Kredibilitní pojistné v pojištění automobilů

Silvie Zlatošová

11. - 16. září 2016, Robust

Obsah

Obsah

Motivace

- **Pojištění motorových vozidel** je pojištění určené pro osobní automobily, nákladní vozy, motocykly a další silniční vozidla.
- Do této kategorie pojištění řadíme **pojištění odpovědnosti** z provozu motorového vozidla a **havarijní pojištění**.
- Konkrétní podmínky pro pojištění vozidel a jeho typy se liší s právními předpisy daných zemí.

Motivace

- Po roce 2005 začíná v ČR i přes nárůst uzavřených smluv klesat předepsané pojistné v obou skupinách tohoto pojištění - **cenová válka**.
- V důsledku toho se pojištění motorových vozidel stalo dokonce ztrátovým.
- V ČR pojišťovny ve snaze získat nové klienty uplatňují spíše slevy na pojistném za bezeškodní průběh, než penalizace u klientů, kteří nahlásili pojistný nárok.

Motivace

- V aktuárské matematice však existují přístupy, pomocí kterých je možné pojistné určit spravedlivěji a díky kterým je možné se vypořádat se ztrátovostí.
- Důležitou roli při určovaní cen pojistného hraje frekvence pojistných nároků.
- Primárně se pro její modelování využívají **zobecněné lineární modely** (GLM).
- Na základě pozorovatelných rizikových faktorů jsou klienti rozděleni do **tarifních (rizikových) skupin**.
- Všichni pojištění v dané skupině pak platí stejné pojistné.
- Problémem tohoto přístupu je existence skrytých rizikových faktorů.
- V důsledku toho zůstávají jednotlivé skupiny stále dosti heterogenní, ačkoli pracují s mnoha proměnnými.

Cíl

- Cílem práce je pomocí teorie kredibility navrhnout systém hodnocení klientů v pojištění motorových vozidel, který umožní spravedlivější rozdělení pojistného.

Obsah

Apriorní klasifikace

- Výše účtovaného pojistného je obvykle závislá na střední hodnotě pojistného plnění.
- Riziko vzniku pojistné události je v celém kmeni **heterogenní** a proto není možné účtovat všem pojištěným stejné pojistné.
- U každého žadatele o pojištění se posuzuje jeho **rizikovost**. Důležitou roli zde hraje počet pojistných událostí v minulosti.
- Pro pojišťovnu je proto důležitým úkolem modelovat frekvenci pojistných nároků v daném portfoliu pojištěných. Ta je počítána jako počet pojistných nároků na jednu smlouvu za dané období (rok).

Apriorní klasifikace

- Frekvence pojistných nároků závisí na mnoha faktorech.
- Mezi tyto faktory pojišťovny řadí **charakteristiky vozu** (typ vozu, stáří vozu) a také **informace o řidiči** (věk, pohlaví, počet pojistných nároků v minulosti).
- Kombinací těchto rizikových faktorů vznikají **tarifní skupiny**, do kterých jsou řidiči rozděleni.

Využití GLM

- Mezi nejčastěji využívaná rozdělení ve spojení s modelováním počtu dopravních nehod patří Poissonovo rozdělení.
- Pro každého klienta i pojištěného na dobu d_i můžeme předpokládat, že jeho počet pojistných nároků N_i splňuje $N_i \sim Poi(d_i\lambda_i)$, kde parametr λ_i závisí na rizikových faktorech (kovariátách) každého klienta.
- Využijeme-li Poissonovu regresi s logaritmickou link funkcí, pak pro parametr λ_i můžeme psát

$$\lambda_i = \exp(\boldsymbol{x_i^T}\beta),$$

kde $\boldsymbol{x_i}$ značí vektor rizikových faktorů a β neznámý vektor parametrů.

Využití GLM

Označíme-li roční frekvenci pojistných nároků každého klienta jako $\frac{N_i}{d_i}$, pak můžeme psát

$$\frac{N_i}{d_i} \sim Poi(\,\mathrm{e}^{\boldsymbol{X}_i'\boldsymbol{\beta}}).$$

Predikce roční frekvence pojistných nároků

- Použitá data se týkají ročního pojištění automobilů, jsou sesbíraná v letech 2004 a 2005 a je možné je najít v publikaci Hellera a Jonga Generalized Linear Models for Insurance Data.
- V datovém souboru je 67 856 pojistek a u 4 624 z nich (6,8%) byl uplatněn alespoň jeden pojistný nárok.
- Celkový počet pojistných nároků v tomto souboru pojistek je 4 937.

Predikce roční frekvence pojistných nároků

Proměnné v datovém souboru

Označení	Název proměnné	Rozsah
expo	Doba trvání pojistky	0–1
clm	Výskyt dopravní nehody	0 (ne), 1 (ano)
numclaims	Počet dopravních nehod	0, 1, 2, 3, 4
agecat	Věková kategorie pojištěného	1 (nejmladší), 2, 3
area	Oblast bydliště řidiče	A, B, C
gender	Pohlaví	muž, žena
veh_body	Typ automobilu	1 (hatchback), 2(sedan), 3 (kombi), 4 (ostatní)
veh_value	Hodnota automobilu v 1000$	1 ($<$ 7,5), 2 (7,5–25), 3 ($>$ 25)

Na základě uvedených rizikových faktorů a jejich hodnot byli řidiči rozděleni do 216 skupin. Pro každou z těchto skupin je vypočítán *počet dopravních nehod ve skupině* (numclaims) a *doba trvání všech pojistek ve skupině* (expo).

Predikce roční frekvence pojistných nároků

- Budeme se snažit najít nejvhodnější model pro očekávanou frekvenci pojistných nároků, tedy očekávanou hodnotu veličiny $\frac{N_i}{d_i}$.
- Porovnáme modely pracující s různými rizikovými faktory.
- Budeme testovat nulovou hypotézu, že přidání dodatečného rizikového faktoru do uvažovaného modelu nemá žádný efekt.
- Pro posuzování vhodnosti daného submodelu je využita analýza deviace (dev).
- Budeme předpokládat, že změna v deviaci mezi uvažovaným modelem a jeho submodelem (Δdev) má χ^2 rozdělení s Δdf stupni volnosti.

Výběr vhodného modelu

Popis modelu	dev	Δ dev	Δ df	$\chi^2_{0,95}(\Delta\text{df})$
1	367,25			
1+*veh_value*	359,35	7,9029	2	5,99
1+*area*	362,45	4,8041	2	5,99
1+*gender*	367,05	0,2005	1	3,84
1+*agecat*	338,29	28,958	2	5,99
1+*veh_body*	298,95	68,297	3	7,81
1+veh_body+*agecat*	271,15	27,799	2	5,99
1+veh_body+*veh_value*	288,10	10,855	2	5,99
1+veh_body+*area*	294,17	4,784	2	5,99
1+veh_body+*gender*	298,71	0,242	1	3,84
1+agecat+*veh_value*	329,87	8,4175	2	5,99
1+veh_body+agecat+*veh_value*	259,24	11,912	2	5,99
1+veh_body+agecat+veh_value+*area*	257,83	1,417	2	5,99
1+veh_body+agecat+veh_value+*gender*	258,71	0,35	1	3,84

Výběr vhodného modelu

Vyhodnotíme-li tedy analýzu deviace, můžeme model se třemi faktory *veh_body*, *agecat*, *veh_value* označit za nejlepší z uvedených modelů.

Odhad parametrů vybraného modelu

Koeficienty:

(Intercept)	veh_body2	veh_body3	veh_body4
-1.840351	-0.021926	-0.231906	-0.284195
veh_value2	veh_value3	agecat2	agecat3
-0.002991	-0.108345	0.117347	0.264975

Na základě vybraného modelu můžeme řidiče rozdělit do 36 tarifních skupin.

Apriorní rozdělení klientů do rizikových skupin

Riziková skupina	Věková kategorie pojištěného			Hodnota automobilu			Typ automobilu				Očekávaná roční frekvence nehod(%)	Váha skupiny (%)
	1	2	3	1	2	3	1	2	3	4		
1	Ano	Ne	Ne	Ano	Ne	Ne	Ano	Ne	Ne	Ne	15.88	1.35
2	Ano	Ne	Ne	Ano	Ne	Ne	Ne	Ano	Ne	Ne	15.53	0.37
3	Ano	Ne	Ne	Ano	Ne	Ne	Ne	Ne	Ano	Ne	12.59	1.04
4	Ano	Ne	Ne	Ano	Ne	Ne	Ne	Ne	Ne	Ano	11.95	0.23
5	Ano	Ne	Ne	Ne	Ano	Ne	Ano	Ne	Ne	Ne	15.83	6.98
6	Ano	Ne	Ne	Ne	Ano	Ne	Ne	Ano	Ne	Ne	15.49	3.24
7	Ano	Ne	Ne	Ne	Ano	Ne	Ne	Ne	Ano	Ne	12.55	5.49
8	Ano	Ne	Ne	Ne	Ano	Ne	Ne	Ne	Ne	Ano	11.91	3.02
9	Ano	Ne	Ne	Ne	Ne	Ano	Ano	Ne	Ne	Ne	14.25	0.10
10	Ano	Ne	Ne	Ne	Ne	Ano	Ne	Ano	Ne	Ne	13.94	1.62
11	Ano	Ne	Ne	Ne	Ne	Ano	Ne	Ne	Ano	Ne	11.30	0.47
12	Ano	Ne	Ne	Ne	Ne	Ano	Ne	Ne	Ne	Ano	10.72	3.54
13	Ne	Ano	Ne	Ano	Ne	Ne	Ano	Ne	Ne	Ne	17.85	2.34
14	Ne	Ano	Ne	Ano	Ne	Ne	Ne	Ano	Ne	Ne	17.47	0.67
15	Ne	Ano	Ne	Ano	Ne	Ne	Ne	Ne	Ano	Ne	14.16	2.48
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	Ne	Ne	Ano	Ano	Ne	Ne	Ne	Ne	Ne	Ano	15.57	0.29
29	Ne	Ne	Ano	Ne	Ano	Ne	Ano	Ne	Ne	Ne	20.63	5.37
30	Ne	Ne	Ano	Ne	Ano	Ne	Ne	Ano	Ne	Ne	20.18	1.81
31	Ne	Ne	Ano	Ne	Ano	Ne	Ne	Ne	Ano	Ne	16.36	7.94
32	Ne	Ne	Ano	Ne	Ano	Ne	Ne	Ne	Ne	Ano	15.53	2.08
33	Ne	Ne	Ano	Ne	Ne	Ano	Ano	Ne	Ne	Ne	18.57	0.06
34	Ne	Ne	Ano	Ne	Ne	Ano	Ne	Ano	Ne	Ne	18.17	0.67
35	Ne	Ne	Ano	Ne	Ne	Ano	Ne	Ne	Ano	Ne	14.72	0.84
36	Ne	Ne	Ano	Ne	Ne	Ano	Ne	Ne	Ne	Ano	13.97	2.48

Obsah

Teorie kredibility

- Klienti patřící do stejné tarifní skupiny platí stejné základní pojistné.
- Ačkoli může být pro toto dělení klientů využito různé množství *apriorních* proměnných, uvnitř jednotlivých tarifních skupin stále zůstává jistá míra heterogenity.
- Působí zde totiž nepozorovatelné faktory, které je možné modelovat pomocí náhodné veličiny Θ.
- Obvykle bývá Θ voleno tak, aby platilo, že $\mathrm{E}(\Theta) = 1$. Pak zůstane zachována střední hodnota celého portfolia.

Teorie kredibility

Obvykle se pro náhodnou veličinu Θ volí Gamma rozložení $\Gamma(a, a)$ s hustotou pravděpodobnosti

$$u(\theta) = \frac{1}{\Gamma(a)} a^a \theta^{a-1} \exp(-a\theta), \quad \theta \in \mathbb{R}^+, \ a \geq 1,$$

kde $\Gamma(a)$ značí gamma funkci definovanou jako

$$\Gamma(a) = \int_0^\infty t^{a-1} \, \mathrm{e}^{-t} \, \mathrm{d}t.$$

Teorie kredibility

- Hodnota rizikového faktoru Θ pro konkrétního klienta není známá.
- Budeme však předpokládat, že jsme schopni určit pravděpodobnostní funkci, případně pravděpodobnostní hustotu $f_\Theta(\theta)$, která udává pravděpodobnost jednotlivých hodnot rizikového faktoru Θ uvnitř tarifní skupiny.
- Předpokládejme, že pro každého klienta i jsme pozorovali po dobu T_i hodnoty $\boldsymbol{N_i} = \boldsymbol{k_i}$, kde $\boldsymbol{N_i} = (N_{i1}, \ldots, N_{iT_i})^T$ a $\boldsymbol{k_i} = (k_{i1}, \ldots, k_{iT_i})^T$.
- Naším cílem je stanovit pojistnou sazbu tak, aby byly kryty pojistné nároky N_{i,T_i+1} v následujícím pojistném období.
- Nechť za podmínky Θ jsou pojistné nároky $N_{i1}, \ldots, , N_{iT_i}, N_{i,T_i+1}$ nezávislé.

Pojistné v optimální teorii kredibility

Definice

Individuální pojistné $\mu(\theta)$ dané rovnicí

$$\begin{aligned}\mu_{T_i+1}(\theta) &= \mathrm{E}(N_{i,T_i+1}|\Theta=\theta)\\ &= \int_0^\infty k_{i,T_i+1} f_{N_{i,T_i+1}|\Theta}(k_{i,T_i+1}|\theta)\,\mathrm{d}k_{i,T_i+1}.\end{aligned}$$

je pojistné, které bychom měli účtovat v případě, že by byla známa riziková úroveň klienta θ. Individuální pojistné je tedy očekávanou hodnotou celkového pojistného nároku daného klienta pro příští období při daném θ.

Protože rizikový faktor Θ je v praxi neznámý, individuální pojistné nemůže být nikdy určeno přesně. Je nutné ho odhadovat z dat.

Pojistné v optimální teorii kredibility

Definice

Kolektivní pojistné μ_{T_i+1} dané rovnicí

$$\mu_{T_i+1} = \mathrm{E}(N_{i,T_i+1}) = \mathrm{E}[\mathrm{E}(N_{i,T_i+1}|\Theta)] = \mathrm{E}[\mu_{T_i+1}(\Theta)],$$

je takové pojistné, které bude účtováno klientovi v případě, že nemáme žádnou informaci o jeho úrovni rizika. Kolektivní pojistné je tedy průměrnou hodnotou všech možných individuálních pojistných.

Kolektivní pojistné nezávisí na θ daného klienta ani na pozorovaných hodnotách $\boldsymbol{k_i}$.

Pojistné v optimální teorii kredibility

Definice

Předpokládejme, že $N_{i1}, \ldots, N_{iT_i}$ označují pojistné nároky za T_i po sobě jdoucích pojistných období. Pak Bayesovské pojistné $B(N_{i1}, \ldots, N_{iT_i})$ je dáno vztahem

$$B(N_{i1}, \ldots, N_{iT_i}) = \arg\min_{\psi(\cdot)} \mathrm{E}\left[\left(\mu_{T_i+1}(\Theta) - \psi(N_{i1}, \ldots, N_{iT_i})\right)^2\right],$$

kde $\psi(\cdot)$ je funkcí $N_{i1}, \ldots, N_{iT_i}$.

Dá se ukázat, že pro argument minima $\psi(\cdot)$ platí

$$\psi(N_{i1}, \ldots, N_{iT_i}) = \mathrm{E}[\mu_{T_i+1}(\Theta) | N_{i1}, \ldots, N_{iT_i}],$$

tedy

$$B(N_{i1}, \ldots, N_{iT_i}) = \mathrm{E}[\mu_{T_i+1}(\Theta) | N_{i1}, \ldots, N_{iT_i}].$$

Pojistné v optimální teorii kredibility

Protože předpokládáme nezávislost N_{it} pro $t \in \{1, 2, \ldots, T_i\}$ při podmíněnosti Θ, platí

$$\mu_{T_i+1}(\Theta) = \mathrm{E}(N_{i,T_i+1}|\Theta) = \mathrm{E}(N_{i,T_i+1}|\Theta, N_{i1}, N_{i2}, \ldots, N_{iT_i}).$$

Na základě tohoto vyjádření pak dostáváme pro Bayesovské pojistné

$$\begin{aligned} B(N_{i1}, \ldots, N_{iT_i}) &= \mathrm{E}\left[\mathrm{E}(N_{i,T_i+1}|\Theta, N_{i1}, \ldots, N_{iT_i})|N_{i1}, \ldots, N_{iT_i}\right] \\ &= \mathrm{E}(N_{i,T_i+1}|N_{i1}, \ldots, N_{iT_i}). \end{aligned}$$

Poissonův kredibilitní model

Definice

Nechť Θ_i je náhodná veličina nabývající kladných hodnot, pro kterou platí $\mathrm{E}(\Theta_i) = 1$. Pak za Poissonův kredibilitní model označíme takový model, kde v portfoliu n klientů je každý klient reprezentován dvojicí $(\Theta_i, \mathbf{N_i})$ a navíc platí:

1. *Pro dané $\Theta_i = \theta$ jsou náhodné veličiny N_{it}, kde $t = 1, 2, \dots, T_i$, vzájemně nezávislé a řídí se Poissonovým rozdělením s parametrem $\lambda_{it}\theta$.*
2. *Pro $i, j = 1, 2, \dots, n$ a $i \neq j$ jsou Θ_i, Θ_j nezávislé a dále také N_{it}, N_{jt} jsou nezávislé.*

Bayesovské pojistné v Poissonově kredibilitním modelu

Bayesovské kredibilitní pojistné hledáme ve tvaru

$$\mathrm{E}(N_{i,T_i+1}|N_{i1} = k_{i1}, N_{i2} = k_{i2}, \ldots, N_{iT_i} = k_{iT_i}).$$

Předpokládáme-li platnost Poissonova modelu, pak

$$\begin{aligned}\mathrm{E}(N_{i,T_i+1}|N_{i1} = k_{i1}, \ldots, N_{iT_i} = k_{iT_i}) &= \int_0^\infty \lambda_{i,T_i+1}\theta \,\mathrm{d}F_{\Theta|N_{i\bullet}}(\theta|k_\bullet) \\ &= \lambda_{i,T_i+1}\,\mathrm{E}(\Theta_i|N_{i1} = k_{i1}, \ldots, N_{iT_i} = k_{iT_i}).\end{aligned}$$

Poisson-Gamma kredibilitní model

Předpokládejme nyní, že pro rizikový faktor Θ_i platí $\Theta_i \sim \Gamma(a, a)$. Dá se ukázat, že **prediktivní distribuce** Θ_i za podmínky $N_{i1} = k_{i1}, N_{i2} = k_{i2} \dots, N_{iT_i} = k_{iT_i}$ je tvaru

$$\frac{e^{-\theta(a+\sum_{t=1}^{T_i} \lambda_{it})} \theta^{a+\sum_{t=1}^{T_i} k_{it}-1}}{\int_0^\infty \mathrm{e}^{-\xi(a+\sum_{t=1}^{T_i} \lambda_{it})} \xi^{a+\sum_{t=1}^{T_i} k_{it}-1} \, \mathrm{d}\xi}$$

$$= \quad \mathrm{e}^{-\theta(a+\sum_{t=1}^{T_i} \lambda_{it})} \theta^{a+\sum_{t=1}^{T_i} k_{it}-1} \frac{\left(a + \sum_{t=1}^{T_i} \lambda_{it}\right)^{a+\sum_{t=1}^{T_i} k_{it}}}{\Gamma\left(a + \sum_{t=1}^{T_i} k_{it}\right)},$$

což je opět tvar **Gamma rozdělení**, tentokrát s parametry $a + N_{i\bullet}$ a $a + \lambda_{i\bullet}$. Tedy

$$(\Theta_i | N_{i1}, N_{i2}, \dots, N_{iT_i}) \sim \Gamma(a + N_{i\bullet}, a + \lambda_{i\bullet}).$$

Bayesovské pojistné v Poisson-Gamma kredibilitním modelu

Kdybychom neměli **žádnou informaci** o klientovi, účtovali bychom mu **kolektivní pojistné.** Za předpokladu Poissonova rozdělení veličiny N_{i,T_i+1} by kolektivní pojistné bylo

$$E(N_{i,T_i+1}) = \lambda_{i,T_i+1}.$$

Jelikož máme k dispozici **data o minulých pojistných nárocích** i-tého klienta, je možné provést ***aposteriorní* korekci** kolektivního pojistného. **Bayesovské kredibilitní pojistné** je tvaru

$$\begin{aligned}
\mathrm{E} \quad & (N_{i,T_i+1} | N_{i1} = k_{i1}, \ldots, N_{iT_i} = k_{iT_i}) \\
= \quad & \lambda_{i,T_i+1} \, \mathrm{E}(\Theta_i | N_{i1} = k_{i1}, N_{i2} = k_{i2}, \ldots, N_{iT_i} = k_{iT_i}) \\
= \quad & \lambda_{i,T_i+1} \frac{a + k_{i\bullet}}{(a + \lambda_{i\bullet})}.
\end{aligned}$$

Obsah

Výpočet aposteriorních korekcí

- Nyní určíme *aposteriorní* korekce $\mathrm{E}(\Theta_i|N_{i\bullet} = k_{i\bullet})$ základního pojistného na základě počtu pojistných nároků nahlášených klientem v minulosti.
- Výpočet provedeme pro různé kombinace délky pojistných období T_i a počtu nahlášených pojistných nároků $k_{i\bullet}$.
- Vybrali jsme dobré řidiče s *apriorní* očekávanou roční frekvencí pojistných nároků $\lambda_i = 0,1072$, průměrné řidiče s *apriorní* očekávanou roční frekvencí pojistných nároků $\lambda_i = 0,1588$ a klienty s *apriorní* očekávanou roční frekvencí pojistných nároků $\lambda_i = 0,2069$.

Korekce pro dobré řidiče

T_i	Počet pojistných nároků $k_{i\bullet}$					
	0	**1**	**2**	**3**	**4**	**5**
1	90,48 %	179,31 %	268,14 %	356,98 %	445,81 %	534,65 %
2	82,61 %	163,72 %	244,83 %	325,94 %	407,05 %	488,16 %
3	76,00 %	150,62 %	225,24 %	299,87 %	374,49 %	449,11 %
4	70,37 %	139,47 %	208,56 %	277,65 %	346,75 %	415,84 %
5	65,52 %	129,85 %	194,18 %	258,51 %	322,84 %	387,17 %
6	61,29 %	121,47 %	181,65 %	241,83 %	302,01 %	362,19 %
7	57,58 %	114,11 %	170,64 %	227,18 %	283,71 %	340,24 %
8	54,29 %	107,59 %	160,89 %	214,19 %	267,50 %	320,80 %
9	51,35 %	101,77 %	152,20 %	202,62 %	253,04 %	303,46 %
10	48,72 %	96,56 %	144,39 %	192,23 %	240,06 %	287,90 %

Srovnání korekcí

$T_i = 1$

Řidič	Počet pojistných nároků $k_{i\bullet}$					
	0	**1**	**2**	**3**	**4**	**5**
dobrý	90,48 %	179,31 %	268,14 %	356,98 %	445,81 %	534,65 %
průměrný	86,51 %	171,45 %	256,39 %	341,33 %	426,27 %	511,21 %
špatný	83,12 %	164,72 %	246,33 %	327,93 %	409,54 %	491,15 %

$T_i = 10$

Řidič	Počet pojistných nároků $k_{i\bullet}$					
	0	**1**	**2**	**3**	**4**	**5**
dobrý	48,72 %	96,56 %	144,39 %	192,23 %	240,06 %	287,90 %
průměrný	39,08 %	77,44 %	115,81 %	154,17 %	192,54 %	230,90 %
špatný	32,99 %	65,38 %	97,77 %	130,15 %	162,54 %	194,93 %

Reference

- DENUIT Michel, et al.:
  Actuarial modelling of claim counts: risk classification, credibility and bonus-malus systems
  Hoboken: Wiley. 2007.
- KAAS Rob:
  Modern Actuarial Risk Theory: Using R.
  Heidelberg: Springer. 2009.
- DOBSON Annette J.:
  An Introduction to Generalized Linear Models
  Boca Raton: CRC Press. 2002.
- HELLER Gillian Z., JONG Piet:
  Generalized Linear Models for Insurance Data
  New York: Cambridge University Press. 2008.

Děkuji za pozornost!