SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Antonio Roić

TEORIJA POVJERENJA U AKTUARSTVU

Diplomski rad

Voditelj rada:
prof. dr. sc. Hrvoje Šikić

Zagreb, rujan, 2018.

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. ____________________________

Mojim roditeljima

Sadržaj

Uvod

Izraz povjerenje je u aktuarstvu originalno bio vezan za formule koje su bile konveksna kombinacija (ponderirani prosjeci) pojedinačnih i grupnih procjena pojedinačne premije rizika. Teorija povjerenja je, dakle, bila grana matematike osiguranja koja je istraživala načine za konstrukciju takvih formula, temeljenu na modelima. Engleski naziv za to je „credibility“, što se obično prevodi kao „vjerodostojnost“, ali kako u hrvatskoj literaturi i polju vjerojatnosti taj pojam ima drugo značenje, u ovom radu će se koristiti pojam „povjerenje“, jer je značenje tog pojma mnogo bliže engleskom. Prvi primjeri korištenja teorije povjerenja dolaze još s početka 20. stoljeća kada su aktuari morali balansirati iskustva o potraživanju pojedinog zaposlenika neke tvrtke s onima od drugih zaposlenika sa sličnim poslovima i uvjetima rada, da bi odredili premije za njihovo osiguranje.

Ovdje se suočavamo sa problemom procjene očekivanog budućeg broja potraživanja i ukupnog broja potraživanja za neki portfelj polica, na osnovi jako malog uzorka ili informacije x, ali gdje nam je druga kolateralna (i moguće korisna) informacija pri ruci. Pretpostavimo da imamo parametar koji nam je od interesa i označimo ga sa θ (što može biti npr. godišnja stopa potraživanja). Često imamo drugu (u nekim slučajevima značajnu) kolateralnu informaciju od tvrtke ili portfelja, koja je slične prirode i koja može biti korisna za procjenu θ. Označimo sa $\hat{\theta}_s$ procjenu od θ na temelju informacije o uzorku x, a sa $\hat{\theta}_c$ procjenu od θ na temelju dostupne kolateralne informacije. U situaciji gdje je θ očekivanje, $\hat{\theta}_s$ može biti uzoračko očekivanje $\bar{x}$, a $\hat{\theta}_c$ neka a priori procjena (recimo μ_0) tog očekivanja. Glavno pitanje u ovakvom slučaju je kako možemo kombinirati ta dva izvora informacija da bi dobili dobru procjenu od θ i koliku težinu $\mathbf{Z}$ treba naš procjenitelj staviti na procjenu uzorka $\hat{\theta}_s$? Vrijednost od $\mathbf{Z}$ treba biti rastuća funkcija veličine informacije o uzorku (koju pribavljamo s vremenom), i, također, treba uzimati u obzir relativne vrijednosti uzorka i dostupnu kolateralnu informaciju. Procjenitelj povjerenja od θ je linearna kombinacija uzoračkog procjenitelja $\hat{\theta}_s$ i kolateralnog procjenitelja $\hat{\theta}_c$, oblika

$$\mathbf{Z}\hat{\theta}_s + (1 - \mathbf{Z})\hat{\theta}_c \tag{1}$$

(npr. $\mathbf{Z}\bar{x} + (1 - \mathbf{Z})\mu_0$), gdje je $\mathbf{Z}$ faktor povjerenja (vjerodostojnosti) koje stavljamo na uzoračkog procjenitelja $\hat{\theta}_s$. Općeniti izraz dan formulom (1) se zove premijska formula povjerenja. Tradicionalno se isticalo korištenje samo procjenitelja $\hat{\theta}_s$ oblika $\sum_1^n a_j x_j$, u premijskoj formuli povjerenja, ali, iako takvi procjenitelji imaju značajnu primjenu, nema teoretskog razloga da ne koristimo ostale uzoračke procjenitelje.

Uvidom u dalje naveden primjer će se dobiti motivacija za daljnja razmatranja. Imamo situaciju gdje su godišnja potraživanja (za krađe i materijalne štete) za 4 građevinske tvrtke iz nekog grada u posljednje 3 godine iznosila u prosjeku 24,000 i nas se traži da odredimo premiju θ (koja predstavlja očekivana godišnja potraživanja) da bismo ih osigurali za sljedeću godinu.Također je dostupna veoma velika količina podataka (u smislu prošlih godina i broja zgrada) o potraživanjima građevinskih tvrtki iz istog grada, gdje je posljednji prosječni broj potraživanja po tvrtki iznosi 8,000. Oslanjajući se samo na (izravno povezane, ali ograničene) informacije od te 4 tvrtke, mogli bismo procijeniti θ s 24,000, dok, koristeći samo kolateralne informacije od velike količine podataka od ostalih tvrtki, možemo procijeniti da θ iznosi $4 \cdot 8,000 = 32,000$. Pristup teorije povjerenja koristio bi oba izvora informacija i procjenio θ s nekim težinskim prosjekom oblika

$$\mathbf{Z}24,000 + (1 - \mathbf{Z})32,000$$

Tako dobijemo premiju koju bi te građevinske tvrtke morale platiti osiguravateljskom društvu sljedeće godine. Taj izračun je posao aktuara i, ako bi izračun bio precizan, osiguravajuće društvo za koje radi bi se zaštitilo od potencijalnih gubitaka izazvanih tim štetama (potraživanjima). Do ideje ovakvog načina računanja premija prvi je došao Whitney[6], koji u svome radu iz 1918. godine spominje upravo faktor povjerenja (vjerodostojnosti) **Z** i formulu (1). On se bavio problemom izračuna premija koje su morale plaćati tvrtke za osiguranje svojih radnika od ozljeda na radu. Predložio je da **Z** bude oblika $n/(n + K)$, gdje je K konstanta. Smatrao je da bi svaki aktuar trebao odrediti vrijednost od K po svojoj vlastitoj procjeni, a ne po nekoj unaprijed poznatoj matematičkoj formuli. Zbog toga je teorija povjerenja stagnirala godinama, sve do pojave Buhlmanna[3] koji je također zagovarao

$$\mathbf{Z} = \frac{n}{n + K},$$

ali ovdje je K (tzv. Buhlmannov parametar povjerenja) omjer očekivanja varijance procesa i varijance hipotetskih očekivanja.

U ovome će radu prvo biti objašnjeni (odnosno ponovljeni) neki osnovni pojmovi iz Bayesovske statistike koja će biti važna za razumijevanje nekih osnovnih pojmova koje koristi teorija povjerenja. Objasnit će se razlika između uobičajene statistike i Bayeosovske statistike te ćemo se upoznati s razlikama i sličnostima između a priori i a posteriori distribucije[5, 1]. Objasnit će se pojam konjugiranih familija i upoznat ćemo se s funkcijama gubitka koje su jako bitne u samoj praksi [2]. Zatim će se preći na temeljni cilj rada, a to je istražiti različite pristupe problemu povjerenja i načine rješavanja istog. Prvi pristup koji ćemo obraditi zove se klasični pristup i on se dijeli na pristup punog povjerenja i pristup parcijalnog povjerenja[8]. U pristupu punog povjerenja gleda se kada postoji dovoljno podataka da se može koristiti $\mathbf{Z} = 1$ (tj. kada se može dati puno povjerenje podacima). Pristup parcijalnog povjerenja gleda situacije u kojima to nije slučaj i daje metode za određivanje razumne vrijednosti za $\mathbf{Z}$. Taj se pristup još naziva i američki pristup teoriji povjerenja. Predloženo je da faktor povjerenja (vjerodostojnosti) $\mathbf{Z}$ bude oblika

$$\mathbf{Z} = \frac{P}{P+K},$$

gdje P predstavlja zarađene premije, a K se određuje procjenom. U Bayesovskom pristupu procjenitelj povjerenja se temelji na a posteriori distribuciji od θ s pomoću koje dobijemo a posteriori očekivanje koja nam predstavlja očekivanu vrijednost budućih premija[7]. Cilj je zapisati a posteriori očekivanje u obliku (1) gdje je faktor povjerenja (vjerodostojnosti) $\mathbf{Z}$ težina koju (bayesovski) a posteriori procjenitelj stavlja na uzorak informacija, a $1 - \mathbf{Z}$ je težina koju stavlja na a priori procjenitelja od θ. Pristup koji su popularizirali Buhlmann i Straub[3, 4] jest zapravo empirijski Bayesovski pristup. Ovdje se koriste dostupni podaci da bi se procijenili prethodni parametri za model te se onda određuju procjenitelji za θ. Buhlmann predlaže već spomenutu formulu $\mathbf{Z} = n/(n+K)$ sa Buhlmannovim parametrom povjerenja K. Buhlmannov i Straubov model poopćava ovaj model na situacije kada neki faktori (npr. godišnji broj ugovaratelja) mogu varirati s vremenom. Na kraju donosimo zaključak o svim pojmovima i metodama koje smo proučili i njihovu primjenu u samoj praksi.

Poglavlje 1

Bayesovska statistika

Ako je θ nepoznati parametar povezan s populacijom ili slučajna varijabla, tada će frenkventistički statističar pokušati napraviti zaključke o θ na osnovi informacije o uzorku x. Bayesovski statističar, pak, uvijek vjeruje da postoji neka dodatna prijašnja informacija o θ, koja se može kombinirati s informacijom o uzorku x, da bi se dobili zaključci o θ. Ime je dobila po svećeniku Thomasu Bayesu i njegovom posmrtno objavljenom radu. Osnovna ideja Bayesovske statistike je da parametri statističkih modela moraju i sami biti modelirani kao slučajni. Pri tome mi i prije prikupljanja podataka imamo neku inicijalnu (više ili manje subjektivnu) ideju o distribuciji nepoznatog parametra. Bayesovski statističar će izraziti prijašnje znanje o nepoznatom parametru θ preko a priori distribucije sa odgovarajućim funkcijama gustoće i distribucije. Svaka informacija o uzorku x se promatra uvjetno na nepoznati parametar θ i koristi se $f_{\mathbf{X}|\theta}$ da se predstavi uvjetna gustoća. Definirajmo prvo uvjetnu gustoću.

Definicija 1.0.1. *Neka je (X, Y) slučajni vektor s funkcijom gustoće $f = f_{X,Y} \colon \mathbb{R}^2 \to \mathbb{R}$. Ako je $y \in \mathbb{R}$ takav da je $f_Y(y) > 0$, onda definiramo* uvjetnu funkciju gustoće

$$f(\cdot|y) = f_{X|Y}(\cdot|y) \colon \mathbb{R} \to \mathbb{R}$$

slučajne varijable X uz uvjet Y=y *formulom*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Da bi shvatili pristup Bayesovskog statističara moramo prvo uvesti neke definicije. Parametru θ pridružujemo vjerojatnosnu distribuciju, te on postaje slučajna varijabla, odnosno slučajni vektor u općenitom smislu. Označimo ga sa Θ. Jedna njegova realizacija bit će $\theta \in \Theta$.

Definicija 1.0.2. *Neka je Θ slučajni vektor. Neka je $\theta \in \Theta$ jedna njegova realizacija. Označimo sa $f_{X|\Theta}(x|\theta)$ uvjetnu funkciju gustoće slučajnog uzorka $\boldsymbol{X}$ i a priori funkciju gustoće $f_\Theta(\theta)$ parametra Θ. Definirajmo sljedeće distribucije:*

1. zajedničku distribuciju *slučajnog vektora $(\Theta, \boldsymbol{X})$ s gustoćom:*

$$f_{\Theta,X}(\theta, x) = f_{X|\Theta}(x|\theta) f_\Theta(\theta),$$

2. marginalnu distribuciju *od $\boldsymbol{X}$ s gustoćom:*

$$f_X(x) = \int f_{X|\Theta}(x|\theta) f_\Theta(\theta)\, d\theta = \int f_{\Theta,X}(\theta, x)\, d\theta,$$

3. a posteriori distribuciju *od Θ, kao uvjetnu distribuciju uz dano $\boldsymbol{X} = x$ s gustoćom:*

$$f_{\Theta|X}(\theta|x) = \frac{f_{\Theta,X}(\theta, x)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta) f_\Theta(\theta)}{f_X(x)}.$$

Cijela Bayesovska statistika se zasniva na a priori funkciji distribucije i njenom odabiru. Kako a priori funkciju distribucije biramo upravo na temelju danih informacija o parametru Θ, očito je da odabir dovoljno precizne a priori funkcije distribucije često neće biti trivijalan. Kako prethodna saznanja možemo na više načina ugraditi u a priori distribuciju, Bayesovska statistika je neizbježno subjektivna u tom odabiru, što dakako stavlja povećanu odgovornost na statističara koji mora pravdati svoj odabir. U praksi moramo voditi računa o tome da je u odabiru a priori distribucije lako napraviti grešku. Osim toga, problem Bayesovske statistike je taj što računanje a posteriori distribucije često može biti naporno i dugotrajno. U nekim slučajevima simulacijske metode mogu biti korisne u generiranju uzoraka iz takve distribucije. Također je važno primjetiti da je

$$f_{\Theta|\mathbf{X}}(\theta|x) \;\propto\; f_{\Theta,\mathbf{X}}(\theta, x) = f_{\mathbf{X}|\Theta}(x|\theta) f_\Theta(\theta) \tag{1.1}$$

gdje $\propto$ indicira „proporcionalno je kao funkcija od θ”. Dakle. $1/f_{\mathbf{X}}(x)$ je jednostavno faktor koji osigurava da $f_{\Theta|\mathbf{X}}(\theta|x)$, kao funkcija od θ, pod integralom daje 1. To znači da u puno slučajeva zapravo nije potrebno računati marginalnu distribuciju $f_{\mathbf{X}}(x)$ da bi se dobio opći oblik a posteriori distribucije.

Pogledajmo sada jedan primjer na kojem možemo vidjeti vezu između a priori i a posteriori distribucije i koji će nam biti uvod u sljedeće podpoglavlje o konjugiranim familijama.

Primjer 1.0.3. *Pretpostavimo da je $X \sim B(n,q)$ broj uspjeha u n nezavisnih ponavljanja slučajnog pokusa. Neka X poprima vrijednost $k \in \{0,\dots,n\}$. Po definiciji binomne distribucije je*

$$f(k|q) = \binom{n}{k} q^k (1-q)^{n-k}.$$

Pretpostavimo da je a priori distribucija parametra q beta distribucija s parametrima $\alpha, \beta > 0$, tj. q ima a priori gustoću

$$f(q) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1}(1-q)^{\beta-1}.$$

Ako pretpostavimo da je $X|p \sim B(n,q)$, zbog (1.1) znamo da a posteriori gustoća zadovoljava

$$f(q|k) \propto q^k(1-q)^{n-k} q^{\alpha-1}(1-q)^{\beta-1} = q^{k+\alpha-1}(1-q)^{n-k+\beta-1}.$$

Dakle, a posteriori distribucija je ponovno beta s parametrima $k+\alpha$ i $n-k+\beta$!

Ovaj primjer ćemo detaljno razraditi u sljedećem potpoglavlju, ovdje nam samo služi da vidimo vezu između a priori i a posteriori distribucije i da nam da motivaciju za dublju analizu te veze.

1.1 Konjugirane familije

Kada a priori i a posteriori distribucija dolaze iz iste parametarske familije kao u Primjeru 1.0.3., kažemo da su konjugirane u odnosu na danu uzoračku distribuciju. Konjugirane familije su nam jako korisne jer je u tom slučaju obično laške pronaći a posteriori distribuciju. Sljedeći primjer je proširenje Primjera 1.0.3. i pokazuje nam da je familija beta distribucija konjugirana za uzorke iz binomne razdiobe. Kasnije će se utvrditi da je familija normalnih distribucija konjugirana za uzorke iz normalne distribucije i da je familija gama distribucija konjugirana za uzorke iz Poissonove distribucije. Matematički je privlačno raditi s konjugiranim familijama, ali je važno napomenuti da su one iznimka i da se u praksi često moraju koristiti simulacije da bi se dobila a posteriori distribucija!

Primjer 1.1.1 (Binomni | Beta model)**.** *Pretpostavimo da je q vjerojatnost događaja koji nam je od interesa. Familija beta distribucija je „bogata“ i često možemo izabrati takvu distribuciju Beta(α,β) koja će predstavljati a priori procjenu nepoznate vrijednosti od q.*

Takva distribucija ima očekivanje μ*, varijancu* σ^2 *i funkciju gustoće* $f(q)$ *dane, respektivno, s*

$$\mu = \frac{\alpha}{\alpha+\beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad i \quad f(q) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1}(1-q)^{\beta-1} \tag{1.2}$$

za $0 < q < 1$*, i* $\alpha, \beta > 0$*. Pretpostavimo sada da možemo promatrati i neku uzoračku informaciju x povezanu sa q, preciznije, da imamo binomno zapažanje koje je broj uspjeha* $\sum x_j$ *u nizu od n nezavisnih pokusa, s vjerojatnošću uspjeha q. Iz jednadžbe (1.1) slijedi da je a posteriori distribucija za q oblika*

$$\begin{aligned} f(q|x) &\propto \binom{n}{\sum x_j} q^{\sum x_j}(1-q)^{n-\sum x_j} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1}(1-q)^{\beta-1} \\ &\propto q^{\alpha-\sum x_j-1}(1-q)^{\beta+n-\sum x_j-1} \end{aligned}$$

Dakle, kada je Beta(α,β) *a priori distribucija za q i kada je uzoračka distribucija* $B(n,q)$*, tada je a posteriori distribucija za q zapravo distribucija Beta*$(\alpha + \sum x_j, \beta + n - \sum x_j)$*. To je upravo ono što smo vidjeli u Primjeru 1.0.3.. Zato se ovaj slučaj zove Binomni | Beta model. Očekivanje a posteriori distribucije za q može biti zapisano u obliku*

$$\begin{aligned} \frac{\alpha + \sum x_j}{\alpha+\beta+n} &= \frac{n}{\alpha+\beta+n}\frac{\sum x_j}{n} + \frac{\alpha+\beta}{\alpha+\beta+n}\frac{\alpha}{\alpha+\beta} \\ &= \frac{n}{\alpha+\beta+n}\frac{\sum x_j}{n} + \left(1 - \frac{n}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta} \\ &= \mathbf{Z}\bar{x} + (1-\mathbf{Z})\mu_0, \end{aligned}$$

gdje je $\mu_0 = \alpha/(\alpha+\beta)$ *a priori očekivanje (znamo iz (1.2)). Dakle, a posteriori očekivanje može biti zapisano kao procjena povjerenja za q s faktorom povjerenja (vjerodostojnosti)* $\mathbf{Z} = n/(\alpha+\beta+n)$*.*

Pogledajmo sada kako to možemo iskoristiti u praksi. Razmotrimo portfelj životnog osiguranja koji se sastoji od 400 *polica, svaka neovisna u odnosu na smrtnost. Pretpostavimo da će vjerojatnost smrti q sljedeće godine biti jednaka za sve ugovaratelje i da se a priori informacija o q može izraziti preko Beta distribucije s očekivanjem* 0.04 *i varijancom* 0.000191*. Koristeći izraze za očekivanje i varijancu Beta distribucije dane u jednadžbi (1.2), dobije se da*

$$\alpha/(\alpha+\beta) = 0.04 \quad \Leftrightarrow \quad \alpha = 4\beta/96\ , \ i$$

$$0.000191 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{4\cdot 96}{100^2(100\beta/96+1)},$$

daje $\beta = 192.045$ *i* $\alpha = 8.0018$. *Ako 36 ugovaratelja umre u sljedećoj godini, tada a posteriori očekivanje za q može biti zapisano u obliku*

$$\frac{\alpha+36}{\alpha+\beta+400} = \frac{400}{\alpha+\beta+400}\frac{36}{400} + \frac{\alpha+\beta}{\alpha+\beta+400}\frac{\alpha}{\alpha+\beta}$$

$$= 0.6666\cdot 0.09 + 0.3334\cdot 0.04 = 0.0733,$$

gdje je težina stavljena na uzoračko očekivanje 0.09 jednaka $\mathbf{Z} = 0.6666$.

Ovo nije jedini primjer konjugirane familije, u poglavlju o Bayesovskoj statistici susrest ćemo se s još nekim primjerima koji će nam biti zanimljivi.

1.2 Funkcije gubitka

U slučaju da želimo izdvojiti jedan procjenitelj nepoznatog parametra, koristimo funkcije gubitka L, koje uspoređuju našu procjenu parametra sa stvarnom vrijednošću parametra. Pretpostavimo da imamo slučajni uzorak $\mathbf{X}$ koji poprima vrijednost x i tražimo $g(x)$ aproksimaciju nepoznatog parametra θ. U tu svrhu procjenu kvalitete našeg procjenitelja označimo sa $l(g(x),\theta)$. Procjenitelja od θ ne možemo naći egzaktno, no možemo se nadati da će prosječni gubitak koji činimo u procjeni biti mali. Zbog toga kvalitetu procjenitelja ocjenjujemo preko funkcije rizika, tj.

$$R(l(x)) = \mathbb{E}[L(g(x),\theta] = \int L(g(x),\theta)f(\theta|x)\,d\theta.$$

Postoje različite funkcije gubitka, a mi ćemo ode spomenuti i koristiti tri: funkcija kvadratnog gubitka, funkcija apsolutnog gubitka i funkcija gubitka sve ili ništa.
Funkcija kvadratnog gubitka se najčešće koristi i definira se kao

$$l_1(g(x),\theta) = (g(x)-\theta)^2$$

i u ovom slučaju rizik minimizira upravo očekivanje od θ u odnosu na a posteriori distribuciju, pa je Bayesovski procjenitelj u ovom slučaju

$$\mathbb{E}_{f(\cdot|x)}(\theta) = \int \theta f(\theta|x)\,d\theta.$$

Funkcija apsolutnog gubitka je još jedna koja se također često koristi, a definira se kao

$$l_2(g(x), \theta) = |g(x) - \theta|\,.$$

Ako pretpostavimo da je $-\infty < \theta < +\infty$ realan parametar i stavimo $g = g(x)$, za funkciju rizika vrijedi

$$R(g) = R(g(x)) = \int_{-\infty}^{g} (g - \theta) f(\theta, x)\, d\theta - \int_{g}^{+\infty} (\theta - g) f(\theta, x)\, d\theta,$$

pa, koristeći pravila deriviranja kompozicije funkcija, nalazimo

$$\frac{\partial}{\partial g} R(g) = \int_{-\infty}^{g} f(\theta|x)\, d\theta - \int_{g}^{+\infty} f(\theta|x)\, d\theta,$$

tako da je kritična točka funkcije R upravo onaj g za koji je

$$\int_{-\infty}^{g} f(\theta|x)\, d\theta = \int_{g}^{+\infty} f(\theta|x)\, d\theta,$$

tj. mora biti $\mathbb{P}(g \geq \theta) = \mathbb{P}(g \leq \theta)$, odnosno Bayesovski procjenitelj za funkciju apsolutnog gubitka je medijan a posteriori distribucije. Kao treću spominjemo *funkciju gubitka sve ili ništa*, koja se definira kao

$$l_3(g(x), \theta) = \begin{cases} 0, & g(x) = 0 \\ 1, & \text{inače.} \end{cases}$$

Kako se ne može koristiti diferenciranje da bi se našao minimum funkcije rizika, uvodi se aproksimativna funkcija gubitka

$$l_3{}^{\varepsilon}(g(x), \theta) = \begin{cases} 0, & |g(x) - \theta| < \varepsilon \\ 1, & \text{inače.} \end{cases}$$

U slučaju da je a posteriori gustoća neprekidna funkcija, za male ε vrijedi aproksimacija

$$R^{\varepsilon}(l) = \mathbb{E}(l_3^{\varepsilon}(g, \theta)) = 1 - \int_{g-\varepsilon}^{g+\varepsilon} f(\theta|x)\, d\theta \approx 1 - 2\varepsilon f(g|x),$$

pa se minimum od R^{ε} dostiže približno za maksimalne vrijednosti $f(g|x)$, a takve točke se zovu mod a posteriori distribucije. Kako gubitak $l_3^{\varepsilon} \leftrightarrow l_3$ za $\varepsilon \leftrightarrow 0$, isto vrijedi i za minimum od R.

Poglavlje 2

Klasična teorija povjerenja

U klasičnom pristupu teoriji povjerenja postavlja se pitanje koliko nam je informacija o uzorku potrebno, da bismo se mogli u potpunosti osloniti na uzorak kod procjenjivanja parametra θ koji nam je od interesa. U takvom slučaju koristili bismo $\mathbf{Z} = 1$ te bi bilo navedeno da se daje puno povjerenje uzorku informacija. Prvi koji se bavio tim pitanjem bio je Mowbray (1914.). Njega je zanimalo koliko pokusa treba biti promatrano da bi se mogli pouzdati u naše podatke. Preciznije, kao jednog aktuara zanimalo ga je koliki minimalan broj radnika mora biti osiguran, tako da bi premiji po svakom radniku mogli dati puno povjerenje, tj. da premije budu takve da osiguravajuće društvo bude zaštićeno od gubitka. Ipak, koristivši jednadžbu (1) on je gledao samo slučajeve u kojma je $\mathbf{Z} = 1$ tj. slučajeve u kojima kolateralna informacija nije nužna za procjenu parametra. Druge aktuare poput Whitneya (1918.) zanimalo je što raditi u slučaju kada nemamo dovoljno informacija o uzorku da bismo dali puno povjerenje, tj. zanimalo ih je kakvo parcijalno povjerenje (vjerodostojnost) $\mathbf{Z}$ trebamo dati informacijama o uzorku (dok će se $\mathbf{Z} = 1$ dati kolateralnoj informaciji)?
Ovaj je pristup očito frenkventistički, a zbog toga što su ga u prijašnja vremena često koristili američki aktuari, još se naziva i Američka teorija povjerenja. Klasični pristup može biti koristan, ali postoje sumnje u vezi osnova na kojima se koncept parcijalnog povjerenja koristi i on se zato rijetko koristi danas.

2.1 Puno povjerenje

Ako je θ parametar od interesa, što zapravo znači da uzorku podataka dajemo puno povjerenje? Npr. uzorački mean $\bar{x}$ je ipak slučajna varijabla i ako se možemo pouzdati u tu vrijednost da bude dobar procjenitelj od θ, moramo se nadati da s velikom vjerojatnošću bude relativno blizu θ. U primjeni je parametar θ većinom ili godišnja stopa potraživanja λ ili premija $\mathbb{E}(S)$ koja predstavlja očekivana ukupna godišnja potraživanja, zato što su to dvije veličine koje su bitne osiguravajućim društvima jer se pravilnom procjenom broja potraživanja i veličine premije mogu zaštititi od gubitka i ostvariti profit. Definirajmo sada puno povjerenje.

Definicija 2.1.1. *Neka je θ nepoznati parametar. Neka su k i $0 < \alpha < 1$ konstante. Kažemo da procjenitelj $\hat{\theta}_s$ od θ zasnovan na uzorku podataka x ima* puno povjerenje (k, α) *za θ, ako vrijedi*

$$\mathbb{P}(|\hat{\theta}_s - \theta| \leq k\theta) = \mathbb{P}(\hat{\theta}_s - k\theta \leq \theta \leq \hat{\theta}_s + k\theta) \geq 1 - \alpha.$$

k i α su obično mali ($k = 0.05$ i $\alpha = 0.10$ su najčešći izbor).

Dakle, ako kažemo da uzorku podataka dajemo puno povjerenje (k,α), zapravo kažemo da su s vjerojatnošću $1 - \alpha$ relativne fluktuacije od $\hat{\theta}_s$ naspram θ ograničene na $k\theta$. Zato se ovaj pristup teoriji povjerenja često naziva Teorija povjerenja sa ograničenim fluktuacijama.

Primjer 2.1.2. *Ukupna godišnja potraživanja su modelirana prema Poissonovoj distribuciji S sa Poissonovim parametrom λ i veličinom potraživanja X. Parametar od interesa je premija $\theta = \mathbb{E}(S) = \lambda\mathbb{E}(X)$. Ukupni podaci o godišnjim potraživanjima $S_1, \ldots, S_r$ su skupljeni kroz r godina. Za dane vrijednosti od (k, α) koliko je aproksimativno potraživanja potrebno, da bismo $\bar{S} = \sum S_i/r$ dali puno povjerenje kao procjenitelju od $\mathbb{E}(S)$?*

$\bar{S}$ je aproksimativno normalan za velike $r\lambda$ i kako je $Var(\bar{S}) = \lambda\mathbb{E}(X^2)/r$ slijedi da je

$$\begin{aligned}\mathbb{P}(|\bar{S} - \theta| \leq k\theta) &= \mathbb{P}\left(\left|\frac{\bar{S} - \theta}{\sqrt{(\lambda\mathbb{E}(X^2)/r)}}\right| \leq \frac{k\theta}{\sqrt{(\lambda\mathbb{E}(X^2)/r)}}\right) \\ &= 2\phi\left(\frac{k\theta}{\sqrt{(\lambda\mathbb{E}(X^2)/r)}}\right) - 1 \\ &\geq 1 - \alpha\end{aligned}$$

$$\Leftrightarrow \quad r\lambda \quad \geq \quad \frac{z^2_{1-\alpha/2}}{k^2}\frac{\mathbb{E}(X^2)}{\mathbb{E}^2(X)} = \frac{z^2_{1-\alpha/2}}{k^2}\Big(1 + cv^2(X)\Big) \tag{2.1}$$

gdje je $\phi(z_{1-\alpha/2}) = 1 - \alpha/2$, *a* $cv(X) = \sqrt{Var(X)/\mathbb{E}^2(X)} = \mu_X/\sigma_X$ *koeficijent varijacije slučajne veličine potraživanja X.*
Pretpostavimo, npr. da je $(k, \alpha) = (0.10, 0.05)$ *i da su potraživanja eksponencijalna. Tada je* $\mathbb{E}(X^2) = 2\mathbb{E}^2(X)$ *i za puno povjerenje* $(0.10, 0.05)$ *potrebno je imati bar* $r\lambda \geq (1.96/0.10)^2 \cdot 2 = 769$ *potraživanja. Dakle, ako je godišnja stopa potraživanja reda* $\lambda = 160$, *onda postoji puno povjerenje* $(0.10, 0.05)$ *za premiju nakon* $r = 5$ *godina. Može se reći da se u teoriji procjenitelj* $\bar{S} = (S_1 + \cdots + S_r)/5$ *razlikuje od* $\theta = \mathbb{E}(S)$ *za najviše* $0.10\mathbb{E}(S)$ *s vjerojatnošću od najmanje* 95%.

Pretpostavimo sada da je parametar od interesa godišnja stopa potraživanja λ *umjesto premije* $\mathbb{E}(S) = \lambda\mathbb{E}(X)$. *Ako nam je namjera koristiti promatranu prosječnu godišnju stopu potraživanja* $\bar{N} = \sum N_i/r$ *da bismo procijenili* λ *s punim povjerenjem* (k, α) *onda je, kroz istu raspravu kao i gore, jasno da imamo*

$$\mathbb{P}(|\bar{N} - \lambda| \leq k\lambda) \geq 1 - \alpha \quad \Leftrightarrow \quad r\lambda \geq \frac{z_{1-\alpha/2}^2}{k^2}.$$

Vrijedi $\mathbb{E}(X^2)/\mathbb{E}^2(X) > 1$ *(osim u slučaju kad je X konstanta,) i iz toga slijedi da uvijek trebamo imati više iskustva o potraživanju da bismo dobro procjenili premiju, nego da bismo procjenili stopu potraživanja. Za eksponencijalno distribuirana potraživanja vrijedi* $\mathbb{E}(X^2)/\mathbb{E}^2(X) = 2$, *dok za lognormalnu distribuciju (s parametrima* μ *i* σ^2*) taj omjer poprima vrijednost* e^{σ^2}.

Primjer 2.1.3. *Razmotrimo situaciju gdje su ukupna godišnja potraživanja* $S = X_1 + \cdots + X_N$ *modelirana distribucijom, ali gdje N nije nužno Poissonov. Koristeći ponovno normalnu aproksimaciju za S i raspravljajući kao u Primjeru 2.1.2., kada je* $r = 1$, *jasno je da možemo postići puno povjerenje* (k, α) *koristeći S za procjenu* θ, *ako je* $\mathbb{E}^2(S)/Var(S) \geq z_{1-\alpha/2}^2/k^2$. *Npr. pretpostavimo da je S binomna gdje je broj potraživanja N iz* $B(m, q)$. *Tada je* $\mathbb{E}^2(S)/Var(S) = mq\mathbb{E}^2(X)/(\mathbb{E}(X^2) - q\mathbb{E}^2(X))$ *pa se puno povjerenje* (k, α) *za* $\theta = \mathbb{E}(S)$ *postiže ako je*

$$mq \geq \frac{z_{1-\alpha/2}^2}{k^2}\left(\frac{\mathbb{E}(X^2)}{\mathbb{E}^2(X)} - q\right).$$

Dakle, pretpostavimo da ukupna potraživanja S za jednu godinu dolaze od izlaganja m jedinica s frekvencijom potraživanja $q = 0.02$ *po izlaganju i da je tipično potraživanje X eksponencijalno. Tada dajemo S puno povjerenje* (k, α) *kao procjenitelju od* $\mathbb{E}(S) = \theta$ *ako je očekivani broj potraživanja* $mq \geq (z_{1-\alpha/2}^2/k^2) \cdot 1.98$. *S druge strane, ako nam je od interesa samo parametar* $\mathbb{E}(N) = mq$, *onda je očekivani broj potraživanja potreban za puno povjerenje* (k, α) *jednak* $mq \geq (z_{1-\alpha/2}^2/k^2) \cdot 0.98$.

U Primjerima 2.1.2. i 2.1.3. smo vdjeli smo kako dobiti očekivani broj potraživanja potreban za puno povjerenje. Pitanje koje se samo nameće je što učiniti ako nemamo dovoljan broj potraživanja? Tako dolazimo do pritupa parcijalnog povjerenja.

2.2 Parcijalno povjerenje

Iako se mogućnost davanja punog povjerenja može činiti kao idealno rješenje kod procjene parametra, često nemamo dovoljno podataka da bismo opravdali korištenje punog povjerenja. Prirodno nam se nameće pitanje koliko nam je tada povjerenja (ili težine) $\mathbf{Z}$ potrebno dati informaciji o uzorku? Najčešće korištena metoda za dodjeljivanje parcijalnog povjerenja je *pravilo drugog korijena*, gdje je faktor povjerenja (vjerodostojnosti) $\mathbf{Z}$ dan sa $\sqrt{n/n_{F(k,\alpha)}}$. n je očekivani (ili, u praksi, promatrani) broj potraživanja za podatke, a $n_{F(k,\alpha)}$ je broj potreban za postići puno povjerenje. Opravdanje za ovu formulu je ono da s ovom vrijednošću $\mathbf{Z}$ ograničavamo fluktuacije od $\mathbf{Z}\bar{X}$ oko $\mathbf{Z}\theta$ na ono koje je potrebno da podaci daju puno povjerenje za θ. Pretpostavimo da želimo procjeniti $\theta = \mathbb{E}(S) = \lambda\mathbb{E}(X)$ za Poissonovu distribuciju i da imamo r godina ukupnih podataka o potraživanju $S_1, \ldots, S_r$. Iz jednadžbe (2.1) znamo da za puno povjerenje (k, α) trebamo aproksimativno

$$n_{F(k,\alpha)} = \frac{{z_{1-\alpha/2}}^2}{k^2} \frac{\mathbb{E}(X^2)}{\mathbb{E}^2(X)}$$

potraživanja. Ako nemamo toliku količinu podataka, koristit ćemo $\mathbf{Z}\bar{S} + (1 - \mathbf{Z})\mu_0$, za procjenu $\theta = \mathbb{E}(S)$, gdje je $0 < \mathbf{Z} < 1$, a μ_0 je kolateralni procjenitelj od θ. Pretpostavimo da smo odlučili koristiti vrijednost od $\mathbf{Z}$ tako da $\mathbf{Z}\bar{S}$ kao procjenitelj od $\mathbf{Z}\theta$ ima jednako povjerenje (k, α) koje bismo zahtijevali od punog povjerenja ili, drugim riječima

$$\mathbb{P}(|\mathbf{Z}\bar{S} - \mathbf{Z}\theta| \leq k\theta) \geq 1 - \alpha.$$

Raspravljanjem kao u jednadžbi (2.1) vidimo da je to ekvivalentno

$$r\lambda = \mathbf{Z}^2 \frac{z^2_{1-\alpha/2}\mathbb{E}(X^2)}{k^2\mathbb{E}^2(X)} = \mathbf{Z}^2\, n_{F(k,\alpha)} \qquad \text{ili} \qquad \mathbf{Z} = \sqrt{\frac{r\lambda}{n_{F(k,\alpha)}}}.$$

Primjenimo sada to na par primjera koji predstavljaju hipotetske situacije s kojima bi se jedan aktuar mogao susresti.

Primjer 2.2.1. *Pretpostavimo da su očekivani ukupni gubici neke tvrtke prošle godine iznosili* $40,000,000$*, ali ukupni promatrani su iznosili* $47,000,000$*. Ovaj ukupni broj zasniva se na* $12,000$ *promatranih potraživanja, dok bi procjena tvrtkinih godišnjih gubitaka kojoj dajemo puno povjerenje zahtjevala oko* $18,500$ *potraživanja.*
Koristeći pravilo drugog korijena za parcijalno povjerenje, možemo procjeniti ukupna potraživanja za sljedeću godinu (u milijunima) s

$$\sqrt{\frac{12,000}{18,500}} \cdot 47 + \left(1 - \sqrt{\frac{12,000}{18,500}}\right) \cdot 40 = 45.638,$$

tj. procjenjujemo da će tvrtkini ukupni gubici iznositi $45,638,00$*. To je dosta velik gubitak za jednu tvrtku i u ovakvim slučaju bi vjerojatno osiguravajuće društvo odustalo od pružanja osiguranja toj tvrtki (i nagradilo aktuara što im je otrkio problem).*

Primjer 2.2.2. *Promatramo tvrtku koja se bavi prodajom nekretnina. Za vrijeme recesije našli su se u problemima i pretpostavimo da su im očekivani ukupni gubici prošle godine iznosili* $12,460,000$*, ali ukupni promatrani su iznosili* $14,445,000$*. Ovaj ukupni broj zasniva se na* $5,000$ *promatranih potraživanja, dok bi procjena njihovih godišnjih gubitaka kojoj dajemo puno povjerenje zahtjevala oko* $7,500$ *potraživanja.*
Kao i u prijašnjem primjeru, koristeći pravilo drugog korijena za parcijalno povjerenje, možemo procjeniti ukupna potraživanja za sljedeću godinu (u milijunima) s

$$\sqrt{\frac{5,000}{7,500}} \cdot 14.445 + \left(1 - \sqrt{\frac{5,000}{7,500}}\right) \cdot 12.46 = 14.08,$$

tj. procjenjujemo da će im ukupni gubici iznositi $14,080,00$*. Ti gubici su manji od ukupnih promatranih prošle godine pa vidimo da se tvrtkino poslovanje ipak poboljšalo.*

Povijesno gledano, ovaj i drugi slični pristupi parcijalnom povjerenju dovode do određe nih pitanja. Primjetimo da je, npr. korištenjem metode drugog korijena za određivanje faktora parcijalnog povjerenja (vjerodostojnosti) $\mathbf{Z}$ ne ograničavamo u vjerojatnosnom smislu varijaciju između $\mathbf{Z}\bar{X}$ i $\mathbf{Z}\theta$. Ipak, ovo nam ne govori koliko je poouzdan $\mathbf{Z}\hat{\theta}_s + (1 - \mathbf{Z})\hat{\theta}_c$ kao procjenitelj od $\mathbf{Z}\theta + (1 - \mathbf{Z})\theta$ i čak se pouzdanost od $\hat{\theta}_c$ kao procjenitelja od θ potpuno ignorira.

Poglavlje 3

Bayesovski pristup teoriji povjerenja

U ovom poglavlju upotrebljavamo znanje koje smo stekli u poglavlju o Bayesovskoj statistici. Kako je tamo spomenuto, ako je θ nepoznati parametar povezan s populacijom ili slučajna varijabla, tada Bayesovski statističar vjeruje da postoji neka dodatna prijašnja informacija o θ koju možemo kombinirati s informacijom o uzorku x, da bismo dobili zaključke o θ. Ovdje bismo htjeli iskoristiti Bayesovski pristup statistici prilikom procjene distribucije budućih šteta u nekom portfelju ili od nekog osiguranika. Odabrana će distribucija naravno utjecati i na visine premija, a njihovo određivanje jedan je od naših centralnih problema. Upravo to je jedna od zadaća aktuara, napraviti što bolju procjenu i tako zaštiti svoje osiguravajuće društvo od potencijalnih gubitaka. Kombinirati ćemo do sada dostupne podatke o samom osiguraniku, sa našim prethodnim znanjem. Kao što smo već spomenuli u poglavlju o Bayesovskoj statistici, odabir dovoljno precizne a priori funkcije distribucije često neće biti trivijalan i zato je pri odabiru a priori distribucije lako napraviti grešku. Upravo je ta subjektivnost pri odabiru jedna od najvećih kritika Bayesovskog pristupa.

3.1 Bayesovsko povjerenje

U Bayesovskom pristupu teoriji povjerenja sumirali bismo kolateralno znanje o nepoznatom parametru θ koristeći a priori distribuciju. Kao što smo već spomenuli, parametru θ pridružujemo tu distribuciju, te on postaje slučajna varijabla Θ. Izbor a priori distribucije za Θ je često subjektivna odluka i zato problem njenog određivanja nije trivijalan. Informativna i precizna informacija obično se ogleda u a priori distribuciji koja je centrirana blizu nepozantog parametra i/ili je precizna (ima jako malu varijabilnost). Ono što nas ovdje zanima je kako to povezati sa samim davanjem povjerenja uzorku podataka. Tu dolazimo do a posteriori distribucije jer za danu informaciju o uzorku x možemo odrediti a posteriori distribuciju za $\Theta|x$ i iskoristiti je za određivanje stvarne vrijednosti od θ (ili neku njenu funkciju). Najčešći procjenitelj je a posteriori očekivanje. Definirajmo sada uvjetno očekivanje jer nam je baš ono potrebno za izraziti a posteriori očekivanje.

Definicija 3.1.1. *Neka su X i Y slučajne varijable i neka slučajna varijabla X ima konačno očekivanje.* Uvjetno (matematičko) očekivanje slučajne varijable X uz dano Y=y *je definirano za sve $y \in \mathbb{R}$ takve da je $f_Y(y) > 0$, formulom*

$$\mathbb{E}(X|Y = y) = \begin{cases} \sum_x x f_{X|Y}(x|y), & \text{za diskretne slučajne varijable,} \\ \\ \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\, dx, & \text{za neprekidne slučajne varijable.} \end{cases}$$

Ako je $g : \mathbb{R} \to \mathbb{R}$ (izmjeriva) funkcija i slučajna varijabla $g(X)$ ima konačno očekivanje, onda za sve $y \in \mathbb{R}$ takve de je $f_Y(y) > 0$ vrijedi

$$\mathbb{E}(g(X)|Y = y) = \begin{cases} \sum_x g(x) f_{X|Y}(x|y), & \text{za diskretne slučajne varijable,} \\ \\ \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x|y)\, dx, & \text{za neprekidne slučajne varijable.} \end{cases}$$

Ako definiramo a priori očekivanje za Θ s $\mu_0 = \mathbb{E}(\Theta)$, tada je $\mathbb{E}(\Theta|\mathbf{X} = x)$ a posteriori očekivanje za Θ. Ono što bi željeli je izraziti to a posteriori očekivanje u obliku procjenitelja povjerenja za θ, tj. kao linearnu kombinaciju statistike θ_s (ovisno o x) i μ_0 , oblika

$$\mathbb{E}(\Theta|\mathbf{X} = x) = \mathbf{Z}\theta_s + (1 - \mathbf{Z})\mu_0. \tag{3.1}$$

Kada bi to mogli, procjenjivanje buduće premije svodilo bi se na određivanje faktora povjerenja $\mathbf{Z}$, koji bi nam u kombinaciji s uzorkom podataka dao traženu procjenu. Važno je napomenuti da nije uvijek moguće izraziti a posteriori očekivanje u tom obliku i to je jedan od nedostataka Bayesovskog pristupa. Mi ćemo proći kroz par primjera u kojima je moguće dobiti taj oblik. Sljedeći primjer je koristan za procjenjivanje očekivanih ukupnih potraživanja, u slučaju kada su normalno distribuirani.

Primjer 3.1.2 (Normalni | Normalni model)**.** *U Normalni | Normalni modelu pretpostavljamo da je $N(\theta, \sigma^2)$ uzoračka distribucija od $\mathbf{X}|\theta$, gdje je σ^2 poznato, i a priori informacija* *za $\theta \equiv \mu$ se može sažeti normalnom distribucijom $N(\mu_0, \sigma_0^2)$. Uzoračka distribucija za $\boldsymbol{X}|\theta$ je*

$$\begin{aligned} f_{\boldsymbol{X}|\Theta}(x|\theta) &= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_j-\theta)^2}{2\sigma^2}} \\ &\propto e^{-(n\theta^2 - 2\theta\sum_{j=1}^{n} x_j)/2\sigma^2} \\ &\propto e^{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2}, \end{aligned}$$

gdje $\propto$ označava "proporcionalno kao funkcija od θ". Zbog (1.1) znamo da je a posteriori distribucija od $\Theta|x$ normalna distribucija s gustoćom

$$\begin{aligned} f_{\Theta|\boldsymbol{X}}(\theta|x) &\propto f_{\boldsymbol{X}|\Theta}(x|\theta) f_{\Theta}(\theta) \\ &\propto e^{-\frac{n}{2\sigma^2}(\theta-\bar{x})^2} \frac{1}{\sqrt{2\pi}\sigma_0} e^{\frac{-(\theta-\mu_0)^2}{2\sigma_0^2}} \\ &\propto e^{-\frac{(\theta-\theta^*)^2}{2\sigma^{*2}}}, \end{aligned}$$

gdje su

$$\theta^* = \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) \Big/ \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right) \quad i \quad \sigma^{*2} = \left(\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2}\right)^{-1}, \tag{3.2}$$

očekivanje i varijanca, respektivno. Dakle, procjenitelj povjerenja za $\theta|x$ može biti zapisan kao

$$\begin{aligned} \theta^* &= \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) \Big/ \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right) \\ &= \left(\frac{n\bar{x}\sigma_0^2 + \sigma^2\mu_0}{\sigma^2\sigma_0^2}\right) \Big/ \left(\frac{n\sigma_0^2 + \sigma^2}{\sigma^2\sigma_0^2}\right) \\ &= \frac{n\bar{x}\sigma_0^2 + \sigma^2\mu_0}{n\sigma_0^2 + \sigma^2} \\ &= \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\bar{x} + \frac{\sigma^2/n}{\sigma_0^2 + \sigma^2/n}\mu_0 \end{aligned}$$

s faktorom povjerenja (vjerodostojnosti)

$$\boldsymbol{Z} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/n} = \frac{n/\sigma_0^2}{n/\sigma_0^2 + 1/\sigma_0^2} = \frac{n}{n + \sigma^2/\sigma_0^2}. \tag{3.3}$$

Vidimo da kao i u Binomni | Beta modelu iz Primjeru 1.1.1. možemo a posteriori očekivanje zapisati u obliku (3.1), ali je faktor povjerenja (vjerodostojnosti) **Z** *drugačijeg oblika. Iz izraza (3.3) jasno je da je u ovom modelu* **Z** *rastuća funkcija od n za fiksirane* σ_0^2 *i* σ^2*, kao i rastuća funkcija od* σ_0^2 *za fiksiran* $\sigma^2/n = Var(\bar{X})$*. Nadalje,* **Z** *je opadajuća funkcija omjera* $K = \sigma^2/\sigma_0^2$ *za fiksiranu veličinu uzorka n.*
Iz (3.2) se lako vidi da je

$$\sigma^{*2} \leq min(\sigma^2/n,\ \sigma_0^2),$$

stoga možemo očekivati da je a posteriori distribucija manje varijabilna od a priori distribucije, kao distribucija od θ*. Kako je* $\sigma^{*2} \leq \sigma^2/n$*, varijanca a posteriori se može napraviti koliko je god potrebno mala, tako što se uzme dovoljno velika veličina uzorka n .*

Primjer 3.1.3. *Prethodnih godina koristila se lognormalna distribucija da se modelira veličina potraživanja X u portfelju polica za osiguranje kućne imovine. Ako je* $Y = \log X$*, onda su do tog trenutka u osiguravajućem društvu zaključili da je* $Y \sim N(6.9, \sigma^2 = 0.1^2)$*. Nedavno su napravljene promjene u proceduri potraživanja i kao rezultat se očekuje da će se očekivanje* μ *za Y u budućnosti smanjiti, ali da će* σ^2 *ostati otprilike ista. Procjenjuje se da će nova vrijednost od* μ *biti otprilike* 6.4 *i da će s pouzdanošću od* 90% *biti iz intervala* $[6.35, 6.45]$*. Ako slučajni uzorak od* 40 *novih potraživanja daje* $\bar{y} = 6.51$*, koji je Bayesovski procjenitelj (s funkcijom kvadratnog gubitka) za* μ *? Kolika veličina uzorka (potraživanja) bi bila potrebna da se dobije a posteriori distribucija za* μ *sa standardnom devijacijom* ≤ 0.01*?*
Pretpostavljamo da μ *ima normalnu a priori distribuciju. pa kako je*

$$0.9 = \mathbb{P}\left(6.4 - 1.645\sigma_0 \leq \mu \leq 6.4 + 1.645\sigma_0\right),$$

slijedi da je $\sigma_0 = 0.05/1.645 = 0.0304$*. Dakle, a priori distribucija za* μ *je* $N(\mu_0 = 6.4, \sigma_0^2 = 0.0304^2)$*. A posteriori distribucija će također biti normalna, a iz Primjera 3.1.2. znamo da ima očekivanje dano s*

$$\begin{aligned}\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\bar{y} + \frac{\sigma^2/n}{\sigma_0^2 + \sigma^2/n}\mu_0 &= \frac{0.0304^2}{0.0304^2 + 0.1^2/40} \cdot 6.51 + \frac{0.1^2/40}{0.0304^2 + 0.1^2/40} \cdot 6.4 \\ &= 0.7870 \cdot 6.51 + 0.2130 \cdot 6.4 \\ &= 6.4866.\end{aligned}$$

Da bismo dobili a posteriori distribuciju sa standardnom devijacijom ≤ 0.01 *u ovoj situaciji, trebamo veličinu uzorka n takvu da vrijedi*

$$\left(\frac{1}{0.1^2/n} + \frac{1}{0.0304^2}\right)^{-1/2} \leq 0.01,$$

ili, ekvivalentno, da vrijedi $n \geq 90$.

Primjer 3.1.4 (Poisson | Gamma model)**.** *Za razliku od prethodnih modela Poisson | Gamma model se često koristi za modeliranje učestalosti potraživanja. Kao i kod Binomni | Beta modela i Normalni | Normalni modela, a posteriori distribucija može biti izražena u obliku (3.1). Pretpostavimo da je godišnji broj potraživanja za portfelj polica modeliran Poissonovom distribucijom s parametrom λ i da se prijašnja informacija o λ može sažeti gamma distribucijom $\Gamma(\alpha,\beta)$. To nam daje a proiori funkciju gustoće*

$$f_\Lambda(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}.$$

Prikupljamo informacije o uzorku za broj potraživanja $x = (x_1, \dots, x_n)$ tijekom n godina, gdje je x_j broj potraživanja u godini j. Uvjetna funkcija gustoće dana je s

$$f_{\boldsymbol{X}|\Lambda}(x|\lambda) \;\propto\; \frac{\prod_{j=1}^n \lambda^{x_j} e^{-\lambda}}{\prod_{j=1}^n x_j!}$$

Zbog (1.1) znamo da a posteriori gustoća od λ poprima oblik

$$\begin{aligned} f_{\Lambda|\boldsymbol{X}}(\lambda|x) &\propto f_{\boldsymbol{X}|\Lambda}(x|\lambda) f_\Lambda(\lambda) \\ &\propto \frac{\prod_{j=1}^n \lambda^{x_j} e^{-\lambda}}{\prod_{j=1}^n x_j!} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \\ &\propto \lambda^{\sum x_j + \alpha - 1} e^{-\lambda(n+\beta)}. \end{aligned}$$

Dakle, a posteriori distribucija od $\lambda|x$ je gamma distribucija $\Gamma(\sum x_j + \alpha, n + \beta)$. Kako za neki $\gamma \sim \Gamma(\alpha,\beta)$ vrijedi $\mathbb{E}(\gamma) = \alpha/\beta$, a posteriori očekivanje (Bayesovski procjenitelj od λ s funkcijom kvadratnog gubitka) može biti izražen u obliku

$$\begin{aligned} \mathbb{E}(\Lambda|\boldsymbol{X} = x) &= \frac{\sum x_j + \alpha}{n + \beta} \\ &= \frac{n}{n+\beta} \frac{\sum x_j}{n} + \frac{\beta}{\beta + n} \frac{\alpha}{\beta} \\ &= \mathbf{Z}\bar{x} + (1 - \mathbf{Z})\mu_0. \end{aligned}$$

Povjerenje (vjerodostojnost) koje dajemo podacima je $\mathbf{Z} = n/(n+\beta)$*, a* $\mu_0 = \alpha/\beta$ *je a priori očekivanje. Primjetimo da je za fiksiranu a priori distribuciju od* λ *, faktor povjerenja* $\mathbf{Z} = n/(n+\beta)$ *rastuća funkcija broja godina tijekom kojih promatramo podatke, dakle što više godina prikupljamo podatke, veće povjerenje možemo dati tim podacima.*
Pogledajmo sada to na jednom konkretnom primjeru. Pretpostavimo da je $\Gamma(3,1)$ *distribucija korištena kao a priori za godišnju stopu potraživanja* λ *i da je broj promatranih godišnjih potraživanja tijekom* 5 *godina dan s* $x = (2,3,6,0,3)$. *Tada je*

$$\Gamma(14+3, 5+1) = \Gamma(17,6)$$

a posteriori distribucija od $\lambda|x$*, a procjenitelj povjerenja od* λ *je*

$$\frac{17}{6} = \frac{5}{6}\cdot\frac{14}{5} + \frac{1}{6}\cdot\frac{3}{1}.$$

Dakle, očekivani broj potraživanja je $17/6 = 2.83$*, a faktor povjerenja (vjerodostojnosti)* $\mathbf{Z}$ *je* 0.83 *pa možemo reći da dajemo veliko povjerenje podacima.*

Primjer 3.1.5. *U ovom primjeru ćemo uzeti jednu hipotetsku situaciju koja nas veže u uz neki grad koji je često pogođen poplavama. Pretpostavimo da su kuće osigurane od štete nastale poplavom. U aktuarstvu je problem prepoznati kako su distribuirana potraživanja zbog nastalih šteta. U praksi se pokazuje da a priori procjenitelji mogu ozbiljno podcijeniti situaciju što kao rezultat daje male premije. One možda neće moći pokriti nastale šete te će tako izazvati velike gubitke osiguravajućem društvu. Ovdje je godišnji broj potraživanja Poissonova slučajna varijabla s* $\lambda = 160$*, ali to nam ovdje (a ni u stvarnoj situaciji obično neće biti) nije poznato. Umjesto toga, početna procjena o mogućoj vrijednosti od* λ *je izražena distribucijom* $\Gamma(120,1)$ *koja očito podcjenjuje realnu situaciju. n godina su prikupljani podaci (n=1, . . . ,12). Iz Primjera 3.1.4. znamo da procjenitelj povjerenja (a posteriori očekivanje) za* λ *može biti izražen kao*

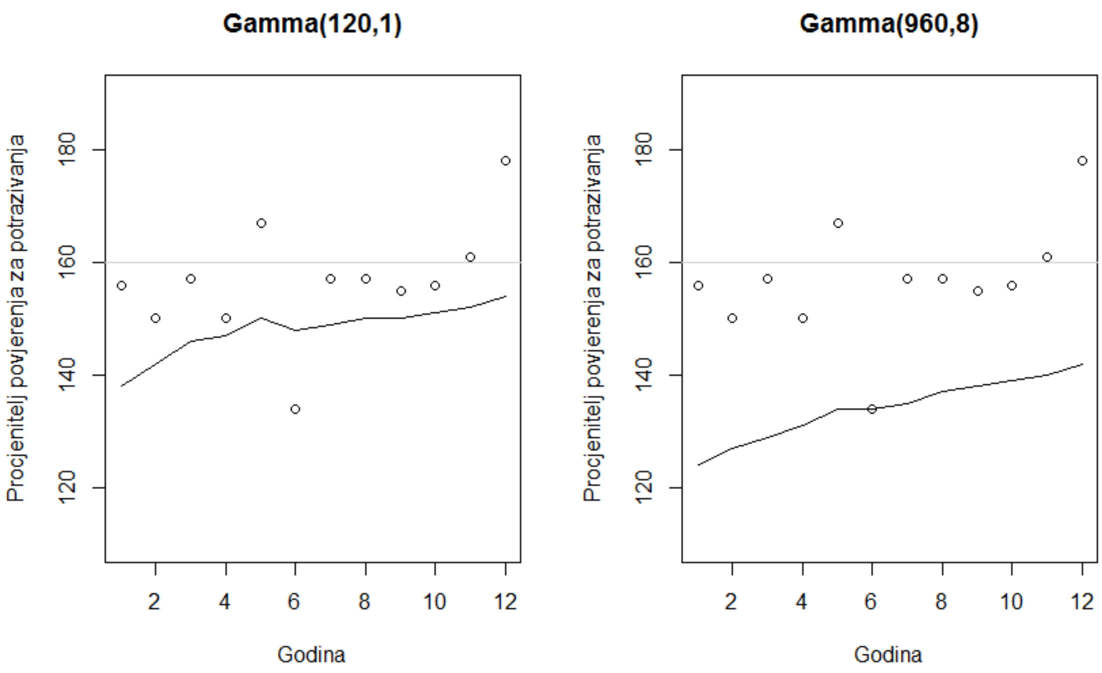
$$\frac{\sum_{j=1}^{n} x_j + \alpha}{n+\beta} = \frac{\sum_{j=1}^{n} x_j + 120}{n+1} = \frac{n}{n+1}\,\bar{x} + \frac{1}{n+1}\,\frac{120}{1},$$

gdje faktor povjerenja (vjerodostojnosti) $\mathbf{Z} = n/(n+1)$ *brzo konvergira prema* 1. *To je u ovom trenutku jako bitno jer je a priori informacija o* λ *daje malu pouzdanost za vrijednosti velike kao što je stvarna vrijednost* 160. *Primijetimo da bismo, da smo koristili* *precizniju a priori distribuciju za λ s istim očekivanjem, konvergencija faktora povjerenja ka* 1 *bila bi sporija. Npr. ako je a priori distribucija* $\Gamma(960, 8)$*, s istim očekivanjem* 120*, tada bi odgovarajući faktor povjerenja nakon n godina promatranja podataka bio* $n/(n+8) < n/n(n+1)$.

Godina	1	2	3	4	5	6	7	8	9	10	11	12
Potraživanja	156	150	157	150	167	134	157	157	155	156	161	178

A priori	Procjenitelji povjerenja											
$\Gamma(120, 1)$	138	142	146	147	150	148	149	150	150	151	152	154
$\Gamma(120, 1)$	124	127	129	131	134	134	135	137	138	139	140	142

Tablica 3.1:



Slika 3.1:

Uzorak stvarnih potraživanja skupljen tijekom 12 *godina prikazan je u Tablici 3.1, zajedno sa odgovarajućim procjeniteljima povjerenja za dvije različite a priori distribucije od* λ. *Podaci su dani u milijunima dolara. Na Slici 3.1 ti su procjenitelji nacrtani da bi pokazali različite brzine konvergencije pravoj vrijednosti* 160. *Kružićima su označena stvarna potraživanja, a linijom procjenitelji tih potraživanja. Na prvome grafu imamo situaciju u kojoj imamo a priori distribuciju* $\Gamma(120, 1)$. *Kao što se vidi na grafu procjenitelji povjerenja su relativno blizu pravih vrijednosti i jako brzo konvergiraju k* 160.
Na drugome grafu imamo drugačiju situaciju. Tu imamo a priori distribuciju $\Gamma(960, 8)$ *i vidimo da u tome slučaju procjenitelji povjerenja veoma podcjenjuju situaciju te da je brzina konvergencije k* 160 *veoma spora. To je pravi primjer "opasnosti" s kojom se aktuar može susresti jer ovako loša procjena može izazvati velike gubitke osiguravajućem društvu!*

Poglavlje 4

Empirijski Bayesovski pristup

Jedna od najboljih situacija u teoriji povjerenja je kada nam je dostupno n godina potraživanja $\mathbf{X} = (X_1, \ldots, X_n)$, a zanima nas X_{n+1} (broj ukupnog broja potraživanja koji ćemo promatrati sljedeće godine). U većini slučajeva to ćemo napraviti procjenjujući uvjetno očekivanje uz dano $\mathbf{X} = x$.
Pretpostavimo da imamo neki nepoznati parametar rizika θ koji određuje distribuciju naših promatranja. Čak i ako nam se čini da imamo malo informacija o θ, Bayesovski statističar će reći da uvijek možemo konstruirati a priori distribuciju za taj parametar. Takva a priori distribucija omogućava nam da vidimo parametar rizika kao slučajnu varijablu Θ. Parametar θ može biti nešto jednostavno kao npr. očekivanje, ali može biti i neka određena karakteristika osigurane grupe (npr. sredovječni vozači velikih vozila). Razlika u odnosu na obični Bayesovski pristup je što ovdje ne pretpostavljamo da nam je a priori distribucija parametra θ poznata. Kako nam i u praksi ta distribucija uglavnom neće biti poznata, ovakav pristup je puno bliži realnim situacijama i samim time veoma koristan u praksi.

U većini osnovnih modela ovog tipa, pretpostavljamo da godišnja promatranja od $(X_1, \ldots, X_n)$ imaju jednaku važnost u donošenju zaključaka o očekivanju od X_{n+1}. Pretpostavljamo da su, uvjetno na $\Theta = \theta$, slučajna opažanja $X_j|\theta, \quad j = 1, \ldots, n+1$ nezavisna i jednako distribuirana. Koristit ćemo $m(\theta)$ da bismo odredili očekivanje bilo kojeg od ovih opažanja uz dano θ, tj. $m(\theta) = \mathbb{E}(X_j|\theta)$ za $j = 1, \ldots, n+1$. Ovdje nas zapravo više zanima $m(\theta) = \mathbb{E}(X_{n+1})$ nego sam θ (iako je u nekim slučajevima, poput modela Normalni | Normalni ili Poisson | Gamma modela, $m(\theta)$ zapravo jednako θ). Za dani parametar θ, imat ćemo varijacije u opažanjima $X_1, \ldots, X_{n+1}$ koje označavamo s $s^2(\theta)$, tj. $Var(X_j|\theta) \equiv s^2(\theta)$. Primjetimo da su u ovom modelu i $m(\theta) = \mathbb{E}(X_j|\theta)$ i $s^2(\theta) = Var(X_j|\theta)$ nezavisni od $j = 1, \ldots, n+1$.

Naravno, kao funkcija nepoznatog parametra rizika $\Theta = \theta$, očekivanje $m(\theta)$ varira između rizika i zato ga možemo gledati kao slučajnu varijablu $m(\Theta)$. Očekivanje i varijaciju od $m(\Theta)$ označavat ćemo s $\mathbb{E}(m(\Theta))$ i $Var(m(\Theta)$, respektivno. Sljedeća svojstva će nam biti od velike koristi u procjenjivanju $\mathbb{E}(X_{n+1}|\mathbf{X})$, $\mathbb{E}(m(\theta)|\mathbf{X})$ ili X_{n+1}:

$$\mathbb{E}(X_j) = \mathbb{E}_\Theta(\mathbb{E}(X_j|\theta)) = \mathbb{E}(m(\Theta)), \tag{4.1}$$

$$\mathbb{E}(X_iX_j) = \mathbb{E}_\Theta(\mathbb{E}(X_iX_j|\theta) = \mathbb{E}_\Theta(\mathbb{E}(X_i|\theta)\mathbb{E}(X_j|\theta)) = \mathbb{E}(m^2(\Theta)), \quad \text{za} \quad i \neq j, \tag{4.2}$$

$$\mathbb{E}(X_j^2) = \mathbb{E}_\Theta(\mathbb{E}(X_j^2|\theta)) = \mathbb{E}_\Theta(s^2(\theta) + m^2(\theta)) = \mathbb{E}(s^2(\Theta)) + \mathbb{E}(m^2(\Theta)) \tag{4.3}$$

Promatranje informacije o uzorku $\mathbf{X} = (X_1, \dots, X_n)$ omogućava nam da proširimo svoju informaciju o θ i tako dobijemo a posteriori distribuciju od θ. Premija je očekivana vrijednost od X_{n+1} i to izražavamo s

$$\begin{aligned}
\mathbb{E}(X_{n+1}|\mathbf{X} = x) &= \mathbb{E}_{\Theta|\mathbf{X}}(\mathbb{E}(X_{n+1}|\Theta = \theta, \mathbf{X} = x)) \\
&= \int m(\theta) f_{\Theta|\mathbf{X}}(\theta|x)\, d\theta \\
&= \int (\int x f_{\mathbf{X}|\Theta}(x|\theta)\, dx) f_{\Theta|\mathbf{X}}(\theta|x)\, d\theta \\
&= \mathbb{E}(m(\theta)|\mathbf{X} = x).
\end{aligned}$$

To je a posteriori očekivanje od $m(\theta)$ uz dano $\mathbf{X} = x$.
$\mathbb{E}_{\Theta|\mathbf{X}}(m(\Theta))$ ćemo koristiti kao oznaku za slučajnu varijablu koja nakon opažanja poprima vrijednost $\mathbb{E}(m(\theta)|\mathbf{X} = x)$.

4.1 Bayesovski i linearni procjenitelj a posteriori očekivanja

U nekim slučajevima se a posteriori očekivanje $\mathbb{E}(m(\theta)|\mathbf{X} = x)$ može izraziti kao linearna funkcija opaženog uzorka x i onoga što zovemo a priori očekivanje $\mu_0 = \mathbb{E}(m(\Theta))$. Ovdje su nam od koristi modeli koje smo prošli kroz dosadašnje primjere. Npr. u Normalni | Normalni modelu iz Primjera 3.1.2, $m(\theta) = \theta$, pa se premija može izraziti kao linearna funkcija opažanja x dana s

$$\begin{aligned}\mathbb{E}(X_{n+1}|\mathbf{X}=x) &= \mathbb{E}(\Theta|\mathbf{X}=x)\\ &= \frac{\sigma_0^2}{\sigma_0^2+\sigma^2/n}\bar{x}+\frac{\sigma^2/n}{\sigma_0^2+\sigma^2/n}\mu_0.\end{aligned}$$

U Poisson | Gamma modelu iz Primjera 3.1.4., premija poprima oblik

$$\begin{aligned}\mathbb{E}t(X_{n+1}|\mathbf{X}=x) &= \mathbb{E}(\Lambda|\mathbf{X}=x)\\ &= \frac{n}{n+\beta}\frac{\sum x_j}{n}+\frac{\beta}{\beta+n}\frac{\alpha}{\beta}\\ &= \mathbf{Z}\bar{x}+(1-\mathbf{Z})\mu_0.\end{aligned}$$

Primjetimo da u oba modela vrijedi još i $m(\theta)=\theta$. U praksi se često događa da je a posteriori distribucija od $m(\theta)$ teška za odrediti, a čak se ni a posteriori očekivanje $\mathbb{E}(m(\theta)|\mathbf{X}=x)$ možda neće moći izraziti kao linearna kombinacija opažanja x i μ_0. Upravo to je bio najveći problem ovoga pristupa i rješenje tog problema dugo je godina okupiralo aktuare.

Do najboljeg rješenja je došao upravo Hans Buhlmann. On je razvio elegantni modela za Teoriju povjerenja, u kojem je formula povjerenja linearna kombinacija opažanja uzorka $\mathbf{X}=(X_1,\dots,X_n)$. Njegov glavni cilj bio je pronaći optimalog procjenitelja za $m(\theta)$ među linearnim funkcijama našeg uzorka tj. među procjeniteljima koji se mogu zapisati kao

$$a_0+a_1X_1+a_2X_2+\cdots+a_nX_n.$$

Kao kriterij optimalnosti ovog linearnog Bayesovskog procjenitelja i dalje je koristio kvadratnu funkciju gubitka, pa je njegov cilj zapravo bio pronaći linearne koeficijente $a_0, a_1, \dots, a_n$ koji minimiziraju $\mathbb{E}(X_{n+1}-(a_0+\sum_{j=1}^n a_jX_j))^2$ u odnosu na a posteriori distribuciju od $\Theta|\mathbf{X}$. Sada određujemo vrijednosti od $a_0, a_1, \dots, a_n$ da bismo minimizirali

$$\mathbb{E}\left(X_{n+1}-(a_0+\sum_{j=1}^n a_jX_j)\right)^2 \tag{4.4}$$

Važno je zapamtiti da su slučajna opažanja $X_j|\theta,\ \ j=1,\dots,n+1$ nezavisna i jednako distribuirana. Zato pretpostavljamo da u minimizaciji možemo promatrati samo slučaj kada je $a_i=a$ za sve $i=1,\dots,n$. Dakle, želimo pronaći a_0 i a koji minimiziraju $\mathbb{E}(X_{n+1}-(a_0+a\sum_{j=1}^n X_j))^2$.

Sljedeći teorem nam je od velike koristi.

Teorem 4.1.1.

$$\mathbb{E}\left(X_{n+1} - (a_0 + a\sum_{j=1}^{n} X_j)\right)^2 = (na^2+1)\mathbb{E}(s^2(\Theta)) + \mathbb{E}\left((na-1)m(\Theta)+a_0\right)^2. \tag{4.5}$$

Dokaz. Uvjetno na $\Theta = \theta$,

$$\mathbb{E}\left((X_{n+1} - (a_0 + a\sum_{j=1}^{n} X_j))^2 \middle| \theta\right) = \mathbb{E}\left((X_{n+1} - m(\theta))^2 \middle| \theta\right) + \mathbb{E}\left((m(\theta) - (a_0 - a\sum_{j=1}^{n} X_j))^2 \middle| \theta\right)$$

$$\left[\text{Kako je } \mathbb{E}\left((X_{n+1} - m(\theta))(m(\theta) - a_0 - a\sum_{j=1}^{n} X_j) \middle| \theta\right) = 0\right]$$

$$= s^2(\theta) + \mathbb{E}\left((-(na-1)m(\theta) - a_0 - a\sum_{j=1}^{n} X_j - m(\theta))^2 \middle| \theta\right)$$

$$= s^2(\theta) + a^2\sum_{j=1}^{n}\mathbb{E}\left((X_j - m(\theta))^2 \middle| \theta\right) + ((na-a)m(\theta)+a_0)^2$$

$$= (na^2+1)s^2(\theta) + ((na-1)m(\theta)+a_0)^2.$$

Uzimanjem očekivanja u odnosu na Θ dobijemo

$$\mathbb{E}_\Theta\left(\mathbb{E}\left((X_{n+1} - (a_0 + a\sum_{j=1}^{n} X_j))^2 \middle| \theta\right)\right) = (na^2+1)\mathbb{E}(s^2(\Theta)) + \mathbb{E}\left((na-1)m(\Theta)+a_0\right)^2$$

$$\mathbb{E}\left(X_{n+1} - (a_0 + a\sum_{j=1}^{n} X_j)\right)^2 = (na^2+1)\mathbb{E}(s^2(\Theta)) + \mathbb{E}\left((na-1)m(\Theta)+a_0\right)^2$$

□

Primjetimo da prvi izraz na desnoj strani jednadžbe (4.5) ne uključuje a_0, pa se zato $\mathbb{E}(X_{n+1} - (a_0 + a\sum_{j=1}^{n} X_j))^2$ minimizira ako je $a_0 = -(na-1)\mathbb{E}(m(\Theta))$. Iz tog razloga se nalaženje optimalne vrijednosti za a svodi na minimiziranje

$$(na^2+1)\mathbb{E}(s^2(\Theta)) + (na-1)^2\mathbb{E}(m(\Theta) - \mathbb{E}(m(\Theta)))^2 = (na^2+1)\mathbb{E}(s^2(\Theta)) + (na-1)^2 Var(m(\Theta)). \tag{4.6}$$

Minimizirajmo sada (4.6):

$$\frac{\partial}{\partial a}\Big((na^2+1)\mathbb{E}(s^2(\Theta)) + (na-1)^2 Var(m(\Theta))\Big) = 0$$

$$2na\mathbb{E}(s^2(\Theta)) + 2n^2 a Var(m(\Theta)) - 2nVar(m(\Theta)) = 0$$

$$na\mathbb{E}(s^2(\Theta)) + n^2 a Var(m(\Theta)) - nVar(m(\Theta)) = 0$$

$$a\Big(n\mathbb{E}(s^2(\Theta)) + n^2 Var(m(\Theta))\Big) = nVar(m(\Theta))$$

$$a = \frac{nVar(m(\Theta))}{n^2 Var(m(\Theta)) + n\mathbb{E}(s^2(\Theta))}$$

$$a = \frac{1}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))}$$

Dakle, optimalna vrijednost od a je

$$a = \frac{1}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))} = \frac{1}{n+K},$$

gdje je $K = \mathbb{E}(s^2(\Theta))/Var(m(\Theta))$. Kako je $s^2(\theta)$ varijanca od $X_j|\theta$, brojnik u izrazu K je prosječna vrijednost u odnosu na a priori distribuciju od Θ, pa se iz tog razloga često naziva očekivana vrijednost varijance procesa. Analogno, nazivnik zovemo varijanca hipotetskog očekivanja. Dakle, zaključujemo da je najbolji linearni procjenitelj od X_{n+1}, oblika $a_0 + a\sum_{j=1}^{n} X_j$ dan s

$$
\begin{aligned}
\mathbb{E}(X_{n+1} \mid \mathbf{X} = x) &= \sum_{j=1}^{n} \frac{1}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))} X_j \\
&\quad + \left(1 - \frac{n}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))}\right) \mathbb{E}(m(\Theta)) \\
&= \frac{1}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))} \bar{\mathbf{X}} + \left(1 - \frac{n}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))}\right) \mathbb{E}(m(\Theta)) \\
&= \mathbf{Z}\bar{\mathbf{X}} + (1 - \mathbf{Z})\, \mathbb{E}(m(\Theta)),
\end{aligned}
$$

gdje je je $\mathbf{Z} = n/(n + K)$.

Procjenitelji povjerenja (vjerodostojnosti) ovog oblika se često zovu Buhlmannovi procjenitelji povjerenja. Primjetimo da se Buhlmannov faktor povjerenja (vjerodostojnosti) $\mathbf{Z} = n/(n + K)$ približava 1 kada se veličina uzorka n povećava. Konačno, moramo uočiti da smo tri parametra u gornjem računu $\mathbb{E}(m(\Theta))$, $\mathbb{E}(s^2(\Theta))$ te $Var(m(\Theta))$ tretirali kao unaprijed poznata iako to u praksi nije slučaj. U Bayesovskim primjerima kao npr. Normalni | Normalni modelu ti parametri zaista jesu poznati jer nam je poznata a priori distribucija. Kako ovdje a priori distribuciju ne znamo, u praksi ove brojeve moramo procijeniti.

Primjer 4.1.2. *Male industrijske tvrtke osigurane su od požara i krađe. Prijašnja iskustva o potraživanjima takvih tvrtki dovela su do modeliranja ukupnih godišnjih potraživanja Poissonovom distribucijom. Tvrtke su su podijeljene na dva tipa i različite pretpostavke za Poissonovu distribuciju nalaze se u Tablici 4.1. Primjetimo da su potraživanja modelirana Pareto distribucijom. To je zato što ona obično dobro opisuje stvarne podatke pa se često koristi u modeliranju vrlo velikih šteta. Otprilike* 60% *tvrtki klasificirano je kao tip* θ_1*, a* 40% *kao tip* θ_2*. Potraživanja tijekom protekle tri godine od jedne relativno nove tvrtke dana su s*

$$\boldsymbol{S} = (S_1, S_2, S_3) = (0, 591 + 790, 740 + 846),$$

Ovdje su npr. u 3 *godine bila* 2 *potraživanja veličine* 740 *i* 846*, respektivno. Uz ovu informaciju, koliko bi sljedeće godine iznosila premija za ovu tvrtku?*

Tvrtka tipa θ	A priori za θ	Poissonova stopa λ	Potraživanje X_j
θ_1	0.6	1	$X_1 \sim Pareto(3,\ 1,200)$
θ_2	0.4	1.5	$X_2 \sim Pareto(3,\ 1,600)$

Tablica 4.1:

Premija od interesa je $(\mathbb{E}(S_4|\boldsymbol{S} = s)$*, a iako ne znamo kojem tipu ova tvrtka pripada, naša informacija* (s_1, s_2, s_3) *omogućava nam da proširimo a priori informaciju. Računamo* $m(\theta_1) = \mathbb{E}(S|\theta_1) = 1 \cdot (1,200/2) = 600$ *i slično* $m(\theta_2) = 1,200$*. Dakle, a priori očekivanje od* $m(\Theta)$ *je*

$$\begin{aligned}\mu_0 = \mathbb{E}(m(\Theta)) &= m(\theta_1) \cdot 0.6 + m(\theta_2) \cdot 0.4\\ &= 600 \cdot 0.6 + 1,200 \cdot 0.4\\ &= 840.\end{aligned}$$

Dobijemo

$$\begin{aligned}s^2(\theta_1) = \lambda_1 \mathbb{E}(X_1^2) &= 1 \cdot (600^2 + 600^2 \cdot 3)\\ &= 4 \cdot 600^2\end{aligned}$$

i slično $s^2(\theta_2) = 6 \cdot 800^2$*. Zbog toga je*

$$\begin{aligned}\mathbb{E}(s^2(\Theta)) &= s^2(\theta_1) \cdot 0.6 + s^2(\theta_2) \cdot 0.4\\ &= 4 \cdot 600^2 \cdot 0.6 + 6 \cdot 800^2 \cdot 0.4\\ &= 2,400,00.\end{aligned}$$

Konačno

$$\begin{aligned}Var(m(\Theta)) &= \mathbb{E}((m(\Theta))^2) - (\mathbb{E}(m(\Theta)))^2\\ &= (m(\theta_1))^2 \cdot 0.6 + (m(\theta_2))^2 \cdot 0.4 - (\mathbb{E}(m(\Theta)))^2\\ &= 600^2 \cdot 0.6 + 1,200^2 \cdot 0.4 - 840^2\\ &= 86,400,\end{aligned}$$

pa je faktor $K = 2,400,00/86,400 = 27.7778$*. S pomoću K izračunamo da je faktor povjerenja (vjerodostojnosti)* $\boldsymbol{Z} = 3/(3 + K) = 0.0975$*. Dakle, procjena tražene premije je*

$$\begin{aligned}\mathbb{E}(S_4|s) &= \boldsymbol{Z} \cdot \bar{s} + (1 - \boldsymbol{Z}) \cdot \mu_0\\ &= 0.0975 \cdot 989 + 0.9025 \cdot 840\\ &= 854.52.\end{aligned}$$

Vidimo da je faktor povjerenja (vjerodostojnosti) **Z** *jako mal, dakle ne dajemo veliko povjerenje prikupljenom uzorku podataka.*

Primjer 4.1.3. *Promatramo tvrtke koje su osigurane od šteta nastalih na zgradama. Prijašnja iskustva o potraživanjima takvih tvrtki dovela su do modeliranja ukupnih godišnjih potraživanja Poissonovom distribucijom. Tvrtke su podijeljene na tri tipa i različite pretpostavke za Poissionovu distribuciju nalaze se u Tablici 4.2. Ovdje se također koristi Pareto distribucija za modeliranje potraživanja. Otprilike* 50% *tvrtki klasificirano je kao tip theta*$_1$*,* 30% *tvrtki klasificirano je kao tip* θ_2*, a* 20% *kao tip* θ_3*. Potraživanja tijekom protekle četiri godine od jedne relativno nove tvrtke dana su s*

$$\mathbf{S} = (S_1, S_2, S_3, S_4) = (300 + 705 + 400, 520 + 635, 475, 702 + 235 + 528),$$

Uz ovu informaciju, koliko bi sljedeće godine iznosila premija za ovu tvrtku?

Tvrtka tipa θ	A priori za θ	Poissonova stopa λ	Potraživanje X_j
θ_1	0.5	1	$X_1 \sim Pareto(4,\ 1,200)$
θ_2	0.3	1.5	$X_2 \sim Pareto(4,\ 1,500)$
θ_3	0.2	2	$X_3 \sim Pareto(4,\ 1,800)$

Tablica 4.2:

Premija od interesa je $\mathbb{E}(S_5 \mid \mathbf{S} = s)$*, a iako ne znamo kojem tipu ova tvrtka pripada, naša informacija* (s_1, s_2, s_3, s_4) *omogućava nam da proširimo a priori informaciju. Računamo* $m(\theta_1) = \mathbb{E}(S|\theta_1) = 1 \cdot (1,200/3) = 400$ *i slično* $m(\theta_2) = 750$ *i* $m(\theta_3) = 1,200$*. Dakle, a priori očekivanje od* $m(\Theta)$ *je*

$$\begin{aligned}\mu_0 = \mathbb{E}(m(\Theta)) &= m(\theta_1) \cdot 0.5 + m(\theta_2) \cdot 0.3 + m(\theta_3) \cdot 0.2\\ &= 400 \cdot 0.5 + 750 \cdot 0.3 + 1,200 \cdot 0.2\\ &= 665.\end{aligned}$$

Dobijemo

$$\begin{aligned}s^2(\theta_1) = \lambda_1 \mathbb{E}(X_1^2) &= 1 \cdot (400^2 + 400^2 \cdot 4)\\ &= 5 \cdot 400^2\end{aligned}$$

i slično $s^2(\theta_2) = 7.5 \cdot 500^2$ *i* $s^2(\theta_3) = 10 \cdot 600^2$*. Zbog toga je*

$$\begin{aligned}
\mathbb{E}(s^2(\Theta)) &= s^2(\theta_1) \cdot 0.5 + s^2(\theta_2) \cdot 0.3 + s^2(\theta_3) \cdot 0.2 \\
&= 5 \cdot 400^2 \cdot 0.5 + 7.5 \cdot 500^2 \cdot 0.3 + 10 \cdot 600^2 \cdot 0.2 \\
&= 1,682,500.
\end{aligned}$$

Konačno

$$\begin{aligned}
Var(m(\Theta)) &= \mathbb{E}((m(\Theta))^2) - (\mathbb{E}(m(\Theta)))^2 \\
&= (m(\theta_1))^2 \cdot 0.5 + (m(\theta_2))^2 \cdot 0.3 + (m(\theta_3))^2 \cdot 0.2 - (\mathbb{E}(m(\Theta)))^2 \\
&= 400^2 \cdot 0.5 + 750^2 \cdot 0.3 + 1,200^2 \cdot 0.2 - 665^2 \\
&= 94,525,
\end{aligned}$$

pa je faktor $K = 1,682,500/94,525 = 17.7995$*. S pomoću K izračunamo da je faktor povjerenja (vjerodostojnosti)* $\mathbf{Z} = 4/(4+K) = 0.1835$*. Dakle, procjena tražene premije je*

$$\begin{aligned}
\mathbb{E}(S_4|s) &= \mathbf{Z} \cdot \bar{s} + (1 - \mathbf{Z}) \cdot \mu_0 \\
&= 0.1835 \cdot 4,204.899 + 0.8165 \cdot 665 \\
&= 1,314.5698.
\end{aligned}$$

Primjetimo da je faktor povjerenja (vjerodostojnosti) $\mathbf{Z}$ *relativno mal, dakle ne dajemo veliko povjerenje prikupljenom uzorku podataka. Također primjetimo da je premija jako velika, što može predstavljati problem za tu tvrtku, a možda čak i odustajanje od budućeg osiguravanja.*

4.2 Modeli u empirijskom Bayesovskom pristupu

U ovome potpoglavlju proučavamo dva modela kod kojih imamo znanje i podatke o N rizika tijekom n godina, sa ciljem da procijenimo premiju za sljedeću godinu za jedan ili sve rizike. Ti modeli su nam jako korisni jer nam omogućavaju procjenu premija za više rizika odjendom, tj. mogućnost grupnog osiguranja. Grupno osiguranje je jedna od opcija koje osiguravajuća društva nude klijentima (ugovarateljima osiguranja) koji žele osigurati više stvari odjendom. Tada ne bi plaćali pojedinačnu premiju, nego grupnu (kolektivnu) premiju što bi im uvelike olakšalo posao. Od dva modela s kojima se susrećemo, Model 1 je jednostavniji jer on ne uzima u obzir varirajući obujam poslovanja koji bi mogao pridonijeti opaženom broju potraživanja. Nepoznati parametar rizika ćemo označavati s θ, a

$m(\theta)$ će biti veličina od interesa. Ta veličina može biti i sam θ, ali postoji i puno drugih mogućnosti. U Modelu 1 koristimo dostupne podatke da procijenimo veličine $\mathbb{E}(m\Theta))$, $Var(m(\Theta))$ i $\mathbb{E}(s^2(\Theta))$ i tako dobijemo „Bulhammovski" procjenitelj povjerenja za pojedini rizik. Dakle, u Modelu 1 koristimo sve do sad naučeno u ovome poglavlju. U Modelu 2 pratimo pristup Buhlmanna i Strauba dopuštajući godišnje varijacije u rizicima i nakon standardizacije godišnjih ukupnih potraživanja preko tih varijacija, dobivamo epimirijske procjenitelje povjerenja za tražene premije.

Model 1

Podaci (slučajna opažanja) su oblika $\{\{X_{ij}\}_{i=1}^{N}\}_{j=1}^{n}$, gdje X_{ij} predstavljaju ukupna potraživanja u j-toj godini od i-tog rizika. Pretpostavimo da imamo nepoznati parametar rizika θ_i za i-ti rizik, što za $i = 1, \ldots, N$ predstavlja realizaciju slučajne varijable Θ. Označavamo sa $m(\theta_i)$ i $s^2(\theta_i)$ očekivanje i varijancu od $X_{ij}|\theta_i$ za $j = 1, \ldots, n$, respektivno. Iako pretpostavljamo da su $X_{ij}|\theta_i$ za $j = 1, \ldots, n$, nezavisne i jednako distribuirane za bilo koji rizik i, bezuvjetne slučajne varijable X_{ij} za $j = 1, \ldots, n$, nisu nužno nezavisne, iako su jednako distribuirane. Ne pretpostavljamo nikakav poseban oblik za distribuciju slučajne varijable $m(\Theta)$, ali ćemo označavati njeno očekivanje i varijancu sa $\mathbb{E}(m(\Theta))$ i $Var(m(\Theta))$, respektivno. Očekivanje varijabilnosti unutar rizika ili očekivana vrijednost varijance procesa označavamo sa $\mathbb{E}(s^2(\Theta))$. Kao što je već spomenuto, u Normalni | Normalni modelu $m(\theta) = \theta$, $s^2(\theta) = \sigma^2$, $\mathbb{E}(m(\Theta)) = \mu_0$, $Var(m(\Theta)) = \sigma_0^2$ i $\mathbb{E}(s^2(\Theta)) = \sigma^2$. Iz prethodnog potpoglavlja znamo da procjenitelj povjerenja za premiju za rizik i (očekivanje ukupnog broja potraživanja $\mathbb{E}(X_{i,n+1} \mid \mathbf{X} = x)$ za sljedeću godinu za rizik i za dane podatke ima oblik

$$\frac{1}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))}\bar{\mathbf{X}}_i + \left(1 - \frac{n}{n + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))}\right)\mathbb{E}(m(\Theta)).$$

U empirijskom Bayesovskom pristupu teoriji povjerenja procjenjujemo tri parametra: $\mathbb{E}(m(\Theta))$, $Var(m(\Theta))$ i $\mathbb{E}(s^2(\Theta))$. Neka je X_{ij} kao prije, $\bar{X}_i = \sum_{j=1}^{n} X_{ij}/n$ i $\bar{X} = \sum_{i=1}^{N} \bar{X}_i/N$, tada se može pokazati da su procjenitelji prikazani u Tablici 4.3 nepristrani.

Procjenitelje $\mathbb{E}(m(\Theta))$ i $\mathbb{E}(s^2(\Theta))$ smo mogli i intuitivno pretpostaviti (procjenitelj za $\mathbb{E}(s^2(\Theta))$ je samo prosjek varijanci uzorka u svakom riziku). Možda bismo mogli očekivati

Model 1: Nepristrani procjenitelji za $\mathbb{E}(m(\Theta))$, $Var(m(\Theta))$ i $\mathbb{E}(s^2(\Theta))$
Procjenitelj
$\mathbb{E}(\widehat{m(\Theta)}) = \bar{X}$
$\mathbb{E}(\widehat{s^2(\Theta)}) = \sum_{i=1}^{N}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_i\right)^2 / (N(n-1))$
$Var\widehat{(m(\Theta)}) = \sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)^2 / (N-1) - \sum_{i=1}^{N}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_i\right)^2 / (nN(n-1))$

Tablica 4.3:

da $\sum_{i=1}^{N}(\bar{X}_i - \bar{X})^2/(N-1)$ bude dobar procjenitelj za $Var(m(\Theta))$, ali faktor ispravka je potreban da bude nepristran.

Primjer 4.2.1. *Tablica 4.4 daje podatke o ukupnim potraživanja vezanim za motorna vozila u svakoj od tri regije neke države, kroz period od pet godina.*

Ukupna potraživanja vezana za motorna vozila, kroz period od pet godina.					
	Godina				
Regija	1	2	3	4	5
1	5,841	7,782	5,373	7,020	7,773
2	5,910	4,491	6,102	5,373	6,651
3	7,011	8,045	7,078	7,266	9,027

Tablica 4.4:

Koristeći Model 1 računamo faktor povjerenja (vjerodostojnosti) i premiju za sljedeću godinu, za svaku od tri regije. Za svaku od njih računamo očekivanje i varijancu, te dobijemo

$$(\bar{X}_1, \bar{X}_2, \bar{X}_3) = (6,757.8,\ 5,705.4,\ 7,685.4)$$

$$(s^2(\theta_1), s^2(\theta_2), s^2(\theta_3)) = (1,226,639.7,\ 669,642.3,\ 732,212.3)$$

Zatim procjenjujemo

$$\mathbb{E}\left(\widehat{m(\Theta)}\right) = \bar{X} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3}$$

$$= \frac{20,148.6}{3}$$

$$= 6,716.2$$

i

$$\mathbb{E}\left(\widehat{s^2(\Theta)}\right) = \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{\left(X_{ij} - \bar{X}_i\right)^2}{N(n-1)}$$

$$= \frac{1}{3(5-1)} \cdot 10,513,977.6$$

$$= 876,164.8.$$

Imamo još i

$$Var\widehat{(m(\Theta))} = \sum_{i=1}^{N} \frac{\left(\bar{X}_i - \bar{X}\right)^2}{N-1} - \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{\left(X_{ij} - \bar{X}_i\right)^2}{nN(n-1)}$$

$$= \sum_{i=1}^{N} \frac{\left(\bar{X}_i - \bar{X}\right)^2}{N-1} - \frac{\mathbb{E}\left(\widehat{s^2(\Theta)}\right)}{n}$$

$$= \frac{1}{3-1} \cdot 1,962,795.8 - \frac{876,164.8}{5}$$

$$= \frac{1}{2} \cdot 1,962,795.8 - 175,232.86$$

$$= 981,397.9 - 175,232.86$$

$$= 806,165.$$

Dakle, faktor povjerenja (vjerodostojnosti) (koji je jednak za svaku regiju) je

$$\mathbf{Z} = \frac{n}{n + \mathbb{E}(\widehat{s^2(\Theta)})/\widehat{Var(m(\Theta))}} = \frac{5}{5 + 1.0868} = 0.8214.$$

Premija za regiju 1 je tada

$$\mathbf{Z}\,\bar{X}_1 + (1 - \mathbf{Z})\bar{X} = 0.8214 \cdot 6,757.8 + 0.1786 \cdot 6,716.2 = 6,750.37,$$

za regiju 2

$$\mathbf{Z}\,\bar{X}_2 + (1 - \mathbf{Z})\bar{X} = 0.8214 \cdot 5,705.4 + 0.1786 \cdot 6,716.2 = 5,885.92,$$

i za regiju 3

$$\mathbf{Z}\,\bar{X}_3 + (1 - \mathbf{Z})\bar{X} = 0.8214 \cdot 7.685.4 + 0.1786 \cdot 6,716.2 = 7,512.35.$$

Primjetimo da je $\mathbf{Z}$ *relativno blizu* 1*, pa su zato i tražene premije za regiju 1,2 i 3 blizu vrijednostima od* $\bar{X}_1$*,* $\bar{X}_2$ *i* $\bar{X}_3$*.*

Primjer 4.2.2. *Tablica 4.5 daje podatke o ukupnim potraživanja vezanih za osiguranje kućanstva za četiri različite grupe ugovaratelja osiguranja, tijekom šest godina. Podaci su dani u milijunima dolara. Koristeći Model 1 izračunajmo procjenitelje premija za sljedeću godinu, za svaku grupu ugovaratelja.*

Ukupna potraživanja vezana za osiguranje kućanstva, kroz period od šest godina.						
	Godina					
Regija	1	2	3	4	5	6
1	206	146	271	178	136	162
2	144	284	310	218	266	301
3	64	57	43	97	132	110
4	204	186	248	222	188	204

Tablica 4.5:

Za svaku od regija računamo očekivanje i varijancu, te dobijemo

$$(\bar{X}_1, \bar{X}_2, \bar{X}_3, \bar{X}_4) = (183.2, 253.8, 83.8, 208.6)$$

$$\left(s^2(\theta_1), s^2(\theta_2), s^2(\theta_3), s^2(\theta_4)\right) = (2,463.4, 3,956.9, 1,191.7, 541.8)$$

Zatim procjenjujemo

$$\mathbb{E}\left(\widehat{m(\Theta)}\right) = \bar{X} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4}{4}$$

$$= \frac{729.4}{4} = 182.35$$

i

$$\mathbb{E}\left(\widehat{s^2(\Theta)}\right) = \sum_{i=1}^{N}\sum_{j=1}^{n}\frac{\left(X_{ij} - \bar{X}_i\right)^2}{N(n-1)}$$

$$= \frac{1}{4(6-1)} \cdot 75,769$$

$$= 3,788.45.$$

Imamo još i

$$Var\widehat{(m(\Theta))} = \sum_{i=1}^{N}\frac{\left(\bar{X}_i - \bar{X}\right)^2}{N-1} - \sum_{i=1}^{N}\sum_{j=1}^{n}\frac{\left(X_{ij} - \bar{X}_i\right)^2}{nN(n-1)}$$

$$= \sum_{i=1}^{N}\frac{\left(\bar{X}_i - \bar{X}\right)^2}{N-1} - \frac{\mathbb{E}\left(\widehat{s^2(\Theta)}\right)}{n}$$

$$= \frac{1}{4-1} \cdot 15,508.83 - \frac{3,788.45}{6}$$

$$= \frac{1}{3} \cdot 15,508.83 - 631.4$$

$$= 5,169.61 - 631.4$$

$$= 4,537.6.$$

Dakle, faktor povjerenja (vjerodostojnosti) (koji je jednak za svaku regiju) je

$$\mathbf{Z} = \frac{n}{n + \mathbb{E}(\widehat{s^2(\Theta)})/Var(\widehat{m(\Theta)})} = \frac{6}{6 + 0.8349} = 0.8778.$$

Premija za regiju 1 je tada

$$\mathbf{Z}\,\bar{X}_1 + (1 - \mathbf{Z})\bar{X} = 0.8778 \cdot 183.2 + 0.1222 \cdot 182.35 = 183.09,$$

za regiju 2

$$\mathbf{Z}\,\bar{X}_2 + (1 - \mathbf{Z})\bar{X} = 0.8778 \cdot 253.8 + 0.1222 \cdot 182.35 = 245.07,$$

za regiju 3

$$\mathbf{Z}\,\bar{X}_3 + (1 - \mathbf{Z})\bar{X} = 0.8778 \cdot 83.8 + 0.1222 \cdot 182.35 = 95.84,$$

i za regiju 4

$$\mathbf{Z}\,\bar{X}_4 + (1 - \mathbf{Z})\bar{X} = 0.8778 \cdot 208.6 + 0.1222 \cdot 182.35 = 205.39.$$

Primjetimo da je kao i u Primjeru 4.3.2. $\mathbf{Z}$ *relativno blizu* 1, *pa su zato i tražene premije za regiju 1,2,3 i 4 blizu vrijednostima od* $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$ *i* $\bar{X}_4$.

Primjer 4.2.3. *Osiguravateljsko društvo ima četiri različite grupe ugovaratelja osiguranja za osiguranje kućanstva. Ukupni boj potraživanja tijekom pet uzastopnih godina izračunat je za svaku grupu zasebno te su ti brojevi dani u Tablici 4.6. Podaci su dani u milijunima dolara. Ovdje nam je* $i \in \{A, B, C, D\}$. *Cilj je procijeniti premiju za grupu A za sljedeću godinu. Dvije metode, Bayesovska i empirijska Bayesovska su predložene od strane dva analitičara.*

Prvi analitičar želi procijeniti premiju za grupu A koristeći empirijski Bayesovski pristup teoriji povjerenja. Drugi analitičar ne smatra da je informacija o potraživanjima grupa B, C i D relevantna, pa ih odlučuje ignorirati. Ali, smatra da ima a priori informaciju za premiju $m(\theta_A)$ *koja je* $N(90, 10^2)$ *i također smatra da se podaci od grupe A skupljeni tijekom pet godina mogu gledati kao slučajni uzorak iz* $N(\theta_A, 35^2)$ *distribucije.*
Označimo sa $m(\hat{\theta}_A^1)$ *i* $m(\hat{\theta}_A^2)$ *procjenitelje od* $m(\theta_A)$ *za prvog i drugog analitičara respektivno. Izračuni za ukupna potraživanja za svaki od četiri rizika tijekom pet godina dani sui u Tablici 4.7. Ovdje* X_{ij} *označava ukupna potraživanja u godini* j *za rizik* i. *To daje*

Ukupna potraživanja vezana za četiri različite grupe ugovaratelja.					
	Godina				
	1	2	3	4	5
Grupa A	58	42	98	130	64
Grupa B	204	186	246	222	186
Grupa C	183	153	215	171	147
Grupa D	78	104	77	116	118

Tablica 4.6:

Rizik	$\bar{X}_i$	$\sum_{j=1}^{5} (X_{ij} - \bar{X}_i)^2/4$
A	78.4	1248.8
B	208.8	655.2
C	173.8	735.2
D	98.6	399.8

Tablica 4.7:

sljedeće procjenitelje

$$\mathbb{E}(\widehat{m(\Theta)}) = \bar{X} = 139.9,$$

$$\mathbb{E}(\widehat{s^2(\Theta)}) = \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{\left(X_{ij} - \bar{X}_i\right)^2}{N(n-1)} = 759.7$$

i

$$
\begin{aligned}
Var\widehat{(m(\Theta))} &= \sum_{i=1}^{N} \frac{\left(\bar{X}_i - \bar{X}\right)^2}{N-1} - \sum_{i=1}^{N}\sum_{j=1}^{n} \frac{\left(X_{ij} - \bar{X}_i\right)^2}{nN(n-1)} \\
&= \sum_{i=1}^{N} \frac{\left(\bar{X}_i - \bar{X}\right)^2}{N-1} - \frac{\mathbb{E}\left(\widehat{s^2(\Theta)}\right)}{n} \\
&= 3,794.7817 - \frac{759.7}{5} \\
&= 3,642.83.
\end{aligned}
$$

Zbog toga nam faktor povjerenja (vjerodostojnosti) iznosi

$$
\mathbf{Z} = \frac{5}{5 + \mathbb{E}(\widehat{s^2(\Theta)})/Var\widehat{(m(\Theta))}} = \frac{5}{5 + 0.2086} = 0.96.
$$

Dakle, procijenjena premija za prvog analitičara je

$$
\begin{aligned}
m(\hat{\theta}_A^1) &= \mathbf{Z}\bar{X}_A + (1 - \mathbf{Z})\bar{X} \\
&= 0.96 \cdot 78.4 + 0.04 \cdot 139.9 \\
&= 75.26 + 5.60 \\
&= 80.86.
\end{aligned}
$$

A posteriori distribucija za drugog analitičara je

$$
N\left(\frac{(8/35^2)\cdot 78.4 + 90/10^2}{5/35^2 + 1/10^2} = 86.64, \left(\frac{1}{35^2/5} + \frac{1}{10^2}\right)^{-1}\right),
$$

pa je njegova procijenjena premija $m(\hat{\theta}_A^2) = 86.64$. *Vidimo je procijenjena premija pvog analitičara manja nego kod drugog analitičara, ali da je bliža* $\bar{X}_A$, *što proizlazi iz činjenice da je faktor povjerenja (vjerodostojnosti)* $\mathbf{Z}$ *koje on daje uzorku podataka jako blizu* 1.

Model 2

U praksi obično nemamo samo informaciju o ukupnim godišnjim potraživanjima nego i o ostalim relevantnim stvarima, kao što je npr. broj polica ugovaratelja osiguranja. Podaci mogu izgledati kao u Tablici 4.8

Godina	1	2	. . .	n
Ukupna potraživanja	Y_1	Y_2	. . .	Y_n
Opseg poslovanja	P_1	P_2	. . .	P_n

Tablica 4.8:

P_j će obično označavati obujam ili veličinu poslovanja u j-toj godini i znajući P_{n+1} željeli bismo procijeniti ukupna potraživanja Y_{n+1} za sljedeću godinu. Definirajmo slučajnu varijablu $X_j = Y_j/P_j$ za $j = 1, \dots, n, n+1$, gdje je θ parametar rizika. $\mathbb{E}(X_j \mid \theta) = m(\theta)$ se može interpretirati kao prosječna veličina potraživanja po jedinici rizika. Tada bismo mogli procijeniti $m(\theta)$ sa $\widehat{m(\theta)}$ pa zatim predvidjeti Y_{n+1} s $P_{n+1}\,\widehat{m(\theta)}$. Naša motivacija za uvođenje X_j (umjesto da koristimo samo Y_j) je dobiti neki oblik standardizirane mjere za veličinu potraživanja u slučaju kada obujam poslovanja P_j varira tijekom godina.
Slično kao u Modelu 1, u Modelu 2 pretpostavljamo da ako nam je dan parametar rizika θ za neku tvrtku, da su onda $X_1, \dots, X_n, X_{n+1}$ nezavisne slučajne varijable (ali nisu nužno jednako distribuirane, kao u Modelu 1). Ipak, pretpostavljamo da imaju jednako očekivanje (tj. da je prosječna veličina potraživanja po jedinici rizika jednaka svake godine) i označavamo ga sa $m(\theta) = \mathbb{E}(X_j \mid \theta)$. Također pretpostavljamo da je $s^2(\theta) = P_j\, Var(X_j \mid \theta)$ nezavisna od j pa zato $Var(X_j \mid \theta) = s^2(\theta)/P_j$ varira tijekom godina zbog variranja obujma poslovanja.
U praksi nam neće biti poznata vrijednost od $m(\theta)$ za pojedinu tvrtku, ali ćemo imati kolateralnu informaciju o drugim tvrtkama. Koristit ćemo oba izvora informacija da bismo procijenili $m(\theta)$ za dani parametar rizika θ. Bit će nam dostupni podaci o N rizika tijekom n godina, gdje su podaci oblika $\{\{Y_{ij}\}_{i=1}^{N}\}_{j=1}^{n}$ i $\{\{P_{ij}\}_{i=1}^{N}\}_{j=1}^{n}$. Ovdje Y_{ij} predstavlja ukupna potraživanja u j-toj godini od i-tog rizika (ili tvrtke), a P_{ij} je odgovarajući obujam poslovanja.
Koristimo sljedeću notaciju:

$$\bar{P}_i = \sum_{i=1}^{n} P_{ij} \tag{4.7}$$

$$\bar{P} = \sum_{j=1}^{N} \bar{P}_i \tag{4.8}$$

$$P^* = (Nn-1)^{-1} \sum_{i=1}^{N} \bar{P}_i (1 - \bar{P}_i / \bar{P}) \tag{4.9}$$

$$X_j = Y_j / P_j \tag{4.10}$$

$$\bar{X}_i = \sum_{i=1}^{n} P_{ij} X_{ij} / \bar{P}_i \tag{4.11}$$

$$\bar{X} = \sum_{i=1}^{N} \sum_{j=1}^{n} P_{ij} X_{ij} / \bar{P} \tag{4.12}$$

Najbolji linearni procjenitelj ukupnih potraživanja sljedeće $(n+1)$ godine $Y_{i\ n+1}$ za rizik i je oblika $P_{i\ n+1}\ X_{i\ n+1}$, gdje je $X_{i\ n+1}$ procijenjen s

$$\begin{aligned} \mathbf{Z}_i\, \bar{X}_i + (1 - \mathbf{Z}_i)\, \mathbb{E}(m(\Theta)) = & \frac{\sum_j P_{ij}}{\sum_j P_{ij} + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))}\, \bar{X}_i \\ & + \frac{\mathbb{E}(s^2(\Theta))/Var(m(\Theta))}{\sum_j P_{ij} + \mathbb{E}(s^2(\Theta))/Var(m(\Theta))}\, \mathbb{E}(m(\Theta)). \end{aligned} \tag{4.13}$$

Primjetimo da bi faktori povjerenja (vjerodostojnosti) $\mathbf{Z}_i$ bili jednaki za sve rizike kada bi ukupni obujmovi poslovanja $(\sum_j P_{ij})$ bili jednaki za sve rizike. Kao i u Modelu 1, trebamo procjenitelje za parametre $\mathbb{E}(m(\Theta))$, $Var(m(\Theta))$ i $\mathbb{E}(s^2(\Theta))$ (Tablica 4.9)

Model 2: Nepristrani procjenitelji za $\mathbb{E}(m(\Theta))$, $Var(m(\Theta))$ i $\mathbb{E}(s^2(\Theta))$
Procjenitelj
$\mathbb{E}\,(\widehat{m(\Theta)}) = \bar{X}$
$\mathbb{E}\,(\widehat{s^2(\Theta)}) = N^{-1} \sum_{i=1}^{N} (n-1)^{-1} \sum_{j=1}^{n} P_{ij}(X_{ij} - \bar{X}_i)^2$
$Var\widehat{(m(\Theta)}) = P^{*-1}((Nn-1)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{n} P_{ij} \left(\bar{X}_{ij} - \bar{X}\right)^2$ $-N^{-1} \sum_{i=1}^{N} (n-1)^{-1} \sum_{j=1}^{n} P_{ij}\left(X_{ij} - \bar{X}_i\right)^2$

Tablica 4.9:

Primjer 4.2.4. *U Tablici 4.10 imamo podatke o ukupnim potraživanjima koja su se dogodila zbog šteta uzrokovanih dimom, za tri rizika, od 2002 do 2006 godine. Ovdje* Y_{ij}

označava ukupna potraživanja za rizik i u j-toj godini. Koristeći Model 1 računamo faktor povjerenja (vjerodostojnosti) i premiju za sljedeću godinu, za svaki od tri rizika.

Ukupna potraživanja od šteta uzrokovanih dimom.							
Godina	2002	2003	2004	2005	2006	$\bar{Y}_i$	$\sum_{j=1}^{5}(Y_{ij}-\bar{Y}_i)^2$
Rizik i							
1	4,560	4,825	4,965	5,325	5,775	5,090	891,136
2	3,425	3,700	3,825	3,940	4,120	3,802	272,484
3	7,200	7,540	7,760	7,810	7,910	7,644	320,356

Tablica 4.10:

$$\mathbb{E}(\widehat{m(\Theta)}) = \bar{Y} = 5,512,$$

$$\mathbb{E}(\widehat{s^2(\Theta)}) = \sum_{i=1}^{N}\sum_{j=1}^{n}\frac{\left(Y_{ij}-\bar{Y}_i\right)^2}{N(n-1)}$$

$$= \frac{891,136 + 272,484 + 320,356}{12}$$

$$= 123,664.67$$

i

$$\widehat{Var(m(\Theta))} = \sum_{i=1}^{N} \frac{\left(\bar{Y}_i - \bar{Y}\right)^2}{N-1} - \sum_{i=1}^{N} \sum_{j=1}^{n} \frac{\left(Y_{ij} - \bar{Y}_i\right)^2}{nN(n-1)}$$

$$= \sum_{i=1}^{N} \frac{\left(\bar{Y}_i - \bar{Y}\right)^2}{N-1} - \frac{\mathbb{E}\left(\widehat{s^2(\Theta)}\right)}{n}$$

$$= 3,823,804 - \frac{123,664.67}{5}$$

$$= 3,799,071.2.$$

Dakle, faktor povjerenja (vjerodostojnosti) $\mathbf{Z}$ *poprima oblik*

$$\mathbf{Z} = \frac{n}{n+K} = \frac{5}{5 + 123,664.6/3,799,071.2} = 0.9935.$$

Primjetimo da bismo, kada bismo koristili Model 1 u ovom slučaju, imali skoro puno povjerenje (vjerodostojnost) ($\mathbf{Z} = 0.9935$*) za podatke za bilo koji od rizika. To je zato što imamo jako malu procjenu za* $K = \mathbb{E}(s^2(\Theta))/Var(m(\Theta))$, 0.03255 *(jer je* $Var(m(\Theta)$ *puno manji od* $\mathbb{E}(s^2(\Theta))$*). Premija za rizik 1 iznosi*

$$\mathbf{Z}\,\bar{Y}_1 + (1-\mathbf{Z})\bar{Y} = 0.9935 \cdot 5,090 + 0.0065 \cdot 5,512 = 5,092.74,$$

za rizik 2 iznosi

$$\mathbf{Z}\,\bar{Y}_2 + (1-\mathbf{Z})\bar{Y} = 0.9935 \cdot 3,802 + 0.0065 \cdot 5,512 = 3,813.12$$

i za rizik 3 iznosi

$$\mathbf{Z}\,\bar{Y}_3 + (1-\mathbf{Z})\bar{Y} = 0.9935 \cdot 7,644 + 0.0065 \cdot 5,512 = 7,630.14.$$

Primjetimo da su sve blizu vrijednosti $\bar{Y}_1$, $\bar{Y}_2$ *i* $\bar{Y}_3$*. Sada pretpostavimo da imamo i mjeru za volumen rizika* P_{ij} *koja odgovara svakom* Y_{ij} *iz Tablice 4.10. Pretpostavimo također da smo koristeći Model 2, premiju po jedinici rizika za sljedeću godinu za rizik 1 već izračunali i da iznosi* 7.10.

Koje bi bile odgovarajuće premije za rizik 2 i 3? Znamo još i

Rizik i	$\sum_{j=1}^{5} P_{ij}$
1	3,760
2	2,448
3	4,624

Sada koristimo (4.7), (4.8), (4.10), i (4.12):

$$\bar{X} = (4.12) = \sum_{i=1}^{N}\sum_{j=1}^{n} P_{ij}X_{ij}/\bar{P}$$

$$(4.10) = \sum_{i=1}^{N} nY_{ij}/\bar{P}$$

$$(4.8) = n\sum_{i=1}^{N} Y_{ij}/\sum_{j=1}^{N}\bar{P}_i$$

$$(4.9) = n\sum_{i=1}^{N} Y_{ij}/\sum_{j=1}^{N}\sum_{i=1}^{n} P_{ij}$$

$$= 5 \cdot \frac{5,090 + 3,802 + 7,644}{3,760 + 2,448 + 4,624}$$

$$= 7.6329.$$

Dakle,

$$7.10 = \mathbf{Z}_1 \cdot 6.7686 + (1 - \mathbf{Z}_1) \cdot 7.6329 \qquad \Rightarrow \qquad \mathbf{Z}_1 = 0.5329/0.8643 = 0.6166.$$

pa, zbog (4.13), dobijemo

$$0.6166 = \frac{3,760}{3,760 + \mathbb{E}(\widehat{s^2(\Theta)})/Var\widehat{(m(\Theta)})} \qquad \Rightarrow \qquad \mathbb{E}(\widehat{s^2(\Theta)})/Var\widehat{(m(\Theta)}) \quad = 2,337.98.$$

Dakle, (opet zbog (4.13))

$$\mathbf{Z}_2 = \frac{2,448}{2,448 + 2,337.98} = 0.5115$$

pa je procijenjena premija po jedinici rizika za rizik 2

$$0.5115 \cdot 5 \cdot (3,802/2,448) + 0.4885 \cdot 7.6329 = 7.70.$$

Slično, za rizik 3,

$$\mathbf{Z}_3 = \frac{4,624}{4,624 + 2,337.98} = 0.6642,$$

s premijom

$$0.6642 \cdot 5 \cdot (7,644/4,624) + 0.3358 \cdot 7,6329 = 8.05.$$

U ovom primjeru možemo vidjeti još jednu razliku između Modela 1 i 2; u Modelu 1 je faktor povjerenja (vjerodostojnosti) jednak za sve rizike, a u Modelu 2 svaki rizik ima svoj faktor povjerenja (vjerodostojnosti).

Primjer 4.2.5. *Podaci o ukupnim godišnjim potraživanjima za sva tri rizika vezana za odgovornost zaposlenika dani su u Tablici 4.11. Ovdje Y_{ij} označava ukupna potraživanja za rizik i u j-toj godini. Koristeći Model 1 računamo faktor povjerenja (vjerodostojnosti) i premiju za sljedeću godinu, za svaki od ta tri rizika.*

Ukupna potraživanja od šteta uzrokovanih odgovornošću zaposlenika.							
Godina	1	2	3	4	5	$\bar{Y}_i$	$\sum_{j=1}^{5} (Y_{ij} - \bar{Y}_i)^2$
Rizik i							
1	3,894	5,188	3,582	4,680	5,182	4,505	2,180,690
2	3,940	2,994	3,582	4,068	4,434	3,804	1,190,476
3	4,382	5,028	4,434	4,844	5,642	4,866	1,049,784

Tablica 4.11:

$$\mathbb{E}(\widehat{m(\Theta)}) = \bar{Y} = 4,391.6,$$

$$\mathbb{E}(\widehat{s^2(\Theta)}) = \sum_{i=1}^{N}\sum_{j=1}^{n} \frac{\left(Y_{ij} - \bar{Y}_i\right)^2}{N(n-1)}$$

$$= \frac{2,180,690 + 1,190,476.1,049,784}{12}$$

$$= 368,412.5$$

i

$$Var\widehat{(m(\Theta))} = \sum_{i=1}^{N} \frac{\left(\bar{Y}_i - \bar{Y}\right)^2}{N-1} - \sum_{i=1}^{N}\sum_{j=1}^{n} \frac{\left(Y_{ij} - \bar{Y}_i\right)^2}{nN(n-1)}$$

$$= \sum_{i=1}^{N} \frac{\left(\bar{Y}_i - \bar{Y}\right)^2}{N-1} - \frac{\mathbb{E}\left(\widehat{s^2(\Theta)}\right)}{n}$$

$$= 291,591 - \frac{368,412.5}{5}$$

$$= 217,908.5.$$

Dakle, faktor povjerenja (vjerodostojnosti) **Z** *poprima oblik*

$$\mathbf{Z} = \frac{n}{n+K} = \frac{5}{5 + 368,412.5/217,908.5} = 0.7473.$$

Premija za rizik 1 iznosi

$$\mathbf{Z}\,\bar{Y}_1 + (1-\mathbf{Z})\bar{Y} = 0.7473 \cdot 4,505 + 0.2527 \cdot 4,391.6 = 4,478.34,$$

za rizik 2 iznosi

$$\mathbf{Z}\,\bar{Y}_2 + (1-\mathbf{Z})\bar{Y} = 0.7473 \cdot 3,804 + 0.2527 \cdot 4,391.6 = 3,952.49$$

i za rizik 3 iznosi

$$\mathbf{Z}\,\bar{Y}_3 + (1-\mathbf{Z})\bar{Y} = 0.7473 \cdot 4,866 + 0.2527 \cdot 4,391.6 = 4,746.12$$

Sada pretpostavimo da imamo i mjeru za volumen rizika P_{ij} *koja odgovara svakom* Y_{ij} *iz Tablice 4.11. Pretpostavimo također da smo koristeći Model 2, premiju po jedinici rizika za sljedeću godinu za rizik 1 već izračunali i da iznosi* 6.46. *Koje bi bile odgovarajuće premije za rizik 2 i 3? Znamo još i*

Rizik i	$\sum_{j=1}^{5} P_{ij}$
1	3,560
2	2,276
3	4,012

Iz Primjera 4.3.4. znamo da vrijedi

$$\begin{aligned}\bar{X} &= \sum_{i=1}^{N} nY_{ij} / \sum_{j=1}^{N} \sum_{i=1}^{n} P_{ij} \\ &= 5 \cdot (4,505 + 3,804 + 4,866)/(3,560 + 2,276 + 4,012) \\ &= 6.6892.\end{aligned}$$

Dakle,

$$6.46 = \mathbf{Z}_1 \cdot 6.3272 + (1 - \mathbf{Z}_1) \cdot 6.6893 \qquad \Rightarrow \qquad \mathbf{Z}_1 = 0.2292/0.362 = 0.6331.$$

pa, zbog (4.13), dobijemo

$$0.6331 = \frac{3,560}{3,560 + \mathbb{E}(\widehat{s^2(\Theta)})/\widehat{Var(m(\Theta))}} \qquad \Rightarrow \qquad \mathbb{E}(\widehat{s^2(\Theta)})/\widehat{Var(m(\Theta))} \quad = 2,063.12.$$

Dakle, zbog (4.13) vrijedi

$$\mathbf{Z}_2 = \frac{2,276}{2,276 + 2,063.12} = 0.5245$$

pa je procijenjena premija po jedinici rizika za rizik 2

$$0.5245 \cdot 5 \cdot (3,804/2,276) + 0.4755 \cdot 6.6892 = 7.56.$$

Slično, za rizik 3 (zbog (4.13)),

$$\boldsymbol{Z}_3 = \frac{4,012}{4,012 + 2,337.98} = 0.6604,$$

s premijom

$$0.6604 \cdot 5 \cdot (4,866/4,012) + 0.3396 \cdot 6.6892 = 6.28.$$

Poglavlje 5

Zaključak

Još od početka 20. stoljeća aktuari se u svom svakodnevnom radu suočavaju sa problemom procjene premija i potraživanja. Njihov posao jedan je od najvažnijih u osiguravajućem društvu i o točnosti njihovih procjena ovisi cijelo poslovanje društva. Zbog toga su trebale biti razvijene nove teorije i metode koje bi im omogućile što točniju procjenu i tako spasile osiguravajuće društvo od potencijalnih gubitaka. Tako je došlo do razvoja Teorije povjerenja, grane matematike osiguranja koja se bavi procjenjivanjem budućih potraživanja na osnovi jako malog uzorka ili informacije, ali gdje nam je druga kolateralna informacija pri ruci. Na osnovi tog broja potraživanja određuju se buduće premije koje ugovaratelji osiguranja moraju plaćati osiguravajućem društvu.

U ovom radu smo je proučili od samih početaka i uvjerili se u njenu korisnost i, što je u aktuarstvu jako važno, primjenjivost u samoj praksi. Obradili smo više pristupa problemu povjerenja i usporedili njihove sličnosti i razlike.

Klasični pristup nam je dao uvid u situacije u kojima smo željeli dati puno povjerenje podacima. Preciznije, željeli smo naći broj podataka koji nam je potreban da bismo se u potpunosti mogli pouzdati u našu procjenu. Tu smo kroz dva primjera (Primjer 2.1.3. i Primjer 2.1.2.) tražili broj potraživanja potreban za postizanje punog povjerenja (vjerodostojnosti) $\mathbf{Z} = 1$. Vidjeli smo kako možemo pronaći traženi broj u situacijama kada potraživnja dolaze iz određenih distribucija. U slučajevima u kojima nije bilo moguće dati puno povjerenje podacima, okrenili smo se parcijalnom povjerenju. Tu nam je $\mathbf{Z}$ bio posebnog oblika i korstili smo ga za procjenu budućih potraživanja, u slučaju da znamo broj prethodnih potraživanja i broj potraživanja koji je potreban za postizanje punog povjerenja. Iz toga smo vidjeli da nam je za korištenje pristupa parcijalnog povjerenja ipak nužno znati pristup punog povjerenja i da zato oba spadaju pod klasični pristup.

Nakon klasičnog proučavali smo Bayesovski pristup. Tu smo koristili Bayesovsku statistiku prilikom procjene budućih potraživanja, a preko njih odrediti i buduće premije. Cilj je bio prikazati (ako je to mogće) a posteriori očekivanje kao linearnu kombinaciju neke sta-

tistike ovisne o uzorku podataka koji promatramo i a priori očekivanja, s pomoću faktora povjerenja **Z**. Tada se procjena buduće premije svodi upravo na određivanje **Z**. Kroz par primjera upoznali smo se sa modelima u kojima se a posteriori očekivanje moglo prikazati baš u tom obliku. Problem kod tog pristupa je što a posteriori očekivanje većinom nije moguće tako prikazati. Ipak, tu smo dobili u uvid neke zanimljive modele i upoznali se sa pojmom konjugiranih familija.
Posljednji pristup koji smo obradili je empirijski Bayesovski pristup. Tu smo proučavali situacije koje su nam jako zanimljive u praksi: imamo određeni broj potraživanja prethodnih godina i zanima nas broj potraživanja sljedeće godine. Tu smo se suočili sa istim problemom kao i u običnom Bayesovskom pristupu, a to je prikaz a posteriori očekivanja u traženom obliku. Ovdje smo riješili taj problem uvođenjem Buhlmmanovskog procjenitelja povjerenja. Zatim smo to znanje koristili na par primjera u kojima smo vidjeli kako izračunati premiju ako znamo broj prethodnih potraživanja, nešto što se često koristi u aktuarskim poslovima. U empirijskom Bayesovskom pristupu smo se upoznali smo se i sa dva jako važna modela, Model 1 i Model 2. Kod ta dva modela imamo znanje i podatke o određenom broju rizika tijekom određenog broja godina i cilj nam je procijeniti premiju za sljedeću godinu za jedan ili sve rizike. Oba modela koriste isti pristup; izraziti a posteriori očekivanje kao linearnu kombinaciju neke statistike ovisne o uzorku podataka koji promatramo i a priori očekivanja, s pomoću faktora povjerenja **Z** koji je u ovom slučaju Buhlmmanovski! Prošli smo kroz par primjer primjera hipotetskih situacija s kojima se aktuari svakodnevno susreću i vidjeli kako možemo dobiti procjene budućih premija koristeći oba modela. Kao što smo mogli vidjeti osnovna razlika između ta dva modela je što Model 2 uzima u obzir i obujam poslovanja, što je naravno nešto što aktuar ne bi smio izostaviti kod svog izračuna. On je zato puno bliži realnoj situaciji i to je razlog zbog kojeg se danas puno više koristi.
Nakon proučavanja svih ovih pristupa i vježbanja stečenog znanja kroz primjere, svakoj osobi koja se bavi aktuarskim poslovima (ali i drugim matematičarima) trebalo bi biti jasno koliko je aktuarstvu potrebna Teorija povjerenja i kolike su mogućnosti njene primjene u praksi.

Bibliografija

[1] *Statistika - vježbe*.

[2] Basrak, B.: *Aktuarska matematika 2 - predavanja*.

[3] Bolland, P. J.: *Statistical and Probabilistic Methods in Actuarial Science*. Chapman and Hall/CRC, 2007.

[4] Buhlmann, P. Gisler, A.: *A Course in Credibility Theory and its Applications*. Springer, 2005.

[5] Congdon, P.: *Applied Bayesian Modelling*. John Wiley and Sons, 2003.

[6] Goulet, V.: *Principles and Application of Credibility Theory*. Journal of Actuarial Practice 1993-2006, 1998.

[7] Herzog, T. N.: *Introduction to Credibility Theory*. ACTEX Publications, Inc., 2010.

[8] Mahler, H. C. Dean, C. G.: *Foundations of Casualty Actuarial Science*. Springer, 2001.

Sažetak

U ovom radu bavimo se Teorijom povjerenja, granom matematike osiguranja koja se bavi procjenjivanjem budućih potraživanja od ugovoratelja osiguranja prema osiguravajućem duštvu, a samim time i budućih premija. U prvome dijelu upoznajemo se sa pojmovima iz Bayesovske statistike kao što su apriori i a posteriori distribucije, konjugirane familije i funkcije gubitka. Zatim krećemo s proučavanjem različitih pristupa Teoriji povjerenja koji se koriste za procjenu potraživanja. Klasični pristup se bavi situacijama u kojima podacima možemo dati puno povjerenje i tu se upoznajemo sa samim pojmom punog povjerenja. Proučavamo i situacije u kojima ne možemo dati puno povjerenje podacima. Bayesovski pristup koristi Bayesovsku statistiku za procjenu budućih potraživanja i premija. Tu se upoznajemo i s različitim modelima poput Poisson | Gamma modela, koji su nam zanimljivi jer kod njih a posteriori očekivanje možemo prikazati u obliku koji nam je pogodan za naše procjene. Empirijski Bayesovski pristup nam je jako zanimljiv jer tu promatramo situacije u kojima imamo određeni broj potraživanja prethodnih godina i zanima nas broj potraživanja sljedeće godine, a to je situacija koju jako često srećemo u praksi. Tu se susrećemo sa dva modela, Model 1 i Model 2, kod kojih imamo znanje i podatke o određenom broju rizika tijekom određenog broja godina i cilj nam je za te rizike procijeniti premiju za sljedeću godinu. Model 2 je bolji od ta dva jer on u obzir uzima i obujam poslovanja, dok ga Model 1 zanemaruje. Na kraju donosimo zaključak o naučenom o Teoriji povjerenja.

Summary

In this work we focused on Credibility theory, line of insurance mathematics, which deals with estimating future claims and premiums. In the first chapter are introduced with some terms from Bayesian statistics such as, prior and posterior distribution, conjugate families and loss functions. Then we study different approaches to Credibility theory which are used in estimating claims. Classical approach deals with situations where we can give full credibility to our data and here we are introduced with term full credibility. We also study situations where we can not give full credibility to our data. Bayesian approach uses Bayesian statistics to estimate future claims and premiums. Here we are introduced with different models such as Poisson | Gamma model, which are of interest to us because their posterior mean can be expressed in our desired form. Empirical Bayesian approach is of interest to us because here we observe situations in which we have possess knowledge about number of claims over the past few years and we want to estimate number of claims for the next year. This is situation we often see in a real world. Here we have two models, Model 1 and Model 2, where we possess knowledge and data on a certain number of risks over a certain number of years. Are main goal is to estimate premiums for those risks, for the next year. Model 2 is better than Model 1 because he takes in account volume of business for those risks. In the end, we draw conclusion about what we have learned about Credibility theory.

Životopis

Rođen sam 18.05.1993. u Splitu. Po nacionalnosti sam Hvat. Završio sam Osnovnu školu Gradac u Gradcu, gdje se već od prvog razreda javila moja ljubav prema matematici. 2009. godine upisao sam Opću gimnaziju u Srednjoj školi fra Andrije Kačića Miošića u Pločama, a 2012. godine Prirodoslovno-matematički fakultet, smjer Matematika u Zagrebu. 2016. godine sam stekao titulu univ. bacc. math. te iste godine upisao Diplomski studij Financijske i poslovne matematike na Prirodoslovno-matematičkom fakultetu u Zagrebu.
Od listopada 2017. godine do srpnja 2018. godine radio sam kao suradnik na aktuarskim poslovima u Wiener osiguranju. U srpnju 2018. godine radio sam u Splitskoj banci kao suradnik ma administrativnim poslovima.