Premija povjerenja

Škobo, Nikolina

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:080471

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: 2024-03-03

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Nikolina Škobo

PREMIJA POVJERENJA

Diplomski rad

Voditelj rada:
prof.dr.sc. Miljenko Huzak

Zagreb, ožujak, 2023

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Prekrasnim ljudima u mom životu na kojima sam beskrajno zahvalna. Vi ste moja radost i snaga. Hvala na svemu.

Sadržaj

Sadržaj **iv**

Uvod **2**

1 Teorija povjerenja **3**
1.1 Individualna i kolektivna premija 3
1.2 Bayesov pristup 5
1.3 Jednostavni model povjerenja 12
1.4 Jednostavni Bühlmannov model 16
1.5 Bühlmann-Straubov Model 18

2 Regresijska stabla odluke **31**
2.1 Stabla odluke 31
2.2 Konstrukcija stabla odluke 32

3 Model povjerenja **36**
3.1 Funkcije gubitka kod particioniranja 37
3.2 Regresijsko stablo povjerenja 40

Bibliografija **47**

Uvod

Život je pun nepredvidljih okolnosti i često nas iznenadi, nekada pozitivno, a nekada, nažalost, i negativno. Prije par godina svjedočili smo razornim potresima u Hrvatskoj koji su uništili domove mnogim obiteljima. U takvim situacijama, na koje ne možemo utjecati, niti se zaštiti u potpunosti od svih rizika, osiguranje nam uvelike može olakšati financijsku situaciju u kojoj smo se našli.

Osiguranje je definirano kao ugovor između osiguravajućeg društva i osiguranika ili grupe osiguranika koji osiguraniku osigurava financijsku zaštitu ili naknadu štete od strane osiguratelja u slučaju štete koju pokriva taj ugovor, koji se naziva polica osiguranja. Da bi osigurateljsko društvo prihvatilo taj rizik, osiguranik mora platiti premiju. Aktuari su ti koji moraju odlučiti kolika će ta premija biti, odnosno koliko "vjeruju" klijentu što ovisi o njegovim karakteristikama i prijašnjem iskustvu. Teorija povjerenja je jedna od ključnih dijelova aktuarske znanosti i ona se bavi određivanjem cijene proizvoda koje osiguravajuće društvo nudi, odnosno određivanjem što točnije premije za rizik. Ideja na kojoj je bazirano osiguranje je svrstati individualne osiguranike u kategorije, ovisno o izloženosti riziku, da bi osiguranici u istoj kategoriji tvorili što homogeniju grupu. Uzmimo na primjer, kolektiv koji je homogen u odnosu na faktor rizika kao što je tip vozila i mjesto rezidentnosti osiguranika. To su faktori rizika koje možemo opažati, no problem su oni faktori rizika koje ne možemo opažati, kao što su karakter ili vještina vozača, koji znatno utječu na percepciju osiguranika kao rizičnog ili nerizičnog. Nakon određenog vremena, imamo puno više informacija o osiguraniku, kao što su broj šteta i veličina šteta koje nam mogu dati uvid u faktore rizičnosti koje nismo mogli opažati prije nego li je osiguranik sklopio policu osiguranja i na osnovu iskustva s osiguranikom možemo ažurirati buduće premije. Ovaj primjer upućuje da klasična teorija povjerenja počiva na bayesovskoj statističkoj teoriji, čija je osnovna ideja da su parametri statističkih modela slučajni i da kao takvi moraju biti modelirani. Prije nego li imamo ikakva saznanja o nepoznatom parametru, potrebna nam je inicijalna ideja o njegovoj distribuciji, tzv. apriorna razdioba, koja nam služi da model uzme u obzir prethodno ili ekspertno znanje. Okosnica bayesovske teorije je Bayesov teorem pomoću kojega ažuriramo uvjerenja koja imamo o distribuciji parametra pomoću novih informacija kojima raspolažemo. Naravno, cilj nam je da osiguravajuće društvo bude što konkuretnije na tržištu što postižemo određivanjem što realnije premije. Prilagođavanje premije bazirano na iskustvu nije jednostavan zadatak. Ukoliko je povjerenje koje osiguravatelj pruža kolektivu visoko, osiguravatelj je u velikom riziku da izgubi velike količine novaca ukoliko se dogodi veći broj šteta, ali također, problem je i ako pretjerano uzimamo u obzir prijašnje iskustvo i često mijenjamo premiju jer je to za osiguravatelja skupo i neisplativo. Pitanja koja si postavljamo su koliko je zapravo bitno prijašnje iskustvo, koliko toga prijašnjeg iskustva zapravo ovisi o individualnom riziku pojedinca (ili grupe osiguranika ukoliko je sklopljeno grupno osiguranje), a koliko je različitost u iskustvu slučajna. Također, zanima nas i u kojoj bismo mjeri trebali uzimati prijašnje iskustvo u računanju buduće premije. U ovom radu ćemo pokušati odgovoriti na sva ova pitanja.

U prvom poglavlju ćemo dati teorijski pregled teorije povjerenja i matematičku osnovu za određivanje premije povjerenja. Objasnit ćemo pojmove poput individualne, kolektivne i Bayesove premije, ali i pojam funkcije gubitka i Bayesovog procjenitelja. Promotrit ćemo različite modele teorije povjerenja, od jednostavnih do kompleksnijih modela, od kojih ćemo se najduže zadržati na Bülhmann-Staubovom modelu koji je najkompleksniji i najkorišteniji model za određivanje premije povjerenja.

U drugom ćemo poglavlju detaljno proučiti metodu statističkog učenja, stabla odluke, pomoću koje ćemo objasniti relativno novi model za procjenu premije povjerenja naziva Regresijsko stablo povjerenja. Model u procjenu premije povjerenja uključuje informacije o kovarijatama.

Spomenuti model ćemo objasniti u trećem poglavlju.

Poglavlje 1

Teorija povjerenja

U ovom poglavlju uvodimo teoriju povjerenja kao matematički alat pomoću kojega možemo opisati heterogene kolektive. Poglavlje sadrži definicije i teoreme potrebne za razumijevanje teorije povjerenja i napisano je po uzoru na [2] i [6].

1.1 Individualna i kolektivna premija

Cilj teorije povjerenja je odgovoriti na pitanje kako osiguravatelj treba obračunavati premiju uzimajući u obzir individualno i kolektivno iskustvo, odnosno individualni i kolektivni rizik.

Individualni rizik i premija

Označimo s Y_j , $j = 1, 2, ..., n$, ukupni iznos šteta pojedinca u razdoblju j. Varijable Y_j, za $j = 1, 2, ..., n$, su slučajne i na temelju opservacija o štetama iz prethodnih razdoblja, $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$, želimo odrediti premiju za očekivani ukupni iznos šteta u idućem razdoblju, Y_{n+1}. Za to su nam potrebne pretpostavke o funkciji distribucije $F(y)$ slučajne varijable Y_j:

P1: slučajne varijable Y_j su jednako distribuirane s funkcijom distribucije $F(y)$

P2: slučajne varijable Y_j su nezavisne (uz danu funkciju distribucije $F(y)$).

U praksi, funkcija distribucije F je nepoznata i varira među rizicima. Formalno, neka je ϑ nepoznati parametar iz apstraktnog parametarskog prostora rizika Θ. Parametar ϑ nazivamo profil rizika i s njime parametriziramo funkciju distribucije F i pišemo F_ϑ. Smatramo da je ϑ nepoznat i da varira među rizicima.

Postoje različiti načini za računanje premije, ali svi oni podrazumijevaju funkciju koja slučajnoj varijabli Y, koja predstavlja slučajnu isplatu osiguraniku od strane osiguravatelja, s distribucijom F pridružuju realan broj, odnosno količinu novca za koju je osiguravatelj spreman osigurati dani rizik, formalno

$$Y \to \mathcal{P}(Y),$$

odnosno

$$F_\vartheta \to \mathcal{P}(F_\vartheta),$$

ako uzmemo u obzir da vrijednost funkcije $\mathcal{P}$ ovisi samo o F_ϑ za profil rizika ϑ. Sada možemo definirati točnu individualnu premiju.

Definicija 1.1.1. *Točna individualna premija rizika s profilom rizika ϑ se definira kao*

$$P^{ind}(\vartheta) = \mathcal{P}[Y_{n+1}] = \mathbb{E}[Y_{n+1}|\vartheta] =: \mu(\vartheta)$$

Da bismo odredili individualnu premiju, potrebno je odrediti veličinu $\mu(\vartheta)$, međutim, obje varijable, i ϑ i $\mu(\vartheta)$ su nepoznate i zbog toga moramo tražiti procjenitelja $\widehat{\mu(\vartheta)}$ za $\mu(\vartheta)$.

Premija i rizik kolektiva

U slučaju da se premija računa uzimajući u obzir iskustvo nastalih šteta heterogenog porfelja, osiguravatelj je izložen većem riziku i postoji veća vjerojatnost da uplaćene premije neće biti dostatne u slučaju nastanka većeg broja šteta ili šteta s većim iznosom. Kako bi osiguravatelj umanjio ovaj rizik, police se grupiraju u kolektive po "sličnim" faktorima rizika, i za svaki se kolektiv posebno računaju premije i stope učestalosti šteta. Da bi se što pravilnije utvrdila premija, rizici kolektiva bi trebali biti što sličniji po nekim osnovnim karakteristikama, no ti rizici će ipak biti različiti jer postoje karakteristike rizika koje nisu mjerljive i zbog toga kolektivi nisu homogeni. Neka je Θ skup svih potencijalnih profila rizika u kolektivu koji se sastoji od individualnih profila rizika ϑ_i. Ukoliko je kolektiv homogen, skup Θ se sastoji od samo jednog elementa. Međutim, u praksi to nije slučaj. Prave vrijednosti ϑ_i nisu poznate osiguravatelju, ali uz prethodno iskustvo i statističke informacije kojima raspolaže, ima uvid u strukturu kolektiva, odnosno vjerojatnosnu funkciju distribucije $U(\vartheta)$ na skupu Θ koju nazivamo **strukturna funkcija kolektiva**.

Sad možemo definirati i premiju kolektiva za koju nam je potrebna strukturna funkcija kolektiva.

Definicija 1.1.2. *Premija kolektiva se definira kao*

$$P^{kol} = \int_\Theta \mu(\vartheta)\, dU(\vartheta) =: \mu_0$$

Iz definicije premije kolektiva možemo vidjeti da je ona zapravo prosjek očekivanog iznosa šteta individualnih rizika za cijeli kolektiv. Iako je ova vrijednost nepoznata osiguravatelju, može je procijeniti pomoću podataka iz prethodnih razdoblja.
Postavlja se pitanje zašto trebamo individualnu premiju ukoliko možemo procijeniti premiju kolektiva, no odgovor je vrlo jednostavan: zbog konkurentnosti na tržištu koja se postiže što boljim procjenjivanjem točne individualne premije. Uzmimo kao primjer dva osiguravatelja A i B. Pretpostavimo da osiguravatelj A uzima u obzir individualno iskustvo i smanjuje premiju za police po kojima u prethodnoj godini nije bilo šteta, dok povećava premiju ukoliko su se štete dogodile. Osiguravatelj B, za razliku od osiguravatelja A, za sve police ima istu premiju, temeljenu na strukturi kolektiva. Posljedično tome, osiguravatelj A će zadržati dobre rizike, a vjerojatno i privući dobre rizike iz portfelja osiguravatelja B, dok će osiguravatelj B privući loše rizike u svoj portfelj i time biti nekonkuretniji na tržištu.

1.2 Bayesov pristup

Okosnica teorije povjerenja je Bayesova statistika i koristeći tu granu statistike, matematičkim jezikom možemo opisati problem određivanja premije.

Da bismo shvatili koncept, promotrimo najprije tzv. **model dvije urne**. Prva urna predstavlja kolektiv s funkcijom distribucije U. Iz nje odabiremo individualni rizik, odnosno njegov profil rizika ϑ. Taj parametar ϑ određuje sadržaj druge urne, odnosno funkciju distribucije F_ϑ. Sada se iz druge urne odabiru vrijednosti slučajnih varijabli $Y_1, Y_2, ...$, koje su nezavisne i jednako distribuirane s funkcijom distribucije F_ϑ. Dakle, svaki rizik je određen svojim individualnim profilom rizika ϑ koji je realizacija slučajne varijable θ i vrijedi sljedeće:

- uvjetno na događaj $\theta = \vartheta$, $Y_1, Y_2, ...$ su nezavisne jednako distribuirane slučajne varijable s funkcijom distribucije F_ϑ
- θ je slučajna varijabla s funkcijom distribucije U.

Sada individualnu premiju možemo promatrati kao slučajnu varijablu čija točna vrijednost nije poznata, ali ipak imamo neke informacije o mogućim vrijednostima $\mu(\theta)$ i o vjerojatnostima pojavljivanja tih vrijednosti pa promatramo i $\mu(\theta)$ kao slučajnu varijablu. Zato je individualna premija u ovoj interpretaciji zapravo uvjetno očekivanje

$$\mu(\vartheta) = \mathbb{E}[Y_{n+1}|\theta = \vartheta].$$

Primijetimo da su a priori svi rizici jednaki. Zaključci o pojedinom riziku mogu se donijeti a posteriori, dakle tek nakon promatranja individualnih rizika.

Kolektivna premija u ovom slučaju je

$$P^{kol} = \mu_0 = \int_{\Theta} \mu(\vartheta)\, dU(\vartheta) = \mathbb{E}[Y_{n+1}]$$

bezuvjetno očekivanje i stoga fiksna vrijednost.

Slučajne varijable $Y_1, Y_2, ...$ su uvjetno nezavisne, uz dani θ. Bezuvjetno, one su pozitivno korelirane što vidimo iz

$$\begin{aligned} Cov(Y_1, Y_2) &= \mathbb{E}[Cov(Y_1, Y_2|\theta)] + Cov(\mathbb{E}[Y_1|\theta], \mathbb{E}[Y_2|\theta]) \\ &= Cov(\mu(\theta), \mu(\theta)) \\ &= \text{Var}[\mu(\theta)] > 0, \end{aligned}$$

zbog čega smo u pretpostavkama o funkciji distribucije F 1.1 pretpostavili uvjetnu nezavisnost slučajnih varijabli $Y_1, Y_2, ...$

Cilj teorije povjerenja je što preciznije procijeniti točnu individualnu premiju $\mu(\theta)$ za svaki rizik. Potencijalni procjenitelj je kolektivna premija μ_0 koja se, u tom slučaju, procjenjuje srednjom vrijednosti očekivanih individualnih rizika za cijeli kolektiv. Taj procjenitelj je prikladan za novi rizik za koji ne raspolažemo informacijama o štetama iz prethodnih razdoblja, no on ne uzima u obzir individualno iskustvo. Zato definiramo Bayesovu premiju koja je temeljena na individualnom iskustvu.

Definicija 1.2.1. *Bayesova premija se definira kao*

$$P^{Bayes} = \tilde{\mu}(\theta) := \mathbb{E}[\mu(\theta)|\boldsymbol{Y}].$$

Bayesova premija

U ovom ćemo potpoglavlju detaljnije proučiti Bayesovu premiju i kako do nje dolazimo koristeći osnovne koncepte Bayesove statistike. Neka je $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ slučajni uzorak s funkcijom distribucije

$$F_\vartheta(\mathbf{y}) = \mathbb{P}_\vartheta[\mathbf{Y} \leq \mathbf{y}]$$

uz vjerojatnost $\mathbb{P}_\vartheta$. Ta funkcija distribucije, odnosno profil rizika ϑ s kojim je parametrizirana, nije poznat. Želimo odrediti što precizniju vrijednost prave premije $\mu(\vartheta)$ uz profil rizika ϑ, što postižemo pronalaskom najboljeg procjenitelja T od $\mu(\vartheta)$. Procjenitelj T je funkcija

$$T : \mathbb{R}^n \to \{\mu(\vartheta) : \vartheta \in \Theta\},$$

gdje je Θ parametarski skup koji sadrži pravu vrijednost od ϑ. Neka je $\mathbf{y} = (y_1, y_2, ...)$ realizacija slučajnog uzorka $\mathbf{Y}$. Tada je $T(\mathbf{y})$ procjena od $\mu(\vartheta)$ na osnovi opaženog uzorka $\mathbf{y}$. Sada definiramo funkciju gubitka $L(\vartheta, T(\mathbf{y}))$ kao

$$L(\vartheta, T(\mathbf{y})) = (\mu(\vartheta) - T(\mathbf{y}))^2.$$

Funkcija gubitka $L(\vartheta, T(\mathbf{y}))$ postiže minimalnu vrijednost kad je vrijednost parametara ϑ pravi profil rizika. Pomoću funkcije gubitka definiramo funkciju rizika za procjenitelj T

$$R_T(\vartheta) := \mathbb{E}_\vartheta[L(\vartheta, T(\mathbf{Y}))] = \int_{\mathbb{R}^n} L(\vartheta, T(\mathbf{y}))dF_\vartheta(\mathbf{y}) = \int_{\mathbb{R}^n} (\mu(\vartheta) - T(\mathbf{y}))^2 dF_\vartheta(\mathbf{y}). \tag{1.1}$$

Cilj nam je pronaći procjenitelj $T \in D$, gdje je D skup svih funkcija kojemu pripada procjenitelj, takav da funkcija rizika $R_T(\vartheta)$ poprima najmanju vrijednost. U Bayesovoj statistici, promatramo očekivanu vrijednost od $R_T(\vartheta)$, promatrajući ϑ kao realizaciju slučajne varijable θ s vjerojatnosnom distribucijom U. Kako smo prethodno specificirali kvadratni oblik funkcije gubitka, uvedimo i sljedeću definiciju.

Definicija 1.2.2. *Procjenitelj T je barem jednako dobar kao procjenitelj T* ukoliko vrijedi*

$$\mathbb{E}[(T(\boldsymbol{Y}) - \mu(\theta))^2] \leq \mathbb{E}[(T^*(\boldsymbol{Y}) - \mu(\theta))^2].$$

$\mathbb{E}[(T(\boldsymbol{Y}) - \mu(\theta))^2]$ *se naziva kvadratni gubitak procjenitelja T.*

Definirajmo i Bayesov rizik procjenitelja T.

Definicija 1.2.3. *Bayesov rizik procjenitelja T uz a priori distribuciju $U(\vartheta)$ definiramo kao*

$$R(T) := \int_\Theta R_T(\vartheta)dU(\vartheta).$$

Napomena 1.2.4. *Uz pretpostavku da je gornji integral dobro definiran, procjenitelje uvijek možemo rangirati prema rastućem riziku, tj. skup svih procjenitelja $T \in D$ je uređen skup.*

Zbog gornje napomene, možemo definirati Bayesov procjenitelj $\tilde{T}$ kao:

Definicija 1.2.5. *Bayesov procjenitelj $\tilde{T}$ definira se kao:*

$$\tilde{T} := \underset{T \in D_1}{\operatorname{argmin}} R(T),$$

gdje je D_1 skup svih procjenitelja s integrabilnom funkcijom rizika. Bayesov procjenitelj $\tilde{T}$ je procjenitelj koji minimizira Bayesov rizik $R(\cdot)$.

Uvrstimo li 1.1 u 1.2.3 dobijemo da je

$$
\begin{aligned}
R(T) &= \int_{\Theta} R_T(\vartheta) dU(\vartheta) \\
&= \int_{\Theta} \mathbb{E}_{\vartheta}[L(\vartheta, T(\mathbf{Y}))] dU(\vartheta) \\
&= \int_{\Theta} \int_{\mathbb{R}^n} L(\vartheta, T(\mathbf{y})) dF_{\vartheta}(\mathbf{y}) dU(\vartheta) \\
&= \int_{\mathbb{R}^n} \int_{\Theta} L(\vartheta, T(\mathbf{y})) dU_{\mathbf{y}}(\vartheta) dF(\mathbf{y}) \\
&= \int_{\mathbb{R}^n} \int_{\Theta} (\mu(\vartheta) - T(\mathbf{y}))^2 dU_{\mathbf{y}}(\vartheta) dF(\mathbf{y})
\end{aligned}
$$

što nam služi za teorem pomoću kojega konstruiramo Bayesov procjenitelj.

Napomena 1.2.6. *S U_y označavamo uvjetnu distribuciju od θ uz dano $\boldsymbol{Y} = \boldsymbol{y}$.*

Teorem 1.2.7. *Za svaku realizaciju $\mathbf{y}$ slučajnog uzorka $\boldsymbol{Y}$, $\tilde{T}(\mathbf{y})$ je Bayesova procjena u odnosu na distribuciju $U_{\mathbf{y}}(\vartheta)$.*

Napomena 1.2.8. *Funkcija distribucije $U(\vartheta)$ od ϑ se naziva a priori distribucija (strukturna distribucija na skupu Θ), dok se funkcija distribucije $U_{\mathbf{y}}(\vartheta)$ naziva a posteriori distribucija od ϑ.*

Sljedeći teorem nam daje oblik Bayesovog procjenitelja u odnosu na kvadratnu funkciju gubitka i objašnjava zašto je Bayesova premija najbolja moguća premija na temelju iskustva.

Teorem 1.2.9. *Bayesov procjenitelj u odnosu na kvadratnu funkciju gubitka je*

$$\tilde{\mu}(\Theta) = \mathbb{E}[\mu(\theta)|\boldsymbol{Y}].$$

Dokaz. Neka je $\widehat{\mu(\theta)}$ proizvoljan procjenitelj od $\mu(\theta)$ i neka je $\tilde{\mu}(\theta)$ a posteriori očekivanje $\mathbb{E}[\mu(\theta)|\mathbf{Y}]$. Da bismo pokazali tvrdnju, dovoljno je pokazati da vrijedi nejednakost

$$\mathbb{E}[(\tilde{\mu}(\theta) - \mu(\theta))^2] \leq \mathbb{E}[(\widehat{\mu(\theta)} - \mu(\theta))^2].$$

Računamo

$$\begin{aligned}
\mathbb{E}[(\widehat{\mu(\theta)} - \mu(\theta))^2] &= \mathbb{E}[\mathbb{E}[(\widehat{\mu(\theta)} - \tilde{\mu}(\theta) + \tilde{\mu}(\theta) - \mu(\theta))^2|\mathbf{Y}]] \\
&= \mathbb{E}[\mathbb{E}[(\widehat{\mu(\theta)} - \tilde{\mu}(\theta))^2|\mathbf{Y}]] + \mathbb{E}[\mathbb{E}[(\tilde{\mu}(\theta) - \mu(\theta))^2|\mathbf{Y}]] \\
&+ 2\mathbb{E}[\mathbb{E}[(\widehat{\mu(\theta)} - \tilde{\mu}(\theta))(\tilde{\mu}(\theta) - \mu(\theta))|\mathbf{Y}]] \\
&= \mathbb{E}[(\widehat{\mu(\theta)} - \tilde{\mu}(\theta))^2] + \mathbb{E}[(\tilde{\mu}(\theta) - \mu(\theta))^2] \\
&+ 2\mathbb{E}[(\widehat{\mu(\theta)} - \tilde{\mu}(\theta))\mathbb{E}[(\tilde{\mu}(\theta) - \mu(\theta))|\mathbf{Y}] \\
&= \mathbb{E}[(\widehat{\mu(\theta)} - \tilde{\mu}(\theta))^2] + \mathbb{E}[(\tilde{\mu}(\theta) - \mu(\theta))^2] \\
&\geq \mathbb{E}[(\tilde{\mu}(\theta) - \mu(\theta))^2]
\end{aligned}$$

što smo trebali i pokazati. □

Napomena 1.2.10. *U prethodnom dokazu, u četvrtoj jednakosti, koristili smo izmjerivost slučajne varijable $\tilde{\mu}(\theta)$ u odnosu na $\boldsymbol{Y}$.*

Da bismo mogli odrediti Bayesovu premiju, potrebno je prethodno specificirati strukturnu funkciju $U(\vartheta)$ i familiju uvjetnih distribucija $\mathcal{F} := \{F_\vartheta(y) : \vartheta \in \Theta\}$, što je u praksi gotovo nemoguće. U nastavku ćemo promotriti jedan od specijalnih slučajeva kada je Bayesovu premiju moguće prikazati u analitičkom obliku, Normalni-Normalni model.

Normalni-normalni model Neka je $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)'$ slučajni uzorak veličine n, gdje je Y_j ukupni iznos šteta u j-toj godini. Želimo procijeniti očekivani ukupni iznos šteta za iduću godinu, Y_{n+1}. Pretpostavke modela su sljedeće:

1. distribucija od Y_j ovisi o fiksnoj, ali nepoznatoj distribuciji parametra ϑ
2. uvjetna distribucija od Y_j uz dani ϑ je $\mathcal{N}(\vartheta, \sigma_1^2)$
3. vrijednost ϑ je nepoznata i tretiramo je kao slučajnu varijablu θ
4. a priori distribucija od θ je normalna, $\theta \sim \mathcal{N}(\mu_0, \sigma_2^2)$
5. vrijednosti μ_0, σ_1 i σ_2 su poznate
6. pretpostavljamo da na raspolaganju imamo realizaciju $\mathbf{y}$ slučajnog vektora $\mathbf{Y}$ za n prethodnih godina.

Iz druge pretpostavke modela slijedi da je

$$P^{ind} = \mathbb{E}[Y_{n+1}|\theta] = \theta,$$

a iz četvrte pretpostavke slijedi da je

$$P^{kol} = \mathbb{E}[\theta] = \mu_0.$$

Želimo odrediti Bayesovu premiju

$$P^{Bayes} = \mathbb{E}[\mu(\theta)|\mathbf{Y} = \mathbf{y}] = \mathbb{E}[\mathbb{E}[Y_{n+1}|\theta]|\mathbf{Y} = \mathbf{y}] = \mathbb{E}[\theta|\mathbf{Y} = \mathbf{y}]$$

uz danu realizaciju $\mathbf{y}$ slučajnog uzorka $\mathbf{Y}$, pretpostavljajući kvadratni gubitak. Dakle, ono što tražimo je a posteriori srednju vrijednost od θ uz dani $\mathbf{y}$, za što nam je potrebno poznavati a posteriori funkciju distribucije $f_{\theta|\mathbf{Y}}(\vartheta|\mathbf{y})$ koju ćemo odrediti koristeći Bayesov teorem čiji je iskaz u nastavku.

Teorem 1.2.11. *Neka je $\mathbf{y}$ realizacija slučajnog uzorka $\mathbf{Y}$ čija je funkcija distribucije $f_{\mathbf{Y}|\theta}(\mathbf{y}|\vartheta)$ uz dani θ. Parametar θ je slučajna varijabla s vjerojatnosnom funkcijom $\pi(\vartheta)$. A posteriori funkcija distribucije slučajne varijable θ uz danu realizaciju $\mathbf{y}$ slučajnog uzorka $\mathbf{Y}$ se računa kao*

$$f_{\theta|\mathbf{Y}}(\vartheta|\mathbf{y}) = \frac{\pi_\theta(\vartheta) f_{\mathbf{Y}|\theta}(\mathbf{y}|\vartheta)}{f_{\mathbf{Y}}(\mathbf{y})},$$

gdje je $f_{\mathbf{Y}|\theta}(\mathbf{y}|\vartheta) = \begin{cases} \int_\theta \pi(\vartheta) f(\mathbf{y}|\vartheta) d\vartheta, & \textit{ako je } \theta \textit{ neprekidna slučajna varijabla} \\ \sum_\theta \pi(\vartheta) f(\mathbf{y}|\vartheta), & \textit{ako je } \theta \textit{ diskretna slučajna varijabla.} \end{cases}$

Napomena 1.2.12. *Budući da funkcija $f(\mathbf{y})$ nije funkcija slučajne varijable θ Bayesov teorem možemo pisati u obliku*

$$f_{\theta|\mathbf{Y}}(\vartheta|\mathbf{y}) \propto \pi_\theta(\vartheta) f_{\mathbf{Y}|\theta}(\mathbf{y}|\vartheta).$$

Bayesov teorem nam govori da a posteriori distribucija ovisi samo o a priori distribuciji slučajne varijable θ i vjerojatnosnoj distribuciji opaženog uzorka uz dani θ. Uočimo da vrijedi

$$\begin{aligned} f_{\mathbf{Y}|\theta}(\mathbf{y}|\vartheta) &= \prod_{j=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma_1}} e^{-\frac{1}{2}\left(\frac{y_j-\vartheta}{\sigma_1}\right)^2} \right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_1}} \right)^n e^{\frac{-\left[\sum_{j=1}^n y_j^2 - 2\vartheta \sum_{j=1}^n y_j + n\vartheta^2\right]}{2\sigma_1}} \\ &\propto e^{\frac{-\left[n\vartheta^2 - 2\vartheta \sum_{j=1}^n y_j\right]}{2\sigma_1}} \\ &\propto e^{-\frac{n}{2\sigma_1^2}(\vartheta-\bar{y})^2} \end{aligned}$$

Sada možemo izvesti a posteriori funkciju distribucije slučajne varijable θ uz danu realizaciju $\mathbf{y}$ slučajnog vektora $\mathbf{Y}$:

$$
\begin{aligned}
f_{\theta|\mathbf{Y}}(\vartheta|\mathbf{y}) &\propto \frac{1}{\sqrt{2\pi\sigma_2}} e^{-\frac{1}{2}\left(\frac{\vartheta-\mu_0}{\sigma_2}\right)^2} \prod_{j=1}^{n} \left(\frac{1}{\sqrt{\pi\sigma_1}} e^{-\frac{1}{2}\left(\frac{y_j-\vartheta}{2\sigma_1}\right)^2} \right) \\
&\propto \frac{1}{\sqrt{2\pi\sigma_2}} e^{-\frac{1}{2}\left(\frac{\vartheta-\mu_0}{\sigma_2}\right)^2} e^{-\frac{n}{2\sigma_1^2}(\vartheta-\bar{y})^2} \\
&= \frac{1}{\sqrt{2\pi\sigma_2}} e^{\frac{-\vartheta^2+\vartheta\mu_0-\mu_0^2}{2\sigma_2^2} - \frac{n}{2\sigma_1^2}(\vartheta-\bar{y})^2} \\
&\propto e^{\frac{1}{2}\left(\frac{-\vartheta^2+2\vartheta\mu_0}{\sigma_2^2} + \frac{-2n\bar{y}+n\vartheta^2}{\sigma_1^2}\right)} \\
&= e^{-\frac{1}{2}\left(\frac{\sigma_1^2\vartheta^2-2\sigma_1^2\vartheta\mu_0-2\sigma_2^2\vartheta n\bar{y}+\sigma_2^2 n\vartheta^2}{\sigma_2^2\sigma_1^2}\right)} \\
&= e^{-\frac{1}{2}\left(\frac{(\sigma_2^2 n+\sigma_1^2)\vartheta^2-2(\sigma_1^2\mu_0+\sigma_2^2 n\bar{y})}{\sigma_2^2\sigma_1^2}\right)} \\
&= e^{-\frac{1}{2}\left(\frac{\vartheta^2-2\vartheta\frac{(\sigma_1^2\mu_0+\sigma_2^2 n\bar{y})}{\sigma_2^2 n+\sigma_1^2}}{\frac{\sigma_2^2\sigma_1^2}{\sigma_2^2 n+\sigma_1^2}}\right)} \\
&\propto e^{-\frac{1}{2}\left(\frac{\left(\vartheta-\frac{\sigma_1^2\mu_0+\sigma_2^2 n\bar{y}}{\sigma_2^2 n+\sigma_1^2}\right)^2}{\frac{\sigma_2^2\sigma_1^2}{\sigma_2^2 n+\sigma_1^2}}\right)},
\end{aligned}
$$

drugim riječima, a posteriori funkcija distribucija od θ uz realizaciju $\mathbf{y}$ slučajnog uzorka $\mathbf{Y}$ je normalna s očekivanjem

$$\frac{\sigma_1^2\mu_0 + \sigma_2^2 n\bar{y}}{\sigma_2^2 n + \sigma_1^2} \tag{1.2}$$

koje možemo pisati u obliku

$$\frac{\sigma_2^2 n}{\sigma_1^2 + n\sigma_2^2}\bar{y} + \frac{\sigma_1^2}{\sigma_1^2 + n\sigma_2^2}\mu_0 \tag{1.3}$$

i varijancom

$$\frac{\sigma_2^2\sigma_1^2}{\sigma_2^2 n + \sigma_1^2}. \tag{1.4}$$

Slijedi da je Bayesova premija P^{Bayes} oblika

$$\begin{aligned}\mathbb{E}[\theta|\mathbf{y}] &= \frac{\sigma_2^2 n}{\sigma_1^2 + n\sigma_2^2}\bar{y} + \frac{\sigma_1^2}{\sigma_1^2 + n\sigma_2^2}\mu_0 \\ &= \alpha\bar{y} + (1-\alpha)\mu_0\end{aligned}$$

gdje je

$$\alpha = \frac{n}{n + \frac{\sigma_1^2}{\sigma_2^2}}$$

kojeg nazivamo koeficijent povjerenja.

1.3 Jednostavni model povjerenja

U ovom potpoglavlju, naš je cilj također pronaći procjenitelja za individualnu premiju $\mu(\theta)$, no procjenitelja ne tražimo među svim mogućim funkcijama procjenitelja, nego se fokusiramo na procjenitelje koji su linearna funkcija slučajnog uzorka $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ i među njima tražimo najboljeg u Bayesovom smislu. Takav procjenitelj nazivamo **procjenitelj premije povjerenja**.
Pretpostavke modela:

1. Slučajne varijable Y_j, $j = 1, 2, ..., n$, uz $\theta = \vartheta$, su nezavisne i jednako distribuirane s funkcijom distribucije F_ϑ i uvjetnim momentima

$$\begin{aligned}\mu(\vartheta) &= \mathbb{E}[Y_j|\theta = \vartheta] \\ \sigma^2(\vartheta) &= \mathrm{Var}[Y_j|\theta = \vartheta]\end{aligned}$$

2. θ je slučajna varijabla s funkcijom distribucije $U(\vartheta)$.

Uočimo da iz druge pretpostavke slijedi da su i $\mu(\theta)$ i $\sigma^2(\theta)$ slučajne varijable.
Vrijedi sljedeće:

$$\begin{aligned}P^{ind} &= \mathbb{E}[Y_j|\theta = \vartheta] \\ P^{kol} &= \mu_0 = \int_\Theta \mu(\vartheta)dU(\vartheta).\end{aligned}$$

Navedimo sada teorem koji nam daje oblik procjenitelja premije povjerenja uz navedene pretpostavke jednostavnog modela povjerenja.

Teorem 1.3.1. *Procjenitelj premije povjerenja pod pretpostavkama jednostavnog modela povjerenja je oblika*

$$\widehat{\widehat{\mu(\theta)}} = \alpha\overline{Y} + (1-\alpha)\mu_0$$

gdje je

$$\mu_0 = \mathbb{E}[\mu(\theta)]$$
$$\alpha = \frac{n}{n + \frac{\sigma^2}{\tau^2}}.$$

Dokaz. Označimo najboljeg linearnog procjenitelja individualne premije s $\widehat{\widehat{\mu(\theta)}}$. On mora biti oblika

$$\widehat{\widehat{\mu(\theta)}} = \hat{a}_0 + \sum_{j=1}^{n} \hat{a}_j Y_j,$$

gdje koeficijenti $\hat{a}_0, \hat{a}_1, ..., \hat{a}_n$ zadovoljavaju jednadžbu

$$\mathbb{E}\left[\left(\mu(\theta) - \hat{a}_0 - \sum_{j=1}^{n} \hat{a}_j Y_j\right)^2\right] = \min_{a_0,...,a_n \in \mathbb{R}} \mathbb{E}\left[\left(\mu(\theta) - a_0 - \sum_{j=1}^{n} a_j Y_j\right)^2\right].$$

Problem minimizacije ćemo riješiti parcijalnim deriviranjem po koeficijentima $a_0, a_1, ..., a_n$ i svaku od dobivenih derivacija izjednačiti s nulom. Prvo parcijalno deriviramo po koeficijentu a_0 i izjednačimo derivaciju s nulom i dobijemo

$$\mathbb{E}\left[\mu(\theta) - a_0 - \sum_{j=1}^{n} a_j Y_j\right] = 0. \tag{1.5}$$

Iz linearnosti očekivanja i

$$\mathbb{E}[Y_j] = \mathbb{E}[\mathbb{E}[Y_j|\theta]] = \mathbb{E}[\mu(\theta)]$$

slijedi

$$a_0 = \mathbb{E}[\mu(\theta)]\left(1 - \sum_{j=1}^{n} a_j\right).$$

Sada parcijalno deriviramo po a_k, $k = 1, 2, ..., n$, i izjednačavanjem s nulom za svaki k dobijemo

$$\begin{aligned} \mathbb{E}\left[Y_k(\mu(\theta) - a_0 - \sum_{j=1}^{n} a_j Y_j)\right] &= 0 \\ \mathbb{E}[Y_k\mu(\theta)] - a_0\mathbb{E}[Y_k] - \sum_{j=1, j\neq k}^{n} a_j\mathbb{E}[Y_k Y_j] - \mathbb{E}[Y_k^2] &= 0 \end{aligned} \tag{1.6}$$

Promotrimo posebno izraze u prethodnoj jednadžbi:

- koristeći definiciju od $\mu(\theta)$ i svojstva matematičkog očekivanja dobijemo

$$\begin{aligned}\mathbb{E}[Y_k\mu(\theta)] &= \mathbb{E}[\mathbb{E}[Y_k\mu(\theta)|\theta]] \\ &= \mathbb{E}[\mu(\theta)\mathbb{E}[Y_k|\theta]] \\ &= \mathbb{E}[\mu(\theta)^2] \\ &= \mathrm{Var}[\mu(\theta)] + (\mathbb{E}[\mu(\theta)])^2\end{aligned}$$

- koristeći činjenicu da su Y_j i Y_k, uz dano θ, nezavisni za $j \neq k$, i svojstva matematičkog očekivanja dobijemo

$$\begin{aligned}\mathbb{E}[Y_jY_k] &= \mathbb{E}[\mathbb{E}[Y_jY_k|\theta]] \\ &= \mathbb{E}[\mathbb{E}[Y_j|\theta]\mathbb{E}[Y_k|\theta]] \\ &= \mathbb{E}[\mu(\theta)\mu(\theta)] \\ &= \mathrm{Var}[\mu(\theta)] + (\mathbb{E}[\mu(\theta)])^2\end{aligned}$$

- posljednji ćemo izraz raspisati koristeći svojstva matematičkog očekivanja i varijance

$$\begin{aligned}\mathbb{E}[Y_k^2] &= \mathbb{E}[\mathbb{E}[Y_k^2|\theta]] \\ &= \mathbb{E}[\mathrm{Var}[Y_k|\theta]] + \mathrm{Var}[\mu(\theta)] + (\mathbb{E}[\mu(\theta)])^2\end{aligned}$$

Označimo:

$$\begin{aligned}\tau^2 &:= \mathrm{Var}[\mu(\theta)] \\ \mu_0 &:= \mathbb{E}[\mu(\theta)] \\ \sigma^2 &:= \mathbb{E}[\mathrm{Var}[Y_k|\theta]]\end{aligned}$$

Sada jednadžbu 1.6 možemo zapisati u obliku

$$\tau^2 + \mu_0^2 - a_0\mu_0 - \sum_{j=1, j\neq k}^{n} a_j(\tau^2 + \mu_0^2) - a_k\sigma^2 + \tau^2 + \mu_0^2 = 0$$

Nakon sređivanja gornje jednadžbe dobijemo

$$a_k\sigma^2 = \left(1 - \sum_{j=1}^{n} a_j\right)(\tau^2 + \mu_0^2) - a_0\mu_0.$$

Iz dobivene jednadžbe slijedi $a_1 = a_2 = ... = a_n$. Označimo s $\alpha = \sum_{j=1}^{n} a_j$ i s $\beta = a_0$. Sada traženi procjenitelj pišemo kao

$$\widehat{\mu(\theta)} = \beta + \alpha\overline{Y},$$

gdje je

$$\overline{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j.$$

Sada iz 1.5 i 1.6 slijedi

$$\beta = (1-\alpha)\mu_0$$
$$\alpha = \frac{n}{n+\frac{\sigma^2}{\tau^2}},$$

iz čega zaključujemo da procjenitelj povjerenja ima oblik

$$\widehat{\mu(\theta)} = \alpha\overline{Y} + (1-\alpha)\mu_0$$

što smo trebali i pokazati. □

Vidimo da je premija povjerenja u modelu težinski prosjek premije kolektiva i aritmetičke vrijednosti slučajnog uzorka **Y**, što i nije iznenađujuće, budući da su jedini izvori informacija kojima osiguravatelj raspolaže a priori znanje i individualno iskustvo. Jedan od procjenitelja koji je temeljen samo na a priori znanju je premija kolektiva i ona je najbolji takav procjenitelj. Njezin kvadratni gubitak je

$$\mathbb{E}[(\mu_0 - \mu(\theta))^2] = \mathrm{Var}(\mu(\theta)) = \tau^2.$$

Procjenitelj $\overline{Y}$ je najbolji linearni i individualni nepristrani procjenitelj, temeljen samo na slučajnom vektoru **Y**. Njegov kvadratni gubitak je

$$\mathbb{E}\left[(\overline{Y} - \mu(\theta))^2\right] = \mathbb{E}\left[\frac{\sigma^2(\theta)}{n}\right] = \frac{\sigma^2}{n}.$$

Sljedeća propozicija nam daje kvadratni gubitak procjenitelja povjerenja $\widehat{\mu(\theta)}$.

Propozicija 1.3.2. *Kvadratni gubitak procjenitelja povjerenja $\widehat{\mu(\theta)}$ je dana s*

$$\mathbb{E}[(\widehat{\mu(\theta)} - \mu(\theta))^2] = (1-\alpha)\tau^2 = \alpha\frac{\sigma^2}{n}.$$

Dokaz. Računamo:

$$
\begin{aligned}
\mathbb{E}[(\mu(\theta) - \widehat{\widehat{\mu(\theta)}})^2] &= \mathbb{E}[\alpha(\mu(\theta) - \overline{Y}) + (1-\alpha)(\mu(\theta) - \mu_0))^2] \\
&= \alpha^2 \frac{\sigma^2}{n} + (1-\alpha)^2 \tau^2 \\
&= \left(\frac{n}{n + \frac{\sigma^2}{\tau^2}} \right)^2 \frac{\sigma^2}{n} + \left(\frac{\frac{\sigma^2}{\tau^2}}{n + \frac{\sigma^2}{\tau^2}} \right)^2 \tau^2 \\
&= \frac{\sigma^2}{n + \frac{\sigma^2}{\tau^2}} = \alpha \frac{\sigma^2}{n} \\
&= \frac{\frac{\sigma^2}{\tau^2}}{n + \frac{\sigma^2}{\tau^2}} \tau^2 = (1-\alpha)\tau^2
\end{aligned}
$$

□

Iz prethodne propozicije slijedi da je kvadratni gubitak procjenitelja povjerenja manji i od kvadratnog gubitka premije kolektiva i od kvadratnog gubitka od $\overline{Y}$.

1.4 Jednostavni Bühlmannov model

U ovom potpoglavlju generaliziramo jednostavni model povjerenja na način da promatramo cijeli porfelj sličnih rizika θ_i, $i = 1, 2, ..., I$. Neka je $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{in})'$ slučajni uzorak za profil rizika θ_i.
Pretpostavke jednostavnog Bühlmannovog modela su:

1. Slučajne varijable Y_{ij}, $j = 1, 2, ..., n$ su nezavisne i jednako distribuirane uz dani $\theta_i = \vartheta$ s istom funkcijom distribucije F_ϑ i prva dva uvjetna momenta

$$
\begin{aligned}
\mu(\vartheta) &= \mathbb{E}[Y_{ij} | \theta_i = \vartheta] \\
\sigma^2(\vartheta) &= \mathrm{Var}[Y_{ij} | \theta_i = \vartheta].
\end{aligned} \tag{1.7}
$$

2. Parovi $(\theta_1, Y_1), ..., (\theta_I, Y_I)$ su nezavisni i jednako distribuirani.

Primijetimo da su profili rizika $\theta_1, \theta_2, ..., \theta_I$ nezavisne slučajne varijable s funkcijom distribucije $U(\vartheta)$ koju smo definirali u prethodnim poglavljima kao strukturnu funkciju distribucije. To nam sugerira da je portfelj heterogen, ali da su a priori svi rizici jednaki. Naš je cilj pronaći najbolji procjenitelj povjerenja $\widehat{\widehat{\mu(\theta_i)}}$ za $\mu(\theta_i)$, za svaki $i = 1, 2, ..., I$ u klasi funkcija

$$
\left\{ \widehat{\mu(\theta_i)} : \widehat{\mu(\theta_i)} = a_0 + \sum_{k=1}^{I} \sum_{j=1}^{n} a_{kj} Y_{kj} \; ; a_0, a_{kj} \in \mathbb{R} \right\}.
$$

Takav procjenitelj povjerenja nazivamo nehomogenim procjeniteljem povjerenja. Definiramo i homogeni procjenitelj povjerenja.

Definicija 1.4.1. *Homogeni procjenitelj od $\mu(\theta_i)$ uz slučajni uzorak* $\mathbf{Y}$ *je najbolji mogući procjenitelj u klasi*

$$\left\{\widehat{\mu(\theta_i)} : \widehat{\mu(\theta_i)} = \sum_{k=1}^{I}\sum_{j=1}^{n} a_{kj}Y_{kj}, a_{kj} \in \mathbb{R}\right\} \tag{1.8}$$

nepristranih procjenitelja takav da vrijedi

$$\mathbb{E}[\widehat{\mu(\theta_i)}] = \mathbb{E}[\mu(\theta_i)] \tag{1.9}$$

i označavamo ga s $\widehat{\widehat{\mu(\theta_i)}}^{hom}$.

U nastavku navodimo teorem za procjenitelje povjerenja u Bühlmannovom modelu.

Teorem 1.4.2. *Uz pretpostavke Bühlmannovog modela vrijedi da je nehomogen procjenitelj premije povjerenja dan s*

$$\widehat{\widehat{\mu(\theta_i)}} = \alpha\overline{Y}_i + (1-\alpha)\mu_0, \tag{1.10}$$

a homogen procjenitelj povjerenja dan s

$$\widehat{\widehat{\mu(\theta_i)}}^{hom} = \alpha\overline{Y}_i + (1-\alpha)\overline{Y}, \tag{1.11}$$

gdje su

$$\begin{aligned}\alpha &= \frac{n}{n+\frac{\sigma^2}{\tau^2}}\\ \overline{Y}_i &= \frac{1}{n}\sum_{j=1}^{n} Y_{ij}\\ \overline{Y} &= \frac{1}{In}\sum_{i=1}^{I}\sum_{j=1}^{n} Y_{ij}.\end{aligned} \tag{1.12}$$

Dokaz. Ovaj teorem nećemo dokazivati budući da je dokaz analogan poopćenom dokazu 1.5 u idućem potpoglavlju. □

1.5 Bühlmann-Straubov Model

Bühlmann-Straubov model je poopćenje jednostavnog Bühlmannovog modela koje uzima u obzir varijacije s obzirom na izloženost ili veličinu kolektiva uvođenjem tzv. **volumena rizika** (eng. risk volume), ω_{ij}, rizika i, $i = 1, 2, ..., I$ u godini j, $j = 1, 2, ..., n$. Volumen rizika osigurava da se u obračunavanju premije uzme u obzir određeno svojstvo osiguranika koje ga čini više ili manje rizičnim. Uzmimo kao primjer osiguranje od požara gdje osiguravajuće društvo kada sklapa policu osiguranja, prvo mora utvrditi kolika je vrijednost objekta kojeg osigurava i vrijednost imovine u njemu, odnosno osigurana svota, jer to utječe na premiju koju će osiguranik platiti. Neki od primjera volumena rizika su i ukupni broj osiguranih ljudi u grupnom zdravstvenom osiguranju ili osiguranju protiv nesreće i broj godina pod rizikom u autoosiguranju.

U ovom modelu definirajmo Y_{ij} kao omjer šteta koji se definira kao ukupni iznos šteta kolektiva S_{ij} s profilom rizika i u godini j u odnosu na volumen rizika, $Y_{ij} := \frac{S_{ij}}{\omega_{ij}}$. Navedimo sada pretpostavke modela.

Pretpostavke Bühlmann-Straubovog modela:

Rizik i je karakteriziran individualnim profilom rizika ϑ_i, koji je realizacija slučajne varijable θ_i i vrijedi

BS1: uvjetno na dani θ_i, slučajne varijable Y_{ij}, $j = 1, 2, ..., n$ su nezavisne s uvjetnim momentima

$$\mathbb{E}[Y_{ij}|\theta_i] = \mu(\theta_i)$$
$$\mathrm{Var}[Y_{ij}|\theta_i] = \frac{\sigma^2(\theta_i)}{\omega_{ij}}.$$

BS2: Parovi $(\theta_1, Y_1), ..., (\theta_I, Y_I)$ su nezavisni i slučajne varijable $\theta_1, \theta_2, ..., \theta_I$ su nezavisne i jednako distribuirane.

Nehomogeni procjenitelj premije povjerenja

Uz navedene pretpostavke modela, uvodimo sljedeći teorem koji će nam dati oblik nehomogenog procjenitelja premije povjerenja u Bühlmann-Straubovom modelu.

Teorem 1.5.1. *(Nehomogeni) procjenitelj premije povjerenja u Bühlmann-Straubovom modelu je dan s*

$$\widehat{\widehat{\mu(\theta_i)}} = \alpha_i \widetilde{Y}_i + (1 - \alpha_i)\mu_0 = \mu_0 + \alpha_i(\widetilde{Y}_i - \mu_0)$$

gdje su

$$\begin{aligned}\widetilde{Y}_i &= \sum_j \frac{\omega_{ij}}{\omega_i} Y_{ij} \\ \omega_i &= \sum_j \omega_{ij} \\ \alpha_i &= \frac{\omega_i}{\omega_i + \frac{\sigma^2}{\tau^2}}.\end{aligned}$$

Dokaz. Dokaz se provodi na sličan način kao i dokaz koji smo proveli za procjenitelja premije povjerenja za jednostavan model.
Prvo minimiziramo izraz

$$h(\mathbf{a}) = \mathbb{E}\left[\left(a_0 + \sum_{k,j} a_{kj} Y_{kj} - \mu(\theta_i)\right)^2\right] \tag{1.13}$$

za svaki $a_0, a_{kj} \in \mathbb{R}$ tako da izračunamo sve parcijalne derivacije po ovim parametrima i izjednačimo ih s nulom. Dobivamo sljedeće jednadžbe

$$\frac{\partial}{\partial a_0} h(\mathbf{a}) = 2\mathbb{E}\left[a_0 + \sum_{k,j} a_{kj} Y_{kj} - \mu(\theta_i)\right] = 0 \tag{1.14}$$

$$\frac{\partial}{\partial a_{st}} h(\mathbf{a}) = 2\mathbb{E}\left[Y_{st}(a_0 + \sum_{k,j} a_{kj} Y_{kj} - \mu(\theta_i))\right] = 0. \tag{1.15}$$

Iz jednadžbe 1.14 slijedi nepristranost nehomogenog procjenitelja premije povjerenja i nakon sređivanja slijedi da je

$$a_0 = \mu_0 \left(1 - \sum_{k,j} a_{kj}\right). \tag{1.16}$$

Uvrstimo li dobiveni a_0 u jednadžbu 1.15 i koristeći ponovno jednadžbu 1.14, vidimo da za sve s, t mora vrijediti

$$Cov\left(Y_{st}, \sum_{k,j} a_{kj} Y_{kj} - \mu(\theta_i)\right) = 0. \tag{1.17}$$

Budući da su rizici neovisni, različiti kolektivi su nekorelirani, tj. vrijedi sljedeće

$$Cov(Y_{st}, \mu(\theta_i)) = \sum_j a_{sj} Cov(Y_{st}, Y_{sj}), \text{ za sve } s, j. \tag{1.18}$$

Raspišemo lijevu stranu

$$Cov(Y_{st}, \mu(\theta_i)) = \mathrm{Var}(\mu(\theta_i))\mathbb{1}_{\{s=i\}} = \tau^2\mathbb{1}_{\{s=i\}}, \tag{1.19}$$

a zatim raspišemo i izraz

$$\begin{aligned} Cov(Y_{st}, Y_{sj}) &= \mathbb{E}[Cov(Y_{st}, Y_{sj}|\theta_s)] + Cov(\mathbb{E}[Y_{st}|\theta_s], \mathbb{E}[Y_{sj}|\theta_k]) \\ &= \frac{1}{\omega_{st}}\mathbb{E}[\sigma^2(\theta_s)]\mathbb{1}_{\{j=t\}} + \mathrm{Var}(\mu(\theta_s)) \\ &= \frac{\sigma^2}{\omega_{st}}\mathbb{1}_{\{j=t\}} + \tau^2 > 0, \end{aligned}$$

što znači da je lijeva strana od 1.18 jednaka 0 za $s \neq i$ i budući da vrijedi da je $Cov(Y_{st}, Y_{sj}) \geq \tau^2 > 0$ slijedi da je $a_{st} = 0$ za sve $s \neq i$. Zato vrijede sljedeće jednakosti

$$a_0 = \mu_0\left(1 - \sum_j a_{ij}\right) = \mu_0(1 - \alpha_i) \tag{1.20}$$

$$\tau^2 = a_{it}\frac{\sigma^2}{\omega_{it}} + \tau^2\sum_j a_{ij} = a_{it}\frac{\sigma^2}{\omega_{it}} + \tau^2\alpha_i, \tag{1.21}$$

pri čemu smo definirali $\alpha_i = \sum_j a_{ij}$. Iz jednakosti 1.21 slijedi da je za svaki t

$$a_{i,t} = \frac{\tau^2}{\sigma^2}(1 - \alpha_i)\omega_{it}. \tag{1.22}$$

Zato vrijedi da je

$$\alpha_i = \sum_t a_{it} = \frac{\tau^2}{\sigma^2}(1 - \alpha_i)\sum_t \omega_{it}. \tag{1.23}$$

Sada, iz gornje jednakosti dobijemo izraz za α_i

$$\alpha_i = \frac{\frac{\tau^2}{\sigma^2}\sum_t \omega_{it}}{\frac{\tau^2}{\sigma^2}\sum_t \omega_{it} + 1} = \frac{\sum_j \omega_{ij}}{\sum_j \omega_{ij} + \frac{\sigma^2}{\tau^2}} = \frac{\omega_i}{\omega_i + \frac{\sigma^2}{\tau^2}}.$$

Koeficijenti a_{it} imaju oblik

$$\begin{aligned} a_{it} &= \frac{\tau^2}{\sigma^2}\left(1 - \frac{\sum_j \omega_{ij}}{\sum_j \omega_{ij} + \frac{\sigma^2}{\tau^2}}\right)\omega_{ij} \\ &= \frac{\tau^2}{\sigma^2} \cdot \frac{\frac{\sigma^2}{\tau^2}}{\sum_j \omega_{ij} + \frac{\sigma^2}{\tau^2}}\omega_{ij} \\ &= \alpha_i\frac{\omega_{it}}{\sum_j \omega_{ij}} \\ &= \alpha_i\frac{\omega_{it}}{\omega_i}. \end{aligned}$$

Konačno, koristeći prethodne jednakosti dobivamo izraz za nehomogeni procjenitelj premije

$$\widehat{\widehat{\mu(\theta_i)}} = \alpha_i \frac{1}{\omega_i} \sum_t \omega_{it} Y_{it} + (1-\alpha_i)\mu_0 = \alpha_i \widetilde{Y}_i + (1-\alpha_i)\mu_0$$

što smo trebali pokazati.

□

Kvadratni gubitak procjenitelja povjerenja u Bühlmann-Straubovom modelu je dan sljedećim teoremom koji navodimo bez dokaza budući da je dokaz analogan dokazu za Jednostavni model povjerenja.

Teorem 1.5.2. *Uz dane pretpostavke Bühlmann-Straubovog modela kvadratni gubitak procjenitelja povjerenja dan je s*

$$L_{1,i} = \mathbb{E}[(\widehat{\widehat{\mu(\theta_i)}} - \mu(\theta_i))^2] = (1-\alpha_i)\tau^2 = \alpha_i \frac{\sigma^2}{\omega_i}. \tag{1.24}$$

Vrijednost τ^2 je kvadratni gubitak kolektivne premije μ_0

$$\mathbb{E}[(\mu_0 - \mu(\theta_i))^2] = \mathrm{Var}(\mu(\theta_i)) = \tau^2.$$

Iz gornjeg teorema vidimo da je korištenjem procjenitelja povjerenja kvadratni gubitak smanjen za faktor $1-\alpha_i$.
Također, za slučajnu varijablu $\widetilde{Y}_i$, koja je najbolji linearni individualni nepristrani procjenitelj koji ovisi samo o slučajnom vektoru $\mathbf{Y}_i$, kvadratni gubitak je dan s

$$\mathbb{E}[(\widetilde{Y}_i - \mu(\theta_i))^2] = \mathbb{E}\left[\mathbb{E}\left[\left(\sum_j \frac{\omega_{ij}}{\omega_i}(Y_{ij} - \mu(\theta_i))\right)^2 |\theta_i\right]\right] = \mathbb{E}\left[\frac{\sigma^2(\theta_i)}{\omega_i}\right] = \frac{\sigma^2}{\omega_i}.$$

Primijetimo da je kvadratni gubitak korištenjem procjenitelja premije povjerenja smanjen za faktor α_i.

Homogeni procjenitelj premije povjerenja

Sljedeći teorem nam daje homogeni procjenitelj premije povjerenja u Bülhmann-Straubovom modelu.

Teorem 1.5.3. *Homogeni procjenitelj u Bülhmann-Straubovom modelu uz dane pretpostavke modela je dan formulom*

$$\widehat{\widehat{\mu(\theta_i)}}^{hom} = \alpha_i \widetilde{Y}_i + (1-\alpha_i)\widehat{\widehat{\mu_0}} = \widehat{\widehat{\mu_0}} + \alpha_i(\widetilde{Y}_i - \widehat{\widehat{\mu_0}}),$$

gdje su

$$\widehat{\widehat{\mu_0}} = \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet}$$
$$\alpha_i = \frac{\omega_i}{\omega_i + \frac{\sigma^2}{\tau^2}}$$
$$\alpha_\bullet = \sum_{i=1}^{I} \alpha_i$$

Dokaz. Dokaz se provodi analogno dokazu teorema za nehomogen procjenitelj u Bühlmann-Straubovom modelu, koristeći Langrangeovu metodu za zamjenu uvjeta 1.14 uvjetom

$$\mu_0 = \mathbb{E}[\hat{\mu}] = \mathbb{E}\left[\sum_{i,j} a_{ij} Y_{ij}\right] = \sum_{i,j} a_{ij} \mathbb{E}[Y_{ij}] = \sum_{i,j} a_{ij} \mu_0$$

što znači da vrijedi

$$\sum_{i,j} a_{ij} = 1.$$

□

Primijetimo da homogeni procjenitelj koristimo za opservacije iz cijelog kolektiva, a ne samo za one za i-tu policu, jer su nam one potrebne za procjenu kolektivne premije μ_0, a procjena od μ_0 je automatski ugrađena u formulu homogenog procjenitelja povjerenja. Jedno od bitnih svojstava homogenog procjenitelja povjerenja je **svojstvo ravnoteže** koje definiramo u sljedećoj propoziciji.

Propozicija 1.5.4. *Uz dane pretpostavke Bühlmann-Straubovog modela za homogeni procjenitelj povjerenja vrijedi*

$$\sum_{i,j} \omega_{ij} \widehat{\widehat{\mu(\theta_i)}}^{hom} = \sum_{i,j} \omega_{ij} Y_{ij}. \tag{1.25}$$

Dokaz. Iz

$$\alpha_i = \frac{\omega_i}{\omega_i + \frac{\sigma^2}{\tau^2}}$$

slijedi

$$\omega_i (1 - \alpha_i) = \frac{\sigma^2}{\tau^2} \alpha_i,$$

iz čega slijedi

$$\begin{aligned}\sum_i \omega_i(\widehat{\widehat{\mu(\theta_i)}}^{hom} - \widetilde{Y}_i) &= \sum_i \omega_i(-(1-\alpha_i)\widetilde{Y}_i + (1-\alpha_i)\widehat{\widehat{\mu(\theta_i)}}) \\ &= \frac{\sigma^2}{\tau^2}\sum_i \alpha_i(\widehat{\widehat{\mu(\theta_i)}} - \widetilde{Y}_i) \\ &= 0.\end{aligned}$$

□

Sljedeći teorem koji navodimo bez dokaza nam daje kvadratni gubitak homogenog procjenitelja.

Teorem 1.5.5. *Uz dane pretpostavke Bühlmann-Straubovog modela kvadratni gubitak homogenog procjenitelja je dan s*

$$L_{1,i}^{hom} = \mathbb{E}[(\widehat{\widehat{\mu(\theta_i)}}^{hom} - \mu(\theta_i))^2] = \tau^2(1-\alpha_i)\left(1 + \frac{1-\alpha_i}{\alpha_\bullet}\right). \tag{1.26}$$

Pogreška predikcije Bühlmann-Straubovog modela

Procjenitelj premije povjerenja $\widehat{\widehat{\mu(\theta_i)}}$ procjenjuje premiju povjerenja $\mu(\theta_i)$ i daje predikciju za omjer šteta $Y_{i,n+1}$ u idućem razdoblju. Zato možemo provjeriti kolika je ukupna greška predikcije

$$Y_{i,n+1} - \widehat{\widehat{\mu(\theta_i)}} \tag{1.27}$$

što možemo napisati u obliku

$$(Y_{i,n+1} - \mu(\theta_i)) + (\mu(\theta_i) - \widehat{\widehat{\mu(\theta_i)}}). \tag{1.28}$$

Kao i ranije, slučajne varijable $Y_{i,1}, Y_{i,2}, ..., Y_{i,n+1}$ su nezavisne uz dani θ_i. Stoga vrijedi

$$\begin{aligned}L_{2,i} &= \mathbb{E}[(Y_{i,n+1} - \widehat{\widehat{\mu(\theta_i)}})^2] \\ &= \mathbb{E}[(Y_{i,n+1} - \mu(\theta_i))^2] + \mathbb{E}[(\mu(\theta_i - \widehat{\widehat{\mu(\theta_i)}})^2] \\ &= \mathbb{E}[\mathrm{Var}[Y_{i,n+1}|\theta_i]] + \mathbb{E}[(\mu(\theta_i - \widehat{\widehat{\mu(\theta_i)}})^2] \\ &= \frac{\sigma^2}{\omega_{i,n+1}} + (1-\alpha_i)\tau^2.\end{aligned}$$

Zadnja jednakost slijedi iz teorema 1.5.2.

Uz isti postupak kao kod nehomogenog procjenitelja premije, dobijemo da je ukupna greška predikcije za homogeni procjenitelj premije povjerenja jednaka

$$L_{2,i}^{hom} = \mathbb{E}[(Y_{i,n+1} - \widehat{\widehat{\mu(\theta_i)}}^{hom})^2] = \frac{\sigma^2}{\omega_{i,n+1}} + (1 - \alpha_{in})\tau^2 \left(1 + \frac{1 - \alpha_i}{\alpha_\bullet}\right).$$

Ukoliko uzmemo da je volumen rizika u budućem razdoblju jednak 1, ukupna greške predikcije za nehomogeni procjenitelj premije povjerenja dana je s

$$L_{2,i} = \mathbb{E}[(Y_{i,n+1} - \widehat{\widehat{\mu(\theta_i)}})^2] = \sigma^2 + (1 - \alpha_i)\tau^2, \tag{1.29}$$

a za homogeni procjenitelj premije povjerenja

$$L_{2,i}^{hom} = \mathbb{E}[(Y_{i,n+1} - \widehat{\widehat{\mu(\theta_i)}}^{hom})^2] = \sigma^2 + (1 - \alpha_i)\tau^2 \left(1 + \frac{1 - \alpha_i}{\alpha_\bullet}\right). \tag{1.30}$$

Procjena strukturnih parametara σ^2 i τ^2

Strukturni parametri σ^2, koji predstavlja očekivanu volatilnost unutar kolektiva, i τ^2, koji predstavlja heterogenost između kolektiva, su u praksi nepoznati i trebamo ih procijeniti iz podataka kojim raspolažemo.
Uzorački procjenitelj za varijancu i-tog razreda rizika je

$$S_i := \frac{1}{n-1} \sum_{j=1}^{n} \omega_{ij}(Y_{ij} - \widetilde{Y}_i)^2.$$

Pokažimo da je svaka slučajna varijabla S_i nepristran procjenitelj za σ^2.

Napišimo S_i u obliku

$$
\begin{aligned}
S_i &= \frac{1}{n-1}\sum_{j=1}^{n}\omega_{ij}\left(Y_{ij}-\mu(\theta_i)+\mu(\theta_i)-\widetilde{Y}_i\right)^2 \\
&= \frac{1}{n-1}\left\{\sum_{j=1}^{n}\omega_{ij}\left[(Y_{ij}-\mu(\theta_i))^2+2(Y_{ij}-\mu(\theta_i))(\mu(\theta_i)-\widetilde{Y}_i)+(\mu(\theta_i)-\widetilde{Y}_i)^2\right]\right\} \\
&= \frac{1}{n-1}\left\{\sum_{j=1}^{n}\omega_{ij}(Y_{ij}-\mu(\theta_i))^2+2\left(\mu(\theta_i)-\widetilde{Y}_i\right)\sum_{j=1}^{n}\omega_{ij}(Y_{ij}-\mu(\theta_i))+\sum_{j=1}^{n}\omega_{ij}(\mu(\theta_i)-\widetilde{Y}_i)^2\right\} \\
&= \frac{1}{n-1}\left\{\sum_{j=1}^{n}\omega_{ij}(Y_{ij}-\mu(\theta_i))^2+2\left(\mu(\theta_i)-\widetilde{Y}_i\right)\left[\sum_{j=1}^{n}\omega_{ij}\widetilde{Y}_i-\sum_{j=1}^{n}\omega_{ij}\mu(\theta_i)\right]+\sum_{j=1}^{n}\omega_{ij}(\mu(\theta_i)-\widetilde{Y}_i)^2\right\} \\
&= \frac{1}{n-1}\left\{\sum_{j=1}^{n}\omega_{ij}(Y_{ij}-\mu(\theta_i))^2+2\left(\mu(\theta_i)-\widetilde{Y}_i\right)\left[\omega_i\widetilde{Y}_i-\omega_i\mu(\theta_i)\right]+\sum_{j=1}^{n}\omega_{ij}(\mu(\theta_i)-\widetilde{Y}_i)^2\right\} \\
&= \frac{1}{n-1}\left\{\sum_{j=1}^{n}\omega_{ij}(Y_{ij}-\mu(\theta_i))^2-2\omega_i\left(\mu(\theta_i)-\widetilde{Y}_i\right)^2+\omega_i(\mu(\theta_i)-\widetilde{Y}_i)^2\right\} \\
&= \frac{1}{n-1}\left\{\sum_{j=1}^{n}\omega_{ij}(Y_{ij}-\mu(\theta_i))^2-\omega_i(\mu(\theta_i)-\widetilde{Y}_i)^2\right\}
\end{aligned}
$$

iz čega slijedi da je

$$\mathbb{E}[S_i|\theta_i]=\sigma^2(\theta_i),$$

pa vrijedi

$$\mathbb{E}[S_i]=\mathbb{E}[\mathbb{E}[S_i|\theta_i]]=\mathbb{E}[\sigma^2(\theta_i)]=\sigma^2. \tag{1.31}$$

Teško je odrediti optimalne težine uzoračkih varijanci S_i u ukupnoj varijanci portfelja pa kao procjenitelj za σ^2 uzimamo njihovu aritmetičku sredinu

$$\widehat{\sigma^2}=\frac{1}{I}\sum_{i=1}^{I}\frac{1}{n-1}\sum_{j=1}^{n}\omega_{ij}(Y_{ij}-\widetilde{Y}_i)^2. \tag{1.32}$$

Ovaj procjenitelj je nepristran što slijedi direktno iz 1.31 i konzistentan što slijedi iz Čebiševljeve nejednakosti.

Za procjenu strukturnog parametra τ^2 najprije definiramo težinsku srednju vrijednost uzorka

$$\overline{Y}=\frac{1}{\sum_{i,j}\omega_{ij}}\sum_{i,j}\omega_{ij}Y_{ij}=\frac{1}{\sum_i\omega_i}\sum_i\omega_i\widetilde{Y}_i. \tag{1.33}$$

Definiramo i

$$T = \frac{1}{I-1} \sum_i \frac{\omega_i}{\sum_{k,l} \omega_{kl}} (\widetilde{Y_i} - \overline{Y})^2. \tag{1.34}$$

Računanjem očekivane vrijednosti od T možemo doći do nepristranog procjenitelja za za τ^2 koji je oblika:

$$\widehat{\tau^2} = c \cdot \left\{ T - \frac{I\widehat{\sigma^2}}{\sum_{k,l} \omega_{kl}} \right\} \tag{1.35}$$

gdje je

$$c = \frac{I-1}{I} \left[\sum_i \frac{\omega_i}{\sum_{k,l} \omega_{kl}} \left(1 - \frac{\omega_i}{\sum_{k,l} \omega_{kl}} \right) \right]^{-1}. \tag{1.36}$$

Problem koji može nastati je taj da procjenitelj $\widehat{\tau^2}$ bude negativan pa koristimo procjenitelj

$$\widehat{\widehat{\tau^2}} = \max\{\widehat{\tau^2}, 0\}.$$

Empirijski procjenitelj povjerenja

Iz homogenog procjenitelja povjerenja dobivamo empirijski procjenitelj povjerenja tako da zamijenimo strukturne parametre σ^2 i τ^2 s njihovim procjeniteljima koje smo odredili u prethodnom odjeljku. Dakle, empirijski procjenitelj povjerenja ima oblik

$$\widehat{\widehat{\mu(\theta_i)}}^{emp} = \widehat{\alpha}_i Y_i + (1 - \widehat{\alpha}_i)\widehat{\widehat{\mu_0}} \tag{1.37}$$

gdje su

$$\widehat{\alpha_i} = \frac{\omega_i}{\omega_i + \frac{\widehat{\sigma^2}}{\widehat{\tau^2}}}$$
$$\widehat{\widehat{\mu_0}} = \frac{\sum_i \widehat{\alpha_i} Y_i}{\sum_i \widehat{\alpha_i}}.$$

Promotrimo sljedeći primjer u kojem ćemo premiju povjerenja procijeniti uz pomoć empirijskog procjenitelja povjerenja.

Primjer 1.5.6. *Osiguravajuće društvo u svojoj ponudi ima autoosiguranje i ugovorilo je police sa pet kompanija za osiguravanje njihovih službenih automobila. U sljedećoj tablici prikazan je ukupni iznos šteta za prethodnih sedam godina za svih pet kompanija.*

Godina Kompanija	1	2	3	4	5	6	7
1	80	56	174	48	162	138	91
2	49	118	91	76	81	84	47
3	38	33	68	57	41	79	59
4	173	92	104	121	153	48	23
5	42	137	117	156	171	124	162

Tablica 1.1: Ukupni iznos šteta za sedam godina za svaku od pet kompanija, S, u tisućama eura

Broj osiguranih automobila u tom istom razdoblju za sve kompanije je dan u sljedećoj tablici. U ovom slučaju broj osiguranih automobila je volumen rizika ω.

Godina Kompanija	1	2	3	4	5	6	7
1	10	8	12	14	15	15	14
2	4	4	4	5	6	6	5
3	2	3	4	3	4	5	4
4	16	19	19	16	18	15	19
5	6	7	6	6	5	5	5

Tablica 1.2: Broj osiguranih automobila za sve kompanije

Prvo želimo normalizirati gubitak tako što ćemo podijeliti ukupne iznose šteta po godinama s volumenom rizika u toj godini za svaku kompaniju. Dobiveni normalizirani gubitci su prikazani u tablici:

Godina Kompanija	1	2	3	4	5	6	7
1	8	7	14.5	3.4286	10.8	9.2	6.5
2	12.25	29.5	22.75	15.2	13.5	14	9.4
3	19	11	17	19	10.25	15.8	14.75
4	10.8125	4.8421	5.4737	7.5625	8.5	3.2	1.2105
5	7	19.5714	19.5	26	34.2	24.8	32.4

Tablica 1.3: Normalizirani gubitci, Y

Zatim izračunajmo ukupni volumen rizika za svaku kompaniju pojedinačno:

$$\omega_1 = \sum_{j=1}^{7} \omega_{1j} = 10 + 8 + 12 + 14 + 15 + 15 + 14 = 88$$
$$\omega_2 = \sum_{j=1}^{7} \omega_{2j} = 34$$
$$\omega_3 = \sum_{j=1}^{7} \omega_{3j} = 25$$
$$\omega_4 = \sum_{j=1}^{7} \omega_{4j} = 122$$
$$\omega_5 = \sum_{j=1}^{7} \omega_{5j} = 40,$$

a zatim i ukupni volumen rizika koji je jednak

$$\omega = \sum_{i=1}^{5} = 309.$$

Sad možemo izračunati težinski prosjek za svaku kompaniju i, $i = 1, 2, ..., 5$ *koji su jednaki*

$$\widetilde{Y}_1 = \sum_{j=1}^{7} \frac{\omega_{ij}}{\omega_i} Y_{ij} = \frac{749}{88} = 8.5114$$
$$\widetilde{Y}_2 = \frac{273}{17} = 16.0589$$
$$\widetilde{Y}_3 = 15$$
$$\widetilde{Y}_4 = \frac{357}{61} = 5.8525$$
$$\widetilde{Y}_5 = \frac{909}{40} = 22.725.$$

Da bismo izračunali koliki je faktor povjerenja pa onda i procjenitelj povjerenja, potrebno je procijeniti strukturne parametre σ^2 *i* τ^2 *što ćemo napraviti po formulama iz prethodnog potpoglavlja 1.32 i 1.35. Sada izračunajmo*

$$\overline{Y} = \frac{1}{\sum_i \omega_i} \sum_i \omega_i \widetilde{Y}_i = \frac{3293}{309} = 10.657.$$

U sljedećoj su tablici izračunati izrazi koji nam trebaju za računanje procjena parametara.

Kompanija	$\sum_{j=1}^{7} \omega_{ij}(Y_{ij} - \widetilde{Y}_i)^2$	$\frac{\omega_i}{\sum_{k,l}\omega_{kl}}(\widetilde{Y}_i - \overline{Y})^2$
1	955.2601	1.3111
2	1249.8824	3.2108
3	237.7	1.5261
4	1103.6482	9.1138
5	2827.9607	18.8528
Ukupno	6374.4514	34.0146

Tablica 1.4: Izrazi potrebni za računanje procjenitelja za σ^2 i τ^2

Sada računamo procjenu za σ^2*:*

$$\begin{aligned}\widehat{\sigma^2} &= \frac{1}{I}\sum_{i=1}^{I}\frac{1}{n-1}\sum_{j=1}^{7}\omega_{ij}(Y_{ij} - \widetilde{Y}_i)^2 \\ &= \frac{1}{5}\cdot\frac{1}{7-1}\cdot 6374.4514 \\ &= 212.4817.\end{aligned} \tag{1.38}$$

Da bismo izračunali procjenu za τ^2 *potrebno je izračunati konstantu* c *po formuli 1.36.*

$$\begin{aligned}c &= \frac{I-1}{I}\left[\sum_{i}^{I}\frac{\omega_i}{\sum_{k,l}\omega_{kl}}\left(1 - \frac{\omega_i}{\sum_{k,l}\omega_{kl}}\right)\right]^{-1} \\ &= \frac{4}{5}\cdot 1.3744 \\ &= 1.0995.\end{aligned} \tag{1.39}$$

Konačno, izračunajmo procjenu za τ^2

$$\begin{aligned}\widehat{\tau^2} &= c\cdot\left\{\frac{1}{I-1}\sum_{i}\frac{\omega_i}{\sum_{k,l}\omega_{kl}}(\widetilde{Y}_i - \overline{Y})^2 - \frac{I\widehat{\sigma^2}}{\sum_{k,l}\omega_{kl}}\right\} \\ &= 1.0995\cdot\frac{1}{5}\cdot 34.0146 - \frac{5\cdot 212.4817}{309} \\ &= 4.0416.\end{aligned} \tag{1.40}$$

Vrijedi da je

$$\widehat{\mathcal{K}} = \frac{\widehat{\sigma^2}}{\widehat{\tau^2}} = \frac{212.4817}{4.0416} = 52.5737. \tag{1.41}$$

Za svaku od kompanija izračunali smo faktor povjerenja α_i i pripadni homogeni procjenitelj premije povjerenja. Rezultati su prikazani u donjoj tablici.

Kompanija	α_i	$\frac{\alpha_i}{\alpha_\bullet}\widetilde{Y_i}$	$\widehat{\widehat{\mu(\theta_i)}}^{emp}$
1	0.626	2.1555	9.9243
2	0.3927	2.5513	13.7696
3	0.3223	1.9556	13.163
4	0.6988	1.6546	7.7912
5	0.4321	3.9723	16.7986
Veličina	$\alpha_\bullet$	$\widehat{\widehat{\mu_0}}$	
Ukupno	2.4719	12.2893	

Tablica 1.5: Empirijski procjenitelj premije povjerenja

Poglavlje 2

Regresijska stabla odluke

U ovom poglavlju ćemo objasniti stablo odluke kao metodu statističkog učenja koja će nam dati novi pristup teoriji povjerenja i procjenjivanju premije povjerenja. Stablo odluke predstavlja hijerarhijski model odluka i posljedica tih odluka. Poglavlje se bazira na literaturi [4] i [3]. Prednost ove metode je da je vrlo jednostavna za implementaciju i lako se može prikazati grafički, što pridonosi lakšem razumijevanju modela. Također, ona ne zahtjeva specificiranje veze između prediktora i varijable odziva prije izvođenja modela što povećava fleksibilnost metode i omogućuje da prepozna nelinearne efekte prediktora na varijablu odziva ili pak interakcijske kovarijacijske efekte prediktora na varijablu odziva. Osim toga, metoda ne zahtijeva a priori selekciju varijabli prediktora da bi se izveo regresijski model što je slučaj s parametarskim regresijskim metodama o kojima detalje pogledati u [2].

2.1 Stabla odluke

Metode statističkog učenja bazirane na stablima odluke imaju svojstvo da dijele prostor prediktora $D \subseteq \mathbb{R}^p$ na (p-dimenzionalne) pravokutnike i zatim se na svakom od tih pravokutnika prilagođava jednostavan model. Pravila po kojima dijelimo prostor prediktora možemo sumirati u stablo pa otuda i naziv **stablo odluke**. Uvedimo osnovne pojmove koje će nam trebati za ovu metodu statističkog učenja.

Osnovni pojmovi

Definicija 2.1.1. ***Stablo*** *T je neprazan konačan skup podataka takav da*

- *postoji jedan istaknuti* ***čvor*** *koji se naziva* ***korijen stabla***

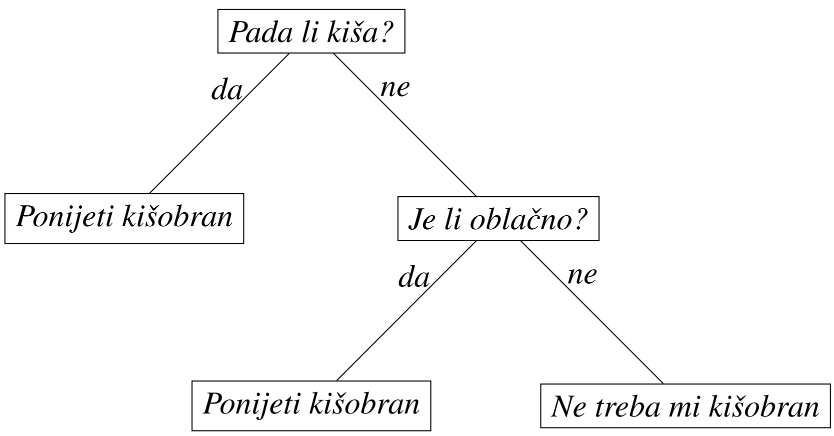
- *skup ostalih čvorova (ako je neprazan) particioniran je na k disjunktnih skupova $T_1, T_2, ..., T_k$ koji su također stabla i ti skupovi se nazivaju* ***podstabla*** *korijena*

Korijen je poveznicama koje nazivamo **grane** spojen s jednim ili više čvorova koje zovemo **djeca**. Čvorovi koji nemaju djecu zovemo **listovi**. **Susjedi** su djeca istog čvora, a **unutarnji čvor** je čvor koji ima najmanje jedno dijete. **Visina stabla** se definira kao udaljenost korijena do najudaljenijeg lista.

Na kraju definirajmo i binarno stablo na koje ćemo se fokusirati u ovom radu.

Definicija 2.1.2. ***Binarno stablo T*** *je konačan skup podataka koji je ili prazan ili ima korijen, a ostali čvorovi su podijeljeni u dva podskupa T_L i T_R od kojih svaki ima strukturu binarnog stabla. Stablo T_L zovemo* ***lijevo podstablo****, a stablo T_R zovemo* ***desno podstablo****.*

Primjer 2.1.3. *Jednostavno binarno stablo odluke na primjeru ponijeti ili ne ponijeti kišobran na izlasku iz kuće*



2.2 Konstrukcija stabla odluke

Postoje dvije vrste stabla odluke:

- klasifikacijska - rješavanje problema klasifikacije
- regresijska - rješavanje regresijskih problema.

U ovom radu se fokusiramo na regresijsko stablo, budući da ćemo pomoću njega procijeniti premiju povjerenja. Za konstrukciju obje vrste stabala, regresijskog i klasifikacijskog, kroz vrijeme su se razvili brojni algoritmi, no u ovom radu ćemo se koncentrirati na najpoznatiji, model CART (Classification And Regression Trees) kojeg ćemo objasniti u sljedećim potpoglavljima na regresijskom stablu odluke.

Regresijska stabla odluke

Pretpostavimo da se naš skup podataka sastoji od p varijabli poticaja $X \in D \subset \mathbb{R}^p$ i varijable odziva $Y \in \mathbb{R}$, i za svaku imamo N opaženih podataka, (x_i, y_i), za $i = 1, 2, ..., N$, gdje je $x_i = (x_{i1}, x_{i2}, ..., x_{ip})$. Da bismo konstruirali regresijsko stablo, prvo dijelimo prostor prediktora kojeg čine sve moguće vrijednosti slučajnih varijabli $X_1, X_2, ..., X_p$ na M disjunktnih višedimenzionalnih pravokutnika $R_1, R_2, ..., R_M$. Zatim za svaku opservaciju koja upada u pravokutnik R_m modeliramo predikciju kao konstantu c_m za $m = 1, 2, ..., M$:

$$f(x) = \sum_{m=1}^{M} c_m \mathbb{1}_{\{x \in R_m\}}.$$

Najbolji izbor za konstantu c_m, za $m = 1, 2, ..., M$, je srednja vrijednost svih opaženih vrijednosti varijable odziva u skupu R_m, po kriteriju minimizacije sume kvadrata razlike $\sum(y_i - f(x_i))^2$, za $(x_i, y_i) \in R_m$. Označimo $\hat{y}_{R_m} := c_m$. Cilj nam je odabrati pravokutnike $R_1, R_2, ..., R_M$ tako da ukupni kvadratni gubitak

$$\sum_{m=1}^{M} \sum_{y_i \in R_m} (y_i - \hat{y}_{R_m})^2, \tag{2.1}$$

bude najmanji, no nemoguće je uzeti u obzir svaku particiju prostora prediktora pa koristimo pohlepni algoritam pod nazivom **rekurzivno binarno dijeljenje** koji u svakom čvoru odabire najbolje moguće dijeljenje prostora prediktora u tom koraku. Algoritam je pohlepan pa ne uzima u obzir što možda u idućem koraku, grananje koje smo prethodno napravili i nije bilo najbolja opcija. Napomenimo, algoritam kreće od korijena prema listovima. Dakle, u početnom koraku promatramo cijeli skup podataka i odabiremo varijablu u odnosu na koju vršimo grananje, odnosno prediktor X_j, i konstantu s pomoću kojih definiramo poluravnine

$$\begin{aligned} R_1(j, s) &= \{X | X_j \leq s\} \\ R_2(j, s) &= \{X | X_j > s\} \end{aligned} \tag{2.2}$$

tako da za odabrane varijable j i s izraz

$$\sum_{i: x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2, \tag{2.3}$$

bude najmanji mogući. Konstanta $\hat{y}_{R_1}$ je srednja vrijednost opaženih vrijednosti varijable odziva iz poluravnine $R_1(j, s)$, a $\hat{y}_{R_2}$ je srednja vrijednost opaženih vrijednosti varijable odziva iz poluravnine $R_2(j, s)$. Za gore navedeni minimizacijski problem 2.3 nije teško pronaći odgovarajuće j i s, pogotovo ako nemamo velik broj prediktora p. Ovaj proces ponavljamo za svaki ovako dobiveni višedimenzionalni pravokutnik. Postavlja se pitanje kada zaustaviti algoritam, budući da, ako ne stanemo na vrijeme, dobijemo previsoko stablo koje dovodi do prenaučenosti (eng. overfitting), dok će stablo s malo točaka grananja imati manju varijancu, ali veću pristranost.

Podrezivanje stabla

Visina stabla je ključna za određivanje kompleksnosti modela i njegovu preciznost predikcije. Stoga je od velikog značaja odabrati optimalnu visinu stabla. CART (Classification and Regression Tree) model, objavljen u [1], predlaže konstrukciju "velikog" stabla T_0 i zatim njegovo **podrezivanje** (eng. tree pruning) kako bismo dobili podstablo $T \subset T_0$ koje minimizira grešku predviđanja na testnom skupu podataka. Za to koristimo metodu **troškovno-kompleksno podrezivanje** (eng. cost-complexity prunning) koje se naziva i **podrezivanje najslabije karike** (eng. weakest link prunning).

Troškovno-kompleksno podrezivanje

Prvi korak ovog algoritma je izrada "velikog" stabla T_0 pomoću gore navedenog rekurzivnog binarnog grananja na temelju podataka za učenje. Zaustavljamo se kad svaki list stabla ima manje od unaprijed zadanog minimalnog broja opservacija. Definiramo podstablo $T \subset T_0$ kao svako podstablo stabla T_0 koje se može dobiti podrezivanjem. Označimo s $|T|$ broj listova u stablu T. Listove indeksiramo indeksom m koji predstavlja pravokutnik R_m. Definiramo troškovno kompleksni kriterij kao

$$C_\nu(T) = \sum_{m=1}^{|T|} \sum_{i:x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \nu|T| \tag{2.4}$$

gdje je

$$\hat{y}_{R_m} = \frac{1}{N_m} \sum_{i:x_i \in R_m} y_i$$
$$N_m = \#\{x_i \in R_m.\}$$

Parametar $\nu \geq 0$ je parametar za podešavanje i on upravlja kompromisom između kompleksnosti podstabla i prilagođenosti podacima za učenje. Ideja je za svaki ν iz spektra vrijednosti koje zadamo pronaći podstablo $T_\nu \subset T_0$ takvo da minimizira 2.4. Može se pokazati da je takvo stablo za svaki ν jedinstveno. Za svaki unutarnji čvor m računamo troškovnu kompleksnost podstabla čiji je korijen taj čvor i troškovnu kompleksnost stabla ako bi podstablo čiji je taj čvor korijen, bilo uklonjeno i zamijenjeno listom. Ako uklanjanje tog podstabla rezultira manjom troškovnom kompleksnosti, uklonimo ga, a ako ne, zadržavamo. Nastavimo podrezivanje dok ne dobijemo stablo koje se sastoji samo od korijena. Na takav način dobivamo konačan niz najboljih podstabala u obliku funkcije parametra za podešavanje ν. Optimalni ν možemo dobiti K-strukom unakrsnom provjerom.

K-struka unakrsna provjera

K-struka unakrsna provjera je tehnika koja se koristi za evaluiranje modela i za optimizaciju hiperparametara. U slučaju regresijskih stabala, K-struku unakrsnu provjeru koristimo za pronalaženje optimalnog parametra za podešavanje ν. Funkcionira tako da se skup podataka podijeli na nasumičan način na K disjunktnih particija $\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_K$ koje sadrže otprilike istu količinu podataka. Model učimo na $K-1$ particija, odnosno preklopa, tako da nad tim podacima ponovimo konstrukciju "velikog" stabla i provedemo troškovno-kompleksno podrezivanje, a testiramo ga na preostaloj, k-toj particiji. Formalno, neka je $\kappa : \{1, ..., N\} \to \{1, ..., K\}$ funkcija indeksiranja koja označava kojoj particiji je pridružen podatak j nakon nasumičnog razvrstavanja. Za svaki $k = 1, ..., K$ dobijemo ν-ti model $\hat{T}^{-k}(\mathbf{X}, \nu)$ koji testiramo na k-toj particiji koju nismo koristili u treniranju modela tako da računamo procjenu greške predikcije koja na k-toj particiji ima oblik

$$CV_k(\hat{T}, \nu) = \sum_{j:\kappa(j)=k} (y_j - \hat{T}^{-k}(\mathbf{x}_j, \nu))^2. \tag{2.5}$$

Na kraju uzmemo ukupnu pogrešku predikcije modela na K particija

$$CV(\hat{T}, \nu) = \sum_{k=1}^{K} CV_k(\hat{T}, \nu) \tag{2.6}$$

i za nju tražimo parametar za podešavanje $\hat{\nu}$ koji je minimizira. Konačno, izaberemo stablo trenirano nad cijelim skupom podataka koje odgovara optimalnom parametru podešavanja $\hat{\nu}$. Najčešće koristimo 5-struku i 10-struku unakrsnu provjeru.

Poglavlje 3

Model povjerenja

U ovom poglavlju ćemo definirati model koji ćemo promatrati po uzoru na znanstveni članak [5]. Neka je $\mathbf{Y}_i = (Y_1, Y_2, ..., Y_{n_i})^T$ slučajni vektor omjera šteta za rizik i, $i = 1, 2, ..., I$ gdje je I broj rizika u portfelju. Pripadni volumeni rizika su dani s vektorom $\omega_i = (\omega_{i1}, \omega_{i2}, ..., \omega_{in_i})$. Također, definiramo i vektor prediktora $\mathbf{X}_i = (X_{i1}, X_{i2}, ..., X_{ip})^T$ za svaki rizik i. Kao i ranije, profil rizika i je karakteriziran skalarom ϑ_i koji je realizacija slučajne varijable θ_i, za svaki $i = 1, 2, ..., I$.
Pretpostavke modela su sljedeće:

1. Trojke $(\theta_1, \mathbf{X}_1, \mathbf{Y}_1), (\theta_2, \mathbf{X}_2, \mathbf{Y}_2), ..., (\theta_I, \mathbf{X}_I, \mathbf{Y}_I)$ su međusobno nezavisne

2. Uz dani $\theta_i = \vartheta_i$ i $\mathbf{X}_i = \mathbf{x}_i$, slučajne varijable $Y_{i,j}$, $j = 1, 2, ..., n$ su nezavisne i vrijedi

$$\begin{aligned} \mathbb{E}[Y_{i,j}|\mathbf{X}_i = \mathbf{x}_i, \theta_i = \vartheta_i] &= \mu(\mathbf{x}_i, \vartheta_i) \\ \mathrm{Var}[Y_{i,j}|\mathbf{X}_i = \mathbf{x}_i, \theta_i = \vartheta_i] &= \frac{\sigma^2(\mathbf{x}_i, \vartheta_i)}{\omega_{ij}}. \end{aligned} \tag{3.1}$$

Ako stavimo $n_i = n$ i $\omega_{ij} = 1$, za $j = 1, 2, .., n_i$ i $i = 1, 2, ..., I$, tada govorimo o modelu uravnoteženih šteta. U ovom ćemo se poglavlju bazirati na metodu particioniranja u svrhu procjenjivanja premije povjerenja. Prvo ćemo se fokusirati na model uravnoteženih šteta. Za njega vrijedi da je koeficijent povjerenja α jednak

$$\alpha := \alpha_i = \frac{n}{n + \frac{\sigma^2}{\tau^2}}. \tag{3.2}$$

3.1 Funkcije gubitka kod particioniranja

U prvom poglavlju smo izveli funkcije gubitka za Bühlmann-Straubov model 1.24, 1.29, 1.26 i 1.30 koje redom u uravnoteženom modelu šteta imaju pojednostavljen oblik

$$L_{1,i} = \frac{\sigma^2}{n + \frac{\sigma^2}{\tau^2}} = \tau^2(1-\alpha) \tag{3.3}$$

$$L_{2,i} = \sigma^2 + \frac{\sigma^2}{n + \frac{\sigma^2}{\tau^2}} = \sigma^2 + \tau^2(1-\alpha) \tag{3.4}$$

$$L_{1,i}^{hom} = \tau^2(1-\alpha) + \frac{1}{I} \cdot \frac{\tau^2(1-\alpha)^2}{\alpha} \tag{3.5}$$

$$L_{2,i}^{hom} = \sigma^2 + \tau^2(1-\alpha) + \frac{1}{I} \cdot \frac{\tau^2(1-\alpha)^2}{\alpha}. \tag{3.6}$$

Budući da je broj rizika I u portfelju u praksi obično velik broj i kako vrijedi $\frac{1}{I} \stackrel{n\to\infty}{\to} 0$, možemo zanemariti drugi član u funkciji gubitka 3.5 i treći član u funkciji gubitka 3.6. Budući da se sad funkcije gubitka 3.5 i 3.6 svode na funkcije gubitka 3.3 i 3.4 njih ćemo koristiti u nastavku. Funkciju gubitka cijelog kolektiva uz funkciju gubitka 3.3 definiramo kao

$$L = \sum_{i=1}^{I} L_{1,i} = \sum_{i=1}^{I} \frac{\sigma^2}{n + \frac{\sigma^2}{\tau^2}} = \frac{I\sigma^2}{n + \frac{\sigma^2}{\tau^2}} = \frac{I}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}. \tag{3.7}$$

Označimo s $\mathcal{I}_1$ i $\mathcal{I}_2$ dva disjunktna podkolektiva od $\mathcal{I}$ takva da vrijedi $\mathcal{I}_1 \cup \mathcal{I}_2 = \mathcal{I}$. Neka je p vjerojatnost da proizvoljan individualni rizik i kolektiva $\mathcal{I}$ upadne u podkolektiv $\mathcal{I}_1$. Tada je $1-p$ vjerojatnost da upadne u podkolektiv $\mathcal{I}_2$. Tada vrijedi da su funkcije gubitka za podkolektive $\mathcal{I}_1$ i $\mathcal{I}_2$ redom

$$L_1 = \frac{Ip}{\frac{n}{\sigma_1^2} + \frac{1}{\tau_1^2}} \tag{3.8}$$

$$L_2 = \frac{I(1-p)}{\frac{n}{\sigma_2^2} + \frac{1}{\tau_2^2}} \tag{3.9}$$

gdje su σ_1^2 i τ_1^2 strukturni parametri podkolektiva $\mathcal{I}_1$, a σ_2^2 i τ_2^2 strukturni parametri podkolektiva $\mathcal{I}_2$ za koje vrijede jednakosti

$$\begin{aligned} \sigma^2 &= p\sigma_1^2 + (1-p)\sigma_2^2 \\ \tau^2 &= p\tau_1^2 + (1-p)\tau_2^2 + p(1-p)(\mu_1 - \mu_2)^2, \end{aligned} \tag{3.10}$$

gdje su μ_1, μ_2 također strukturni parametri redom podkolektiva $\mathcal{I}_1$ i $\mathcal{I}_2$. Za detalje pogledati članak [5]. Sljedeći teorem daje glavnu prednost metode particioniranja u procjenjivanju premije povjerenja u odnosu na standardne metode.

Teorem 3.1.1. *Za kvadratne gubitke definirane s L_1, L_2 i L vrijedi*

$$L_1 + L_2 \leq L. \tag{3.11}$$

Dokaz. Za dokaz ovog teorema nam je potrebna lema koja kaže da je funkcija definirana s

$$h(x, y) = \frac{1}{\frac{1}{x} + \frac{1}{y}} : (0, \infty) \times (0, \infty) \to (0, \infty) \tag{3.12}$$

konkavna funkcija. Za dokaz leme pogledati Dodatak A članka [5]. Kvadratne gubitke L, L_1 i L_2 možemo pisati redom kao

$$\begin{aligned} L &= I \cdot h\left(\frac{\sigma^2}{n}, \tau^2\right) \\ L_1 &= pI \cdot h\left(\frac{\sigma_1^2}{n}, \tau_1^2\right) \\ L_2 &= (1-p)I \cdot h\left(\frac{\sigma_2^2}{n}, \tau_2^2\right) \end{aligned}$$

Računamo

$$\begin{aligned} L_1 + L_2 &= pI \cdot h\left(\frac{\sigma_1^2}{n}, \tau_1^2\right) + (1-p)I \cdot h\left(\frac{\sigma_2^2}{n}, \tau_2^2\right) \\ &\leq I \cdot h\left(\frac{p\sigma_1^2}{n} + \frac{(1-p)\sigma_2^2}{n}, p\tau_1^2 + (1-p)\tau_2^2\right) \\ &= I \cdot h\left(\frac{\sigma^2}{n}, \widetilde{\tau}^2\right) \\ &= \frac{I}{\frac{n}{\sigma^2} + \frac{1}{\widetilde{\tau}^2}} \\ &\leq \frac{I}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \\ &= L, \end{aligned}$$

gdje je $\widetilde{\tau}^2 := p\tau_1^2 + (1-p)\tau_2^2$. U prvoj nejednakosti smo koristili konkavnost funkcije h, a u drugoj činjenicu da je $\widetilde{\tau}^2 = p\tau_1^2 + (1-p)\tau_2^2 \leq p\tau_1^2 + (1-p)\tau_2^2 + p(1-p)(\mu_1 - \mu_2)^2 = \tau^2$. □

Ovim teoremom smo pokazali da se ukupni gubitak ne povećava primjenom metode particioniranja prostora prediktora, no treba imati na umu da teorem vrijedi pod pretpostavkom da su strukturni parametri izračunati bez statističke pogreške. Problem je što to u praksi često nije slučaj, jer su nam ti parametri nepoznati. Zato ne vršimo previše podjela prediktorskog prostora kako bismo pronašli kompromis između pogreške procjene i prednosti particioniranja. Kvadratni gubitak za cijeli portfelj koristeći funkciju $L_{2,i}$ definiranu u 3.4 dobivamo na sličan način i ima oblik

$$\widetilde{L} = I\sigma^2 + \frac{I}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} = I\sigma^2 + L. \tag{3.13}$$

Tada su kvadratni gubitci za podkolektive definirane na isti način kao gore uz funkciju $L_{2,i}$ redom:

$$\begin{aligned} \widetilde{L_1} &= Ip\sigma_1^2 + \frac{Ip}{\frac{n}{\sigma_1^2} + \frac{1}{\tau_1^2}} = Ip\sigma_1^2 + L_1 \\ \widetilde{L_2} &= I(1-p)\sigma_2^2 + \frac{I(1-p)}{\frac{n}{\sigma_2^2} + \frac{1}{\tau_2^2}} = I(1-p)\sigma_2^2 + L_2 \end{aligned} \tag{3.14}$$

I za ovaj slučaj vrijedi da je $\widetilde{L_1} + \widetilde{L_2} \leq \widetilde{L}$ što dokazujemo na analogan način. Budući da smo pokazali da particioniranje prostora podataka može smanjiti grešku predikcije, aproksimirat ćemo slučajne varijable $\mu(\mathbf{X}_i, \theta_i)$ i $\sigma^2(\mathbf{X}_i, \theta_i)$ s

$$\begin{aligned} \mu(\mathbf{X}_i, \theta_i) &= \sum_{k=1}^{K} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \mu_k(\theta_i) \\ \sigma^2(\mathbf{X}_i, \theta_i) &= \sum_{k=1}^{K} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \frac{\sigma_k^2(\theta_i)}{\omega_{ij}}, \end{aligned}$$

gdje su $A_1, A_2, ..., A_K$, $K \in \mathbf{N}$, particije prostora prediktora, a varijable $\mu_k(\theta_i)$ i $\sigma_k^2(\theta_i)$ predstavljaju premiju i varijancu za individualni rizik i za k-ti podkolektiv s profilom rizika θ_i i jednake su

$$\begin{aligned} \mu_k(\theta_i) &= \mathbb{E}[Y_{ij} | \mathbf{X}_i \in A_k, \theta_i = \vartheta_i] \\ \sigma_k^2(\theta_i) &= \mathrm{Var}[Y_{ij} | \mathbf{X}_i \in A_k, \theta_i = \vartheta_i]. \end{aligned} \tag{3.15}$$

Sada na svakoj particiji prediktorskog prostora $A_1, A_2, ..., A_K$, odnosno podkolektivu, primijenimo Bühlmann-Straubov model procjenjivanja premije povjerenja. Nehomogeni procjenitelj premije povjerenja za i-ti rizik u k-tom podkolektivu se računa kao

$$\widehat{\widehat{\mu^k(\theta_i)}} = \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \alpha_i^k \overline{Y_i} + (1 - \alpha_i^k)\mu_k, \tag{3.16}$$

gdje je $\mu_k = \mathbb{E}[Y_{ij} | \mathbf{X}_i \in A_k] = \mathbb{E}[\mu^k(\theta_i)]$ premija k-tog podkolektiva, a $\alpha_i^k = \frac{\omega_i}{\omega_i + \frac{\sigma_k^2}{\tau_k^2}}$.

Kao i gore σ_k^2 i τ_k^2 su strukturni parametri k-tog podkolektiva. Za rizik i, ukupni nehomogeni procjenitelj premije povjerenja računa se kao

$$\widehat{\widehat{\mu(\theta_i)}} = \sum_{k=1}^{K} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \widehat{\widehat{\mu^k(\theta_i)}}. \tag{3.17}$$

Homogeni procjenitelj premije povjerenja se računa na isti način i za rizik i je jednak

$$\widehat{\widehat{\mu(\theta_i)}}^{hom} = \sum_{k=1}^{K} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \widehat{\widehat{\mu^k(\theta_i)}}^{hom}, \tag{3.18}$$

gdje je

$$\widehat{\widehat{\mu^k(\theta_i)}}^{hom} = \alpha_i^k \overline{Y_i} + (1 - \alpha_i^k)\overline{Y_k},$$

gdje je $\overline{Y_k}$, slično kao u definiciji homogenog procjenitelja za Bühlman-Straubov model, jednak prosjeku individualnih iskustava u k-tom podkolektivu. Kvadratni gubitak za nehomogeni procjenitelj za k-ti podkolektiv $\mathbb{E}[(\mu_k(\theta_i) - \widehat{\widehat{\mu^k(\theta_i)}})^2]$ je dan s

$$L_{1,k} = \sum_{i=1}^{I} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \cdot \frac{\sigma_k^2}{\omega_i + \frac{\sigma_k^2}{\tau_k^2}}, \tag{3.19}$$

a za homogeni procjenitelj je jednak

$$L_{1,k}^{hom} = \sum_{i=1}^{I} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \tau_k^2 (1 - \alpha_i^k) \left(1 + \frac{1 - \alpha_i^k}{\alpha_\bullet^k}\right). \tag{3.20}$$

Ukupna greška predikcije u k-tom podkolektivu $\mathbb{E}[(Y_{i,n_i+1} - \widehat{\widehat{\mu^k(\theta_i)}})^2]$ za nehomogen procjenitelj premije povjerenja je dana s

$$L_{2,k} = \sum_{i=1}^{I} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \cdot \left(\sigma_k^2 + \frac{\sigma_k^2}{\omega_i + \frac{\sigma_k^2}{\tau_k^2}}\right), \tag{3.21}$$

a za homogen procjenitelj s

$$L_{2,k}^{hom} = \sum_{i=1}^{I} \mathbb{1}_{\{\mathbf{X}_i \in A_k\}} \left[\sigma_k^2 + \tau_k^2 (1 - \alpha_i^k) \left(1 + \frac{1 - \alpha_i^k}{\alpha_\bullet^k}\right)\right]. \tag{3.22}$$

3.2 Regresijsko stablo povjerenja

Autori znanstvenog rada [5] predlažu algoritam regresijskog stabla povjerenja za određivanje procjenitelja premije povjerenja. Postupak se svodi na regresijsko stablo odluke, ali se standardna funkcija gubitka u CART algoritmu zamjenjuje s jednom od funkcija gubitaka 3.19, 3.20, 3.21 i 3.22. Također, ne koristi se standardna unakrsna provjera nego longitudinalna. Longitudinalna unakrsna provjera duplicira omjere šteta za neke individualne rizike tako da svi individualni rizici imaju jednak broj omjera šteta. Ne koristimo unakrsnu provjeru iz CART algoritma budući da je ona primjenjiva samo na nove osiguranike koji još nemaju prijavljene štete.

Longitudinalna unakrsna provjera

Promotrimo detaljnije proces longitudinalne 5-struke unakrsne provjere.
U prvom koraku za svaki rizik $i = 1, 2, ..., I$ stavimo $k_i = \lfloor n_i/K \rfloor$ i $l_i = (n_i \bmod K)$, gdje je K broj disjunktnih particija na koje dijelimo kolektiv $\mathcal{I}$ slučajnim odabirom. Definiramo i niz $S_i = \{\mathcal{K}, \mathcal{K}, ..., \mathcal{K}, 1, 2, ..., l_i\}$, gdje je $\mathcal{K} = \{1, 2, ..., K\}$ i taj podniz u nizu S_i ponavljamo k_i puta.
U drugom koraku za svaki $i = 1, 2, ..., I$ nezavisno uzorkujemo bez zamjene iz niza S_i i dobijemo uzorak $\{W_{ij}, j = 1, 2, ..., n_i\}$.
Zatim označimo skup indeksa $\mathcal{J}_{i,k} = \{j \in \{1, 2, ..., n_i\} : W_{ij} = k\}$, a onda i skup $\mathbf{Y}_i^k = \{Y_{ij} : j \in \mathcal{J}_{i,k}\}$, za $k = 1, 2, ..., 5$. Nakon toga skup podataka podijelimo na pet podskupova:

$$\{(\mathbf{Y}_i^k, \mathbf{X}_i) : \mathbf{Y}_i^k \neq \emptyset, i \in \mathcal{I}\}$$

za $k = 1, 2, ..., 5$.
U trećem koraku provodimo troškovno-kompleksno podrezivanje kao i kod standardne metode CART za koje dobijemo niz stabala u ovisnosti o parametru podešavanja i računamo ukupnu grešku unakrsne provjere kao

$$\sum_{k=1}^{K} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \mathbb{1}_{j \in \mathcal{J}_{i,k}} \left(Y_{ij} - \hat{T}^{-k}(\mathbf{X}_i, \nu) \right)^2, \tag{3.23}$$

gdje je $\hat{T}^{-k}(\mathbf{X}_i, \nu)$ predviđena vrijednost premije za i-tog osiguranika pomoću modela stabla koje je učeno na svim particijama osim k-te uz parametar za podešavanje ν.
U četvrtom koraku izaberemo stablo koje odgovara optimalnom parametru $\hat{\nu}$ i učeno je nad cijelim skupom podataka, kao i kod obične 5-struke unakrsne provjere.

Praktični primjer

Promotrimo na praktičnom primjeru model regresijskog stabla povjerenja. Zbog nedostatka dostupnih podataka, simulirat ćemo ih po uzoru na znanstveni članak [5].
U našem ćemo modelu za svaki rizik i simulirati vektor kovarijata $\mathbf{X}_i = (X_1, X_2, ..., X_8)$ iz diskretne uniformne distribucije na skupu prirodnih brojeva $\{10, ..., 120\}$. Radi jednostavnosti, radit ćemo s homogenim modelom šteta, dakle svaki osiguranik i će imati jednak

broj šteta n i njihove težine će biti jednake $w_{ij} = 1, j = 1, 2, ..., n$.
U nastavku ćemo simulirati štete za svakog osiguranika $i = 1, ..., I$ na tri različita načina.

U prvom načinu simulacije šteta, za svaki rizik $i = 1, ..., I$ simuliramo n šteta kao

$$Y_{i,j} = e^{f(\mathbf{X}_i)} + \epsilon_{i,j}, \; j = 1, ..., n, \tag{3.24}$$

gdje je funkcija f proizvoljno zadana s

$$f(\mathbf{X}_i) = 0.01(2X_{i,1} + X_{i,3} + 3\sqrt{X_{i,5}} - 2\sqrt{X_{i,1}X_{i,4}}),$$

a vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ nezavisno simulirane iz gama distribucije $\Gamma\left(\frac{3}{2}, \frac{2}{3}e^{\frac{1}{2}}\right)$. Potom na isti način za svaki rizik $i = 1, ..., I$ ponovno simuliramo štete, ali sada vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ nezavisno simuliramo iz log-normalne distribucije $LN(0, 1)$. Za štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz gama distribucije $\Gamma\left(\frac{3}{2}, \frac{2}{3}e^{\frac{1}{2}}\right)$ vrijedi da su $\mu(\mathbf{X}_i, \theta_i)$ i $\sigma^2(\mathbf{X}_i, \theta_i)$ redom jednaki

$$\mu(\mathbf{X}_i, \theta_i) = \mathbb{E}[e^{f(\mathbf{X}_i)} + \epsilon_i] = e^{f(\mathbf{X}_i)} + \mathbb{E}[\epsilon_i] = e^{f(\mathbf{X}_i)} + e^{\frac{1}{2}}$$

i

$$\sigma^2(\mathbf{X}_i, \theta_i) = \mathrm{Var}[e^{f(\mathbf{X}_i)} + \epsilon_i] = \mathrm{Var}[\epsilon_i] = \frac{2}{3}e,$$

a za štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz log-normalne distribucije $LN(0, 1)$ vrijedi

$$\mu(\mathbf{X}_i, \theta_i) = \mathbb{E}[e^{f(\mathbf{X}_i)} + \epsilon] = e^{f(\mathbf{X}_i)} + \mathbb{E}[\epsilon] = e^{f(\mathbf{X}_i)} + e^{0.5}$$

i

$$\sigma^2(\mathbf{X}_i, \theta_i) = \mathrm{Var}[e^{f(\mathbf{X}_i)} + \epsilon] = \mathrm{Var}[\epsilon] = (e - 1) \cdot e.$$

U drugom načinu simulacije šteta, štete za svaki rizik $i = 1, ..., 350$ simuliramo kao u prethodnom slučaju 3.24, ali vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ nezavisno simuliramo prvo iz distribucije $\Gamma\left(1.5, \frac{2}{3}e^{g(\mathbf{X}_i)/2}\right)$, a zatim i iz log-normalne distribucije $LN(0, g(\mathbf{X}_i))$ gdje je funkcija g proizvoljno zadana s

$$g(\mathbf{X}_i) = \frac{1}{139.75}(X_{i,1} + 0.5X_{i,3} + 0.01X_{i,1}X_{i,3}) \tag{3.25}$$

Za ovu funkciju vrijedi da je

$$\begin{aligned}\mathbb{E}[g(\mathbf{X}_i)] &= \frac{1}{139.75}\left(\mathbb{E}[X_{i,1}] + 0.5\mathbb{E}[X_{i,3}] + 0.01\mathbb{E}[X_{i,1}] \cdot \mathbb{E}[X_{i,3}]\right) \\ &= \frac{1}{139.75}\left(65 + 32.5 + 0.01 \cdot 65 \cdot 65\right) = 1\end{aligned}$$

zbog nezavisnosti slučajnih vektora $X_{i,1}$ i $X_{i,3}$. Za ovaj postupak simulacije za štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz gama distribucije $\Gamma\left(\frac{3}{2}, \frac{2}{3}e^{g(\mathbf{X}_i)/2}\right)$ vrijedi

$$\mu(\mathbf{X}_i, \theta_i) = e^{f(\mathbf{X}_i)} + \mathbb{E}[\epsilon_i] = e^{f(\mathbf{X}_i)} + \frac{3}{2} \cdot \frac{2}{3} e^{g(\mathbf{X}_i)/2} = e^{f(\mathbf{X}_i)} + e^{g(\mathbf{X}_i)/2}$$
$$\sigma^2(\mathbf{X}_i, \theta_i) = \mathrm{Var}[\epsilon_i] = \frac{3}{2} \cdot \frac{4}{9} e^{g(\mathbf{X}_i)} = \frac{2}{3} e^{g(\mathbf{X}_i)},$$

a za one štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz log-normalne distribucije $LN(0, g(\mathbf{X}_i))$ vrijedi

$$\mu(\mathbf{X}_i, \theta_i) = e^{f(\mathbf{X}_i)} + \mathbb{E}[\epsilon_i] = e^{f(\mathbf{X}_i)} + e^{g(\mathbf{X}_i)/2} = e^{f(\mathbf{X}_i)} + e^{g(\mathbf{X}_i)/2}$$
$$\sigma^2(\mathbf{X}_i, \theta_i) = \mathrm{Var}[\epsilon_i] = \left(e^{g(\mathbf{X}_i)} - 1\right) \cdot e^{g(\mathbf{X}_i)},$$

Vidimo da za oba slučaja, za razliku od prethodnog načina simulacije šteta, varijanca nije jednaka za svaki rizik i već ovisi o varijablama odziva.

Naposljetku, promotrimo simulaciju u kojoj osiguranici, iako im se vrijednosti varijabli odziva podudaraju, mogu imati različite profile rizika. Za svaki rizik $i, i = 1, ..., I$ simuliramo štete kao u prethodnom slučaju, s jednom razlikom, a to je multipliciranje iznosa šteta s varijablom θ_i što omogućuje različite distribucije šteta za osiguranike s istim karakteristikama. Dakle, za svaki i, simuliramo n šteta kao

$$Y_{i,j} = \theta_i(e^{f(\mathbf{X}_i)} + \epsilon_{i,j}), j = 1, ..., n,$$

gdje je θ_i varijabla koju nezavisno simuliramo iz uniformne distribucije $U(0.9, 1.2)$. U ovom slučaju, za štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz gama distribucije $\Gamma\left(\frac{3}{2}, \frac{2}{3}e^{g(\mathbf{X}_i)/2}\right)$ vrijedi

$$\mu(\mathbf{X}_i, \theta_i) = \theta_i(e^{f(\mathbf{X}_i)} + \mathbb{E}[\epsilon_i]) = \theta_i(e^{f(\mathbf{X}_i)} + \frac{3}{2} \cdot \frac{2}{3} e^{g(\mathbf{X}_i)/2}) = \theta_i(e^{f(\mathbf{X}_i)} + e^{g(\mathbf{X}_i)/2})$$
$$\sigma^2(\mathbf{X}_i, \theta_i) = \mathrm{Var}[\epsilon_i] = \theta_i^2 \cdot \frac{3}{2} \cdot \frac{4}{9} e^{g(\mathbf{X}_i)} = \frac{2}{3} \theta_i^2 e^{g(\mathbf{X}_i)},$$

a za štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz log-normalne distribucije $LN(0, g(\mathbf{X}_i))$ te varijable su jednake

$$\mu(\mathbf{X}_i, \theta_i) = \theta_i(e^{f(\mathbf{X}_i)} + \mathbb{E}[\epsilon_i]) = \theta_i(e^{f(\mathbf{X}_i)} + e^{g(\mathbf{X}_i)/2})$$
$$\sigma^2(\mathbf{X}_i, \theta_i) = \mathrm{Var}[\epsilon_i] = \theta_i^2 \left(e^{g(\mathbf{X}_i)} - 1\right) \cdot e^{g(\mathbf{X}_i)}.$$

Za svaki od navedenih postupaka simulacije podataka, simuliramo 1000 uzoraka šteta za $I = 350$ osiguranika i za različite brojeve šteta $n = 5, 10, 20$ po osiguraniku. Za svaki uzorak gradimo regresijsko stablo povjerenja uz funkcije gubitka 3.19 i 3.20 i biramo najbolje stablo longitudinalnom unakrsnom provjerom. Metrika koju ćemo promatrati je relativna predikcijska pogreška koju definiramo kao

$$\mathrm{RPE} = \frac{\mathrm{PE}}{\mathrm{PE}_0}, \tag{3.26}$$

gdje je PE greška predikcije za model regresijskog stabla povjerenja, a PE_0 je greška predikcije nad cijelim kolektivom bez particioniranja. Grešku predikcije za model regresijskog stabla povjerenja definiramo kao

$$\mathrm{PE} = \frac{1}{I}\sum_{i=1}^{I}\sum_{k=1}^{K} A_k\left\{\mathbf{X}_i \in A_k\right\}\left(\widehat{\mu^k(\theta_i)} - \mu\left(\mathbf{X}_i, \theta_i\right)\right)^2, \tag{3.27}$$

gdje je $\widehat{\mu^k(\theta_i)}$ definirana u 3.16, a greška predikcije nad cijelim kolektivom bez particioniranja se definira se kao

$$\mathrm{PE}_0 = \frac{1}{I}\sum_{i=1}^{I}\left(\widehat{\mu(\theta_i)} - \mu\left(\mathbf{X}_i, \theta_i\right)\right)^2, \tag{3.28}$$

gdje je $\widehat{\mu(\theta_i)}$ procjenitelj premije povjerenja u Bühlmann-Straubovom modelu definiran u prvom poglavlju.

U sljedećoj tablici su za svaki način simulacije za štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz gama distribucije prikazane prosječne greške predikcije za 1000 simulacija.

	$n = 5$			$n = 10$			$n = 20$		
	SIM1	SIM2	SIM3	SIM1	SIM2	SIM3	SIM1	SIM2	SIM3
$\mathbf{PE_0}$	0.3358	0.3815	0.4257	0.1744	0.1959	0.2187	0.0893	0.0996	0.1111
$\mathbf{PE_1}$	0.2390	0.3151	0.3640	0.1363	0.1730	0.1975	0.0753	0.0926	0.1050
$\mathbf{PE_2}$	0.2396	0.3156	0.3650	0.1366	0.1735	0.1979	0.0756	0.0928	0.1052

Tablica 3.1: Prosječna greška predikcije za 1000 simulacija za $n = 5, 10, 20$, gama

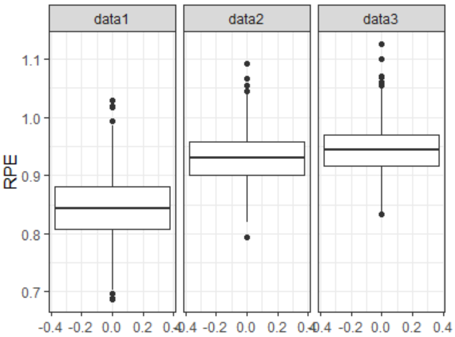
Primijetimo da se prosječna greška predikcije za oba regresijska stabla povjerenja smanjuje kako povećavamo broj šteta n po osiguraniku, bez obzira na način simulacije podataka. Promotrimo istu tablicu i za štete za koje su vrijednosti $\{\epsilon_{i,j}, j = 1, ..., n\}$ simulirane iz log-normalne distribucije.

	$n = 5$			$n = 10$			$n = 20$		
	SIM1	SIM2	SIM3	SIM1	SIM2	SIM3	SIM1	SIM2	SIM3
$\mathbf{PE_0}$	0.7739	1.2599	1.3992	0.4223	0.7325	0.7978	0.2230	0.4032	0.4426
$\mathbf{PE_1}$	0.4348	0.7986	0.9077	0.2670	0.5019	0.5664	0.1594	0.3061	0.3435
$\mathbf{PE_2}$	0.4354	0.9008	0.9077	0.2678	0.5030	0.5673	0.1610	0.3067	0.3439

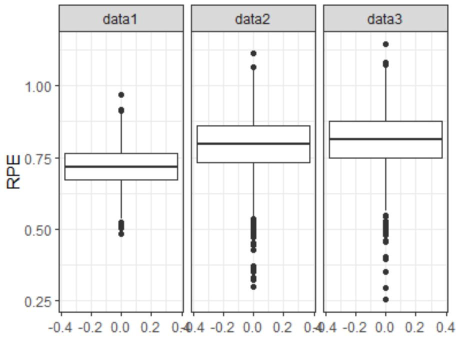
Tablica 3.2: Prosječna greška predikcije za 1000 simulacija za $n = 5, 10, 20$, log-normal

Vidimo da isti zaključak možemo izvesti i iz ove tablice.

Za kraj prikažimo dijagrame pravokutnika za relativnu grešku predikcije za sva tri načina simulacije podataka gdje je broj šteta po osiguraniku $n = 20$ za obje distribucije. Prvo prikažimo relativnu grešku predikcije za funkciju gubitka 3.19 za obje distribucije.

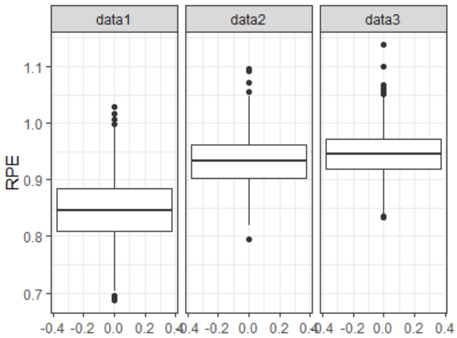


Slika 3.1: RPE uz funkciju gubitka 3.19 i gama distribuciju

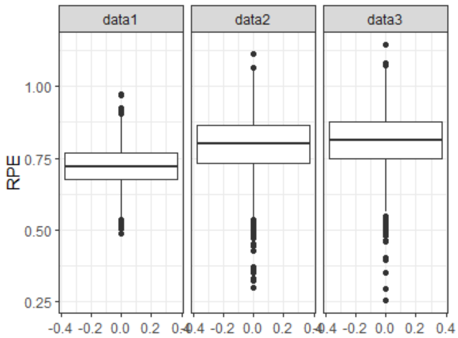


Slika 3.2: RPE uz funkciju gubitka 3.19 i log-normalnu distribuciju

Vidimo da je za obje distribucije i podatke za sva tri načina simulacije šteta pravokutnik koji prikazuje relativnu grešku predikcije od donjeg do gornjeg kvartila ispod 1. To znači da model regresijskog stabla povjerenja uz funkciju gubitka 3.19 daje bolje rezultate od Bühlmann-Straubovog modela koji ne uzima u obzir informacije o kovarijatama, odnosno ne particionira kolektiv. Primijetimo i da je relativna greška predikcije manja za log-normalnu distribuciju nego za gama distribuciju. Promotrimo i dijagrame pravokutnika za relativnu grešku predikcije za regresijsko stablo uz funkciju gubitka 3.20.



Slika 3.3: RPE uz funkciju gubitka 3.20 i gama distribuciju



Slika 3.4: RPE uz funkciju gubitka 3.20 i log-normalnu distribuciju

Iz gornjih dijagrama pravokutnika možemo zaključiti da metoda regresijskog stabla uz funkciju gubitka 3.20 daje slične rezultate kao uz funkciju gubitka 3.19. Osim toga, možemo primijetiti i da se relativna greška predikcije za podatke simulirane na treći način, gdje smo uveli slučajni efekt, ne razlikuje bitnije od podataka simuliranih na drugi način, gdje nam je varijanca ovisila samo o kovarijatama.

Bibliografija

[1] Olshen R.A. Breiman L., Friedman J.H. i Stone C.J., *Classification and Regression Trees, 1st Edition*, Chapman Hall, 1984.

[2] H. Bühlmann i A. Gisler, *A Course in Credibility Theory and its Applications*, Springer, 2005.

[3] Witten D. Hastie T. Gareth, J. i Tibshirani R., *An Introduction to Statistical Learning: with Applications in R, 2nd Edition*, Springer, 2021.

[4] Tibshirani R. Hastie T. i Friedman J., *The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition*, Springer, 2016.

[5] Diao L. i Weng C., *Regression Tree Credibility Model*, North American Actuarial Journal (2019), br. 23:2, 169–196.

[6] Tse Yiu-Kuen, *Nonlife Actuarial Models Theory, Methods and Evaluation*, Cambridge University Press, 2009.

Sažetak

U ovom diplomskom radu predstavljamo teoriju povjerenja, granu aktuarske matematike koja ima ključnu ulogu u industriji osiguranja. Ona nam daje statistički okvir za procjenjivanje budućih potraživanja osiguranika prema osiguravatelju, odnosno budućih premija. U prvom dijelu ovog diplomskog rada upoznajemo se s osnovnim pojmovima teorije povjerenja kao što su individualna i kolektivna premija te uvodimo koncepte, pojmove i teoreme iz bayesovske statistike koja čini okosnicu teorije povjerenja. Pokazujemo da je Bayesova premija najbolji procjenitelj premije povjerenja, ali da ju je u praksi teško izračunati. Zbog toga za procjenu premije povjerenja uvodimo linearne procjenitelje premije povjerenja koje obrađujemo u detalje. U drugom poglavlju uvodimo metodu statističkog učenja, stablo odlučivanja, koju koristimo u trećem poglavlju da bismo uključili informacije o kovarijatama u procjenu premije povjerenja. Tu metodu nazivamo Regresijsko stablo povjerenja i detaljno je objašnjavamo u trećem poglavlju. Rad završavamo s praktičnim primjerom, gdje koristimo model regresijskog stabla povjerenja na simuliranim podacima.

Summary

In this Master's thesis, we present Credibility theory, a branch of actuarial mathematics that plays a crucial role in the insurance industry. It gives us a statistical framework for estimating the future claims of the insured and, therefore, future premiums. In the first part of this thesis, we get acquainted with the basic terms of the Credibility theory, such as individual and collective premium, and we introduce concepts, terms and theorems from Bayesian statistics, which form the backbone of the Credibility theory. We show that the Bayesian premium is the best estimator of the credibility premium, but it is difficult to calculate it explicitly. For this reason, we introduce linear estimators of credibility premium, which we discuss in detail. In the second chapter, we introduce a statistical learning method, the decision tree, which we use later to incorporate covariate information into estimating credibility premium. This method is called the Regression tree credibility model, which we explain in detail in the third chapter. We finish the paper with a practical example using this model on simulated data.

Životopis

Rođena sam 25.veljače 1998. godine u Dubrovniku. Nakon završetka Prirodoslovno-matematičke gimnazije u Metkoviću upisujem Preddiplomski studij Matematike na Prirodoslovno-matematičkom fakultetu u Zagrebu. Preddiplomski studij završavam 2019. godine te iste godine upisujem Diplomski studij Financijske i poslovne matematike na istom fakultetu.