můžeme rozšířit o další sloupce, ve kterých provedeme nejprve výpočet průměrné hodnoty z předchozích let. V prvním roce bude tato hodnota nulová, protože nemáme k dispozici žádné informace z předešlých let. V dalších letech sledování se hodnota průměru bude rovnat $\bar{x}_2 = \frac{2112000}{1} = 2112000$ pro n = 2, pro n = 3 dostáváme $\bar{x}_2 = \frac{2112000+2140000}{2} =$ $= 2126000$ a stejným způsobem budeme pokračovat až po n = 6. Dále provedeme výpočet faktoru kredibility Z. Díky znalosti rozptylů obou rozdělení vypočítáme faktor kredibility v jednotlivých letech $(n+1) = 2, \dots, 8$ podle vzorce (4.17) $Z = \frac{n}{n+\frac{135\,000^2}{150\,000^2}}$. Při dosazení v prvním roce je hodnota Z opět nulová. ve druhém roce $Z_2 = \frac{1}{1+\frac{135\,000^2}{150\,000^2}} =$ $= 0{,}5524862$. Takto pokračujeme až do 8. roku. Výsledky jsou zaznamenány v tab. 7.

Díky znalosti Z můžeme najít bayesovský odhad θ_B v jednotlivých letech podle vzorce (4.16) uvedeného v předchozí kapitole. Při dosazení za rok 2004 nám vyjde kredibilní bayesovský odhad $E(\theta/x_1) = 2\,100\,000$. V roce 2005 je odhad ve výši $E(\theta/x_2) =$ $= 0{,}5524862 \,.\, 2112000 + 0{,}4475138 \,.\, 2100000 = 2106629{,}834$. Výpočet provedeme pro každý rok a výsledky zaznamenáme do tab. 7. V posledním řádku je odhad pojistného plnění v roce 2011, který byl získán díky dodatečné znalosti skutečného počtu pojistných plnění v předcházejících 7 letech. Pojišťovna tedy může předpokládat, že výše pojistného plnění bude asi 2 145 070 v roce, pro který potřebujeme stanovit netto pojistné.

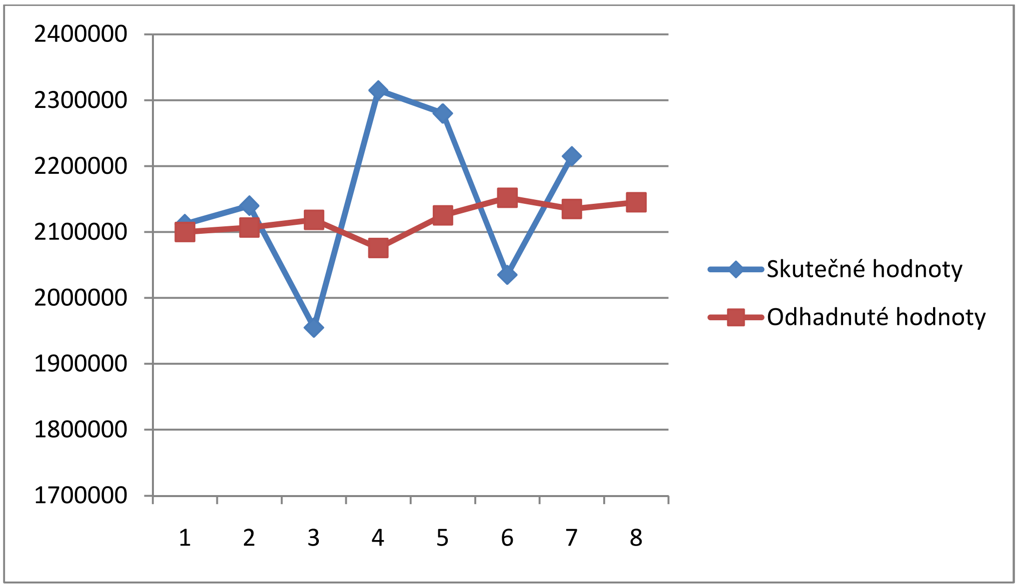
Tabulka 7: Výpočetní tabulka bayesovského odhadu pojistného plnění

n	x_i	$\bar{x}$	Z	1-Z	θ_B
1	2112000	0	0	1	**2100000**
2	2140000	2112000	0,5524862	0,4475138	**2106629,834**
3	1955000	2126000	0,7117438	0,2882562	**2118505,338**
4	2315000	2069000	0,7874016	0,2125984	**2075590,551**
5	2280000	2130500	0,8316008	0,1683992	**2125363,825**
6	2035000	2160400	0,8605852	0,1394148	**2151979,346**
7	2215000	2139500	0,8810573	0,1189427	**2134801,762**
8		2150285	0,8962868	0,1037132	**2145070,423**

[zdroj: vlastní zpracování]

Opět provedeme grafické znázornění skutečné a odhadnuté výše pojistného plnění. I zde platí, že čím více dat z předchozích let známe, tím více se odhad a realita budou přibližovat.

Graf 3: Porovnání skutečné a odhadnuté výše pojistného plnění u modelu normální/normální



[zdroj: vlastní zpracování]

5. Empirická bayesovská teorie kredibility

Tato teorie byla publikována na konci 60. let minulého století. Hlavními iniciátory byli švýcarští aktuáři Hans Bühlmann a Erwin Straub. Hans Bühlmann je švýcarský profesor, je zaměstnancem Swiss Re Insurance Co a také prezident prestižní Swiss Federal Institute of Technology (ETH), což je jedna ze světově nejlepších technických univerzit. Je významný aktuár, jenž je hlavním zastáncem teorie kredibility, především myšlenky odhadnout budoucí hodnotu pomocí metody minimalizace průměrné čtvercové chyby vycházející z apriorních dat. Jeho velký přínos do aktérské teorie a praxe spočívá v tom, že rozvinul nynější Bayesovský přístup k teorii kredibility. Hans Bühlmann je velkým kritikem americké teorie. Je to z toho důvodu, že americká teorie používá nepřesvědčivé matematické odvození, které jsou vázané klasickou statistikou a tak ignorují apriorní údaje. V bayesovské teorie zase na druhou stranu shledával problém při nutné znalosti apriorního rozdělení.

Pokud bychom chtěli aplikovat bayesovskou teorii krediblity na řešení problémů v pojistné praxi, narazíme na dva základní problémy. Prvním z nich je ten, že parametry apriorního rozdělení nejsou obvykle známé. K jejich určení je vyžadován určitý stupeň subjektivity, který však nemusí být ostatními akceptován. Druhý problém spočívá v tom, že ne vždy je možné upravit bayesovský odhad neznámého parametru do tvaru $P_k = Z\, P_r + (1 - Z).P_K$. Při modelech binomické/beta, Poissonovo/gama a normální/normální se jednoduchými úpravami k tomuto tvaru dostaneme, avšak při jiných typech modelů s jinými předpoklady o rozdělení f (X / θ) a $f(\theta)$ se k tomuto tvaru nemusíme dostat. Z toho tedy vyplývá, že tento postup můžeme užít jen pro některé parametry některých typů rozdělení. Empirická bayesovská teorie se snaží tyto omezení překonat.[60]

Pokud bychom to měli shrnout, tak empirická bayesovská teorie kredibility, která se také označuje jako evropská teorie kredibility, na rozdíl od bayesovské teorie kredibility má tyto vlastnosti:[61]

- vyžaduje méně matematických schopností,
- vyžaduje méně výpočetních zdrojů,
- nevyžaduje výběr apriorního rozdělení, což může být někdy složité rozhodování.

[60] Pacáková, V.: Aplikovaná poistná statistika [5], s. 205
[61] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 13

5.1. Bühlmannův model kredibility

Tento model byl publikován roku 1967 v „*Experience Rating and Credibility*“ prof. H. Bühlmannem, po kterém je i pojmenován. Model může být v některé literatuře pojmenován jako empirický bayesův model kredibility 1, můžeme se často setkat s označením EBCT 1. Pomocí tohoto modelu odhadujeme výšku čistého pojistného anebo počet pojistných plnění pro dané riziko v následujícím roce.

Předpokládáme, že portfolio je rozdělené do rizikových tříd nebo do I rizik, pro které chceme stanovit pojistné. Označíme $X_1, X_2, \dots X_j, \dots X_n$ celkový počet pojistných plnění n po sobě následujících obdobích. Pomocí těchto údajů můžeme odhadnout střední hodnotu celkového pojistného plnění $E(X)$, tedy čisté pojistné X_{n+1}, resp. průměrný počet pojistných událostí v následujícím roce. Předpokladem je, že platí:[62]

- X_j jsou identicky rozdělené, ne však nevyhnutelně nezávislé náhodné proměnné.
- Rozdělení X_j závisí od fixní hodnoty neznámého parametru θ, stejného pro všechny X_j.
- Pro pevně zvolené θ jsou X_j / θ nezávislé a identické rozdělené.

Kredibilní odhad v Bühlmannovém modelu tedy nezávisí na průměrech ostatních rizikových tříd, ale pouze na výběrovém průměru $\overline{X_\iota}$. Úpravami (odvození můžeme nalézt např. v Aplikovaná poistná štatistika od V. Pacákové) poté dostaneme vzorec pro kredibilní odhad pojistného:[63]

$$E(m(\theta) \,/\, \mathbf{X}) = Z\overline{X_\iota} + (1 - Z) \cdot E\big(m(\theta)\big) \tag{5.1}$$

Když střední hodnotu odhadujeme jako průměrnou výši pojistných plnění za celé portfolio, dostaneme ve výsledku homogenní kredibilní odhad.

Hodnota Z se nazývá faktor kredibility a vztah pro jeho výpočet byl navrhnut Albertem W. Whitneym v roce 1918. Faktor kredibility měří výšku věrohodnosti vztahující se na individuální zkušenosti. Nabývá hodnot z intervalu $\langle 0, 1\rangle$. Je vyjádřením věrohodností zkušeností s uvažovanou třídou pojištění při stanovení pojistného. Hodnota 0 značí to, že

[62] Pacáková, V.: Aplikovaná poistná statistika [5], s. 206
[63] Pacáková, V.: Aplikovaná poistná statistika [5], s. 206

údaje o individuálním riziku nemají žádné použití při stanovení pojistného. Naopak hodnota 1 se přiřazuje údajům o individuálním riziku, které jsou naprosto věrohodné při stanovení pojistného. Z toho tedy vyplývá, že pokud platí Z = 0, tak při určování pojistného využíváme jen údaje z porovnatelných rizik. Pokud hodnota Z = 1, tak se plně spoléháme na informace o vlastním riziku. Faktor kredibility se vypočítá podle vztahu:[64]

$$Z = \frac{n}{n + \frac{E(s^2(\theta))}{D(m(\theta))}} \tag{5.2}$$

Počet pozorování o vlastním riziku značíme *n*, podíl odhadu a vnitroskupinové variability (variabilita uvnitř rizikových skupin) a meziskupinového rozptylu (variabilita mezi rizikovými třídami) se nazývá koeficient kredibility.

Hodnota Z roste, pokud:[65]

- počet pozorovaných roků *n* roste,
- heterogenita portfolia roste,
- variabilita uvnitř rizikových tříd klesá.

V Bühlmannovém modelu se uvažuje se stejným počtem pozorování *n* pro každou třídu pojištění. To je příčina toho, že faktor kredibility je jednotný pro všechny rizikové třídy. Realita je však rozdílná, protože pojišťovny mají k dispozici různé množství informací z různých tříd pojištění. Na tomto modelu však můžeme dobře vidět principy empirické bayesovské teorie kredibility, ilustruje společné vlastnosti i rozdíly v přístupu k bayesovské teorii kredibility.

5.1.1. Praktická ukázka použití EBCT 1

Následující tabulka obsahuje data o výši celkového pojistného plnění u povinného ručení za čtyři předcházející roky pojišťoven ČSOB, Kooperativa, Allianz a Generali. Údaje jsou získané z výročních zpráv jednotlivých pojišťoven. Pomocí těchto údajů vypočítáme kdredibilní čisté pojistné v následujícím roku pro každou pojišťovnu. Počet pojišťoven značíme i = 1, 2 … N, počet roků značíme j = 1, 2 … n.

[64] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 17
[65] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 17

Tabulka 8: Údaje o celkové výšce pojistného plnění u povinného ručení

Pojišťovna	Roky			
	2006	2007	2008	2009
ČSOB	938 113	1 066 877	716602	589012
Kooperativa	3 776 108	2 872 894	2 156 929	2 969 002
Allianz	1 382 500	1 201 100	1 186 000	1 456 300
Generali	836 221	1 520 014	1 671 119	1 311 745

[zdroj: výroční zprávy pojišťoven]

Celkové pojistné plnění jednotlivých pojišťoven v jednotlivých letech značíme X_{ij}. Tab. 8 rozšíříme o dva sloupce. V prvním vypočítáme průměrné plnění jednotlivých pojišťoven za všechny čtyři roky $\overline{X}_{\iota}$ a ve druhém provedeme pomocný výpočet, který budeme potřebovat pro odhad E(s^2(θ)).

Tabulka 9: Pomocné výpočty pro odhad E(s2(**θ**))

Pojišťovna	Roky					
	2006	2007	2008	2009	$\overline{X}_{\iota}$	$\frac{1}{n-1}\sum_{j=1}^{n}(X_{ij}-\overline{X}_{\iota})$
ČSOB	938 113	1 066 877	716602	589 012	827 651	46 237 128 414
Kooperativa	3 776 108	2 872 894	2 156 929	2 969 002	2 943 733	439 188 453 74
Allianz	1 382 500	1 201 100	1 186 000	1 456 300	1 306 475	17 948 482 500
Generali	836 221	1 520 014	1 671 119	1 311 745	1 334 775	132 175 748 24
Celkem $\sum_{i=1}^{4}\frac{1}{4-1}\sum_{j=1}^{4}\left(X_{ij}-\overline{X}_{\iota}\right)^2$						**635 549 813 112**

[zdroj: vlastní zpracování]

Dále provedeme pomocné výpočty potřebné pro odhad D (m(θ)).

Tabulka 10: Pomocné výpočty pro odhad D (m(**θ**))

Pojišťovna	$(\overline{X}_{\iota}-\overline{X})^2$
ČSOB	601 411 882 556
Kooperativa	1 797 140 660 338
Allianz	88 021 099 172
Generali	72 029 837 264
$\frac{1}{N-1}\sum_{j=1}^{n}(\overline{X}_{\iota}-\overline{X})^2$	**852 867 826 443**

[zdroj: vlastní zpracování]

Nyní z těchto hodnot můžeme odhadnout charakteristiky E $(m(\theta))$, D $(m(\theta))$ a $E(s^2(\theta))$. Tyto charakteristiky (v některé literatuře jsou označeny jako strukturální parametry) odhadujeme podle vztahů:[66]

- $est\ E(m(\theta)) = \bar{X}$
- $est\ E(s^2(\theta)) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_\iota\right)^2$
- $est\ D(m(\theta)) = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{X}_\iota - \bar{X})^2 - \frac{1}{Nn}\sum_{i=1}^{N}\frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_\iota\right)^2$

První odhadneme střední hodnotu pomocí průměru za všechny pojišťovny a všechny roky.

$$est\ E\big(m(\theta)\big) = \bar{X} = \frac{1}{N}\sum_{i=1}^{N}\bar{X}_\iota = \frac{1}{4}\cdot(827\,651 + 2\,943\,733 + 1\,306\,475\ + 1\,334\,775)$$
$$= 1\,603\,159$$

Jakmile známe odhad průměrného pojistného plnění, můžeme odhadnout vnitroskupinový rozptyl.

$$est\ E\big(s^2(\theta)\big) = \frac{1}{4}\cdot 635\,549\,813\,112\ = 158\,887\,453\,278$$

Jako poslední charakteristiku můžeme odhadnout meziskupinový rozptyl.

$$est\ D\big(m(\theta)\big) = \frac{1}{3}\cdot 852\,867\,826\,443 - \frac{158\,887\,453\,278}{4} = 813\,145\ \ 963\,124$$

V případě, že známe odhad základních charakteristik, lze vypočítat faktor kredibility. Podle vzorce (5.2), který je uveden v kapitole 5.1, dostaneme:

$$Z = \frac{4}{4 + \frac{158\,887\,453\,278}{813\,145\ \ 963\,124}} = 0{,}953425534$$

Vypočítali jsme všechny potřebné údaje a nyní již zbývá jen vypočítat odhad čistého kredibilního pojistného. Pro tento odhad platí podle vztahu (5.1):

$$Z\cdot\bar{X}_\iota + (1-Z)\cdot E\big(m(\theta)\big) = 0{,}953425534\cdot\bar{X}_\iota + 0{,}0465745\cdot 1\,603\,159$$

[66] Pacáková, V.: Aplikovaná poistná statistika [5], s. 214

Pomocí všech vypočítaných hodnot nyní můžeme tedy určit odhady čistého kredibilního pojistného pro každou pojišťovnu, které jsou uvedeny v tab. 11.

Tabulka 11: Odhad čistého kredibilního pojistného pro každou pojišťovnu

Pojišťovna	Čisté kredibilní pojistné
ČSOB	863 769,85
Kooperativa	2 881 296,70
Allianz	1 320 292,88
Generali	1 347 274,58

[zdroj: vlastní zpracování]

5.2. Bühlmannův-Straubův model kredibility

Tento model má základ v Bühlamnnovým modelu, je jeho určitým zevšeobecněním. Řekli jsme, že Bühlmannův model nezohledňuje různé množství informací z různých tříd pojištění, kdežto Bühlmannův-Straubův model již tyto informace částečně bere na vědomí, například můžeme zohlednit různý počet uzavřených pojistek. V porovnání s Bühlmannovým modelem je však teoreticky a výpočetně složitější.

V literatuře se můžeme setkat s názvem empirický Bayesův model krediblity 2, nebo i s názvem EBCT 2. Byl publikovaný v roce 1970 v článku „*Glaubwürdigkeit für Schadensätze*“.[67] Hans Bühlmann a jeho doktorand Erwin Straub tímto zdokonalili Bühlmannův přístup. Jejich tzv. Bühlmannův-Straubův model je jedním z nejdůležitějších modelů v teorii kredibility. Význam spočívá především v tom, že se dá aplikovat v aktuárské praxi neživotního pojištění, životního pojištění i v zajištění. Také tvoří základ pro další specifické kredibilní modely, např. regresní kredibilní modely.

Bühlmannův-Straubův model slouží k odhadu čistého pojistného (resp. odhad celkového počtu pojistných plnění v následujícím roce), můžeme ho však využít i ke kredibilnímu odhadu škodové frekvence a ke kredibilnímu odhadu průměrné škody. V tomto případě pak lze odhad pojistného zjistit jako součin kredibilního odhadu škodové frekvence a kredibilního odhadu průměrné škody.[68]

[67] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 24
[68] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 24

Předpokládáme, že portfolio je rozdělené do I rizikových tříd. Poté označme Y_1, Y_2, …, Y_j, …, Y_n známé hodnoty celkových pojistných plnění v letech 1, 2, …, j, … n. Také známe parametr P_1, P_2, …, P_j, …, P_n charakterizující rozsah rizika ve sledovaných letech. Důležité je zmínit, že to jsou konstanty a ne náhodné proměnné.[69] Dále uvažujeme normované proměnné $X_j = \frac{Y_i}{P_j}$.

Model vychází z toho, že:[70]

- Náhodné proměnné X_j, $j = 1, 2, \ldots, n$, mají rozdělení stejné pro každé $j = 1, 2, \ldots, n$ (rozdělení závisí na parametru, který je fixní, ale neznámý**)**.
- Proměnné X_1/θ , X_2/θ , …, X_n/θ jsou nezávislé, ale nemusí být identicky rozdělené.
- $E(X_j\ /\ \theta) = m(\theta)$ nezávisí od *j*.
- $D(X_j\ /\ \theta) \cdot P_j = s^2(\theta)$ nezávisí od *j*.

Zevšeobecnění oproti předchozímu modelu můžeme vidět v tom, že neexistují podmínky, aby X_j / θ, resp. X_j nabývaly stejného rozdělení.

Pro výpočet kredibilního odhadu netto pojistného využíváme vztahy, jejíchž odvození můžeme nalézt např. v knize Aplikovaná poistná statistika:[71]

$$E(m(\theta)\ /\boldsymbol{X}) = Z_i\overline{X_\iota} + (1 - Z_i)\ E\big(m(\theta)\big) = Z_i\overline{X_\iota} + (1 - Z_i)\bar{X} \qquad (5.3)$$

V různých letech může být uzavřen různý počet uzavřených různých pojistek, proto je odhad individuálního pojistného $\bar{X}_i$.

V tomto modelu platí, že hodnota Z roste, když:

- Počet pojištění P_i pro uvažovanou rizikovou třídu roste, tj. pro $P_i \rightarrow \infty$ platí $Z_i \rightarrow \infty$,
- koeficient kredibility $\frac{E(s^2(\theta))}{D(m(\theta))}$ (tedy podíl vnitroskupinové variability a meziskupinové variability) klesá, tj. pro $\frac{E(s^2(\theta))}{D(m(\theta))} \rightarrow 0$ platí $Z \rightarrow 1$.[72]

---

[69] Pacáková, V.: Aplikovaná poistná statistika [5], s. 216
[70] Pacáková, V.: Aplikovaná poistná statistika [5], s. 217
[71] Pacáková, V.: Aplikovaná poistná statistika [5], s. 220

Míra variability $E(s^2(\theta))$ je vyjádřením průměrné variability průměrných ročních výšek pojistných plnění uvnitř tarifních tříd. Je to taková variabilita, která neuvažuje ratingové faktory a náhodné činitelé. $E(s^2(\theta))$ se rovná 0, pokud by byly tarifní třídy homogenní. Čím více jsou tarifní třídy heterogenní, tím více $E\big(s^2(\theta)\big)$ roste. Na druhou stranu $D(m(\theta))$ je vyjádřením variability způsobené uvažováním ratingových faktorů použitých při segmentaci portfolia. Čím menší bude podíl vnitroskupinové a meziskupinové variability, tím věrohodnější budou údaje z jednotlivých tříd pojištění při stanovení pojistného.

Při těchto známých informacích můžeme říci, že pro odhad charakteristik platí:[73]

$$est\, E(m(\theta)) = \bar{X} \tag{5.4}$$

$$est\, E(s^2(\theta)) = \frac{1}{N(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{n} P_{ij}\, (X_{ij} - \bar{X}_i)^2 \tag{5.5}$$

$$est\, D(m(\theta)) = \frac{1}{P^*} \left\{ \frac{1}{Nn-1} \sum_{i=1}^{N} \sum_{j=1}^{n} P_{ij}\, (X_{ij} - \bar{X})^2 - \frac{1}{N(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{n} P_{ij}\, (X_{ij} - \bar{X}_i)^2 \right\}. \tag{5.6}$$

Dále platí vztahy:[74]

$$P_i = \sum_{j=1}^{n} P_{ij} \tag{5.7}$$

$$Y_i = \sum_{j=1}^{n} Y_{ij} \tag{5.8}$$

$$P = \sum_{i=1}^{N} P_i \tag{5.9}$$

[72] Šoltés, E.: Modely kredibility na výpočet poistného [6], s. 26
[73] Pacáková, V.: Aplikovaná poistná statistika [5], s. 221
[74] Pacáková, V.: Aplikovaná poistná statistika [5], s. 220

$$P^* = \frac{1}{Nn-1}\sum_{i=1}^{N} P_i \cdot \left(1-\frac{P_i}{P}\right) \tag{5.10}$$

$$\bar{X}_i = \frac{1}{P_i}\sum_{j=1}^{n} P_{ij}\, X_{ij} = \frac{1}{P_i}\sum_{i=1}^{n} Y_{ij} \tag{5.11}$$

$$\bar{X} = \frac{1}{P}\sum_{i=1}^{N}\sum_{j=1}^{n} P_{ij}\, X_{ij} = \frac{1}{P}\sum_{i=1}^{N} P_i \bar{X}_i \tag{5.12}$$

$$Z_i = \frac{P_i}{P_i + \frac{E(s^2(\theta))}{D(m(\theta))}} \tag{5.13}$$

5.2.1. Praktická ukázka použití

Ukážeme aplikaci modelu Bühlmanna-Strauba na pojištění odpovědnosti z provozu motorového vozidla. Uvažujeme pojišťovnu, u které chceme odhadnout výši čistého kredibilního pojistného pro sedm leasingových společností pro rok 2011. Nalezneme odhad čistého pojistného pro každou společnost, když víme, že každá pojistka je uzavřena klientem na jeden rok. Tab. 12 obsahuje celkové pojistné plnění u pojištění odpovědnosti za provoz motorového vozidla v předešlých 5 letech (Y) a v tab. 13 jsou uvedeny počty automobilů (P), u kterých byla pojistka uzavřena. Počet automobilů nám bude složit jako váhy. Díky tomu jsme schopni vzít v potaz „velikost“ pojišťovny, což nebylo v EBCT 1 zohledněno.

Tabulka 12: Celkové pojistné plnění u povinného ručení v tis. Kč (Yij)

Společnost (i)	**Roky (j)**				
	2006	2007	2008	2009	2010
Společnost 1	100	57	20	180	38
Společnost 2	2	2	0	7	1
Společnost 3	4	12	15	0	2
Společnost 4	5	0	1	1	4
Společnost 5	0	3	3	0	2
Společnost 6	6	0	0	0	12
Společnost 7	14	4	43	7	10

[zdroj: WATERS, H. R. Credibility theory [21], s. 35]

Tabulka 13: Počet pojištěných automobilů u povinného ručení v tis. Kč (P_{ij})

Společnost (i)	**Roky (j)**				
	2006	2007	2008	2009	2010
Společnost 1	80	80	83	85	85
Společnost 2	5	5	5	5	5
Společnost 3	20	20	20	23	26
Společnost 4	10	10	9	9	10
Společnost 5	15	20	25	25	30
Společnost 6	10	11	11	11	12
Společnost 7	30	29	29	30	30

[zdroj: WATERS, H. R. Credibility theory [21], s. 35]

V prvním kroku vyjádříme celkový počet pojištěných automobilů. Pro Společnost 1 (tedy i = 1) podle (5.7) platí: $\mathbf{P_1 = \sum_{j=1}^{5} P_{ij} = 80 + 80 + 83 + 85 + 85 = 413}$, ekvivalentní výpočet provedeme i pro ostatní společnosti (i = 2, …, 7). Dále uvedeme součet celkového pojistného plnění u společností podle vzorce (5.8), pro i = 1 platí $\mathbf{Y_1 = \sum_{j=1}^{n} Y_{ij} = 100 + 57 + 20 + 180 + 38 = 395}$. Můžeme tedy konstatovat, že Společnost 1 za roky 2006 až 2009 vyplatila 395 tis. Kč na pojistném plnění u pojištění odpovědnosti z provozu motorového vozidla. Výše pojistného plnění u dalších pojišťoven je uvedena v tab. 11.

Pro výpočet průměrného pojistného plnění v období předchozích 5 let využijeme vztah (5.11). Opět uvedeme jako příklad Společnost 1 (i = 1): $\bar{X}_1 = \frac{1}{P_1}\sum_{i=1}^{n} Y_{ij} = \frac{1}{413} 395 = 0{,}9564$. Tato hodnota říká, že Společnost 1 vyplatila v průměru 956,4 Kč na jednu pojistku. Pro i = 2, …, 7 použijeme stejný postup a všechny hodnoty zaznačíme v tab.14.

Vypočítané průměry $\bar{X}_i$ představují individuální pojistné. Tab. 14 obsahuje vypočítané charakteristiky a v součtovém řádků vidíme průměr za celé portfolio a za všechny roky, který je podle (5.12) roven $\bar{X} = \frac{1}{913}555 = \mathbf{0{,}6079}$. Celkově za celých 5 let a za všechny společnosti bylo vyplaceno na pojistném plnění průměrně 608 Kč na jednu pojistku.

Tabulka 14: Průměrné pojistné plnění společností

Společnost (i)	**Y_i**	**P_i**	**$\bar{x}_i$**
	$\sum_{j=1}^{5} Y_{ij}$	$\sum_{j=1}^{5} P_{ij}$	$\frac{1}{P_i}\sum_{i=1}^{n} Y_{ij}$
Společnost 1	395	413	0,9564
Společnost 2	12	25	0,4800
Společnost 3	33	109	0,3028
Společnost 4	11	48	0,2292
Společnost 5	8	115	0,0696
Společnost 6	18	55	0,3273
Společnost 7	78	148	0,5270
Celkem	555	913	**0,6079**

[zdroj: vlastní zpracování]

Pro výpočet odhadu výše čistého kredibilního pojistného je nutné hodnoty standardizovat $X_{ij} = \frac{Y_{ij}}{P_{ij}}$.[75]

Tabulka 15: Standartizované hodnoty X_{ij}

Společnost (i)	**Roky (j)**				
	2006	2007	2008	2009	2010
Společnost 1	1,2500	0,7125	0,2410	2,1176	0,4471
Společnost 2	0,4000	0,4000	0,0000	1,4000	0,2000
Společnost 3	0,2000	0,6000	0,7500	0,0000	0,0769
Společnost 4	0,5000	0,0000	0,1111	0,1111	0,4000
Společnost 5	0,0000	0,1500	0,1200	0,0000	0,0667
Společnost 6	0,6000	0,0000	0,0000	0,0000	1,0000
Společnost 7	0,4667	0,1379	1,4828	0,2333	0,3333

[zdroj: vlastní zpracování]

Dále provedeme dílčí výpočet $P_{ij}(X_{ij} - \bar{X}_i)^2$ a součet hodnot $\sum_{i=1}^{7}\sum_{j=1}^{5} P_{ij}\,(X_{ij} - \bar{X}_i)^2$, který budeme využívat pro odhad $est\,E(s^2(\theta))$ podle vzorce (5.5). Do tab. 16 zaznamenáme hodnoty pro všechny společnosti i = 1, 2, ..7 a roky j = 1, 2, … 5.

[75] Pacáková, V.: Aplikovaná poistná statistika [5], s. 224

Tabulka 16: Pomocné výpočty $\sum_{i=1}^{7}\sum_{j=1}^{5}\mathbf{P_{ij}}(\mathbf{X_{ij}}-\bar{\mathbf{X}}_{\iota})^2$

Společnost (i)	**Roky (j)**					**Celkem**
	2006	2007	2008	2009	2010	$\sum_{j=1}^{5} P_{ij}(X_{ij}-\bar{X}_{\iota})^2$
Společnost 1	6,8953	4,7596	42,4854	114,6188	22,0528	190,8120
Společnost 2	0,0320	0,0320	1,1520	4,2320	0,3920	5,8400
Společnost 3	0,2112	1,7671	4,0006	2,1082	1,3260	9,4130
Společnost 4	0,7335	0,5252	0,1254	0,1254	0,2918	1,8014
Společnost 5	0,0726	0,1294	0,0636	0,1210	0,0003	0,3868
Společnost 6	0,7438	1,1782	1,1782	1,1782	5,4307	9,7091
Společnost 7	0,1093	4,3905	26,4893	2,5877	1,1255	34,7022
Celkem ($\sum_{i=1}^{7}\sum_{j=1}^{5} P_{ij}\,(X_{ij}-\bar{X}_{\iota})^2$)						**252,6645**

[zdroj: vlastní zpracování]

Pro odhad meziskupinové variability (5.6) budeme potřebovat znát hodnotu P^*, kterou vypočítáme podle vztahu (5.10). Hodnoty P_i a $P = 913$ jsou již vyjádřeny v prvním kroku (viz. Tabulka 14). Díky dílčím výsledkům v Tabulce 17 můžeme vyjádřit $P^* = \frac{1}{7.5-1}\cdot 668{,}1665 == \mathbf{19{,}6520}$.

Tabulka 17: Pomocné výpočty pro hodnotu $\boldsymbol{P}^*$

Společnost (i)	$P_i\cdot\left(1-\frac{P_i}{P}\right)$
Společnost 1	226,1774
Společnost 2	24,3154
Společnost 3	95,9869
Společnost 4	45,4765
Společnost 5	100,5148
Společnost 6	51,6867
Společnost 7	124,0088
$\sum_{i=1}^{7} P_i\cdot\left(1-\frac{P_i}{P}\right)$	**19,6520**

[zdroj: vlastní zpracování]

Pro nalezení odhadu kredibilního pojistného potřebujeme znát hodnoty $P_{ij}(X_{ij}-\bar{X})^2$ u každé společnosti za všechny roky a jejich součet $\sum_{i=1}^{N}\sum_{j=1}^{n} P_{ij}\,(X_{ij}-\bar{X})^2$, jak je uvedeno ve vzorci (5.6).

Tabulka 18: Pomocné hodnoty pro výpočet hodnoty $\mathbf{est\ D(m(\theta))}$

Společnost (i)	**Roky (j)**					**Celkem**
	2006	2007	2008	2009	2010	$\sum_{j=1}^{5} P_{ij}(X_{ij}-\bar{X})^2$
Společnost 1	32,9848	0,8755	11,1744	193,7471	2,1986	240,9805
Společnost 2	0,2161	0,2161	1,8476	3,1372	0,8319	6,2489
Společnost 3	3,3274	0,0012	0,4039	8,4991	7,3300	19,5616
Společnost 4	0,1164	3,6953	2,2211	2,2211	0,4322	8,6860
Společnost 5	5,5429	4,1932	5,9508	9,2381	8,7876	33,7126
Společnost 6	0,0006	4,0648	4,0648	4,0648	1,8450	14,0400
Společnost 7	0,5983	6,4049	22,1967	4,2087	2,2614	35,6699
Celkem ($\sum_{i=1}^{7}\sum_{j=1}^{5} P_{ij}\,(X_{ij}-\bar{X})^2$)						**358,8894**

[zdroj: vlastní zpracování]

Máme veškeré potřebné údaje pro odhadnutí charakteristik podle (5.4, 5.5, 5.6).

$$est\ E\big(m(\theta)\big) = \bar{X} = 0{,}6079$$

$$est\ E(s^2(\theta)) = \frac{1}{7(5-1)} \cdot 252{,}6645 = 9{,}0237$$

$$est\ D\big(m(\theta)\big) = \frac{1}{19{,}6520} \cdot \left(\frac{358{,}8894}{7 \cdot 5 - 1} - \frac{252{,}6645}{7(5-1)}\right) = 0{,}0780$$

Hodnota faktoru kredibility se vypočítá pro každou pojišťovnu podle (5.13) pomocí odhadnutých charakteristik $est\ E(s^2(\theta))$ a $est\ D\big(m(\theta)\big)$. Podíl těchto dvou charakteristik nám vychází 115,7317. Platí, že čím menší je tento podíl, tím věrohodnější jsou údaje z jednotlivých tříd pojištění při stanovení pojistného. Pro Společnost 1 nyní vypočítáme faktor kredibility následovně: $Z_i = \frac{413}{413+\frac{9{,}0237}{0{,}00780}} = 0{,}7811$. Hodnoty pro Společnost 2 – 7 jsou uvedeny v tab. 19. Platí, že čím více se hodnota faktoru kredibility blíží jedné, tím více se společnost spoléhá na své údaje. Můžeme říci, že Společnost 1 se spoléhá na vlastní údaje ze všech společností nejvíce, protože hodnota faktoru kredibility je rovna 0,7811. Nejméně se na své vlastní údaje spoléhá Společnost 2, u které faktor kredibility je pouze 0,1776.

Nyní máme všechny potřebné hodnoty pro výpočet odhadu kredibilního pojistného v roce 2011 pro různé společnosti. K tomuto využijeme vztah (5.3), pro Společnost 1 platí $\mathrm{E}(\mathrm{m}(\theta)\ /\mathbf{X}) = 0{,}7811 \cdot 0{,}9564 + (1-0{,}7811)\ \cdot 0{,}6079 = 0{,}8801$. Stejným způsobem pokračujeme pro každou společnost. Hodnoty odhadu kredibilního pojistného jsou uvedeny v tab. 19. Tento odhad zohledňuje data z předchozích let a také „velikost“ jednotlivých společností s ohledem na počet pojištěných automobilů.

Tabulka 19: Kredibilní odhad pojistného

Společnost (i)	Z_i	Kredibilní pojistné
Společnost 1	0,7811	0,8801
Společnost 2	0,1776	0,5852
Společnost 3	0,4850	0,4599
Společnost 4	0,2932	0,4969
Společnost 5	0,4984	0,3396
Společnost 6	0,3221	0,5175
Společnost 7	0,5612	0,5625

[zdroj: vlastní zpracování]

Pokud bychom předpokládali, že společnosti zamýšlí pojistit 90, 5, 28, 12, 38, 15 a 35 aut v roce 2011, můžeme odhadnout čisté pojistné celkem. Kredibilní pojistné na jednu jednotku vynásobíme počtem zamýšlených pojištěných aut a získáme hodnoty uvedeny v tab. 20.

Tabulka 20: Kredibilní pojistné celkem

Společnost (i)	Kredibilní pojistné (na jedno auto)	Kredibilní pojistné (celkem)
Společnost 1	0,8801	79,2115
Společnost 2	0,5852	2,9258
Společnost 3	0,4599	12,8769
Společnost 4	0,4969	5,9623
Společnost 5	0,3396	12,9040
Společnost 6	0,5175	7,7623
Společnost 7	0,5625	19,6878

[zdroj: vlastní zpracování]

Závěr

Diplomová práce je zaměřena na popsání důležitých aspektů bayesovské statistiky a na vysvětlení hlavního principu a významu teorie kredibility pro pojistnou praxi. Vzhledem k tomu, že pojistná matematika tvoří základ každé pojišťovny, je pojistný matematik povinen neustále hledat nové a přesnější metody pro stanovení pojistného, pro odhad počtu škod atd. Jedním z těchto nových přístupů mohou být bayesovské odhady.

Vzhledem k tomu, že statistika a teorie pravděpodobnosti jsou teoretickým základem pro moderní pojistné vědy, je první kapitola práce zaměřena na vysvětlení těchto pojmů. Byly uvedeny různé definice statistiky a také vysvětleny pojmy pro pochopení výpočtů provedených v diplomové práci.

Bayesovskou statistikou se zabývá kapitola 2, ve které je tento pojem podrobně vysvětlen. Přehledně jsou uvedeny hlavní rozdíly mezi klasickým a bayesovským přístupem, které jsou interpretovány do tabulky. Následující kapitola popisuje teorii kredibility. Je uvedeno, co si lze pod tímto slovem představit, uvedena je definice a také stručně popsána historie vývoje této teorie.

Důležitou částí práce jsou nejen teoretické poznatky, ale i jejich aplikace na praktických ukázkách. V kapitole 4 a v kapitole 5 byl ukázán a vysvětlen princip jednotlivých modelů. Kapitola 4 je zaměřena na Bayesovské odhady, převážně tedy na model binomické/beta, Poissonovo/gama a normální/normální. Po vysvětlení teoretického principu těchto modelů se přešlo k ukázce aplikace. K výpočtům byl použit program Microsoft Excel. Jednotlivé kroky výpočtu jsou podrobně popsány a mezivýsledky byly uvedeny pro přehlednost v tabulkách. V grafu byly porovnány výsledné odhady a skutečně zjištěné hodnoty. I přes kritiku byesovských metod bylo ukázáno, že mohou být užitečně použity v praktických situacích, kdy máme známá data z minulých let.

Program Microsoft Excel byl použit i pro výpočty u Bühlmannova modelu a u modelu Bühlmann-Straubova v poslední kapitole práce. Opět byly modely nejdříve vysvětleny a poté na příkladu aplikovány. Na výpočtu byl ukázán způsob odhadu kredibiliního pojistného u obou modelů krok za krokem. Bühlmann-Straubův model je jedním z nejdůležitějších v teorii kredibility, proto mu byla věnována značná pozornost. Je možné ho aplikovat v různých oblastech neživotního pojištění a také tvoří základ pro různé specifické kredibilní modely.

Vzhledem k tomu, že literatura k tématu diplomové práce je v České republice omezená, bylo potřeba hledat informace o problematice týkající se teorie kredibiliy a bayesovských odhadů převážně v zahraničích publikacích. V dané práci byla názorně popsána a vysvětlena podstata bayesovských metod v oblasti neživotního pojištění a byly využity veškeré dostupné zdroje a materiály.

Seznam použité literatury a ostatních zdrojů

[1] HÁBEK, Petr. *Texty k bayesovské statistice* . Praha : Vysoká škola ekonomická v Praze, 1990. 139 s. ISBN 80-7079-862-9.

[2] KOTLEBOVÁ, Eva. *Bayesovská štatistická indukcia v ekonomických aplikáciách*. Bratislava : EKONÓM, 2009. 80 s. ISBN 978-80-225-2788-0.

[3] KUBANOVÁ, Jana. *Statistické metody pro eknomickou a technickou praxi*. Bratislava : Statis, 2004. 253 s. ISBN 80-85659-37-9.

[4] KUBANOVÁ, Jana; LINDA, Bohdan. *Sbírka příkladů z pravděpodobnosti*. Bratislava : Statis, 2004. 139 s. ISBN 80-85659-36-0.

[5] PACÁKOVÁ, Viera. *Aplikovaná poistná štatistika*. Bratislava : EKONÓM, 2004. 264 s. ISBN 80-8078-004-8.

[6] ŠOLTÉS, Erik. *Modely kredibility na výpočet poistného*. Bratislava : EKONÓM, 2009. 152 s. ISBN 978-80-225-2798-9.

[7] TSE, Yiu-Kuen . Nonlife actuarial models : Theory, methods and evaluation. Cambridge : Cambridge University press, 2009. 524 s. ISBN 978-0-521-76465-0.

[8] WALLER, Rya A. ; MARTZ, Harry F. *Bayesian reiability analysis*. New York : J. Wiley, 1982. 745 s.

[9] WATERS, Howard R. *Credibility theory.* Edinburgh : Heriot-Watt University, Dept. of Actuarial Maths and Statistics, 1993. 136 s.

[10] *Náhodné jevy. Pravděpodobnost a její základní vlastnosti. Podmíněná pravděpodobnost. Nezávislost* [online]. 2011 [cit. 2011-01-08]. Dostupné z WWW: <http://www.informatika.xcars.cz/nahodnejevy.html>.

[11] *ISBA* [online]. 2011 [cit. 2011-01-09]. Dostupné z WWW: <http://www.bayesian.org/>.

[12] *Bayesova věta - Wikipedie* [online]. 2010 [cit. 2011-01-08]. Dostupné z WWW: <http://cs.wikipedia.org/wiki/Bayesova_v%C4%9Bta>.

[13] *Http://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=16870* [online]. 2011 [cit. 2011-01-08]. Dostupné z WWW: <http://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=16870>.

[14] *Http://www.karlin.mff.cuni.cz/~huskova/backup_files/huskova_bayes.pdf* [online]. 2011 [cit. 2011-01-08]. Dostupné z WWW: <http://www.karlin.mff.cuni.cz/~huskova/backup_files/huskova_bayes.pdf>.

[15] *Http://www.wyomingbioinformatics.org/~achurban/docs/presentationBayesian.pdf* [online]. 1. 2. 2011 [cit. 2011-02-01]. Dostupné z WWW: <http://www.wyomingbioinformatics.org/~achurban/docs/presentationBayesian.pdf>.

[16] *Http://www.1kspa.cz/kladno/stud_materialy/matematika/Pravdepodobnost.pdf* [online]. 2011 [cit. 2011-03-15]. Dostupné z WWW: <http://www.1kspa.cz/kladno/stud_materialy/matematika/Pravdepodobnost.pdf>.

[17] *http://www.insurancehalloffame.org/laureateprofile.php?laureate=78* [online]. 2011 [cit. 2011-02-27]. Dostupné z WWW: <http://www.insurancehalloffame.org/laureateprofile.php?laureate=78>.

[18] *CREDIBILITY THEORY: THE CORNERSTONE OF ACTUARIAL SCIENCE* [online]. 2011 [cit. 2011-04-18]. Dostupné z WWW: <http://www.soa.org/library/pdftest/journals/north-american-actuarial-journal/1999/april/naaj9904_10.pdf>.

[19] *Http://custom.kbbarko.cz/e+m/02_2009/09_pacakova_soltes_ova.pdf* [online]. 2011 [cit. 2011-03-15]. Dostupné z WWW: <http://custom.kbbarko.cz/e+m/02_2009/09_pacakova_soltes_ova.pdf>.

[20] *Http://www.behan.ws/Credibility_Presentation.pdf* [online]. 2011 [cit. 2011-03-15]. Dostupné z WWW: <http://www.behan.ws/Credibility_Presentation.pdf>.

[21] *Matematický aparát – statistika* [online]. 2011-04-11 [cit. 2011-04-11]. Dostupné z WWW: <http://docs.google.com/viewer?a=v&q=cache:VbKjfZE6mPEJ:artemis.osu.cz:8080/artemis/uploaded/171_UVMET%252005.doc+stredni+hodnota+diskr%C3%A9tn%C3%AD+prom%C4%9Bnn%C3%A9&hl=cs&gl=cz&pid=bl&srcid=ADGEEShOmyGZVu

rcZRcMJsnOR6FcV7jjoIi9_f8E-
5L8aLonksI740JAAQ9OD85S05MIYBMTWs6c8iJBBw5hqVQ7Yn47gtoqF6kwqyW
xzKMUJ_PWJIjxeWl5AZYNgLra3YSzPKv7aVGq&sig=AHIEtbTxu4KxZPVZbcpyI
DFLoJstZ5YeUw>

[22] Pacáková, V. *Teorie rizika (cvičení)*. Pardubice, Univerzita Pardubice, 9. 4. 2010.

Seznam příloh

Příloha č. 1 **Model EBCT 1 - výpočty**

1	2	3	4	$\bar{X}_i$	$(\bar{X}_i - \bar{X})^2$
938 113	1 066 877	716602	589012	827 651	601 411 882 556
3 776 108	2 872 894	2 156 929	2 969 002	2 943 733	1 797 140 660 338
1 382 500	1 201 100	1 186 000	1 456 300	1 306 475	88 021 099 172
836 221	1 520 014	1 671 119	1 311 745	1 334 775	72 029 837 264
celkem				1 603 159	2 558 603 479 330
$\frac{1}{N-1}\sum_{j=1}^{n}(\bar{X}_i - \bar{X})^2$					852 867 826 443

1	2	3	4	$\sum_{j=1}^{n}(\bar{X}_i - \bar{X})^2$	$\frac{1}{n-1}\sum_{j=1}^{n}(\bar{X}_i - \bar{X})^2$
12201853444	57229079076	12331880401	56948572321	1,38711E+11	46 237 128 414,00
6,92848E+11	5018199341	6,19061E+11	638509726,6	1,31757E+12	439 188 453 774,25
5779800625	11103890625	14514225625	22447530625	53845447500	17 948 482 500,00
2,48556E+11	34313579741	1,13127E+11	530369385,1	3,96527E+11	132 175 748 424,25
celkem					635 549 813 112,50

estE(s^2(Θ))	158 887 453 278,13
estE(s^2(Θ))/4	39721863320
estD(m(Θ))	813 145 963 124
Z	0,953425534
1-Z	0,046574466

Θ_B
863769,8474
2881296,697
1320292,875
1347274,58

Příloha č. 2 **Model EBCT 2 – výpočty**

Y_{ij}	1	2	3	4	5	Y_i	$\overline{X_i}$
1	100	57	20	180	38	395	0,956
2	2	2	0	7	1	12	0,48
3	4	12	15	0	2	33	0,303
4	5	0	1	1	4	11	0,229
5	0	3	3	0	2	8	0,07
6	6	0	0	0	12	18	0,327
7	14	4	43	7	10	78	0,527
celkem						555	0,608

P_{ij}	1	2	3	4	5	P_i	$P_i \cdot \left(1 - \frac{P_i}{P}\right)$
1	80	80	83	85	85	413	226,2
2	5	5	5	5	5	25	24,32
3	20	20	20	23	26	109	95,99
4	10	10	9	9	10	48	45,48
5	15	20	25	25	30	115	100,5
6	10	11	11	11	12	55	51,69
7	30	29	29	30	30	148	124
celkem						913	668,2
P*	**19,65**						

X_{ij}	1	2	3	4	5
1	1,25	0,713	0,241	2,118	0,447
2	0,4	0,4	0	1,4	0,2
3	0,2	0,6	0,75	0	0,077
4	0,5	0	0,111	0,111	0,4
5	0	0,15	0,12	0	0,067
6	0,6	0	0	0	1
7	0,467	0,138	1,483	0,233	0,333

	1	2	3	4	5	$\sum_{j=1}^{5} P_{ij}(X_{ij} - \overline{X}_i)^2$
1	6,895	4,76	42,49	114,6	22,05	190,8
2	0,032	0,032	1,152	4,232	0,392	5,84
3	0,211	1,767	4,001	2,108	1,326	9,413
4	0,734	0,525	0,125	0,125	0,292	1,801
5	0,073	0,129	0,064	0,121	0,0003	0,387
6	0,744	1,178	1,178	1,178	5,431	9,709
7	0,109	4,39	26,49	2,588	1,126	34,7
celkem						252,7

	1	2	3	4	5	$\sum_{j=1}^{5} P_{ij}(X_{ij} - \overline{X})^2$
1	32,98	0,876	11,17	193,7	2,1986	241
2	0,216	0,216	1,848	3,137	0,8319	6,249
3	3,327	0,001	0,404	8,499	7,33	19,56
4	0,116	3,695	2,221	2,221	0,4322	8,686
5	5,543	4,193	5,951	9,238	8,7876	33,71
6	0,0006	4,065	4,065	4,065	1,845	14,04
7	0,598	6,405	22,2	4,209	2,2614	35,67
celkem						358,9

Xpr	0,6079	
E(s^2)	9,0237	
D(m)	10,556	0,078
	1,5321	

Z_i	Kred.pojistné
0,781	0,8801
0,178	0,5852
0,485	0,4599
0,293	0,4969
0,498	0,3396
0,322	0,5175
0,561	0,5625