UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA
MATEMATIKU I INFORMATIKU

Bojana Bobar

Bejzova statistika i njena primena u aktuarskom kredibilitetu

-MASTER RAD-

Mentor:

Prof.dr Zagorka Lozanov-Crvenković

Novi Sad, 2016.

Sadržaj

Predgovor

Bejzova statistika, nazvana po Tomasu Bejz (1701-1761.), je teorija u oblasti statistike u kojoj se verovatnoća ishoda nekog događaja izražava u vidu stepena verovanja (uverenja) da će se dati događaj realizovati, tzv. Bejzovih verovatnoća. Jedna od ključnih ideja Bejzove statistike je da se verovatnoća smatra mišljenjem, koje odražava znanje, informacije i verovanje koje posedujemo o nekom događaju, a da analiza podataka nije ništa drugo no revizija takvog mišljenja u svetlu novih relevantnih informacija. Bejzova analiza je pristup statističke analize koji je drugačiji od frekventističke analize.

Bejzove ideje su predstavljene u aktuarskoj nauci u godinama nakon 1960. u vidu metoda empirijskog kredibiliteta za uspostavljanje premija za polise osiguranja. Razvoj Bejzove metodologije je bio spor zbog njegove subjektivne prirode, koja je ujedno jedan od njegovih osnovnih kvaliteta, i zbog poteškoća u vezi sa proračunima u Bejzovoj analizi.

Statistička teorija kredibiliteta obezbeđuje metode za ocenjivanje parametara podskupa date populacije, tako što kombinuje rezultate za određeni podskup sa rezultatima populacije kao celine, koja je veća, i mnogo statistički stabilnija. Aktuari retko koriste teoriju kredibiliteta u čistom matematičkom maniru. Podskupovi cele populacije imaju karakteristike koje nisu u potpunosti poznate, stoga je neophodno uvesti rasuđivanje aktuara kako bi se odredila upotreba ovih karakteristika.

U ovom radu želim da ukratko predstavim osnovnu Bejzovu paradigmu, uz osvrt na široku primenu Bejzove analize. Takođe, želim da, kroz primere, objasnim teoriju kredibiliteta, kako bi čitalac imao razumevanje o stvarnoj prirodi kredibiliteta.

Glavna motivacija za obradu ove teme u master tezi, esencijalno leži u važnosti obrađene teorije. Naime, teorija kredibiliteta i Bejzova analiza predstavljaju jedan od temelja aktuarske nauke, bez kog moderno osiguranje ne bi postojalo u obliku u kom ga danas poznajemo. Odatle sledi i moja lična ambicija da istražim i bolje razumem teoriju tako široke primene. Drugo, jednako motivišuće, za mene predstavlja činjenica da, iako je tema veoma popularna u svetu, ne raspolažemo opsežnom literaturom na srpskom jeziku, koja na sveobuhvatan način obrađuje obe teme.

Rad će biti podeljen u tri tematske celine.

U uvodnom delu biće navedeni osnovni pojmovi Bejzove statističke analize bazirane na Bejzovoj teoremi. Takođe, biće navedene glavne razlike „frekventističke“ i „Bejzove“ statistike kao i oblasti primene Bejzove analize.

Centralni deo rada će biti posvećen primeni Bejzove statistike u aktuarstvu, konkretnije biće dati osnovni pojmovi teorije kredibiliteta.

Teoriju kredibiliteta u aktuarstvu čini skup alata koji omogućavaju osiguravaču da, na osnovu istorijskih podataka, odredi buduću premiju za određeni rizik ili riziko grupu. Ključno pitanje na koje teorija daje odgovor je koliki značaj pri određivanju premije polise osiguranja ima iskustvo pojedinačnog osiguranika ili jedne riziko grupe u odnosu na prosečno iskustvo date klase osiguranika. Drugim rečima, koliko verujemo ličnom iskustvu pojedinačnog osiguranika. Kvantitativno formulišući gore navedeni problem, biće predstavljene teorija kredibiliteta ograničene fluktuacije i teorija kredibiliteta visoke preciznosti.

Ako pretpostavimo da nivo rizika svakog osiguranika karakteriše parametar rizika θ sa raspodelom verovatnoće $\pi(\theta)$, onda iskustvo, odnosno štetu osiguranika sa datim parametrom θ dobijamo iz uslovne raspodele $f_{X|\Theta}(x|\theta)$. Određivanje vrednosti parametara $\mu(\theta) = E(X_j|\Theta = \theta)$, $\nu(\theta) = Var(X_j|\Theta = \theta)$, $\mu = E[\mu(\Theta)]$, $\nu = E[\nu(\Theta)]$ i $a = Var[\mu(\Theta)]$, gde je $j = 1,\ldots,n$ prošli vremenski period, moguće je ukoliko se pretpostavlja da su $f_{X|\Theta}(x|\theta)$ i $\pi(\theta)$ poznate. Iako je to zgodno pretpostaviti zarad ilustracije metodologije, u praksi često nemamo predznanje o tim raspodelama. Praktična upotreba teorije zahteva da se nepoznati parametri ocene na osnovu podataka, kako bi model što bolje oslikavao stvarnost. Stoga će se razmotriti empirijsko Bejzovo ocenjivanje parametara kroz neparametarsko, semiparametarsko i parametarsko ocenjivanje.

Veliku zahvalnost dugujem svojoj mentorki, Zagorki Lozanov-Crvenković, na pruženim savetima i smernicama tokom pisanja ovog rada.

Takođe, zahvalila bih se članovima komisije, Ljiljani Gajić i Ivani Štajner-Papuga, na svom prenetom znanju i lepoj saradnji.

I na kraju, najveću zahvalnost dugujem svojoj porodici, ocu Bošku i majci Radmili, a naročito sestri Mini, na nesebičnoj podršci tokom čitavog školovanja.

Novi Sad, oktobar 2016. Bojana Bobar

Poglavlje 1

1. Bejzova statistika

Po mišljenju velikog broja statističara, postoje dva međusobno isključujuća pristupa analizi podataka. „Klasična“ ili „frekventistička“ teorija, koja se sastoji od intervala poverenja i testiranja hipoteza, se najviše koristi i najrasprostranjenija je u tipičnim statističkim tekstovima. S druge strane, „Bejzova“ statistika, analiza zasnovana na Bejzovoj teoremi, privukla je malu grupu strastvenih pristalica. Debata se odvija u mnogim radovima, od kojih je jedan „Zašto nismo svi Bejzovci?“[1], i već godinama ne prestaje da jenjava.[2]

Ovo poglavlje predstavlja uvod u Bejzovu statistiku sa osvrtom na oblasti primene.

1.1. Šta je Bejzova analiza?

“Današnja predviđanja su sutrašnja priorna[3] znanja.”[4]

Naučno istraživanje je iterativni process integracije i akumuliranja informacija. Istraživači ocenjuju trenutno znanje o problem koji se izučava, sakupljaju nove podatke da bi našli odgovore na preostala pitanja, i potom ažuriraju i oplemenjuju svoje razumevanje da bi pripojili i nove i stare podatke. Bejzova analiza daje logičan i kvantitativan okvir za ovaj proces.

Izraz "Bejzov" se odnosi na sveštenika Tomasa Bejza[5]. Razvoj teorije verovatnoće na početku 18. veka je iznedrio odgovor na pitanja u vezi sa kockanjem, i podupreo nove i slične ideje u vezi sa osiguranjem. Javio se problem, poznatiji kao pitanje obrnute verovatnoće: matematičari tog vremena su znali kako da pronađu verovatnoću da će, recimo, četvoro ljudi, iz skupa od 60 ljudi starosti 50 godina, umreti u datoj godini, ukoliko je poznata verovatnoća umiranja za svakog od njih. Ali nisu znali kako da izračunaju verovatnoću da će jedan pedesetogodišnjak umreti, na osnovu opservacije da je umrlo četvoro od njih 60. Odgovor na ovo pitanje je dao Tomas Bejz u radu koji je objavljen 1763. godine, godinu dana nakon njegove smrti. Kao i veliki broj obrazovanih ljudi tog vremena, Bejz je bio sveštenik i samouki naučnik i

[1] "Why Isn't Everyone a Bayesian?", Efron, 1986
[2] Vidi [4].
[3] “A priori“ se prevodi kao onaj koji se odnosi na ili je dobijen rasuđivanjem iz očiglednih propozicija.
[4] Vidi [16].
[5] Thomas Bayes

matematičar. Njegovo rešenje, poznato kao Bejzova teorema, predstavlja temelj modernog Bejzovog pristupa analizi svih vrsta podataka, koje po njemu nosi ime.[6]

1.2. Ko je Bejz?

Slika 1.1 Tomas Bejz (rođen u Londonu 1702. – umro u Tanbridž Velsu 1761.)

Tomas Bejz je prvi matematičar koji je induktivno koristio verovatnoću i ustanovio matematičku osnovu za analizu verovatnoće (sredstva analize, verovatnoću da će se događaj realizovati u budućim ponavljanjima na osnovu broja ponavljanja u kojima se događaj nije realizovao).[7]

Svoje pronalaske u vezi sa verovatnoćom je naveo u eseju „Na putu rešenja problema iz doktrine verovatnoća" (1763), koji je posthumno objavljen u radu „Filozofske transakcije Londonskog kraljevskog udruženja".

Jedini radovi koje je za života objavio su u originalu nazvani „*Divine Benevolence, or an attempt to Prove That the Principal End of the Divine Providence and Government is the happiness of His Creatures*" (1731) i „*An Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of the Analyst*" (1736), koji je više puta Bišop Berkli[8] napadao na logičkim osnovama Njutnove analize.

Bejz je bio sin sveštenika Džošue Bejza[9], nekonformističkog ministra, a sam je od 1731. bio prezbiterijanski ministar u Tanbridž Velsu[10]. Misli se da je primljen u Kraljevsko društvo zbog

[6] Vidi [17].
[7] Vidi [17].
[8] Bishop Berkeley
[9] Rev. Joshua Bayes
[10] Tunbridge Wells

rasprave u kojoj je 1736. branio stavove i filozofiju Sera Isaka Njutna. Sahranjen je u Engleskoj u poljima Banhil[11] londonskog parka.

Doprinosi Tomasa Bejza su dobili status besmrtnosti, nakon što je fundamentalna propozicija verovatnoće po njemu nazvana Bejzovo pravilo.

1.3. Bejzova paradigma

Bejzov pristup pretpostavlja da su samo trenutno posmatrani podaci relevantni za analizu i da je raspodela populacije ta koja je promenljiva. Nepoznati parameter koji posmatramo je θ, a ono što je poznato je funkcija raspodele $\pi(\theta)$, koja izražava naše trenutno relativno mišljenje o verodostojnosti da su razne moguće vrednosti za θ stvarne vrednosti. Ovo se zove priorna raspodela, zato što predstavlja stanje našeg znanja pre nego što smo izveli eksperiment ili pre nego što smo videli podatke iz opservacija.

Za ocenu parametara, sledeće definicije opisuju proces, a potom Bejzova teorema daje rešenje za opisani problem.[12]

■Definicija 1.1
Priorna raspodela *je funkcija raspodele verovatnoća za prostor svih mogućih vrednosti parametra* θ. *Označava se sa* $\pi(\theta)$ *i predstavlja naše mišljenje u vezi sa relativnom šansom da razne vrednosti za* θ *budu stvarne vrednosti parametra*.[13]

Kada je nejasno ili ako slučajna promenljiva može biti apsolutno neprekidna, diskretna ili mešovita, koristimo izraz funkcija raspodele. U formulama, u skladu sa tipom raspodele, integrale treba zameniti sa sumama.

Određivanje priorne raspodele je oduvek predstavljalo široko rasprostranjenu barijeru za prihvatanje Bejzovog metoda. Gotovo je sigurno da vaše iskustvo daje neki uvid u moguće vrednosti za parametar, pre nego što se prvi podatak opazi. Ono što je komplikovano, jeste prevođenje ovakvog znanja u raspodelu verovatnoće.
Zbog poteškoće pronalaska priorne raspodele koja je ubedljiva (moraćete ubediti druge da su vaša priorna mišljenja validna) i mogućnosti da zapravo nemate formirano nikakvo priorno mišljenje, definicija o priornoj raspodeli se može popustiti.

■Definicija 1.2
Neispravna priorna raspodela *je ona čije su verovatnoće nenegativne, ali im je suma (ili integral) beskonačan.*

[11] Bunhill Fields
[12] Vidi [1].
[13] Parametar θ može biti skalar ili vektor.

Velika količina istraživanja se bazirala na određivanju takozvane neinformativne ili nejasne priorne raspodele. Njena svrha je da reflektuje minimalno znanje. Ne postoji univerzalni dogovor za formiranje nejasne priorne raspodele. Međutim, postoji dogovor da je odgovarajuća neinformativna priorna raspodela za skalirni parametar $\pi(\theta) = 1/\theta, \theta > 0$. Primetimo da je ovo neispravna priorna raspodela.

Drugi pojam je funkcija raspodele $f(x|\theta)$. Ona opisuje relativnu verodostojnost da će se razne vrednosti x realizovati izvođenjem eksperimenta, u slučaju da je θ stvarna vrednost parametra. Ona se još naziva raspodela modela i zajednička je i za Bejzovu i za klasičnu analizu. Treba napomenuti da je moguće da i x i θ budu vektori, gde su x vrednosti iz uzorka, a θ skup nepoznatih parametara.

■Definicija 1.3

Raspodela modela *je funkcija raspodele prikupljenih podataka, ako je data određena vrednost parametra. Međutim, konzistentno sa Bejzovom notacijom, raspodela modela se označava sa*

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

gde je korišćena vektorska notacija za **x** *da nas podseti da se ovde pojavljuju svi podaci.*

Dakle, kao i kod ocene maksimalne verodostojnosti, moramo biti u mogućnosti da napišemo funkciju verodostojnosti za datu situaciju.

Prisetimo se sada koncepta višedimenzionalne statističke analize.

Ako vektor sa opservacijama $\boldsymbol{x} = (x_1, \cdots, x_n)^T$ čine međusobno nezavisne, jednako raspodeljene slučajne promenljive, onda

$$f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) = f_{X|\Theta}(x_1|\theta) \cdots f_{X|\Theta}(x_n|\theta).$$

■Definicija 1.4

Zajednička funkcija raspodele *ima funkciju raspodele*

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

■Definicija 1.5

Marginalna raspodela za **x** *ima funkciju raspodele*[14]

$$f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)d\theta.$$

[14] U slučaju da imamo više od jednog parametra, ovo će biti višestruki integral.

Sledeća dva pojma su ključni rezultati Bejzove analize. Posteriorna raspodela nam govori kako se menjalo naše mišljenje o parametru, predstavlja naše revidirano mišljenje o θ, nakon što smo videli rezultate tekućeg eksperimenta. Sada ova raspodela sadrži svo naše predznanje o nepoznatom parametru, baš kao što je i priorna raspodela ranije sadržala naše predznanje. Prediktivna raspodela nam govori kako može izgledati sledeća opservacija, uz date informacije sadržane u podacima.

■Definicija 1.6
Posteriorna raspodela *je uslovna raspodela od parametara, za date posmatrane podatke. Označava se sa*

$$\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}).$$

.
Od najvećeg interesa je buduća vrednost opservacije.

■Definicija 1.7
Prediktivna raspodela *je uslovna raspodela za novu opservaciju* y*, uz date podatke* $\mathbf{x}$*. Označava se sa*

$$f_{Y|\mathbf{X}}(y|\mathbf{x}).$$

Primetimo da ova raspodela ne mora da se poklapa sa modelom koji je dao vrednosti x, već mora da zavisi od istog parametra θ.

U sledećem koraku koristimo Bejzovu teoremu da izračunamo posteriornu (prilagođenu) raspodelu za θ.

■Teorema 1.1
Posteriorna raspodela se dobija formulom

$$\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)}{\int f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)d\theta}, \tag{1.1}$$

dok se prediktivna raspodela može izračunati sa

$$f_{Y|\mathbf{X}}(y|\mathbf{x}) = \int f_{Y|\Theta}(y|\theta)\,\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta, \tag{1.2}$$

gde je $f_{Y|\Theta}(y|\theta)$ *funkcija raspodele od nove opservacije, za datu vrednost parametra*.

Prediktivna raspodela se može tumačiti kao mešovita raspodela, gde se mešovito odnosi na posteriornu raspodelu.

Konačni korak je upotreba posteriorne raspodele da se izvuku adekvatni zaključci za razmatrani problem.

1.4. Analiza i predviđanje

U jednom smislu, analiza je gotova. Počnemo sa distribucijom koja kvantifikuje naše znanje o parametru i/ili sledećoj opservaciji, i završimo sa revidiranom (prilagođenom) distribucijom. Ali, pretpostavimo da naš šef neće biti zadovoljan ako odredimo raspodelu, kao odgovor na njegov zahtev. Nema sumnje, traži se specifičan broj, možda čak i sa marginom za grešku. Uobičajeno Bejzovo rešenje jeste da se uvede funkcija gubitaka(štete).

■Definicija 1.8

Funkcija štete $l_j(\hat{\theta}_j, \theta_j)$ opisuje penale koje plaća investitor kada je $\hat{\theta}_j$ *ocena a* θ_j *stvarna vrednost j-og parametra*.

Možemo imati i višedimenzionalnu funkciju štete $l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$, koja omogućuje da šteta simultano zavisi i od grešaka u raznim ocenama parametara.

■Definicija 1.9

Bejzova ocena *za datu funkciju štete je ona ocena koja minimizuje očekivani gubitak, za datu posteriornu raspodelu razmatranih parametara.*

Tri najčešće korišćene funkcije štete su definisane na sledeći način.

■Definicija 1.10

Za ***štetu kvadratne greške***, *funkcija štete je (svi indeksi su izostavljeni zbog jednostavnosti)*

$$l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2.$$

Za ***apsolutnu štetu***, *funkcija štete je*

$$l(\hat{\theta}, \theta) = |\hat{\theta} - \theta|.$$

Za ***nula-jedan štetu***, *funkcija štete je*

$$l(\hat{\theta}, \theta) = 0 \; ako \; \hat{\theta} = \theta,$$

a inače je 1.

Sledeća teorema pokazuje Bejzove ocene za tri najčešće funkcije štete.

■Teorema 1.2

Za štetu kvadratne greške,Bejzova ocena je sredina posteriorne raspodele;za apsolutnu grešku, to je medijan; a za nula-jedan štetu, to je mod.

Primetimo da nema garancije da će sredina posteriorne raspodele postojati, niti da će medijan ili mod posteriorne raspodele biti jedinstveni. Dalje primetimo da, ukoliko se koristi neispravna priorna raspodela $\pi(\theta) = 1$ i ocena je mod posteriorne raspodele, onda će ocena biti jednaka oceni maksimalne verodostojnosti. Ukoliko se drugačije ne naglasi, pojam Bejzova ocena će se odnositi na sredinu posteriorne raspodele.

Često se traži očekivana vrednost prediktivne raspodele u svrhu prognoziranja. O njoj možemo razmišljati kao o tačkastoj oceni za $(n + 1)$. opservaciju za datu priornu raspodelu i za date prvih n opservacija:

$$\begin{aligned} \mathrm{E}(Y|\mathbf{x}) &= \int y f_{Y|\mathbf{X}}(y|\mathbf{x})dy \\ &= \int y \int f_{Y|\Theta}(y|\theta)\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta dy \\ &= \int \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) \int y f_{Y|\Theta}(y|\theta)dyd\theta \\ &= \int \mathrm{E}(Y|\theta)\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta. \end{aligned} \tag{1.3}$$

Jednačina (1.3) se može tumačiti kao težinska sredina, gde kao težine koristimo posteriornu raspodelu.

Bejzov ekvivalent intervala poverenja se lako računa uz pomoć sledeće definicije.

■Definicija 1.11

Tačke $l < u$ *definišu* $100(1-\alpha)\%$ ***kredibilni interval*** *za* θ_j *za bilo koji par* (l, u) *koji zadovoljava*[15]

$$\Pr\left(l \leq \Theta_j \leq u \middle| \mathbf{x}\right) \geq 1 - \alpha.$$

Ovaj kredibilitet ne treba mešati sa aktuarskim kredibilitetom, o kome ćemo pričati u Poglavlju 2.

Nejednakost je prisutna u slučaju kada je posteriorna raspodela za θ_j diskretna, ali tada postoji mogućnost da verovatnoća ne bude tačno $1-\alpha$. Ova definicija ne daje jedinstveno rešenje. Sledeća teorema nam daje način za određivanje jedinstvenog intervala i možda najpoželjnijeg izbora za krajnje vrednosti intervala.

■Teorema 1.3

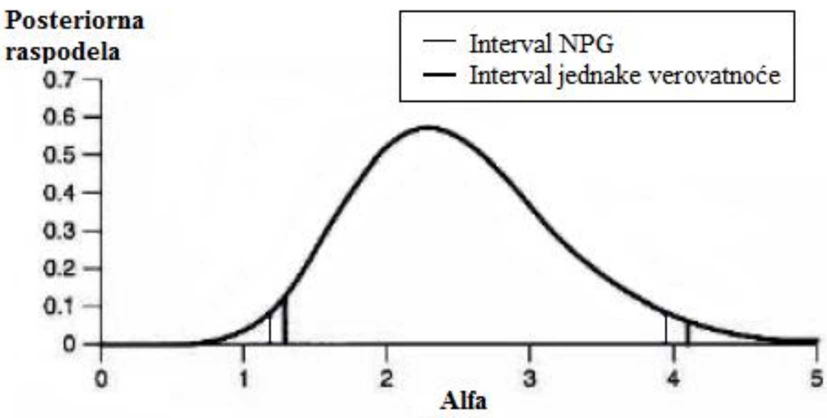
Ako je posteriorna slučajna promenljiva $\theta_j|\boldsymbol{x}$ *neprekina i unimodalna*[16]*, onda je* $100(1-\alpha)\%$ *kredibilni interval sa najmanjom razlikom* $u - l$ *jedinstveno rešenje za:*

$$\int_l^u \pi_{\Theta_j|\mathbf{X}}\left(\theta_j \middle| \mathbf{x}\right)d\theta_j = 1 - \alpha,$$
$$\pi_{\Theta|\mathbf{X}}(l|\mathbf{x}) = \pi_{\Theta|\mathbf{X}}(u|\mathbf{x}).$$

Ovaj interval je specijalan slučaj kredibilnog skupa najveće posteriorne gustine (NPG)[17].

---

[15] *l potiče od reči „lower“ što znači donji, a u potiče od „upper“ što znači gornji.*
[16] Raspodela je unimodalna ako postoji samo jedan mod.
[17] highest posterior density (HPD)



Slika 1.2 Dva Bejzova kredibilna intervala.

Sledeća definicija nam daje ekvivalentan rezultat za bilo koju posteriornu raspodelu.

■Definicija 1.12

Za bilo koju posteriornu raspodelu, $100(1-\alpha)\%$ ***kredibilnog skupa NPG*** *predstavlja skup vrednosti parametara iz C takvih da je*

$$\Pr\left(\theta_j \in C\right) \geq 1-\alpha \tag{1.4}$$

i

$$C = \left\{\theta_j : \pi_{\Theta_j|\mathbf{X}}\left(\theta_j \middle| \mathbf{x}\right) \geq c\right\} \text{ za neko } c,$$

gde je c najveća vrednost za koju važi nejednakost (1.4).

Ovaj skup može sačinjavati i unija više intervala, što se može desiti kod višemodalne posteriorne rapodele. Ova definicija daje skup intervala minimalne širine koji imaju traženu posteriornu verovatnoću. Skup se konstruiše tako što se počne od velike vrednosti za c , koja se potom smanjuje. Kako se c smanjuje, skup C se povećava zajedno sa verovatnoćom. Ovaj proces se nastavlja sve dok verovatnoća ne dostigne vrednost $1-\alpha$.

Nekad se može desiti da je lakše izračunati momente posteriorne raspodele, nego njene verovatnoće. Tada možemo koristiti Bejzovu centralnu graničnu teoremu.

■Teorema 1.4[18]

Ako su $\pi(\theta)$ *i* $f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta)$ *dvaputa diferencijabilne u* θ *i ako važe uobičajeno zadovoljene pretpostavke, onda je posteriorna raspodela od* Θ *za dato* $\mathbf{X} = \mathbf{x}$ *asimptotski normalna.*

Bejzova centralna granična teorema tvrdi da se, pod odgovarajućim uslovima, posteriorna raspodela može aproksimovati sa normalnom raspodelom, pa je kredibilni interval približno jednak[19]

[18] Berger, J. (1985), *Bayesian Inference in Statistical Analysis*, 2nd ed., New York: Springer-Verlag.
[19] Vidi [4].

$$\mathrm{E}(\theta_j|\mathbf{x}) \mp z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{Var}(\theta|\mathbf{x})}.$$

Kao i kod uobičajene centralne granične teoreme, aproksimacija se poboljšava kada broj opservacija raste.

Kada su nam potrebni samo momenti, onda je korisno upotrebiti formule dvostrukog očekivanja.

Ako momenti postoje za bilo koje slučajne promenljive X i Y, onda su

$$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(\mathrm{Y}|\mathrm{X})], \tag{1.5}$$

$$\mathrm{Var}(Y) = \mathrm{E}[\mathrm{Var}(\mathrm{Y}|\mathrm{X})] + \mathrm{Var}[\mathrm{E}(\mathrm{Y}|\mathrm{X})]. \tag{1.6}$$

Za prediktivnu raspodelu važi:

$$\begin{aligned}\mathrm{E}(Y|\mathbf{x}) &= \mathrm{E}_{\Theta|\mathbf{x}}[\mathrm{E}(\mathrm{Y}|\Theta,\mathbf{x})]\\ &= \mathrm{E}_{\Theta|\mathbf{x}}[\mathrm{E}(\mathrm{Y}|\Theta)]\end{aligned}$$

i

$$\begin{aligned}\mathrm{Var}(Y|\mathbf{x}) &= \mathrm{E}_{\Theta|\mathbf{x}}[\mathrm{Var}(\mathrm{Y}|\Theta,\mathbf{x})] + \mathrm{Var}_{\Theta|\mathbf{x}}[\mathrm{E}(\mathrm{Y}|\Theta,\mathbf{x})]\\ &= \mathrm{E}_{\Theta|\mathbf{x}}[\mathrm{Var}(\mathrm{Y}|\Theta)] + \mathrm{Var}_{\Theta|\mathbf{x}}[\mathrm{E}(\mathrm{Y}|\Theta)].\end{aligned}$$

1.5. Konjugovane (srodne) priorne raspodele i linearno eksponencijalna familija

Sledi definicija velike familije raspodela, koja obuhvata mnoge raspodele značajne za aktuare i koja ima svoju primenu u Bejzovoj analizi i teoriji kredibiliteta.

■Definicija 1.13

Slučajna promenljiva X (*diskretna ili apsolutno neprekidna*) *je član* ***linearno[20] eksponencijalne familije*** *raspodela, ukoliko se njena raspodela može parametrizovati sa parametrom θ i izraziti na sledeći način*

$$f(x;\theta) = \frac{p(x)e^{-\theta x}}{q(\theta)}. \tag{1.7}$$

Funkcija p(x) zavisi samo od x (ne i od θ), dok je funkcija q(θ) normalizaciona konstanta. Takođe, oslonac slučajne promenljive ne sme zavisiti od θ. Parametar θ se naziva ***prirodan parametar*** *raspodele.*

■Primer 1.1

Pokaži da eksponencijalna raspodela ima oblik (1.7).

[20] „linearno“ označava da je stepen kojim se eksponira zapravo linearna funkcija od *x*.

Rešenje:
Raspodela je data sa

$$f(x;\beta) = \beta^{-1}e^{-\beta^{-1}x}.$$

Neka je $\theta = 1/\beta$, tada je raspodela jednaka

$$f(x;\theta) = \frac{1e^{-\theta x}}{\theta^{-1}},$$

što je oblika (1.7) sa $p(x) = 1$ i $q(\theta) = 1/\theta$. □

■Primer 1.2
Pokaži da je Poasonova raspodela član linearno eksponencijalne familije.

Rešenje:
Raspodela je data sa

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(1/x!)e^{-(-ln\lambda)x}}{e^{\lambda}}.$$

Neka je $\theta = -ln\lambda$, tada je raspodela jednaka

$$f(x;\theta) = \frac{(1/x!)e^{-\theta x}}{e^{e^{-\theta}}},$$

što je oblika (1.7) sa $p(x) = 1/x!$ i $q(\theta) = e^{e^{-\theta}}$. Obratite pažnju da je u ovakvoj parametrizaciji sredina Poasonove raspodele jednaka $e^{-\theta}$ □

■Primer 1.3
Pokaži da je normalna raspodela sa sredinom μ i poznatom varijansom υ član linearno eksponencijalne familije.

Rešenje:
Raspodela je data sa

$$\begin{aligned} f(x;\mu,\upsilon) &= (2\pi\upsilon)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\upsilon}(x-\mu)^2\right] \\ &= (2\pi\upsilon)^{-\frac{1}{2}}\exp\left(-\frac{x^2}{2\upsilon}+\frac{\mu}{\upsilon}x-\frac{\mu^2}{2\upsilon}\right) \\ &= \frac{(2\pi\upsilon)^{-\frac{1}{2}}\exp\left(-\frac{x^2}{2\upsilon}\right)\exp\left(\frac{\mu}{\upsilon}x\right)}{\exp\left(\frac{\mu^2}{2\upsilon}\right)}. \end{aligned}$$

Neka je $\theta = -\mu/\upsilon$, tada je raspodela jednaka

$$f(x;\theta,v) = \frac{(2\pi v)^{-\frac{1}{2}}\exp\left(-\frac{x^2}{2v}\right)\exp(-\theta x)}{\exp\left(\frac{\theta^2 v}{2}\right)},$$

što je oblika (1.7) sa $p(x) = (2\pi v)^{-\frac{1}{2}}\exp\left(-\frac{x^2}{2v}\right)$ i $q(\theta) = \exp\left(\frac{\theta^2 v}{2}\right)$. □

Nađimo sada sredinu i varijansu rasopodele zadate sa (1.7).
Primetimo da je

$$\ln f(x;\theta) = \ln p(x) - \theta x - \ln q(\theta).$$

Diferencirajmo sada po θ da bismo dobili

$$\frac{\partial}{\partial\theta} f(x;\theta) = \left[-x - \frac{q'(\theta)}{q(\theta)}\right] f(x;\theta). \tag{1.8}$$

Sada integralimo (ili sumiramo) po vrednostima za x (za koje znamo da ne zavise od θ) da bismo dobili

$$\int \frac{\partial}{\partial\theta} f(x;\theta)dx = -\int xf(x;\theta)dx - \frac{q'(\theta)}{q(\theta)}\int f(x;\theta)dx.$$

Sa leve strane jednakosti, obrnimo redosled integracije i diferenciranja (ili sumiranja) da bismo dobili

$$\frac{\partial}{\partial\theta}\left[\int f(x;\theta)dx\right] = -\int xf(x;\theta)dx - \frac{q'(\theta)}{q(\theta)}\int f(x;\theta)dx.$$

Znamo da je $\int f(x;\theta)dx = 1$ i da je $\int xf(x;\theta)dx = \mathrm{E}(X)$, stoga je

$$\frac{\partial}{\partial\theta}(1) = -\mathrm{E}(X) - \frac{q'(\theta)}{q(\theta)}.$$

Drugim rečima, sredina je

$$\mathrm{E}(X) = \mu(\theta) = -\frac{q'(\theta)}{q(\theta)} = -\frac{\partial}{\partial\theta}\ln q(\theta). \tag{1.9}$$

Da bismo odredili varijansu, (1.8) možemo drugačije zapisati

$$\frac{\partial}{\partial\theta} f(x;\theta) = -[x-\mu(\theta)]f(x;\theta).$$

Opet diferenciramo u odnosu na θ da bismo dobili

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2} f(x;\theta) &= \mu'(\theta)f(x;\theta) - [x-\mu(\theta)]\frac{\partial}{\partial\theta}f(x;\theta)\\ &= \mu'(\theta)f(x;\theta) + [x-\mu(\theta)]^2 f(x;\theta).\end{aligned}$$

Opet integralimo po skupu vrednosti x da bismo dobili

$$\int \frac{\partial^2}{\partial\theta^2} f(x;\theta)dx = \mu'(\theta)\int f(x;\theta)dx + \int [x-\mu(\theta)]^2 f(x;\theta)dx.$$

Ovo drugačije zapisujemo

$$\int [x-\mu(\theta)]^2 f(x;\theta)dx = -\mu'(\theta) + \frac{\partial^2}{\partial\theta^2}\int f(x;\theta)dx.$$

Pošto je $\mu(\theta)$ sredina, leva strana jednakosti je varijansa (po definiciji), i pošto je drugi sabirak sa desne strane jednakosti jednak nula, dobijamo

$$\mathrm{Var}(X) = \upsilon(\theta) = -\mu'(\theta) = \frac{\partial^2}{\partial\theta^2}\ln q(\theta). \tag{1.10}$$

■Definicija 1.14

Za priornu raspodelu se kaže da je ***konjugovana priorna raspodela*** *za dati model, ako je rezultujuća posteriorna raspodela iz iste familije kao i priorna raspodela (ali verovatno sa drugačijim parametrima).*

Sledeća teorema tvrdi da, ako je model član iz linearno eksponencijalne familije raspodela, konjugovana priorna raspodela se lako pronalazi.

■Teorema 1.5

Neka je dato $\Theta = \theta$ *za međusobno nezavisne i jednako raspodeljene slučajne promenljive* $X_1, \dots, X_n$ *sa raspodelom verovatnoća*

$$f_{X_j|\Theta}(x_j|\theta) = \frac{p(x_j)e^{-\theta x_j}}{q(\theta)},$$

gde Θ *ima raspodelu*

$$\pi(\theta) = \frac{[q(\theta)]^{-k}e^{-\theta\mu k}}{c(\mu,k)},$$

gde su k *i* μ *parametri raspodele, a* $c(\mu,k)$ *je normalizaciona konstanta. Tada je posteriorna raspodela* $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$ *istog oblika kao i* $\pi(\theta)$.

Dokaz:

Posteriorna distribucija je

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \frac{\left[\prod_{j=1}^{n} p(x_j)\right]e^{-\theta\sum x_j}}{q(\theta)^n}\frac{[q(\theta)]^{-k}e^{-\theta\mu k}}{c(\mu,k)}\\ &\propto [q(\theta)]^{-(k+n)} exp\left[\left(-\theta\frac{\mu k+\sum x_j}{k+n}\right)(k+n)\right]\\ &\propto [q(\theta)]^{-k^*}\exp(-\theta\mu^* k^*),\end{aligned}$$

što je istog oblika kao $\pi(\theta)$ sa parametrima

$$k^* = k + n$$
$$\mu^* = \frac{\mu k + \sum x_j}{k + n} = \frac{k}{k + n}\mu + \frac{n}{k + n}\bar{x}. \qquad \square$$

■Primer 1.4

Pokaži da je Gama raspodela, konjugovana priorna raspodela za Poasonov model, definisana kao u Teoremi 1.5.

Rešenje:

Iz Primera 1.2 imamo da je $q(\theta) = e^{e^{-\theta}}$ i $\lambda = \exp(-\theta)$. Na osnovu teoreme, priorna raspodela je data sa

$$\pi(\theta) \propto \left[\exp\left(e^{-\theta}\right)\right]^{-k} \exp(-\theta\mu k).$$

Tada je priorna gustina za λ

$$\pi(\lambda) \propto [\exp(\lambda)]^{-k}\lambda^{\mu k}\lambda^{-1} = \lambda^{\mu k - 1}e^{-\lambda k},$$

što je gama distribucija sa $\alpha = \mu k$ i $\theta = \frac{1}{k}$. Izraz λ^{-1} potiče od $|d\theta/d\lambda|$, što je potrebno za smenu promenljive. $\square$

Drugi poznati primeri članova linearno eksponencijalne familije raspodela uključuju binomnu i negativnu binomnu raspodelu sa beta konjugovanom priornom raspodelom. Slično, gama je konjugovana priorna raspodela za eksponencijalnu raspodelu.

1.6. Poteškoće pri računanju

Postoje tri glavna problema u vezi sa Bejzovom paradigmom.

Jedan je odabir priorne raspodele. Ovo je aspekt Bejzove analize koji se najčešće kritikuje, sa prigovorom da lična priroda ove raspodele uklanja auru objektivnosti, koja se očekuje oko naučnog istraživanja.

Drugo je odabir modela. Ovaj problem nije ništa drugačiji od onog sa kojim se susreću svi analitičari.

Nakon što se odrede prethodna dva elementa, analizu čini evaluacija odgovarajućih izraza kao što je dato u prethodnim odeljcima. Iako formule izgledaju jednostavno, korišćeni integrali su često veoma komplikovani za izračunavanje. Za jednog Bejzovog analitičara je od esencijalne važnosti da ume da izvede višestruku numeričku integraciju. Za ovaj problem, do sada je razvijeno veliki broj tehnika, od kojih su neke prezentovane u [4]. Međutim, jedna tehnika koja je danas široko rasprostranjena se zove Markovi lanci Monte Karlo simulacije.[21]

1.7. Frekventistička i Bejzova statistika

U ovom odeljku nećemo favorizovati jedan statistički pristup u odnosu na drugi, već ćemo predstaviti određene karakteristike koje odlikuju razlike između ova dva pristupa.

Ako verovatnoću ishoda nekog događaja karakterišemo kao **verodostojnost** sa kojom će se taj posmatrani događaj realizovati, onda se osnovna razlika ova dva pristupa ogleda u različitom tumačenju pojma verodostojnosti.[22]

Frekventističko shvatanje pojma verovatnoće daje okvir i smatra se osnovom klasične statističke metodologije. Prema ovom shvatanju, koncept verovatnoće treba da bude objektivan i oslobođen uticaja bilo kakvih subjektivnih faktora i verovatnoća se može definisati i primeniti u situacijama koje se mogu iznova i iznova ponavljati pod identičnim uslovima. Frekventističke metode su još poznate i kao metode najveće preciznosti.[23] Frekventistički pristup definiše verodostojnost kao relativnu učestalost pozitivnih ishoda posmatranog događaja, odnosno broj ponavljanje u n pokušaja (izvođenja eksperimenta).

Prema Bejzovoj statistici pojam verovatnoće se tumači na subjektivan način, gde verovatnoća predstavlja stepen uverenja koji se koriguje u trenutku kada informacije ili podaci postanu raspoloživi. Bejzov pristup pretpostavlja da su samo trenutno posmatrani podaci relevantni za analizu i da je raspodela populacije ta koja je promenljiva. Ovaj pristup odlikuje upotrebna priornih informacija i izražava verodostojnost u vidu stepena verovanja (uverenja) da će se dati događaj realizovati. Priorne informacije postoje u izobilju i potiču iz prethodnih studija, objavljenih radova, intuicije istraživača, mišljenja nezavisnih eksperata, iz pogodnosti, srodnosti, neodređenosti itd. Bejzova statistika ima veliku primenu u aktuarstvu, o čemu ćemo više pričati u narednom odeljku.

U tabeli koja sledi navodimo neke oprečne stavove, koji odlikuju različite statističke pristupe.[24]

Frekventistički pristup	Bejzov pristup
Tradicionalan stav da je uzorkovanje beskonačno i da pravila odlučivanja mogu biti precizna.	Nepoznati kvantiteti se tretiraju probabilistički i stanje sveta se uvek može ažurirati.

[21] Dobra diskusija o ovom metodu se može naći u radu Scollnik, D. (2001), "Actuarial Modeling with MCMC and BUGS," *North American Actuarial Journal*, **5**, 96–124.
[22] Vidi [12].
[23] Vidi [13].
[24] Vidi [15].

Nepoznati parametri su fiksni i nepromenljivi pod svim realnim okolnostima.	Svet se pre posmatra probabilistički, nego kao fiksan fenomen koji je ili poznat ili nepoznat.
Nema a priori informacija za specifikaciju modela.	Priorne informacije postoje u obilju, od velike su pomoći i važno je koristiti ih.
Statistički rezultati pretpostavljaju da podaci potiču iz kontrolisanog eksperimenta.	Pretpostavke se postavljaju veoma pažljivo i često se moraju braniti.
Ponavljanje eksperimenta je važno, bez obzira na troškove koje iziskuje.	Svaki statistički model koji je ikada napravljen od postojanja sveta je subjektivan i to se mora priznati.
Studije su ponovljive.	Studije su fiksne.
Parametri su fiksni i ostaju konstantni tokom ponavljanja procesa (eksperimenta).	Pararametri su nepoznati i opisani probabilistički.
Posmatrani podaci potiču iz ponovljenih slučajnih uzorkovanja.	Posmatrani podaci potiču iz realizovanog uzorka.
Javlja se učestalost.	Podaci su fiksni.
Verovatnoće su jednake za svaku studiju.	Verovatnoće se ažuriraju sa svakom studijom.
Kreira se interval poverenja.	Kreira se kredibilni interval.
Tumačenje 90% intervala poverenja: kolekcija intervala, od kojih 90% njih sadrži stvarnu vrednost parametra.	Tumačenje 90% kredibilnog intervala: interval koji ima 90% šanse da sadrži stvarnu vrednost parametra.
U ponovljenom uzorkovanju, 90% realizovanih intervala pokriva stvarnu vrednost parametra.	Na osnovu posmatranih podataka, parametar će biti u intervalu sa verovatnoćom od 90%.

Tabela 1.1 Oprečni stavovi dva statistička pristupa

Čini se prikladnim da ovde parafraziramo delove diskusije, koju je u svom radu objavio Artur Bejli[25], o razlikama u filozofiji aktuara i statističara u uobičajenim poljima statističkih studija. Artur je bio aktuar, koji je više od bilo koga drugog doprineo znanju o teoriji kredibiliteta.[26]

- **Prvo**, postoji verovanje da aktuari neživotnog osiguranja nikada nisu lišeni predznanja pre nego što pristupe statističkoj analizi. Kada se formira nova linija proizvoda osiguranja, ili se uspostavlja nova klasifikacija, postoje značajna odstupanja u mišljenju pojedinačnih aktuara o tome koliko treba da iznosi nova premija, ali na kraju koncenzus mišljenja proizvodi premiju. Ova premija se potom ugrađuje u mišljenje aktuara kao prava premija. Kasnije, kada se konačno prikupe stvarni podaci o štetama iz nove klase ili nove linije proizvoda, problem više nije koliko treba da iznosi nova premija, već na koji način je potrebno izmeniti već postojeću premiju zbog prikupljenih opservacija.
- Statističke metode, razvijene od strane matematičara, se bave evaluacijom indikacija grupe opservacija, pod prećutnom ili implicitnom pretpostavkom da nije postojalo nikakvo predznanje koje je prethodilo stvaranju tih određenih opservacija. Kredibilnosne procedure, korišćene pri revidiranju premija, su razvili aktuari da bi dali konzistentnu težinu dodatnom znanju u kombinaciji sa već postojećim znanjem.
- **Drugo** verovanje aktuara neživotnog osiguranja jeste da su oni u kontinualnom poslu. Rizici nastaju i šire se stalno. Stoga aktuar ima na hiljade premija koje treba revidirati u

[25] Diskusija Arthur L. Bailey, Journal of the American Teachers of Insurance, Vol. 17, p. 24 (1950)
[26] Vidi [5].

relativno kratkim i čestim intervalima. Da bi mogao da napravi veliki broj ocena, odriče se uslova, nametnutog od strane drugih statističara, da svaka ocena bude nepristrasna. Umesto njega, može nametnuti manje restriktivan uslov, da određena grupa ocena agregatno bude nepristrasna. Na ovaj način se značajno smanjuje varijansa u greškama, u odnosu na one koje bi nastale primenom prezentovanom metodom statističkog ocenjivanja.

- **Treća** naročitost jeste da aktuari smatraju da su svi osiguranici međusobno različiti. Na primer, svaki vozač ima određene navike koje su samo njemu svojstvene, za svaki kamion se pretpostavlja da ima svoju rutu putovanja koja ga izlaže riziku drugačijem od ostalih. Posedovanje ove pretpostavke je potvrđeno toliko mnogo puta u praksi, da je razlika u rizicima postala osnovni koncept ili aksiom.
- Uprkos jedinstvenosti svojstvenih rizika različitih osiguranika, svi su oni izloženi ćudljivosti šanse i slučajnoj grešci klasifikacije i mere uobičajene svim statistikama. Statističke metode, generalno podučavane i objavljivane u udžbenicima, se bave populacijama za koje je ukupna varijacija rezultat ćudljivosti šanse i slučajne greške merenja. Populacije u neživotnom osiguranju se, međutim, sastoje od individua sa varijansom u očekivanoj vrednosti nastalom usled nekih drugačijih razloga od gore navedenih. Mora se pretpostaviti da se njihovi svojstveni rizici razlikuju, čak i ako je nemoguće precizno meriti te razlike. Ovo bavljenje heterogenim populacijama daje neke veoma interesantne rezultate, čemu se statističari podsmehuju kao da je nemoguće, ali je bez obzira na to, sasvim ispravno i opravdano.

Bez obzira da li se neko smatra pristalicom jednog ili drugog pristupa, svi su oni statističari, i u tom smislu, čini se razumnim da bi za rešenje problema trebalo koristiti sve moguće raspoložive alate.

1.8. Primena Bejzove statistike

> „Analiza u verovatnoći je fascinantna tema, kojoj je suđeno da igra veliku ulogu u aktuarskoj nauci; i možda će doći dan kada će se istinski tvrditi da je aktuar uspeo da spoji teoriju finansija sa teorijom verovatnoće.“[27]

Generalno, u istoriji Bejzove statistike možemo razlikovati određene revolucionarne periode. Pre pojave Markovih lanaca Monte Karlo simulacije, krajem osamdesetih i početkom devedesetih godina prošlog veka, to je bilo malo filozofsko polje u kome su se ljudi bavili manjim problemima. Onog momenta kada smo na raspolaganje dobili kompjuterske alate, teorija je prosto doživela ekspanziju i primenu u raznim oblastima. Takva situacija se s godinama menja, kako podaci postaju sve obimniji (upotreba koncepta Big Data), da tradicionalni algoritmi korišćeni devedesetih i ranih godina 21. veka više nisu toliko korisni. Jedna od veoma interesantnih stvari u Bejzovoj statistici, jeste pokušaj da se dizajniraju potpuno novi algoritmi,

[27] Phillips E. W., A Biometry of the Measurement of Mortality--privately published, p. 5 (1935)

novi načini analize i novi tipovi modela za stvarne visokodimenzionalne komplikovane podatke, uz dozvoljavanje nesigurnosti i očuvanje teorijskih garancija da su ovi modeli zaista dobri.[28]

Bejzova paradigma se već dugi niz godina smatra najpogodnijom za primenu u analizi različitih modela iz oblasti aktuarstva, osiguranja i upravljanja rizicima. Po ugledu na aktuarska društva, koja su tokom vremena koristila Bejzov i Bulmanov kredibilitet za optimizaciju preciznog određivanja premija, i finansijski analitičari su počeli da koriste ovakav kredibilitet kako bi optimizovali svoje procene beta koeficijenata i kretanja akcija.[29]

Kao što smo već pomenuli, do početka devedesetih godina prošlog veka, Bejzovi metodi nisu bili toliko vidljivi zbog poteškoća u računanju vrednosti višestrukih integrala. Međutim, primena ovih metoda je svakako bila veoma rasprostranjena, što potvrđuje istraživanje na internet pretraživaču za zdravstvenu zaštitu Medlajn[30]. U istraživanju je upotreba ključne reči „Bejzov“, pokrivajući period od 1966. do 2001., proizvela 2,120 citata. Pretraga sa kombinacijom reči „Bejzov“ sa naslovom časopisa iz oblasti statističke medicine je proizvela 178 citata. Ovih 178 citata obuhvala članke koji se bave kliničkim ispitivanjem metodologija, rezultata, ekonomskom evaluacijom troškova, donošenjem odluka u kliničkoj praksi, lekovima, meta-analizom, sistemom prismotre, ekranizacijom i mapiranjem bolesti i metodama preživljavanja.[31]

Treba još pomenuti i primenu u analizi neuronskih mreža(uz upotrebu brain-imaging tehnologije) i njihovim vezama sa načinom ponašanja i formiranjem ličnosti.

Po mišljenju Džima Bergera[32](Berger Jim), profesora statističkih nauka sa Djuk univerziteta u SAD (Duke University), postoje četiri vrste problema za koje je primena Bejzove statistike veoma značajna:

- Bejzova statistika, sa uvođenjem subjektivnosti pri određivanju priorne raspodele i upotrebom njenog priornog znanja, od velike je pomoći kada sprovodimo istraživanje u oblasti u kojoj nemamo dovoljno raspoloživih informacija o posmatranom problemu. Džim se bavio problemom automobilske industrije u Mičigenu, gde su pokušali da ocene koliko je ekonomske dobiti moguće ostvariti u vezi sa efikasnošću goriva i regulacijama koje nameće vlada. Izgradili su veliki komplikovani hijerarhijski model uključujući različite proizvođače, različita vozila, različite delove automobila koji utiču na efikasnost goriva itd. Ustanovili su da za određeni deo problema nemaju raspoložive informacije, te su se obratili inženjerima kako bi iz razgovora sa njima, na osnovu njihovog ekspertskog znanja, prikupili neophodne informacije.
- Druga vrsta problema su oni problemi koji se jednostavnije rešavaju uz primenu Bejzovog pristupa. Na primer, jedan od strateških modela za radne mašine u industriji, naime modeli komponenata varijanse. Kod ovog modela, primena Bejzovog pristupa je

[28] Preuzeto iz intervjua sa Donson Dejvidom(Dunson David), profesorom statističkih nauka sa Djuk univerziteta u SAD (Duke University). Vidi [12].
[29] Vidi [13].
[30] Medlajn (Medline) je kreiran od strane Nacionalne medicinske bibliteke u SAD.
[31] Diskusija Rozenberga (Marjorie A. Rosenberg), profesora sa Univerziteta u Viskonsinu u SAD (University of Wisconsin). Vidi [14].
[32] Preuzeto iz intervjua sa Berger Džimom. Vidi [12].

veoma jednostavna, dok alternative mogu biti veoma komplikovane. Na primer, metoda maksimalne verodostojnosti je veoma česta alternativa koju koriste nebejzovci.

- U treću klasu spada upotreba statističkih metoda od strane nestatističara. Nestatističari nemaju najbolje razumevanje značenja uobičajenih statističkih testova, kao npr. značenje p vrednosti, čije pogrešno tumačenje u nekim industrijama može biti veoma skupo i opasno. Tada je Bejzovo rešenje, kao što je posteriorna verovatnoća da je hipoteza tačna, mnogo jednostavnije i lakše za razumevanje. Upotreba Bejzove analize pruža mogućnost nestatističarima da se s razumevanjem bave statističkom analizom podataka iz svoje oblasti.
- Zatim postoje cela polja istraživanja koja su apsolutno Bejzova. Jedno polje, kojim se Džim bavio, je klasifikacija nesigurnosti simulacionih modela. Ovaj model nastaje iz potrebe da se u inženjerstvu fizički eksperimenti zamene kompjuterskim modelima. Upotrebom kompjuterskih modela, na primer u automobilskoj industriji, se na ovaj način smanjuju troškovi testiranja prototipa. Postavlja se pitanje, koliko dobro ovakav kompjuterski model predstavlja realnost? Odgovor na ovo pitanje uključuje razne vrste podataka i upotrebu raznih statističkih koncepata, koji se ne mogu izvesti na nebejzovski način.

Kompanija Gugl, koja je specijalizovana za servise i proizvode u vezi sa internetom, takođe koristi metode Bejzove statistike.[33] Bejzov model uprosečavanja i Bejzova selekcija promenljivih ima veliku upotrebu kod uprosečavanja linarnih modela i modela logističke regresije. Ove metode se posebno koriste u pozadini Gugl anketiranja potrošača, kada je potrebno proširiti grupu ljudi na koju je anketa primenljiva. Najveća primena ovih modela je u proizvodu Gugl analitike, koji se koristi za optimizaciju internet stranica.

Bejzov model uprosečavanja (Bayesian model averaging) se primenjuje u određivanju nesigurnosti modela, gde se umesto jednog selektovanog modela, u analizi koristi ansambl modela u vezi sa posmatranim kvantitetom δ. Posteriorna gustina se dobija kao težinska prosečna vrednost gustina od δ, pod svakim od individualnih modela, gde su težine zapravo posteriorne verovatnoće posmatranih modela.

[33] Preuzeto iz intervjua sa Skot Stivom (Scott Steve), direktorom departmana za statističko istraživanje u kompaniji Gugl (Google). Vidi [12].

Poglavlje 2

2. Teorija kredibiliteta

> „U ovom momentu diskusije pojedinac mora da prizna da, iako se čini da iza tvrdnji aktuara stoji maglovita logika, ovo je previše nejasno za njega da bi razumeo. Iskusni statističar viče: „Apsurd! To je u direktnoj suprotnosti sa bilo kojom prihvaćenom teorijom statističkog ocenjivanja.“ Aktuari sami moraju da priznaju da su otišli izvan bilo čega što je matematički dokazano, da su sve korišćene vrednosti i dalje birane na osnovu rasuđivanja i da je jedina demonstracija koju mogu da naprave jeste, da u stvarnoj praksi, to zaista funkcioniše. Nemojmo zaboraviti, međutim, da su ovu demonstraciju napravili mnogo puta. Zaista funkcioniše!“[34]

Teorija kredibiliteta je jedan od temelja aktuarske nauke, bez kog moderno osiguranje ne bi postojalo u obliku u kom ga danas poznajemo. Ovo poglavlje služi kao uvod u tematiku i čini se razumnim da započnemo sa diskusijom o značenju reči kredibilitet.

2.1. Značenje kredibiliteta

Kada je originalno predstavljena u teoriji aktuarstva, reč kredibilitet označavala je *meru vere* koju aktuar veruje da treba dati određenim istorijskim podacima, radi određivanja premije polise osiguranja. Stoga, kada kažemo da je šteta nastala u okviru nove klase osiguranja “i dalje previše mala“ da bi bila kredibilnosna (verodostojna), impliciramo da iskustvo koje će nastati u budućnosti može biti veoma različito od onoga što smo do sada iskusili, te da se više uzdamo u naše predznanje bazirano na drugim podacima, kao što su premije iz sličnih klasa. U istom maniru, kada kažemo da su istorijski podaci automobilskog vozača jedne regije „potpuno verodostojni za određivanje premije“, impliciramo da su podaci, prilagođeni za trend faktor, adekvatni za uspostavljanje premije važeće za celu datu regiju, bez upućivanja na predhodne premije ili istorijske podatke iz drugih regija.

Kada koristimo termin kredibilitet, mi ne dovodimo u pitanje tačnost podataka, već dajemo parcijalnu statističku težinu određenim podacima u različitim formulama za određivanje premije.

Često je skup istorijskih podataka previše mali da bi bio potpuno verodostojan, ali je pak dovoljno veliki da bi imao određeni nivo kredibiliteta. Posledično, uspostavljena je skala

[34] Bailey, A. (1950), “*Credibility Procedures*,” Proceedings of the Casually Actuarial Society, XXXVII, 7–23, 94–115.

kredibilnosti, koja daje 0 kredibiliteta podacima koji su „previše mali“ da bi bili korisni za određivanje premije, i koja daje kredibilitet 1 onim podacima koji su „potpuno verodostojni“. U nastavku ovog poglavlja videćete da se teorija kredibiliteta bavi uspostavljanjem mera verodostojnosti i standarda za potpuni kredibilitet.

Artur Bejli[35] je odao posebno priznanje predznanju u teoriji kredibiliteta:

> „U ovom momentu, praktično sve metode statističkog ocenjivanja, koje se pojavljuju u udžbenicima o statističkim metodama ili se predaju na američkim univerzitetima, se baziraju na pretpostavci da su bilo koja informacija i sve srodne informacije ili *a priori* znanje bezvredni. ... Čini se da je samo u polju aktuarstva organizovan revolt protiv odbacivanja svog predznanja, kada se pravi ocena koristeći novostečene podatke.“

Međutim, količina kredibiliteta data istorijskim podacima nije u potpunosti unutrašnja karakteristika podataka. Na primer, svrha sa kojom se podaci koriste je uvek navedena ili implicirana u bilo kojoj meri kredibiliteta. Stoga, istorijski podaci vozača u jednoj državi mogu imati veći kredibilitet za uspostavljanje različitih premija za osiguranja od odgovornosti u datoj državi, dok isti podaci mogu imati manji kredibilitet za određivanje različitih premija za osiguranje od odgovornosti u susednoj državi.

Za razliku od standardne devijacije, ili neke druge mere efekta slučajne varijacije, kredibilitet nije jednostavna karakteristika koja se može izračunati nekom matematičkom formulom. Iako postoji određeni nivo korelacije između kredibiliteta i statističke varijanse, kredibilitet ima značaj samo uz impliciranu svrhu za koju se podaci koriste i uz razmatranje vrednosti dostupnog predznanja.

Bez obzira što je reč kredibilitet tokom godina značila različite stvari različitim ljudima, i što se obično bazirala na ograničenjima postavljenim datom rešenju, sve kredibilnosne procedure imaju zajednički problem koji treba da reše:

> „Na osnovu opservacija o članovima razmatrane klase i opservacija o članovima drugih klasa, oceni distribuciju (ili njene karakteristike) budućih zahteva za odštetu članova date klase.“[36]

2.2. Potreba za matematičkim modelom

Teorija kredibiliteta obezbeđuje set kvantitativnih alata, koji omogućuju osiguravaču da se izbori sa slučajnošću podataka, koji se koriste za predviđanje budućih događaja ili šteta. Drugim

[35] Bailey, A. (1950), “*Credibility Procedures*,” Proceedings of the Casually Actuarial Society, XXXVII, 7–23, 94–115.
[36] Vidi [4].

rečima teorija omogućuje osiguravaču da, na osnovu istorijskih podataka, koriguje buduću premiju za određeni rizik ili riziko grupu.

Mnogi faktori će uticati na istorijske podatke, kao na primer, individualne karakteristike osiguranika, karakteristike veće grupe kojoj osiguranik pripada, ekonomski i drugi spoljni faktori i slučajna priroda samog osiguranog događaja. Međutim, u ovoj studiji mi ćemo ignorisati ekonomske i slične faktore, i pretpostaviti da su oni uzeti u razmatranje izvan analize kredibiliteta.

Da bismo stvorili okruženje za teoriju kredibiliteta, moramo razmotriti različite karakteristike koje razlikuju osiguranike jednog od drugog, klasifikaciju varijabli koje su deo bilo kog procesa. Ne moramo se fokusirati na same karakteristike, već na činjenicu da se svaki osiguranik može svrstati u *k različitih klasa* na osnovu posmatranih karakteristika. Cilj je da se za svaku klasu odredi premija koja je konzistentna sa sklonošću te klase da proizvede štetu. Svaka klasa je jedan osiguranik, gde jedan osiguranik može biti poslodavac sa više individua pokrivenim grupnom polisom osiguranja.

Napomena: kada koristimo izraz „rizik“, generalno iste komentare možemo primeniti i na grupu rizika, gde je grupa kolekcija rizika sa nekom sličnom karakteristikom. Takođe, stvarna opservacija za određeni rizik ili riziko grupu tokom vremena t, može da predstavlja broj zahteva za odštetu tokom vremena t , racio štete(gubitaka)[37] u godini t, izloženost riziku u godini t itd.

Ako je posmatrani osiguranik konstantno bolji (ima manje istorijske štete) od pretpostavljene manualne (teorijske) premije, onda osiguranik može da traži umanjenje svoje premije. Ista logika je primenljiva i u situaciji sa lošijim osiguranikom. Veći rizik traži povećanje premije, međutim, u takvoj situaciji osiguranik sigurno neće sam tražiti povećanje premije.

Manualna premija[38] je dizajnirana tako da reflektuje očekivano ponašanje (prošlo ili buduće) cele riziko klase i implicitno pretpostavlja da su rizici homogeni. Međutim, svaki rizik je i dalje jedinstven i ne sasvim isti kao drugi rizici u klasi, stoga uvek postoji određeni nivo heterogenosti među rizicima. Sledi da će neki osiguranici imati veći, a neki manji, rizik u odnosu na pretpostavljenu manualnu premiju. Štete osiguranika nastaju kao slučajni događaji, te istorisjki podaci mogu biti loš ocenjivač za buduće štete. Umesto da se oslonimo samo na istorijske opservacije, bolje ocene možemo dobiti ukoliko kombinujemo date opservacije sa drugim informacijama (manualna premija).

U ovom slučaju, osiguravač je primoran da odgovori na veoma važno pitanje: Koliko je varijacija u iskustvu datog osiguranika objašnjena slučajnom varijacijom štete (zahteva za odštetu), a koliko činjenicom da je osiguranik zaista bolji ili lošiji od proseka? Drugim rečima, koliko verujemo ličnom iskustvu pojedinačnog osiguranika? Dok tražimo balans, moramo obratiti pažnju na sledeće:

[37] Prim. prev. Racio gubitaka=gubici/premije (Loss ratio=Losses/Premiums).
[38] Manualna premija je premija iz priručnika osiguravajuće kompanije(manual se sa engleskog prevodi sa priručnik).

- *Da li su klase homogene?* Ukoliko su svi rizici jedne klase jednaki sa istim očekivanim štetama, da li uopšte treba razmatrati lično iskustvo? U ovom slučaju mogli bismo samo koristiti manualnu premiju za datu klasu. U suprotnom, značajne varijacije u očekivanim štetama za rizik date klase, sugerišu da bi trebalo dati veću težinu(važnost) istorijskim podacima za individualni rizik.
- *Da li postoji varijacija u istorijskim podacima?* Što je veća varijacija u stvarnim štetama za individualni rizik, veća je i razlika između stvarnog posmatranog iskustva i njegove očekivane vrednosti, te je manja težina tih podataka, t.j. manji kredibilitet se pripisuje individualnom iskustvu.

Nastavimo li dalje, dve činjenice se moraju uzeti u obzir:

- Lično iskustvo osiguranika je verodostojnije, ukoliko osiguravač poseduje više istorijskih podataka o datom osiguraniku. Slično je i u grupi osiguranika, iskustvo veće grupe je verodostojnije od iskustva manje grupe.
- Uslovi konkurencije mogu prisiliti osiguravača da, zarad očuvanja posla, da nekom osiguraniku potpun kredibilitet (uzimajući u obzir samo istorijske podatke osiguranika i zanemarujući manualnu premiju) ili skoro potpun kredibilitet.

Drugo okruženje za teoriju kredibiliteta je formiranje premija za sistem klasifikacija. U ovom problemu svaka grupa je sačinjena od više individua, na primer, u osiguranju automobila klasu definišu određene karakteristike osiguranika: godine, pol, lokacija, bračni status itd. Pretpostavimo da smo zainteresovani za osiguranje zaštite na radu, gde mogu postojati hiljade profesija, a samim tim i hiljadu klasa, od kojih neke obezbeđuju veoma mali broj podataka. Da bismo precizno ocenili očekivane štete, zgodno je kombinovati ograničeno stvarno iskustvo sa drugim informacijama, kao što su prošle premije, ili iskustvo blisko povezane profesije. Teorija kredibiliteta omogućava osiguravaču da kvantitativno formuliše gore navedeni problem.

U narednom segmentu predstavićemo *teoriju kredibiliteta ograničene fluktuacije*, koja se drugačije naziva i *klasična teorija kredibiliteta*. Originalni rad na ovu temu je 1914. predstavio Mowbray[39]. Ova teorija obezbeđuje mehanizam za pripisivanje potpunog ili parcijalnog kredibiliteta iskustvu osiguranika (istorijskim podacima), i posledično pokušava da ograniči efekat koji će slučajne fluktuacije u opservacijama imati na ocene.

Drugi pristup se zove *teorija kredibiliteta visoke preciznosti*. Ovaj pristup je formalizovao Bulman u svom klasičnom radu iz 1967[40], gde je predstavio statistički okvir u kome je teorija kredibiliteta doživela svoj razvoj i procvat. Bulmanov pristup se drugačije naziva *kredibilitet najmanjih kvadrata*, zato što je cilj pristupa da se minimizuje srednjekvadratna greška između ocene i stvarne vrednosti parametra koji se ocenjuje.

[39] Mowbray, A. H. (1914), "*How Extensive a Payroll Exposure Is Necessary to Give a Dependable Pure Premium*?" Proceedings of the Casualty Actuarial Society, I, 24–30.
[40] Bühlmann, H. (1967), "*Experience Rating and Credibility*," ASTIN Bulletin, 4, 199–207.

Ako pretpostavimo da su se n zahteva za odštetu sa štetama $X_1, X_2, \dots, X_n$ javili tokom vremena razmatranja, sledeće vrednosti mogu biti korisne pri analizi troškova osiguranja rizika ili grupe rizika:

$$\text{Agregatna šteta} = X_1 + X_2 + \cdots + X_n$$
$$\text{Neto premija} = (X_1 + X_2 + \cdots + X_n)/\text{Izloženost riziku}$$
$$\text{Racio štete} = (X_1 + X_2 + \cdots + X_n)/\text{Zarađene premije}$$

U nastavku ćemo raditi sa neto premijom, ali napominjemo da je izvođenje primenljivo i na druge dve vrednosti. Neto premija se definiše kao ukupna šteta podeljena sa izloženošću[41], na primer ukoliko 300 automobila tokom godine proizvede štetu u ukupnom iznosu od €90,000, tada posmatrana neto premija iznosi €90,000/300 ili €300 po auto-godini.

Treba primetiti da je neto premija zapravo proizvod učestalosti i prosečnog iznosa štete:

$$\text{Neto premija} = \frac{\text{Ukupna šteta}}{\text{Izloženost}} = \left(\frac{\text{Broj zahteva za odštetu}}{\text{Izloženost}}\right)\left(\frac{\text{Ukupna šteta}}{\text{Broj zahteva za odštetu}}\right)$$
$$= (\text{Učestalost})(\text{Iznos štete}).$$

S obzirom da zavisi od broja i iznosna nastalih šteta, neto premija ima više razloga da varira u odnosu na bilo koju od navedenih vrednosti, učestalosti i iznosa štete.

Pre nego što izložimo teoriju kredibiliteta ograničene fluktuacije, razmotrimo primer koji demonstrira način na koji teorija kredibiliteta proizvodi bolje ocene.

■Primer 2.1

U posmatranoj populaciji automobilskih vozača, prosečni vozač doživi nesreću na svakih pet godina, ili ekvivalentno, prosečni vozač godišnje doživi nesreću sa 0.2 učestalosti. Pretpostavimo da je slučajno izabrani vozač iz ove populacije imao tri nesreće u proteklih pet godina, t.j. godišnje doživi nesreću sa 0.6 učestalosti. Oceni buduću stopu učestalosti za ovog vozača. Da li ona oznosi 0.2, 0.6 ili nešto između?[42]

Rešenje:

U slučaju da nemamo nikakve informacije o ovom vozaču,, osim činjenice da dolazi iz pomenute populacije, koristili bismo 0.2. Međutim, mi znamo, na osnovu istorijskih podataka, da učestalost nastanka nesreće za ovog vozača iznosi 0.6. Da li ovo treba da bude naša ocena buduće učestalosti nesreća za datog vozača? Verovatno ne. Postoji korelacija između prošle i buduće učestalosti, ali one nisu u savršenoj korelaciji. Nesreća je slučajni događaj koji se može dogoditi čak i dobrom vozaču sa malom očekivanom vrednosti za učestalost nesreće. S druge strane, loši vozači mogu da učestvuju u saobraćaju i po nekoliko godina bez da im se dogodi

[41] Izloženost se može različito definisati u različitim linijama osiguranja, kao na primer, vreme tokom kojeg je osiguranik izložen riziku, godina pokrivenosti za auto osiguranje(auto-godina), godina pokrivenosti za kuću, platni spisak, itd. Godina pokrivenosti za kuću predstavlja jednu kuću koja je osigurana za period od cele godine, ili dve kuće koje su osigurane na period od po pola godine, ili n kuća koje su osigurane za period od po 1/n godine.
[42] Primeri predstavljeni u poglavlju 2 potiču iz [1] i [2].

nesreća. Na osnovu navedene analize, možemo zaključiti da je očekivana buduća vrednost najverovatnije negde između 0.2 i 0.6. □

Ključ za rešenje ovog problema se nalazi u odgovarajućoj ravnoteži između istorijski posmatranih podataka i drugih informacija. Problem koji posmatramo je formulisan u sledećem izrazu sa težinskim koeficijentima:

$$Ocena = Z[Opservacije] + (1 - Z)[Druge\ informacije], 0 \leq Z \leq 1.$$

Z predstavlja *faktor kredibiliteta* i intuitivno izražava koliko su verodostojni posmatrani podaci. Posledično, 1-Z predstavlja *komplement kredibiliteta*. Ukoliko je skup istorijskih podataka dovoljno veliki i ne varira mnogo iz jednog u drugi period, tada faktor kredibiliteta Z teži jedinici. Suprotno, ukoliko su naše opservacije ograničen skup podataka, onda Z teži nuli i veća težina se daje drugim podacima. Izraz „druge informacije“ predstavlja ocenu premije, ili a priori hipotezu o premiji koja se naplaćuje u slučaju nedostatka istorijskih podataka (iskustva, istorijskog ponašanja). U momentu kada informacije o iskustvu postanu dostupne, možemo izračunati korigovanu ocenu premije, kombinujući istorijske podatke i a priori hipotezu. Ovakva formulacija nagoveštava da, kredibilitet podrazumeva linearnu ocenu stvarnog očekivanja, uz kompromis između opservacija i priorne hipoteze. Ukoliko se setimo prethodnog primera sa ovim modelom, za posmatranog vozača možemo izračunati očekivanu učestalost kao $Z0.60 + (1 - Z)0.20$.

2.3. Teorija kredibiliteta ograničene fluktuacije

Teorija kredibiliteta ograničene fluktuacije predstavlja prvi pokušaj da se kvantifikuje problem kredibilnosti, tako što će se ograničiti efekat koje slučajne fluktuacije u opservacijama imaju na ocene. Ovu teoriju je predstavio Mowbray[43] početkom dvadesetog veka u svom radu u vezi sa osiguranjem za zaštitu na radu.

Problem koji posmatramo:

Neka je slučajno izabrani osiguranik imao štetu X_j u godini j, gde je $j \in \{1,2,3,\dots,n\}$. Alternativno neka X_j predstavlja štetu j-te polise iz grupe ili štetu j-tog člana određene riziko klase. Treba napomenuti da, iako izraz „šteta“ označava iznos nastale štete, a „zahtev za odštetu“ broj zahteva za odštetu, u većini slučajeva ideja je podjednako primenljiva, bez obzira da li brojimo štete ili zahteve za odštetu.

- Sada pretpostavimo da je srednja vrednost stabilna tokom vremena, ili za članove iste grupe(ili klase) $E(X_j) = \xi$. Kada bismo znali srednju vrednost, ona bi bila naša najbolja ocena buduće premije (nakon obračunatih troškova, profita i provizije).
- Takođe, pretpostavimo da je varijansa jednaka za sve j, $Var(X_j) = \sigma^2$.

[43] Mowbray, A. H. (1914), “How Extensive a Payroll Exposure Is Necessary to Give a Dependable Pure Premium?” *Proceedings of the Casualty Actuarial Society*, **I**, 24–30.

- Istorijsko ponašanje možemo sumirati sa uzoračkom sredinom $\bar{X} = n^{-1}(X_1 + \cdots + X_n)$. Za $\bar{X}$ znamo da je $E(\bar{X}) = \xi$, a ukoliko su X_j međusobno nezavisni, onda je $Var(\bar{X}) = \frac{\sigma^2}{n}$.

Cilj je da se odredi vrednost ξ. Da bismo ocenili očekivanu vrednost ξ, možemo ignorisati istorijske podatke (bez kredibiliteta) i koristiti samo druge informacije M za određivanje premije, gde je M dobijeno iz iskustva sličnih ali ne identičnih osiguranika. Ova vrednost se zove *manualna premija* zato što potiče iz priručnika(prim. prev. manual se sa engleskog prevodi kao priručnik) sa premijama. U suprotnom, možemo ignorisati M i dati potpun kredibilitet istorijskim podacima, te kao premiju naplatiti prosečan iznos štete $\bar{X}$. Konačno, za premiju možemo naplatiti kombinaciju M i $\bar{X}$, te dati samo parcijalni kredibilitet istorijskim podacima.

Kao što smo već napomenuli u uvodu ovog poglavlja, što je iskustvo „stabilnije“ (manje varijabilno, σ^2 je malo), to više verujemo $\bar{X}$ i možemo ga koristiti kao ocenu rezultata sledeće godine. Suprotno, što je iskustvo volatilnije, manje ćemo koristiti uzoračku sredinu kao ocenu budućih vrednosti i ima više smisla verovati da će M odlično opisati sklonost ka nastanku štete. Dodatno, dajemo veću težinu $\bar{X}$ ukoliko verujemo da postoji velika verovatnoća da je rizik posmatranog osiguranika drugačiji od rizika ostalih osiguranika, čije iskustvo je korišćeno za određivanje manualne premije M. S druge strane, postojanje osiguranika sa sličnim srednjim vrednostima ξ , implicira da nema svrhe oslanjati se na (možda ograničeno) iskustvo bilo kog pojedinačnog osiguranika.

Napomena: za datu godinu j, šteta X_j može nastati od jednog osiguranika (npr. broj zahteva za odštetu podnetih u toku jedne godine na osnovu jedne auto polise osiguranja), od klase osiguranika sa sličnim karakteristikama (u braku, žene, ispod 35 godina starosti, žive u urbanoj zoni, pređu preko 10,000km godišnje), ili od grupe nekoliko osiguranika okupljenih zbog nekog određenog razloga (prosečan iznos štete za vozilo iz voznog parka jednog lanca veleprodaje).

Klasična teorija kredibiliteta nam omogućava da odredimo koliko je istorijskih podataka potrebno da bismo im pripisali verodostojnost od 100%. Količina neophodnih podataka se naziva *kriterijum za potpuni kredibilitet* ili *standard za potpuni kredibilitet*. Ukoliko količina podataka zadovoljava kriterijum za potpuni kredibilitet, onda je $Z = 1$, u suprotnom, posmatrana količina podataka nije dovoljna, te $0 \leq Z < 1$ (parcijalni kredibilitet).

„Kada aktuar kaže da su određeni podaci potpuno verodostojni, on ne implicira da, ukoliko je moguće prikupiti drugi uzorak istog obima pod jednakim uslovima, da će rezultat biti praktično identičan, već da je obim uzorka adekvatan za određivanje premije, bez upućivanja na drugo iskustvo i druge prethodno naplaćene premije.“[44] Interesantno je znati koliko je podataka potrebno da se prođe pomenuti test, ali test često zahteva toliko veliki skup podataka, koji realno nije dostupan, da upit postaje čisto teorijski.

Podaci iz polja osiguranja su drugačiji u odnosu na podatke nekih drugih studija, gde je moguće obezbediti uzorak praktično bilo kog obima, kada se utroši neophodno vreme i novac. Obično se razmatraju podaci određene klase i drugi jednaki podaci nisu raspoloživi. Samo u

[44] Vidi [5].

slučaju razmatranja podataka jedne kompanije, dostupan je veći skup podataka iz svih kompanija koje posluju na sličan način. Međutim, i u ovom slučaju veoma je bitno naglasiti reči „posluju na sličan način“.

U nastavku poglavlja će se razmatrati dva osnovna koncepta teorije klasičnog kredibiliteta, t.j. kako odrediti:

- kriterijum za *potpun kredibilitet,* kada ocenjujemo učestalost zahteva, iznos štete ili izloženost riziku
- koliko iznosi *parcijalni kredibilitet* u slučaju da raspolažemo sa manje podataka nego što je neophodno za potpun kredibilitet.

2.3.1. Potpun kredibilitet

Na momenat razmotrimo sledeći problem.

Pretpostavimo da u toku jedne godine broj zahteva za odštetu prati Poasonovu raspodelu sa sa parametrom $\lambda = 500$. Kako su kod Poasonove raspodele srednja vrednost i varijansa jednake parametru λ, posmatrani broj zahteva za odštetu će iz godine u godinu varirati oko srednje vrednosti *500*. Ovaj Poasonov proces možemo aproksimovati Normalnom distribucijom da bi ocenili koliko često će posmatrani rezultat odstupati od srednje vrednosti.

Na primer, koliko često očekujemo da će biti podneto više od 550 zahteva za odštetu? Standardna devijacija iznosi $\sqrt{500} = 22.36$, te je 550 zahteva za $\frac{50}{22.36} = 2.24$ standardne devijacije veće od proseka. Stoga $\Phi(2.24) = 0.9875$ implicira da postoji približno 1.25% šanse da ćemo registrovati više od 550 zahteva, t.j. postoji 0.0125 verovatnoća da će posmatrani broj zahteva biti veći od očekivanog broja zahteva za 10% ili više. Analogno, šansa da će broj zahteva biti manji od 450 je približno 1.25%. Ovakva formulacija problema ukazuje da postoji šansa od 2.5% da registrovani broj zahteva za odštetu bude izvan intervala od -10% ispod do +10% iznad srednje vrednosti. Drugim rečima, verovatnoća da broj zahteva bude u ±10% očekivane vrednosti od broja zahteva za odštetu, u ovom slučaju, iznosi 0.975.

Pristupimo ovom problemu algebarski.

Da bismo kvantifikovali stabilnost $\bar{X}$, možemo zaključiti da je $\bar{X}$ stabilna ukoliko je razlika između $\bar{X}$ i ξ relativno mala u odnosu na ξ sa velikom verovatnoćom. Da se izrazimo statističkim terminima, stabilnost možemo definisati tako što ćemo izabrati dva broja $r > 0$ i $0 < p < 1$ (u prethodnom primeru ove vrednosti su $r = 0.1$ i $p = 0.975$) i dodeliti potpun kredibilitet ako

$$\Pr(-r\xi \leq \bar{X} - \xi \leq r\xi) \geq p. \tag{2.1}$$

Zgodno je prethodni izraz (2.1) zabeležiti kao

$$\Pr\left(\left|\frac{\bar{X} - \xi}{\sigma/\sqrt{n}}\right| \leq \frac{r\xi\sqrt{n}}{\sigma}\right) \geq p$$

Sada definišimo y_p

$$y_p = \inf_y \left\{ \Pr\left(\left| \frac{\bar{X} - \xi}{\sigma/\sqrt{n}} \right| \leq \frac{r\xi\sqrt{n}}{\sigma} \right) \geq p \right\}. \tag{2.2}$$

Ovde y_p predstavlja najmanju vrednost y za koju funkcija verovatnoće zadovoljava nejednačinu u vitičastim zagradama (2.2). Ako $\bar{X}$ ima neprekidnu raspodelu, onda se znak $\geq$ u (2.2) može zameniti sa = i y_p zadovoljava jednačinu

$$\Pr\left(\left| \frac{\bar{X} - \xi}{\sigma/\sqrt{n}} \right| \leq y_p \right) = p, \tag{2.3}$$

Tada je uslov za potpuni kredibilitet dat sa $r\xi\sqrt{n}/\sigma \geq y_p$,

$$\frac{\sigma}{\xi} \leq \frac{r}{y_p}\sqrt{n} = \sqrt{\frac{n}{\lambda_0}}\,, \tag{2.4}$$

gde je $\lambda_0 = (y_p/r)^2$. Uslov iz (2.4) tvrdi da se potpuni kredibilitet dodeljuje ako koeficijent varijacije σ/ξ nije veći od $\sqrt{n/\lambda_0}$, što je intuitivan rezultat.

Interesantno je primetiti da se (2.4) može drugačije zabeležiti, tako da ukazuje da se potpuni kredibilitet javlja kada je

$$Var(\bar{X}) = \frac{\sigma^2}{n} \leq \frac{\xi^2}{\lambda_0}. \tag{2.5}$$

Alternativno, rešavajući (2.4) po n , dobijamo neophodan broj za *potpuni kredibilitet po jedinici izloženost*

$$n \geq \lambda_0 \left(\frac{\sigma}{\xi}\right)^2. \tag{2.6}$$

Često je opravdano aproksimovati $\bar{X}$ normalnom distribucijom sa sredinom ξ i varijansom σ^2/n. Uz primenu centralne granične teoreme za veliko n, $(\bar{X} - \xi)/(\sigma/\sqrt{n})$ ima standardizovanu normalnu raspodelu. Sada (2.3) možemo napisati kao (gde Z ima standardizovanu normalnu raspodelu, a $\Phi(y)$ je njena funkcija verovatnoće)

$$\begin{aligned} p &= \Pr(|Z| \leq y_p) \\ &= \Pr(-y_p \leq Z \leq y_p) \\ &= \Phi(y_p) - \Phi(-y_p) \\ &= \Phi(y_p) - 1 + \Phi(y_p) \\ &= 2\Phi(y_p) - 1. \end{aligned}$$

Tada je $\Phi(y_p) = (1 + p)/2$, a y_p je $(1 + p)/2$ kvantil standardizovane normalne raspodele.

Da sumiramo, ukoliko želimo da budemo u $\pm r\%$ sredine sa verovatnoćom najmanje p, onda *standard za potpuni kredibilitet u odnosu na jedinicu za izloženost riziku* glasi

$$n \geq \lambda_0 \left(\frac{\sigma}{\xi}\right)^2, \text{ gde } \lambda_0 = (y_p/r)^2 \text{ i } y_p \text{ je takvo da } \Phi(y_p) = (1+p)/2.$$

Na primer, ako je $p = 0.9$, onda iz tablice za normalnu raspodelu znamo da $y_{0.9} = 1.645$. Ako dodatno znamo da $r = 0.05$, onda $\lambda_0 = (1.645/0.05)^2 = 1{,}082.41$ i iz (2.6) sledi $n \geq 1{,}082.41\ \sigma^2/\xi^2$, uz pretpostavku da znamo koeficijent varijacije za *Xj*.

Napomena: kada koristimo (2.6), upotrebljavamo koeficijent varijacije ocenjivača kojim ocenjujemo željenu vrednost. Na desnoj strani izraza imamo standard za potpuni kredibilitet izražen u jedinici za izloženost riziku. Druge merne jedinice se mogu izračunati, tako što će se obe strane izraza pomnožiti sa odgovarajućom vrednosti. Na primer, često se ovaj standard prevodi u broj zahteva za odštetu, tako što se obe strane pomnože sa (približnom) očekivanom vrednosti učestalosti zahteva. Dakle, s obzirom da se sve količine ocenjuju na osnovu podataka, implicitno važi da se pitanje kredibiliteta može postaviti na više različitih načina.

Potrebno je napraviti adekvatan izbor za vrednosti p i r. Iako je rasuđivanje koristi u ovom procesu, u praksi se standard za potpuni kredibilitet generalno bira u okviru sličnog intervala. Primećujemo da se isti tip rasuđivanja koristi u izboru nivoa greške prve vrste za statističko ocenjivanje parametara. Obično se bira ±2 standardne devijacije, što približno odgovara 95% intervalu poverenja, međutim, to ne znači nužno da je je takav izbor bolji od, na primer, ±1.5 ili ±2.5 standardne devijacije. Činjenica da je rasuđivanje korišćeno u teoriji klasičnog kredibiliteta donekle proizvoljno, nije sprečila njenu višedecenijsku primenu.

Ovde imamo tabelu[45] sa vrednostima koje označavaju *standard za potpuni kredibilitet u odnosu na učestalost n*, za razne vrednosti p i r:

Nivo verovatnoće *p*	*r*=30%	*r*=20%	*r*=10%	*r*=7.5%	*r*=5%	*r*=2.5%	*r*=1%
80.00%	18	41	164	292	657	2,628	16,424
90.00%	30	68	271	481	**1,082**	4,329	27,055
95.00%	43	96	384	683	1,537	6,146	38,415
96.00%	47	105	422	750	1,687	6,749	42,179
97.00%	52	118	471	837	1,884	7,535	47,093
98.00%	60	135	541	962	2,165	8,659	54,119
99.00%	74	166	664	1,18	2,654	10,616	66,349
99.90%	120	271	1,083	1,925	4,331	17,324	108,276
99.99%	168	378	1,514	2,691	6,055	24,219	151,367

Tabela 2.1 Standard za potpuni kredibilitet u odnosu na učestalost (broj zahteva za odštetu)

[45] Vidi [5]

Sa izborom *p=90%* i *r=5%*, mi želimo da sa verovatnoćom od 90% budemo udaljeni najviše ±5% od srednje vrednosti. Drugim rečima, spremni smo da u 5% slučajeva budemo izvan na svakom repu, a u totalu da sa verovatnoćom 0.1 budemo izvan prihvatljivog intervala.

Standard koji odgovara vrednostima *p*=90% i *r* =5%, naime 1,082, se najčešće koristi kao standard za kredibilitet. Ovakva tablica sa standardima je dostupna i aktuari je koriste već dugi niz godina.

Sada pretpostavimo da nas interesuje standard za potpuni kredibilitet izražen u drugim mernim jedinicama.

■Primer 2.2[46]

Pretpostavimo da su nam za određenog osiguranika dostupni istorijski podaci o štetama $X_1, \dots, X_n$ i da se očekivana vrednost $E(X_j) = \xi$ može oceniti uzoračkom sredinom. Ako pretpostavimo da imamo 10 opservacija: 253, 0, 0, 0, 398, 0, 439, 0, 0, 756, odredi standard za potpuni kredibilitet u slučaju da je r = 0.05, a p = 0.9.

Rešenje:

Podsetimo se uslova za potpuni kredibilitet (2.6)

$$n \geq \lambda_0 \left(\frac{\sigma}{\xi}\right)^2.$$

U konkretnom slučaju sredina i standardna devijacija se mogu oceniti iz podataka sa 184.6 i 267.89 (gde za ocenu varijanse koristimo centriranu ocenu sa $n - 1$). Sa $\lambda_0 = 1082.41$, standard iznosi

$$n \geq 1082.41 \left(\frac{267.89}{184.6}\right)^2 = 2279.51,$$

Količina istorijskih podataka (10 opservacija) ne ispunjava uslov za potpuni kredibilitet. □

U daljem tekstu razmotrimo primer sa opservacijama, koje potiču iz određenog tipa raspodele.

■Primer 2.3

Neka su

- *za određenog osiguranika dostupni istorijski podaci $X_1, \dots, X_n$*
- *svi X_j međusobno nezavisni i isto raspodeljeni sa složenom Poasonovom distribucijom $X_j = Y_{j1} + \dots + Y_{jN_j}$, gde N_j ima Poasonovu raspodelu sa parametrom λ, a iznosi štete Y su raspodeljeni sa sredinom θ_Y i varijansom $\sigma^2{}_Y$.*

Odredi standard za potpuni kredibilitet u odnosu na očekivani broj zahteva za odštetu po polisi osiguranja i standard za potpuni kredibilitet u odnosu na očekivani iznos šteta po polisi osiguranja. Proveri da li podaci iz Primera 2.2 zadovoljavaju ove standarde, s tim da ćemo

[46] Za opsežan skup zadataka za vežbu, pogledajte [1] i [2].

smatrati da prve tri nenegativne isplate potiču od jednog zahteva za odštetu, a poslednji iznos potiče od dva zahteva, jedan u vrednosti od 129 i drugi 627.

Rešenje:

Prvi slučaj: Prvo ćemo odrediti standard za potpuni kredibilitet u odnosu na očekivani broj zahteva za odštetu po polisi osiguranja. Posmatramo $N_j s$, za koje imamo $\xi = E(N_j) = \lambda$ i $\sigma^2 = \mathrm{Var}(N_j) = \lambda$. Pozovimo se na uslov za potpuni kredibilitet iz (2.6), i odredimo

$$n \geq \lambda_0 \left(\frac{\lambda^{\frac{1}{2}}}{\lambda} \right)^2 = \frac{\lambda_0}{\lambda}.$$

Dakle, *standard u odnosu na broj polisa (izloženost)* će morati da bude veći od λ_0/λ da bi bio potpuno kredibilnosan, a λ se mora oceniti na osnovu podataka. Standard u odnosu na očekivani broj zahteva iznosi $n\lambda$. Ako pomnožimo obe strane sa λ dobijamo standard

$$n\lambda \geq \lambda_0.$$

Čini se da nije potrebno nikakvo ocenjivanje za ovaj standard, ali potrebno je u odnosu na očekivani broj zahteva. U praksi, standard se određuje u odnosu na stvaran očekivani broj zahteva, efektivno zamenjujući $n\lambda$ na levoj strani sa njegovom ocenom $N_1, \dots, N_n$.

Da li su ovi standardi ispunjeni za raspoložive podatke? U Primeru 2.2 smo imali 5 zahteva za odštetu, te ocena za učestalost λ iznosi 0.5 po polisi. Standard iznosi

$$n \geq \frac{1{,}082.41}{0.5} = 2{,}164.82,$$

i 10 polisa nisu dovoljne da bi se zadovoljio standard. Ako sada uporedimo 5 stvarnih zahteva sa $\lambda_0 = 1{,}082.41$, dolazimo do istog rezultata.

Drugi slučaj: Drugo po redu određujemo *standard za potpuni kredibilitet u odnosu na prosečnu ukupnu isplatu štete*. Prisetimo se, kada koristimo uslov za potpuni kredibilitet, koristi se koeficijent varijacije za ocenjivača kvantiteta koji ocenjujemo. Za složenu Poasonovu raspodelu imamo $\xi = E(X_j) = \lambda\theta_Y$ i $\mathrm{Var}(X_j) = \lambda({\theta_Y}^2 + {\sigma_Y}^2)$. *U odnosu na obim uzorka*, standard iznosi

$$n \geq \lambda_0 \frac{\lambda\left({\theta_Y}^2 + {\sigma_Y}^2\right)}{\lambda^2 {\theta_Y}^2} = \frac{\lambda_0}{\lambda}\left[1 + \left(\frac{\sigma_Y}{\theta_Y}\right)^2\right].$$

Da bismo dobili *standard u odnosu na očekivani broj zahteva za odštetu* moramo pomnožiti obe strane sa λ

$$n\lambda \geq \lambda_0 \left[1 + \left(\frac{\sigma_Y}{\theta_Y}\right)^2\right].$$

Konačno, da bismo postavili *standard u odnosu na očekivani iznos ukupne isplaćene odštete*, množimo obe strane sa θ_Y

$$n\lambda\theta_Y \geq \lambda_0 \left(\theta_Y + \frac{{\sigma_Y}^2}{\theta_Y}\right).$$

Ako se prisetimo podataka iz Primera 2.2, pet zahteva imaju sredinu 369.2 i standardnu devijaciju 189.315, te

$$n \geq \frac{\lambda_0}{\lambda}\left[1 + \left(\frac{\sigma_Y}{\theta_Y}\right)^2\right] = \frac{1{,}082.41}{0.5}\left[1 + \left(\frac{189.315}{369.2}\right)^2\right] = 2{,}734.02.$$

Sa 10 raspoloživih opservacija, opet nemamo dovoljno podataka da ispunimo standard. Ako želimo da postavimo standard u odnosu na broj zahteva (kojih imamo pet), množimo obe strane sa 0.5 i dobijamo standard u iznosu od 1,367.01. Konačno, da bismo odredili standard u odnosu na ukupan iznos isplaćene štete, množimo obe strane sa 369.2 i dobijamo standard 504,701. Primetimo da je racio posmatranog kvantiteta u odnosu na odgovarajući standard nepromenjen:

$$\frac{10}{2{,}734.02} = \frac{5}{1{,}367.01} = \frac{1{,}846}{504{,}701} = 0.003658. \qquad \square$$

Videli smo da standard za potpuni kredibilitet nije zadovoljen ni u jednom od tri razmatrana primera. Pošto uzoračka sredina nije dovoljno precizna za ocenu očekivane vrednosti, potreban nam je drugačiji metod da se izborimo sa pomenutom situacijom.

2.3.2. Parcijalni kredibilitet

> "Koliko će se težine dati određenoj količini podataka u poslovanju neživotnog osiguranja, ostaje da bude stvar ličnog rasuđivanja."[47]

Ako osiguravač ima najmanje broj zahteva neophodnih za potpuni kredibilitet, onda može tim istorijskim podacima pripisati 100% kredibiliteta. Međutim, kada mu nedostaju podaci, poseduje manje podataka nego što je traženo standardom za potpuni kredibilitet, onda istorijskim podacima pripisuje manje od 100% kredibiliteta.

Prisetimo se još jednom uslova za potpuni kredibilitet (2.6). Ukoliko proračun implicira da potpuni kredibilitet nije prikladan u datoj situaciji, onda u našoj oceni buduće neto premije želimo da reflektujemo i neke druge informacije. To činimo tako što kombinujemo istorijsko

[47] Arthur L. Bailey – Some Notes on Credibility – Proceedings of Casualty Actuarial Society XIX, p.65, 1932

ponašanje $\bar{X}$ sa eksterno dobijenom sredinom, M, koristeći težinsku sredinu, odnosno, kredibilnosnu premiju

$$P_c = Z\bar{X} + (1 - Z)M, \quad 0 \leq Z \leq 1 \tag{2.7}$$

Sada je neophodno izabrati faktor kredibiliteta $Z \in [0, 1]$. Aktuarska literatura nam sugeriše različite formule za Z, koje se obično više zasnivaju na intuiciji nego na teoriji. U odeljku posvećenom teoriji kredibiliteta visoke preciznosti, pokazaćemo da je specijalan izbor za Z,

$$Z = \frac{n}{n + k}, \tag{2.8}$$

gde je neophodno odrediti k, teorijski opravdan na osnovu statističkog modela. Ali za sada ćemo se fokusirati na *pravilo korena*, izbor baziran na istoj ideji na kojoj je zasnovan potpun kredibilitet.

Rezimirajmo da, standard za potpuni kredibilitet ima za cilj da obezbedi da razlika između neto premije koju posmatramo ($\bar{X}$) i one koju bi trebalo da koristimo (ξ) bude mala sa velikom verovatnoćom. Drugim rečima, cilj je da se kontroliše varijansa predložene neto premije, $\bar{X}$, imajući u vidu da je nepristrasan ocenjivač iz normalne raspodele. Iz (2.5) možemo videti da ne postoji garancija da će varijansa od $\bar{X}$ biti dovoljno mala, ali s druge strane možemo kontrolisati varijansu kredibilnosne premije, P_c.

Neka je $\bar{X}_{parcijalno}$ dobijeno iz parcijalno verodostojnih podataka, a $\bar{X}_{potpuno}$ dobijeno iz podataka koji zadovoljavaju standard za potpuni kredibilitet. U slučaju potpuno verodostojnih podataka,

$$\text{Ocena} = \bar{X}_{potpuno},$$

dok parcijalno verodostojni podaci utiču na ocenu sa težinom Z:

$$Ocena = Z\bar{X}_{parcijalno} + (1 - Z)M.$$

Sa svakim izborom $Z < 1$ varijansa od $Z\bar{X}_{parcijalno}$ se može smanjiti, čime želimo da ograničimo očekivanu varijaciju u $Z\bar{X}_{parcijalno}$ varijacijom dozvoljenom u oceni potpunog kredibiliteta $\bar{X}_{potpuno}$.

Pretpostavimo da želimo da ocenimo očekivani broj zahteva za odštetu po jedinici za izloženost riziku. Kao ocene koristićemo uzoračke sredine $\bar{X}_{parcijalno}$ and $\bar{X}_{potpuno}$, formirane na osnovu dva različita uzorka uzeta iz homogene populacije sa sredinom ξ i standardnom devijacijom σ, gde su X_i za svako i međusobno nezavisni. Oba uzorka će imati istu očekivanu vrednost, ali zbog manjeg obima uzorka, $\bar{X}_{parcijalno}$ će imati veću standardnu devijaciju $\sigma_{parcijalno}$ od standardne devijacije $\sigma_{potpuno}$ za ocenu potpunog kredibiliteta $\bar{X}_{potpuno}$. Želeći da ograničimo fluktuacije u $\bar{X}_{parcijalno}$ i $\bar{X}_{potpuno}$ na $\pm r\xi$ $(r > 0)$ od sredine populacije, cilj da ograničimo fluktuacije u izrazu $Z\bar{X}_{parcijalno}$ sa onim iz $\bar{X}_{potpuno}$ formulišemo na sledeći način:

$$\Pr\left(-r\xi \le \bar{X}_{potpuno} - \xi \le r\xi\right) = \Pr\left(-r\xi \le Z\bar{X}_{parcijalno} - Z\xi \le r\xi\right).$$

Ako podelimo sa odgovarajućim standardnim devijacijama, dobijamo:

$$\Pr\left(\frac{-\mathrm{r}\xi}{\sigma_{\mathrm{potpuno}}} \le \frac{\bar{\mathrm{X}}_{\mathrm{potpuno}} - \xi}{\sigma_{\mathrm{potpuno}}} \le \frac{\mathrm{r}\xi}{\sigma_{\mathrm{potpuno}}}\right) = \Pr\left(\frac{-\mathrm{r}\xi}{\mathrm{Z}\sigma_{\mathrm{parcijalno}}} \le \frac{\mathrm{Z}\bar{\mathrm{X}}_{\mathrm{parcijalno}} - \mathrm{Z}\xi}{\mathrm{Z}\sigma_{\mathrm{parcijalno}}} \le \frac{\mathrm{r}\xi}{\mathrm{Z}\sigma_{\mathrm{parcijalno}}}\right),$$

gde je $Z\xi$ sredina za $Z\bar{X}_{parcijalno}$, a njena standardna devijacija je $Z\sigma_{parcijalno}$.

Ako pretpostavimo da $(\bar{X}_{potpuno} - \xi)/\sigma_{potpuno}$ i $(Z\bar{X}_{parcijalno} - Z\xi)/Z\sigma_{parcijalno}$ možemo aproksimovati normalnom raspodelom, onda su desna i leva strana izraza jednake ako:

$$\frac{r\xi}{\sigma_{potpuno}} = \frac{r\xi}{Z\sigma_{parcijalno}}.$$

Rešavajući ovo po *Z* dobijamo:

$$Z = \frac{\sigma_{potpuno}}{\sigma_{parcijalno}}.$$

Takva formulacija implicira da je parcijalni kredibilitet Z obrnuto proporcionalan standardnoj devijaciji parcijalno verodostojnih podataka. Znamo da je $(\bar{X}) = \frac{\sigma^2}{n}$ za uzorak obima *n*, gde su X_i međusobno nezavisni za svako *i* , i potiču iz populacije sa varijansom σ^2. Kako oba uzorka potiču iz iste populacije sa varijansom σ^2 , uzoračka sredina za uzorak potpunog kredibiliteta $\bar{X}_{potpuno}$ obima n_0 ima standardnu devijaciju $\sigma_{potpuno} = \sigma/\sqrt{n_0}$, a uzoračka sredina za uzorak parcijalnog kredibiliteta $\bar{X}_{parcijalno}$ obima *n* ima standardnu devijaciju $\sigma_{parcijalno} = \sigma/\sqrt{n}$. Ako u formuli uvrstimo vrednosti za $\sigma_{potpuno}$ i $\sigma_{\mathrm{partial}}$,

$$Z = \frac{\sigma_{potpuno}}{\sigma_{parcijalno}} = \frac{\sigma/\sqrt{n_0}}{\sigma/\sqrt{n}} = \sqrt{\frac{n}{n_0}}$$

dobijamo faktor kredibiliteta za manji uzorak.

Napominjemo da smo usput pretpostavili normalnu aproksimaciju.

Ako je n (očekivani) broj zahteva za dati obim uzorka, i ako je n_0 standard za potpuni kredibilitet, onda je parcijalni kredibilitet koji se pripisuje:

$$Z = \sqrt{\frac{n}{n_0}},$$

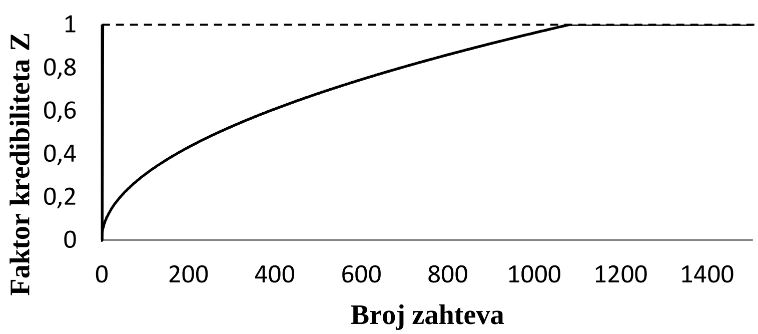
Ako je $n \ge n_0$, onda je $Z = 1,$ što se može izraziti formulom:

$$Z = min\left\{\sqrt{\frac{n}{n_0}}, 1\right\}. \tag{2.9}$$

Iz očiglednih razloga, formula se zove *pravilo korena za parcijalni kredibilitet*, bez obzira na mernu jedinicu vrednosti koju ocenjujemo. Takođe, primetimo da je Z uvek kvadratni koren racia stvarnog broja podataka i broja neophodnog za potpuni kredibilitet.

Kako je često očekivani broj zahteva nepoznat, kao aproksimacija se koristi posmatrani broj zahteva. Ukoliko, međutim, znamo kolika je izloženost riziku, kao i očekivanu učestalost zahteva, onda možemo izračunati očekivani broj zahteva tako što ćemo pomnožiti izloženost sa učestalosti.

Na primer, ako je 1,082.41 zahteva potrebno za potpuni kredibilitet, onda Slika 2.1 prikazuje koje vrednosti faktora kredibiliteta Z pripisujemo različitom broju raspoloživih zahteva za odštetu.



Slika 2.1 Parcijalni kredibilitet

■Primer 2.4

Prisetimo se Primera 2.2 i pretpostavimo da manualna premija M iznosi 225. Odredi kredibilnosnu ocenu.

Rešenje:

Prosečna isplata štete je 184.6. Koristeći pravilo korena, dobijamo faktor kredibiliteta

$$Z = \sqrt{\frac{10}{2{,}279.51}} = 0.06623.$$

Sada možemo izračunati kredibilnosnu premiju

$$P_c = 0.06623(184.6) + 0.93377(225) = 222.32. \qquad \square$$

■Primer 2.5

Prisetimo se Primera 2.3 i pretpostavimo da manualna premija M iznosi 225. Odredi kredibilnosnu ocenu koristeći oba slučaja.

Rešenje:

U prvom slučaju, koristeći standard za ocenjivanje učestalosti nastanka štete, dobijamo faktor kredibiliteta

$$Z = \sqrt{\frac{5}{1{,}082.41}} = 0.06797$$

koji primenjujemo u kredibilnosnoj premiji

$$P_c = 0.06797(184.6) + 0.93203(225) = 222.25.$$

Ovaj standard je uspostavljen u odnosu na očekivanu učestalost, te se može činiti neprikladnim da ga koristimo za agregatne zahteve. Činjenica je da češće razlikujemo osiguranike na osnovu razlika u učestalosti sa kojom podnose zahteve za odštetu nego na osnovu razlika u isnosima štete po zahtevu, dakle, ovaj faktor obuhvata najvažniju karakteristiku.

U drugom slučaju, koristeći standard za ocenjivanje iznosa štete, možemo koristiti bilo koju od sledeće tri kalkulacije:

$$Z = \sqrt{\frac{10}{2{,}734.02}} = \sqrt{\frac{5}{1{,}367.01}} = \sqrt{\frac{1{,}846}{504{,}701}} = 0.06048.$$

Dakle,

$$P_c = 0.06048(184.6) + 0.93952(225) = 222.56. \qquad \square$$

Koristimo $\bar{X}$ i pretpostavljamo da varijansa može obuhvatiti varijaciju od $\bar{X}$ u pravom smislu. Problem je u cilju. Iako je $\bar{X}$ centrirana ocena za ξ, P_c to nije. Jedna od glavnih karakteristika, koja omogućava teoriji kredibiliteta da funkcioniše, jeste upotreba pristrasnih ocenjivača. Međutim, za pristrasne ocenjivače, prihvatljiva mera kvaliteta nije varijansa već srednjekvadratna greška, koja zahteva znanje o pristrasnosti i odnosu između ξ i M. Međutim, mi ne znamo ništa o tom odnosu, a prikupljeni podaci su od male koristi. Ovo nije samo problem našeg odabira za Z, već problem koji generalno karakteriše pristup ograničene fluktuacije, o čemu ćemo više pričati u odeljku 2.3.3.

■Primer 2.6

U grupnom zdravstvenom osiguranju analiza podataka, u nekoliko godina unazad, je pokazala da godišnji iznosi šteta po čoveku imaju srednju vrednost 175 i standardnu devijaciju 140. Jedna određena grupa osiguranika je osigurana na dve godine. Tokom prve godine osigurano je 100 ljudi, a tokom druge godine 110 ljudi. Tokom te dve godine prosečan iznos *nastalih šteta po čoveku iznosio je 150. Odredi da li se može primeniti potpuni ili parcijalni kredibilitet i izračunaj premiju za ovu grupu osiguranika za sledeću godinu, ako će naredne godine u grupi biti 125 ljudi. U kalkulaciji koristi λ_0 = 1,082.4.*

Rešenje:

Odredićemo kredibilitet po osiguranoj osobi. Tokom prethodne dve godine imali smo 100 + 110 = 210 jedinica za izloženost riziku i $\bar{X} = 150$. Takođe, pretpostavimo da su štete međosubno nezavisne za različite ljude i godine u kojima se javljaju. Sada je M = 175, i pretpostavimo da će σ biti 140 za ovu grupu osiguranika. Iako razmatramo situaciju sa agregatnim štetama (svaka osoba ima sučajan broj zahteva sa slučajnim iznosima šteta), nemamo informacija ni o broju zahteva(učestalosti) ni o njihovim iznosima. Ovde primenjujemo (2.6), gde ξ ocenjujemo sa uzoračkom sredinom 150 i dobijamo standard za potpuni kredibilitet

$$n \geq 1{,}082.41\left(\frac{140}{150}\right)^2 = 942.90$$

i računamo

$$Z = \sqrt{\frac{210}{942.90}} = 0.472$$

Prisetimo se da je $\bar{X}$ srednja vrednost za 210 zahteva, pa se normalna aproksimacija može pretpostaviti zarad primene centralne granične teoreme. Neto premija po jednom osiguraniku iznosi

$$P_c = 0.472(150) + 0.528(175) = 163.2.$$

Neto premija za celu grupu osiguranika iznosi 125(163.2) = 20,400. □

■Primer 2.7

Posmatramo 715 zahteva za određenu grupu osiguranika i pretpostavimo da koristimo kredibilitet u odnosu na broj zahteva za odštetu. Ako broj zahteva prati Poasonovu raspodelu i $\lambda_0 = 1{,}082.41$, *odredi odgovarajući faktor kredibiliteta.*

Rešenje:

Prisetimo se prvog dela Primera 2.3. Standard za potpun kredibilitet u odnosu na broj zahteva sa Poasonovom raspodelom iznosi $n\lambda \geq \lambda_0 = 1{,}082.41$. Dakle

$$Z = \sqrt{\frac{715}{1{,}082.41}} = 0.813.$$ □

■Primer 2.8

Za određenu grupu osiguranika, istorijski podaci su $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$, *gde su* X_j *međusobno nezavisni i jednako raspodeljeni sa složenom Poasonovom raspodelom, gde su iznosi* *šteta eksponencijalno raspodeljeni. Odredi odgovarajući faktor kredibiliteta u odnosu na ukupne štete, ukoliko se zna da je faktor kredibiliteta u odnosu na broj zahteva 0.8.*

Rešenje:

Učestalost zahteva ima Poasonovu raspodelu, a faktor kredibiliteta u odnosu na broj zahteva iznosi $Z = 0.8$, što implicira da $n\lambda/\lambda_0 = (0.8)^2 = 0.64$, gde je $n\lambda$ posmatrani broj zahteva (iz prvog dela Primera 2.3). Iznosi šteta imaju eksponencijalnu raspodelu, pa je $\sigma_Y{}^2 = \theta_Y{}^2$. Iz drugog dela Primera 2.3, standard za potpuni kredibilitet u odnosu na broj zahteva iznosi

$$n\lambda \geq \lambda_0 \left[1 + \left(\frac{\sigma_Y}{\theta_Y}\right)^2\right] = 2\lambda_0.$$

Dakle,

$$Z = \sqrt{\frac{n\lambda}{2\lambda_0}} = \sqrt{0.32} = 0.566. \qquad \square$$

Individualne premije su deo obrasca koji se ne može ignorisati prilikom revizije premija, i baš zbog tog razloga se sistem parcijalnih kredibiliteta pokazao kao zadovoljavajući u praksi.[48]

2.3.3. Teorijske poteškoće pristupa

U aktuarskoj literaturi možemo naći mnoge argumente za odabir i određivanje vrednosti faktor Z. Svi oni imaju mane, uključujući i pristup koji je predstavljen u ovom poglavlju. Neke procedure teorije kredibiliteta su nastale davno, u vremenu kada su proračuni predstavljali veću poteškoću, pa su zbog toga bile previše pojednostavljene. Metod ograničenih fluktuacija nudi jednostavno rešenje problema kredibiliteta podataka, ali zato ima i teorijskih poteškoća:

- Metod ignoriše sve druge uzorke i samim tim i vrednost M.[49] Nejasno je zašto je premija (2.7) adekvatna i bolja od M, kada ne postoji model koji teorijski opisuje raspodelu podataka X_j. U ovom slučaju bi zgodno bilo razmotriti ocenjivanje ξ na osnovu kolekcije homogenih osiguranika i svim osiguranicima naplatiti jednaku premiju. Očigledno je da postoji praktičan razlog za korišćenje premije iz (2.7), ali bez modela koji opravdava ovu formulu, izbor Z (a time i izbor P_c) se čini potpuno proizvoljnim.
- Pokušavamo da utvrdimo „koliko je $\bar{X}$ pouzdano“, umesto da se pitamo „koliko je $\bar{X}$ pouzdanije od M“.[50] Sa (2.7) tvrdimo da je vrednost M tačna reprezentacija očekivane vrednosti, bez raspoloživih informacija o datom osiguraniku. U većini slučajeva M je takođe ocena, dakle i sama je nepouzdana, a odnos ξ i M nije ispitan.
- Čak i da je (2.7) opravdano određenim modelom, ne postoji nikakav vodič za odabir vrednosti r i p. Dodatno, kako je vrednost r i p proizvoljna, teško je opravdati rezultate pred regulatorima i sudijama.

[48] Vidi [5]
[49] Vidi [9]
[50] Vidi [1]

- Ako je broj opservacija, broj podnetih zahteva za odštetu, jednak nuli ili je jako mali, centralna granična teorema nije primenljiva.[51] Pretpostavljamo normalnu aproksimaciju, a ona ne mora biti dobra pretpostavka u praksi, pogotovo u ovoj situaciji, kada raspolažemo ograničenim skupom podataka. Tada bi logičniji i praktičniji pristup bio upotreba Poasonove distribucije.

U narednom odeljku ćemo predstaviti sistematski modelovan pristup za istorijsko ponašanje šteta određenog osiguranika, koji sugeriše da su istorijski podaci relevantni za određivanje buduće premije.

2.4. Teorija kredibiliteta visoke preciznosti

> "Određivanje premija na statističkim osnovama je uvek kompromis između dva suprotna razmatranja o blagovremenoj reakciji na indikacije o skorom prošlom iskustvu i stabilnosti koja je dovoljna da se izbegnu česte i neprikladne smetnje na terenu."[52]

U ovom odeljku predstavićemo drugi oblik teorije kredibiliteta. Ova teorija, zasnovana na matematičkom modelu rešenja problema kredibiliteta, zove se *teorija kredibiliteta visoke preciznosti*, a predstavljena je u klasičnom Bulmanovom[53] spisu iz 1967. Faktor kredibiliteta se određuje formulom

$$Z = \frac{n}{n+k},$$

gde, kada broj opservacija n raste, faktor kredibiliteta Z teži ka 1 kao asimptotskoj vrednosti i nikad je ne dostiže. Da bismo mogli primeniti Bulmanov kredibilitet na razne realne situacije, moramo izračunati ili oceniti takozvani Bulmanov parametar kredibiliteta k, koji zahteva analizu varijanse: proračun očekivane vrednosti procesne varijanse i varijanse hipotetičkih sredina. Vratimo se, na momenat, na osnovni problem za pojedinačnog osiguranika:

- imamo n opservacija o istorijskim štetama $\mathbf{X} = (X_1, \cdots, X_n)^T$
- imamo manualnu premiju μ primenljivu za datog osiguranika
- iz istorijskih podataka se može zaključiti da je μ neprikladna, odnosno da su $E(X)$ i $\bar{X} = n^{-1}(X_1 + \cdots + X_n)$ prilično različite od μ
- Koliko treba da iznosi neto premija po jedinici za izloženost riziku za narednu godinu? Da li ocenjena vrednost treba da se bazira na μ, $\bar{X}$ ili na kombinaciji obe vrednosti?

[51] Vidi [10]
[52] T. O. Carlson and L. H. Longley-Cook, Multiple Line Insurance, Michaelbacher p.98, McGraw-Hill, 1957
[53] Bühlmann, H., "Experience Rating and Credibility," *ASTIN Bulletin*, **4**, 199–207, 1967

Osiguravač mora da odgovori na sledeće pitanje: Da li je razlika između μ i $\overline{\mathrm{X}}$ uzrokovana slučajnom varijacijom ili je dati osiguranik zaista drugačiji od onoga što je pretpostavljeno proračunom μ?

Jasno je da ni jedan sistem formiranja polisa osiguranja nije savršen. Bez obzira koliko je procedura detaljna, neće baš svi rizici u klasi biti homogeni. Dakle, uvek će postojati određeni nivo heterogenosti među karakteristikama rizika jedne riziko klase i zaista je moguće da dati osiguranik bude različit od onoga što pretpostavljamo. Dolazimo do pitanja, kako odrediti odgovarajuću premiju za osiguranika?

Sada nastavljamo sa problemom i pretpostavljamo:

- nivo rizika svakog osiguranika date riziko klase karakteriše parametar rizika θ (moguće je da bude vektorska vrednost),
- vrednost θ varira za svakog osiguranika i može se posmatrati kao predstavnik reziduala, neobuhvaćenih faktora koji utiču na nivo rizika,
- θ postoji, ali ne znamo njegovu pravu vrednost.

Kako θ varira za svakog osiguranika, postoji raspodela verovatnoća zadata sa $\pi(\theta)$ za sve vrednosti iz riziko klase. $\Pi(\theta) = \Pr(\Theta \leq \theta)$ predstavlja verovatnoću da slučajno izabrani osiguranik iz riziko klase ima parametar rizika jednak ili manji od θ.

Iako je nivo rizika vezan za pojedinačnog osiguranika nepoznat, možemo pretpostaviti da je struktura karakteristika rizika u populaciji poznata, t.j. $\pi(\theta)$ je poznato.

Dodatno, jasno je da iskustvo slučajnog osiguranika potiče iz složenog procesa i da sistematski varira sa parametrom θ. Parametar rizika θ potiče iz raspodele $\pi(\theta)$, a zatim iznosi štete X potiču iz uslovne raspodele za X za dato θ, $f_{X|\Theta}(x|\theta)$. Stoga raspodela zahteva za odštetu varira od osiguranika do osiguranika i oslikava razliku u parametrima rizika.

■Primer 2.9

U posmatranoj populaciji 75% vozača spada u dobre vozače, a preostalih 25% u loše vozače. Tokom jedne godine, dobar vozač ima 0 zahteva sa verovatnoćom 0.7, 1 zahtev sa verovatnoćom 0.2, i 2 zahteva sa verovatnoćom 0.1. Loš vozač ima 0, 1 ili 2 zahteva za odštetu sa verovatnoćama 0.5, 0.3, i 0.2, respektivno. Opiši ovaj proces i vezu sa nepoznatim parametrom rizika.

Rešenje:

Kada vozač kupuje polisu osiguranja, mi ne znamo da li on spada u dobre ili loše vozače. Stoga parametar rizika Θ može uzeti dve vrednosti, $\Theta = \mathrm{D}$ za dobrog vozača i $\Theta = \mathrm{L}$ za lošeg vozača. U Tabeli 2.2 je formulisan zakon verovatnoća za broj zahteva, X, i parametar rizika Θ.□

x	$\Pr(X = x \mid \Theta = \mathrm{D})$	$\Pr(X = x \mid \Theta = \mathrm{L})$	θ	$\Pr(\Theta = \theta)$
0	0.7	0.5	D	0.75
1	0.2	0.3	L	0.25
2	0.1	0.2		

Tabela 2.2 Verovatnoće za Primer 2.9.

■**Primer 2.10**

Iznos zahteva za odštetu ima eksponencijalnu raspodelu sa sredinom $1/\Theta$. *Parametar* Θ *varira u klasi osiguranika i potencijalnih osiguranika sa Gama distribucijom sa parametrima* $\alpha = 4$ *i* $\beta = 0.001$. *Matematički opiši ovaj model.*

Rešenje:

Broj zahteva ima uslovnu raspodelu

$$f_{X|\Theta}(x|\theta) = \theta e^{-\theta x}, x, \theta > 0,$$

a parametar rizika ima raspodelu

$$\pi_\Theta(\theta) = \frac{\theta^3 e^{-1.000\theta} 1.000^4}{6}, \theta > 0. \qquad \square$$

2.4.1. Bejzova metodologija

Prisetimo se pretpostavke da se distribucija karakteristika rizika u populaciji može predstaviti sa raspodelom $\pi(\theta)$, i da štete određenog osiguranika sa parametrom rizika θ imaju uslovnu raspodelu $f_{X|\Theta}(x|\theta)$ za dato θ.

Za određenog osiguranika sa nepoznatim parametrom rizika θ imamo opservacije $\boldsymbol{X} = \boldsymbol{x}$, gde $\boldsymbol{X} = (X_1, \cdots, X_n)^T$ i $\boldsymbol{x} = (x_1, \cdots, x_n)^T$. Želimo da odredimo premiju za pokriće štete X_{n+1}. Pretpostavimo da su istorijski podaci osiguranika $X_1, \cdots, X_n, X_{n+1}$, koji odgovaraju različitim periodima izloženosti riziku, međusobno nezavisni (pri uslovu $\Theta = \theta$), ali ne i nužno jednako raspodeljeni sa uslovnim raspodelama

$$f_{X_j|\Theta}(x_j|\theta), \qquad j = 1, \dots, n, n+1.$$

Primetimo da ukoliko bi X_j bili jednako raspodeljeni, onda $f_{X_j|\Theta}(x_j|\theta)$ ne bi zavisilo od j. Da je parametar rizika θ poznat, mogli bismo koristiti $f_{X_{n+1}|\Theta}(x_{n+1}|\theta)$, ali kako znamo $\mathbf{x}$ za istog osiguranika, možemo odrediti uslovnu raspodelu za X_{n+1} pri uslovu $\boldsymbol{X} = \boldsymbol{x}$, koja se naziva *prediktivna raspodela*.

Prediktivna raspodela za X_{n+1} pri uslovu $\boldsymbol{X} = \boldsymbol{x}$ predstavlja relevantnu distribuciju za analizu rizika, menadžment, donošenje odluka, zato što kombinuje neizvesnost u vezi sa brojem nastalih šteta sa parametrima povezanim sa procesom rizika.

Kako smo pretpostavili da su X_j međusobno nezavisni pri uslovu $\Theta = \theta$, imamo[54]

$$f_{X,\Theta}(\boldsymbol{x},\theta) = f(x_1,\cdots,x_n|\theta)\pi(\theta) = \left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\right]\pi(\theta).$$

Da bismo dobili zajedničku raspodelu za $\boldsymbol{X}$, a onda i marginalnu raspodelu, integralimo po θ

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int \left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\right]\pi(\theta)\, d\theta. \tag{2.10}$$

Analogno, zajednička raspodela za $X_1,\cdots,X_{n+1}$ se dobija kada u izrazu sa desne strane jednakosti u (2.10) u proizvodu umesto *n* stavimo *n*+1. Uslovna gustina za X_{n+1} pri uslovu $\boldsymbol{X} = \boldsymbol{x}$ je jednaka zajedničkoj gustini za $(X_1,\cdots,X_{n+1})$ podeljeno sa zajedničkom gustinom za $\boldsymbol{X}$,

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\boldsymbol{x}) = \frac{1}{f_{\boldsymbol{X}}(\boldsymbol{x})}\int \left[\prod_{j=1}^{n+1} f_{X_j|\Theta}(x_j|\theta)\right]\pi(\theta)\, d\theta. \tag{2.11}$$

Koristeći matematičku strukturu u (2.11) možemo dobiti *posteriornu raspodelu* za Θ pri uslovu $\mathbf{X}$

$$\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{f_{\mathbf{X},\Theta}(\mathbf{x},\theta)}{f_{\mathbf{X}}(\mathbf{x})} = \frac{1}{f_{\mathbf{X}}(\mathbf{x})}\left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\right]\pi(\theta). \tag{2.12}$$

Ovo drugačije možemo napisati

$$\left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\right]\pi(\theta) = \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})$$

i ako uvrstimo brojilac u (2.11) dobijamo

$$f_{\mathrm{X}_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \int f_{X_{n+1}|\Theta}(x_{n+1}|\theta)\,\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta. \tag{2.13}$$

Iz (2.13) zaključujemo da se uslovna raspodela za X_{n+1} pri uslovu $\mathbf{X}$ može posmatrati kao mešovita raspodela, gde je mešovita raspodela posteriorna raspodela $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$, koja kombinuje i sumira informacije o θ sadržane u priornoj raspodeli i verodostojnost.

[54] Napomena: Ukoliko Θ ima diskretnu raspodelu, integrali se zamenjuju sumama.

■Primer 2.11

Prisetimo se Primera 2.9. Pretpostavimo da, za određenog osiguranika, imamo dve opservacije $x_1 = 0$ *i* $x_2 = 1$. *Odredi prediktivnu raspodelu za* $X_3|X_1 = 0, X_2 = 1$ *i posteriornu raspodelu za* $\Theta|X_1 = 0, X_2 = 1$.

Rešenje:

Koristeći (2.10) dobijamo marginalnu raspodelu

$$\begin{aligned} f_{\mathbf{X}}(0,1) &= \sum_{\theta} f_{X_1|\Theta}(0|\theta) f_{X_2|\Theta}(1|\theta)\,\pi(\theta) \\ &= 0.7(0.2)(0.75) + 0.5(0.3)(0.25) \\ &= 0.1425. \end{aligned}$$

Analogno, zajednička raspodela za sve tri varijable je

$$f_{\mathbf{X},X_3}(0,1,x_3) = \sum_{\theta} f_{X_1|\Theta}(0|\theta) f_{X_2|\Theta}(1|\theta)\, f_{X_3|\Theta}(x_3|\theta)\pi(\theta).$$

Stoga,

$$\begin{aligned} f_{\mathbf{X},X_3}(0,1,0) &= 0.7(0.2)(0.7)(0.75) + 0.5(0.3)(0.5)(0.25) = 0.09225, \\ f_{\mathbf{X},X_3}(0,1,1) &= 0.7(0.2)(0.2)(0.75) + 0.5(0.3)(0.3)(0.25) = 0.03225, \\ f_{\mathbf{X},X_3}(0,1,2) &= 0.7(0.2)(0.1)(0.75) + 0.5(0.3)(0.2)(0.25) = 0.01800. \end{aligned}$$

Prediktivna raspodela je onda

$$\begin{aligned} f_{X_3|\mathbf{X}}(0|0,1) &= \frac{0.09225}{0.1425} = 0.647368, \\ f_{X_3|\mathbf{X}}(1|0,1) &= \frac{0.03225}{0.1425} = 0.226316, \\ f_{X_3|\mathbf{X}}(2|0,1) &= \frac{0.01800}{0.1425} = 0.126316. \end{aligned}$$

Sada iz (2.12) dobijamo posteriornu raspodelu

$$\begin{aligned} \pi(D|0,1) &= \frac{f(0|D)f(1|D)\pi(D)}{f(0,1)} = \frac{0.7(0.2)(0.75)}{0.1425} = 0.736842, \\ \pi(L|0,1) &= \frac{f(0|L)f(1|L)\pi(L)}{f(0,1)} = \frac{0.5(0.3)(0.25)}{0.1425} = 0.263158. \end{aligned}$$

Napomena: od sada pa nadalje ćemo izostavljati indekse za f i π, osim u slučaju kada su nužni za razumevanje izraza. Takođe smo mogli odrediti prediktivnu raspodelu koristeći (2.13). Ovaj metod je obično jednostavniji za računanje:

$$\begin{aligned} f(0|0,1) &= \sum_{\theta} f(0|\theta)\,\pi(\theta|0,1) \\ &= 0.7(0.736842) + 0.5(0.263158) = 0.647368, \\ f(1|0,1) &= 0.2(0.736842) + 0.3(0.263158) = 0.226316, \\ f(2|0,1) &= 0.1(0.736842) + 0.2(0.263158) = 0.126316, \end{aligned}$$

što se poklapa sa prethodnom kalkulacijom. □

■Primer 2.12

Prisetimo se Primera 2.10 Sada pretpostavimo da je osiguranik imao štete u iznosu 100, 950, i 450 novčanih jedinica. Odredi prediktivnu raspodelu za četvrtu štetu i posteriornu raspodelu za Θ.

Rešenje:

Marginalna gustina za posmatrane opservacije je

$$\begin{aligned} f(100,950,450) &= \int_0^{\infty} \theta e^{-100\theta}\theta e^{-950\theta}\theta e^{-450\theta}\,\frac{1{,}000^4}{6}\,\theta^3 e^{-1{,}000\theta}\,d\theta \\ &= \frac{1{,}000^4}{6}\int_0^{\infty}\theta^6 e^{-2{,}500\theta}\,d\theta = \frac{1{,}000^4}{6}\,\frac{720}{2{,}500^7}. \end{aligned}$$

Slično,

$$\begin{aligned} f(100,950,450,x_4) &= \int_0^{\infty} \theta e^{-100\theta}\theta e^{-950\theta}\theta e^{-450\theta}\theta e^{-\theta x_4}\,\frac{1{,}000^4}{6}\,\theta^3 e^{-1{,}000\theta}\,d\theta \\ &= \frac{1{,}000^4}{6}\int_0^{\infty}\theta^7 e^{-(2{,}500+x_4)\theta}\,d\theta = \frac{1{,}000^4}{6}\,\frac{5{,}040}{(2{,}500+x_4)^8}. \end{aligned}$$

Prediktivna gustina je sada

$$f(x_4|100,950,450) = \frac{\frac{1{,}000^4}{6}\,\frac{5{,}040}{(2{,}500+x_4)^8}}{\frac{1{,}000^4}{6}\,\frac{720}{2{,}500^7}} = \frac{7(2{,}500)^7}{(2{,}500+x_4)^8},$$

što ima oblik Pareto gustine raspodele sa parametrima 7 i 2,500.

Razmotrimo sada posteriornu distribuciju. U imeniocu se nalazi integral koji generiše broj koji ćemo zanemariti u ovom momentu. Brojilac zadovoljava

$$\pi(\theta|100{,}950{,}450)\left(=\frac{f(100{,}950{,}450,\theta)}{f(100{,}950{,}450)}\right)\propto \theta e^{-100\theta}\theta e^{-950\theta}\theta e^{-450\theta}\frac{1{,}000^4}{6}\theta^3 e^{-1{,}000\theta},$$

što je izraz koji smo integralili pri određivanju marginalne gustine. Takođe ćemo zanemariti konstante i u brojiocu, pa je

$$\pi(\theta|100{,}950{,}450)\propto \theta^6 e^{-2{,}500\theta}.$$

Da bi ovaj izraz predstavljao gustinu raspodele, neophodno je integraliti izraz i odrediti konstantu. Međutim, umesto da integralimo, prepoznaćemo da je ovo raspodela za Gama distribuciju sa parametrima 7 i 1/2,500. Dakle,

$$\pi(\theta|100{,}950{,}450)=\frac{\theta^6 e^{-2{,}500\theta}2{,}500^7}{\Gamma(7)}.$$

Sada ćemo na alternativan način izračunati prediktivnu gustinu i videti da se rezultat poklapa sa prethodno dobijenim rezultatom

$$\begin{aligned} f(x_4|100{,}950{,}450) &= \int_0^\infty \theta e^{-\theta x_4}\frac{\theta^6 e^{-2{,}500\theta}2{,}500^7}{\Gamma(7)}d\theta \\ &= \frac{2{,}500^7}{6!}\int_0^\infty \theta^7 e^{-(2{,}500+x_4)\theta}d\theta \\ &= \frac{2{,}500^7}{6!}\frac{7!}{(2{,}500+x_4)^8}. \end{aligned}$$

Vidimo da je posteriorna raspodela istog tipa (Gama) kao i priorna raspodela. □

Vratimo se sada na originalni problem, gde imamo opservacije **X** = **x** za određenog osiguranika i želimo da predvidimo vrednost X_{n+1}. Kada bismo znali θ, logičan izbor bi bila *hipotetička sredina (ili individualna premija)*

$$\mu_{n+1}(\theta)=\mathrm{E}(X_{n+1}\,|\Theta=\theta)=\int x_{n+1}f_{X_{n+1}|\Theta}\,(x_{n+1}|\theta)\mathrm{d}\,x_{n+1}\,. \qquad (2.14)$$

Ako zamenimo θ za Θ u (2.14) i uzmemo očekivanje, dobijamo

$$\mu_{n+1}=\mathrm{E}(X_{n+1})=\mathrm{E}[\mathrm{E}(X_{n+1}|\Theta)]=\mathrm{E}[\mu_{n+1}(\Theta)],$$

neto, ili *kolektivnu*, *premiju* kao bezuslovno očekivanje hipotetičkih sredina.

Međutim, kako je parametar θ nepoznat, naš najbolji pokušaj jeste da iskoristimo prikupljene individualne podatke (istorijske podatke), **x**, i Bejzovu premiju (sredinu prediktivne raspodele)

$$\mathrm{E}(X_{n+1} \,|\mathbf{X} = \mathbf{x}) = \int x_{n+1} f_{X_{n+1}|\mathbf{X}}\,(x_{n+1}|\mathbf{x})\mathrm{d}\,x_{n+1}. \tag{2.15}$$

Koristeći (2.13) i (2.14), izraz možemo drugačije napisati

$$\begin{aligned}\mathrm{E}(X_{n+1} \,|\mathbf{X} = \mathbf{x}) &= \int x_{n+1} f_{X_{n+1}|\mathbf{X}}\,(x_{n+1}|\mathbf{x})\mathrm{d}\,x_{n+1} \\ &= \int x_{n+1} \left[\int f_{X_{n+1}|\Theta}(x_{n+1}|\theta)\,\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta\right] d\,x_{n+1} \\ &= \int \left[\int x_{n+1} f_{X_{n+1}|\Theta}(x_{n+1}|\theta)\,dx_{n+1}\right] \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta \\ &= \int \mu_{n+1}(\theta)\,\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta. \end{aligned} \tag{2.16}$$

Ovakva formulacija implicira da je Bejzova premija očekivana vrednost hipotetičkih sredina za posteriornu raspodelu $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$.[55]

■Primer 2.13

Prisetimo se Primera 2.11. Odredi Bejzovu premiju koristeći (2.15) i (2.16).

Rešenje:

Hipotetičke sredine su

$$\mu_3(D) = (0)(0.7) + 1(0.2) + 2(0.1) = 0.4,$$
$$\mu_3(L) = (0)(0.5) + 1(0.3) + 2(0.2) = 0.7.$$

Na osnovu opservacija $x_1 = 0$ i $x_2 = 1$, možemo izračunati Bejzovu premiju direktno iz (2.15)

$$\mathrm{E}(X_3 \,|0{,}1\,) = 0(0.647368) + 1(0.226316) + 2(0.126316) = 0.478948,$$

i bezuslovnu neto premiju

$$\mu_3 = \mathrm{E}(X_3) = \sum_{\theta} \mu_3(\theta)\pi(\theta) = (0.4)(0.75) + (0.7)(0.25) = 0.475.$$

Da bismo potvrdili (2.16) sa opservacijama $x_1 = 0$ i $x_2 = 1$, koristimo posteriornu raspodelu $\pi(\theta|0{,}1)$ već izračunatu u Primeru 2.11 i dobijamo

$$\mathrm{E}(X_3 \,|0{,}1\,) = 0.4(0.736842) + 0.7(0.263158) = 0.478947,$$

sa razlikom nastalom usled zaokruživanja. □

[55] Još jednom napominjemo da se kod diskretnih slučajnih promenjivih, integrali zamenjuju sumama.

Generalno, drugi pristup koji smo koristili (2.16) je zgodniji za računanje od direktnog pristupa sa uslovnom raspodelom za $X_{n+1}|\mathbf{X} = \mathbf{x}$.

I kao što smo očekivali, ocena vrednosti bazirana na opservacijama je između priorne vrednosti ($\mu_3 = 0.475$) određene bez obzira na istorijske podatke, i vrednosti bazirane samo na istorijskim podacima ($\bar{X} = 0.5$).

■Primer 2.14

Prisetimo se Primera 2.12. Odredi Bejzovu premiju.

Rešenje:

Iz Primera 2.12 imamo $\mu_4(\theta) = \theta^{-1}$. Koristeći (2.16) dobijamo

$$\begin{aligned} \mathrm{E}(X_4|100{,}950{,}450) &= \int_0^\infty \theta^{-1} \frac{\theta^6 e^{-2{,}500\theta} 2{,}500^7}{720} d\theta \\ &= \frac{2{,}500^7}{720} \frac{120}{2{,}500^6} = 416.67. \end{aligned}$$

Iz priorne raspodele, koristeći formulu za momente Gama distribucije dobijamo

$$\mu = \mathrm{E}(\Theta^{-1}) = \frac{1{,}000}{3} = 333.33$$

i opet je Bejzova ocena između priorne ocene i one bazirane samo na istorijskim podacima (uzoračka sredina od 500). Iz (2.15)

$$\mathrm{E}(X_4|100{,}950{,}450) = \frac{2{,}500}{6} = 416.67,$$

dobijamo sredinu prediktivne Pareto distribucije. □

■Primer 2.15

Za proizvoljni uzorak obima n i proizvoljnu Gama distribuciju sa parametrima α i β, gde je β recipročna vrednost uobičajenog parametra, generališi rezultat iz Primera 2.12.

Rešenje:

Posteriorna raspodela se može odrediti iz

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \left(\prod_{j=1}^{n} \theta e^{-\theta x_j}\right) \frac{\theta^{\alpha-1} e^{-\beta\theta} \beta^\alpha}{\Gamma(\alpha)} \\ &\propto \theta^{n+\alpha-1} e^{-(\sum x_j + \beta)\theta}, \end{aligned}$$

Gde su izostavljeni svi činioci koji ne sadrže θ. Umesto da integralimo da bismo dobili vrednost konstante, prepoznajemo da je ova posteriorna distribucija zapravo Gama distribucija sa parametrima $n+\alpha$ i $\left(\sum x_j+\beta\right)^{-1}$. Bejzova ocena za X_{n+1} je očekivana vrednost od Θ^{-1} koristeći posteriornu raspodelu:

$$\frac{\sum x_j+\beta}{n+\alpha-1}=\frac{n}{n+\alpha-1}\bar{x}+\frac{\alpha-1}{n+\alpha-1}\frac{\beta}{\alpha-1}.$$

Napomenimo da je dobijena ocena težinska sredina istorijskih podataka i bezuslovne sredine, i da ima oblik kao u (2.7). □

2.4.2. Kredibilnosna premija

Najveći problem Bejzovog sistematičnog pristupa je poteškoća pri numeričkoj evaluaciji Bejzove premije $\mathrm{E}(X_{n+1}\,|\mathbf{X}=\mathbf{x})$. Primeri navedeni u prethodnoj sekciji su bili jednostavni i teško da možemo očekivati da mogu obuhvatiti esenciju karakteristika realnog scenarija u osiguranju. Realističniji model može uvesti analitičke poteškoće, s obzirom da zahteva numeričku integraciju.

Razmotrimo sada alternativu koju je Bulman[56] predložio 1967. Da bismo ocenili zahteve za oštetu za narednu godinu, potrebna nam je uslovna raspodela $f_{X_{n+1}|\Theta}\,(x_{n+1}|\theta)$ ili hipotetička sredina $\mu_{n+1}(\theta)$. Predlog je da se $\mu_{n+1}(\theta)$ aproksimira linearnom funkcijom istorijskih podataka $\mathbf{x}$, sa ocenjivačem u obliku $\alpha_0+\sum_{j=1}^{n}\alpha_j X_j$, gde se vrednosti $\alpha_0,\alpha_1,\dots,\alpha_n$ moraju izabrati. Mi biramo one vrednosti α_j koje minimizuju srednjekvadratnu štetu, odnosno,

$$Q=\mathrm{E}\left\{\left[\mu_{n+1}(\Theta)-\alpha_0-\sum_{j=1}^{n}\alpha_j X_j\right]^2\right\}, \tag{2.17}$$

a kvadratna greška je uprosečena preko zajedničke distribucije svih mogućih opservacija $X_1,\dots,X_n$ i svih mogućih vrednosti Θ. Q minimizujemo diferenciranjem:

$$\frac{\partial Q}{\partial \alpha_0}=\mathrm{E}\left\{2\left[\mu_{n+1}(\Theta)-\alpha_0-\sum_{j=1}^{n}\alpha_j X_j\right](-1)\right\}.$$

Sa $\tilde{\alpha}_0,\tilde{\alpha}_1,\dots,\tilde{\alpha}_n$ ćemo označiti vrednosti $\alpha_0,\alpha_1,\dots,\alpha_n$ koje minimizuju (2.17). Sada izjednačavamo $\partial Q/\partial\alpha_0=0$ da bismo dobili:

[56] Bühlmann, H., “Experience Rating and Credibility,” *ASTIN Bulletin*, **4**, 199–207., 1967

$$\mathrm{E}[\mu_{n+1}(\Theta)] = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j \mathrm{E}(X_j).$$

Ali $\mathrm{E}(X_{n+1}) = \mathrm{E}[\mathrm{E}(X_{n+1}\,|\Theta)] = \mathrm{E}[\mu_{n+1}(\Theta)]$ i $\partial Q/\partial\alpha_0 = 0$, pa je

$$\mathrm{E}(X_{n+1}) = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j \mathrm{E}(X_j). \quad (2.18)$$

Ova jednačina se naziva **jednačina centriranosti**, zato što zahteva da kredibilnosna ocena $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j$ bude centrirana za $\mathrm{E}(X_{n+1})$. Međutim, ona može biti pristrasni ocenjivač za $\mu_{n+1}(\Theta) = \mathrm{E}(X_{n+1}\,|\Theta)$, gde prihvatajući pristrasnost možemo smanjuti ukupnu srednjekvadratnu grešku (MSE). Za svako $i = 1, \dots, n$, imamo

$$\frac{\partial Q}{\partial \alpha_i} = \mathrm{E}\left\{2\left[\mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right](-X_i)\right\}.$$

Sada iz $\partial Q/\partial\alpha_i = 0$ dobijamo:

$$\mathrm{E}[\mu_{n+1}(\Theta)X_i] = \tilde{\alpha}_0 \mathrm{E}(X_i) + \sum_{j=1}^{n} \tilde{\alpha}_j \mathrm{E}(X_i X_j).$$

Ako pretpostavimo nezavisnost X_i i X_{n+1} uslovno za dato Θ, možemo levu stranu jednačine drugačije napisati kao

$$\begin{aligned}
\mathrm{E}[\mu_{n+1}(\Theta)X_i] &= \mathrm{E}\{\mathrm{E}[(X_i\,\mu_{n+1}(\Theta)|\Theta)]\}\\
&= \mathrm{E}\{\,\mu_{n+1}(\Theta)\mathrm{E}[X_i|\Theta]\}\\
&= \mathrm{E}[\mathrm{E}(X_{n+1}\,|\Theta)\mathrm{E}(X_i|\Theta)]\\
&= \mathrm{E}[\mathrm{E}(X_{n+1}X_i\,|\Theta)]\\
&= \mathrm{E}(X_i X_{n+1}),
\end{aligned}$$

Dakle, iz $\partial Q/\partial\alpha_i = 0$ dobijamo

$$\mathrm{E}(X_i X_{n+1}) = \tilde{\alpha}_0 \mathrm{E}(X_i) + \sum_{j=1}^{n} \tilde{\alpha}_j \mathrm{E}(X_i X_j). \quad (2.19)$$

Ako sada pomnožimo (2.18) sa $\mathrm{E}(X_i)$ i dobijenu vrednost oduzmemo od (2.19) dobijamo

$$\mathrm{Cov}(X_i, X_{n+1}) = \sum_{j=1}^{n} \tilde{\alpha}_j \mathrm{Cov}(X_i, X_j),\ i = 1, \dots, n. \quad (2.20)$$

Tačno jedan sabirak sa desne strane jednačine predstavlja varijansu, $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$ a preostalih $n - 1$ sabiraka predstavljaju kovarijanse.

Jednačine (2.18) i (2.20) zajedno nazivamo **sistem normalnih jednačina**. Ove jednačine rešene za $\tilde{\alpha}_0, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_n$ daju *kredibilnosnu premiju*

$$\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j \, . \qquad (2.21)$$

Primetimo da vrednosti $\tilde{\alpha}_0, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_n$ takođe minimizuju i

$$Q_1 = \mathrm{E}\left\{\left[\mathrm{E}(X_{n+1}\,|\mathbf{X}) - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right]^2\right\} \qquad (2.22)$$

i

$$Q_2 = \mathrm{E}\left[\left(X_{n+1} - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right)^2\right], \qquad (2.23)$$

što se može pokazati diferenciranjem izraza Q_1 i Q_2 u odnosu na $\alpha_0, \alpha_1, \ldots, \alpha_n$. Rešenja i dalje zadovoljavaju sistem normalnih jednačina, pa zaključujemo da je kredibilnosna premija najbolji linearni ocenjivač za hipotetičku sredinu $\mathrm{E}(X_{n+1}\,|\Theta)$, Bejzovu premiju $\mathrm{E}(X_{n+1}\,|\mathbf{X})$ i za X_{n+1}.

■Primer 2.16

Odredi kredibilnosnu premiju $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j$ *,ako je f* $E(X_j) = \mu$, $Var(X_j) = \sigma^2$, *i, za* $i \neq j$, $Cov(X_i, X_j) = \rho\sigma^2$, *gde je koeficijent korelacije* $-1 < \rho < 1$.

Rešenje:

Iz jednačine centriranosti (2.18) dobijamo

$$\mu = \tilde{\alpha}_0 + \mu \sum_{j=1}^{n} \tilde{\alpha}_j,$$

ili drugačije zapisano

$$\sum_{j=1}^{n} \tilde{\alpha}_j = 1 - \frac{\tilde{\alpha}_0}{\mu}.$$

Jednačine (2.20) postaju, za $i = 1, \ldots, n$,

$$\rho = \sum_{\substack{j=1 \\ j \neq i}}^{n} \tilde{\alpha}_j \, \rho + \tilde{\alpha}_i,$$

odnosno

$$\rho = \sum_{j=1}^{n} \tilde{\alpha}_j \rho + \tilde{\alpha}_i (1-\rho), \text{for } i = 1, \dots, n.$$

Stoga, koristeći (2.18)

$$\tilde{\alpha}_i = \frac{\rho\left(1 - \sum_{j=1}^{n} \tilde{\alpha}_j\right)}{1-\rho} = \frac{\rho \tilde{\alpha}_0}{\mu(1-\rho)}.$$

Sumirajući ovaj izraz po *i* od 1 do *n*

$$\sum_{i=1}^{n} \tilde{\alpha}_i = \sum_{j=1}^{n} \tilde{\alpha}_j = \frac{n\rho\tilde{\alpha}_0}{\mu(1-\rho)},$$

i kombinujući ga sa jednačinom centriranosti, dobijamo jednačinu za $\tilde{\alpha}_0$

$$1 - \frac{\tilde{\alpha}_0}{\mu} = \frac{n\rho\tilde{\alpha}_0}{\mu(1-\rho)}.$$

Ako rešimo ovo za $\tilde{\alpha}_0$ dobijamo

$$\tilde{\alpha}_0 = \frac{(1-\rho)\mu}{1-\rho+n\rho}.$$

Dakle,

$$\tilde{\alpha}_j = \frac{\rho\tilde{\alpha}_0}{\mu(1-\rho)} = \frac{\rho}{1-\rho+n\rho}.$$

Kredibilnosna premija sada glasi

$$\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j = \frac{(1-\rho)\mu}{1-\rho+n\rho} + \sum_{j=1}^{n} \frac{\rho X_j}{1-\rho+n\rho} = (1-Z)\mu + Z\bar{X},$$

gde je $Z = n\rho/(1-\rho+n\rho)$ i $\bar{X} = n^{-1}\sum_{j=1}^{n} X_j$. Pa, ako $0 < \rho < 1$, onda $0 < Z < 1$, i kredibilnosna premija je oblika (2.7), kao težinska sredina od $\mu = E(X_{n+1})$ i $\bar{X}$.□

2.4.3. Bulmanov model

Najjednostavniji kredibilnosni model, Bulmanov model, pretpostavlja da je za posmatrani rizik, svako X_j jednako raspodeljeno za svaki vremenski period, odnosno, konstantno je u odnosu na vremensku odrednicu, kako je to Hans Bulman opisao, pretpostavlja se „homogenost u vremenu“. Drugim rečima, za svakog osiguranika (uslovno za Θ), istorijske štete (slučajne promenljive u prošlom posmatranom periodu $X_1, \dots, X_n$) imaju jednaku sredinu i varijansu, međusobno su nezavisne i jednako uslovno raspodeljene pri uslovu Θ:

$$\text{hipotetička sredina}^{57}\text{:}\;\; \mu(\theta) = \mathrm{E}\big(X_j \,\big|\Theta = \theta\big)$$

i

$$\text{procesna varijansa:}\;\; \upsilon(\theta) = \mathrm{Var}\big(X_j \,\big|\Theta = \theta\big).$$

Ova pretpostavka se lako može narušiti u praksi, pošto se karakteristike rizika mogu promeniti zbog raznih razloga: inflacija može povećati iznose šteta; mladi vozač sa iskustvom postaje bolji vozač; gustina saobraćaja može narasti u nekoj oblasti što dovodi do povećanja broja saobraćajnih nesreća; u poslovanju se mogu uvesti procedure za kontrolu rizika i time smanjiti gubici.

Da bismo primenili Bulmanov kredibilitet, potrebno je odrediti srednje vrednosti ovih kvantiteta za celu populaciju rizika (preko svih riziko klasa), kao i varijansu hipotetičkih sredina te populacije:

$$\text{očekivana vrednost hipotetičkih sredina:}\;\; \mu = \mathrm{E}[\mu(\Theta)], \tag{2.24}$$

$$\text{očekivana vrednost procesne varijanse:}\;\; \upsilon = \mathrm{E}[\upsilon(\Theta)], \tag{2.25}$$

i

$$\text{varijansa hipotetičkih sredina:}\;\; a = \mathrm{Var}[\mu(\Theta)]. \tag{2.26}$$

μ, se zove *kolektivna premija* (sredina populacije), i daje ocenu očekivane vrednosti za X_j u odsustvu bilo kakve a priori informacije o riziku θ (i posledično u odsustvu bilo kakve a priori informacije o $\mu(\theta)$). υ je mera za heterogenost unutar riziko klasa, dok je a mera za heterogenost između riziko klasa.

Odredimo sada sredinu, varijansu i kovarijansu za sve X_j. Prvi korak je

$$\mathrm{E}\big(X_j\big) = \mathrm{E}\big[\mathrm{E}\big(X_j \,\big|\Theta\big)\big] = \mathrm{E}[\mu(\Theta)] = \mu. \tag{2.27}$$

U sledećem koraku određujemo

$$\begin{aligned}\mathrm{Var}\big(X_j\big) &= \mathrm{E}\big[\mathrm{Var}\big(X_j \,\big|\Theta\big)\big] + \mathrm{Var}\big[\mathrm{E}\big(X_j \,\big|\Theta\big)\big] \\ &= \mathrm{E}[\upsilon(\Theta)] + \mathrm{Var}[\mu(\Theta)] \\ &= \upsilon + a\,.\end{aligned} \tag{2.28}$$

Ova formulacija implicira da je

$$\begin{gathered}\text{ukupna varijansa} = \\ \text{očekivana vrednost procesne varijanse} + \text{varijansa hipotetičkih sredina.}\end{gathered}$$

Poslednji korak, za $i \neq j$,

$$\begin{aligned}\mathrm{Cov}\big(X_i, X_j\big) &= \mathrm{E}\big(X_i X_j\big) - \mathrm{E}\big(X_i\big)\mathrm{E}\big(X_j\big) \\ &= \mathrm{E}\big[\mathrm{E}\big(X_i X_j \,\big|\Theta\big)\big] - \mu^2 \\ &= \mathrm{E}\big[\mathrm{E}(X_i \,|\Theta)\mathrm{E}\big(X_j \,\big|\Theta\big)\big] - \{\mathrm{E}[\mu(\Theta)]\}^2 \\ &= \mathrm{E}\{[\mu(\Theta)]^2\} - \{\mathrm{E}[\mu(\Theta)]\}^2 \\ &= \mathrm{Var}[\mu(\Theta)] \\ &= a\,.\end{aligned} \tag{2.29}$$

[57] Očekivana šteta unutar riziko klase Θ.

Ovaj rezultat ima oblik kao u Primeru 2.16 sa parametrima μ, $\sigma^2 = \upsilon + a$ i $\rho = a/(\upsilon + a)$, dakle kredibilnosna premija je

$$\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j = Z\bar{X} + (1 - Z)\mu, \tag{2.30}$$

gde

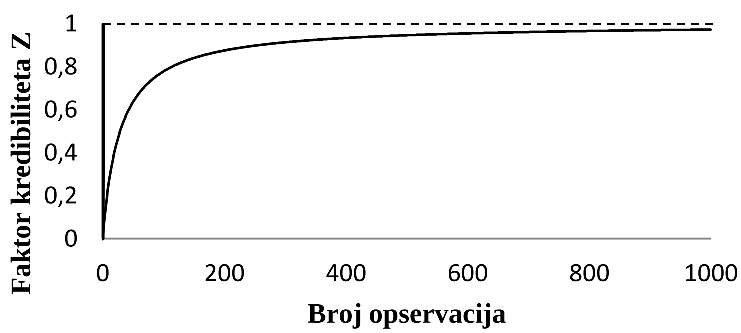
$$Z = \frac{n}{n + k} \tag{2.31}$$

i

$$k = \frac{\upsilon}{a} = \frac{\mathrm{E}\left[\mathrm{Var}\left(X_j \,|\Theta\right)\right]}{\mathrm{Var}\left[\mathrm{E}\left(X_j \,|\Theta\right)\right]}. \tag{2.32}$$

Faktor kredibiliteta Z u (2.31) sa k zadatim u (2.32) se zove *Bulmanov faktor kredibiliteta.*

Kredibilnosna premija formulisana kao u (2.30) daje težinsku sredinu uzoračke sredine $\bar{X}$ i kolektivne premije μ. Kako n raste, akumuliranjem istorijskih podataka ($n \to \infty$), Z asimptotski teži ka 1 ($Z \to 1$), dajući veći kredibilitet $\bar{X}$ nego kolektivnoj premiji μ. Slika 2.2 pokazuje Bulmanov kredibilitet (za k=18.5926 iz Primera 2.18) gde Z nikada ne uzima vrednost 1, za razliku od klasičnog kredibiliteta.



Slika 2.2 Bulmanov kredibilitet

Dalje, u populaciji homogenoj u odnosu na parametar rizika Θ, hipotetičke sredine $\mathrm{E}\left(X_j \,|\Theta\right)$ ne variraju puno sa promenom parametra Θ, što implicira da je a malo u odnosu na υ, dakle, k je veliko i Z je blizu nule. S druge strane, u heterogenoj populaciji, hipotetičke sredine $\mathrm{E}\left(X_j \,|\Theta\right)$ su više volatilne, te je a veliko, k je malo i Z je bliže jedinici. Ovakve opservacije se slažu sa našom intuicijom: za homogenu populaciju, ukupna sredina μ je značajnija za predviđanje budućih šteta za određenog osiguranika, ali je zato u heterogenoj populaciji iskustvo drugih osiguranika manje važno.

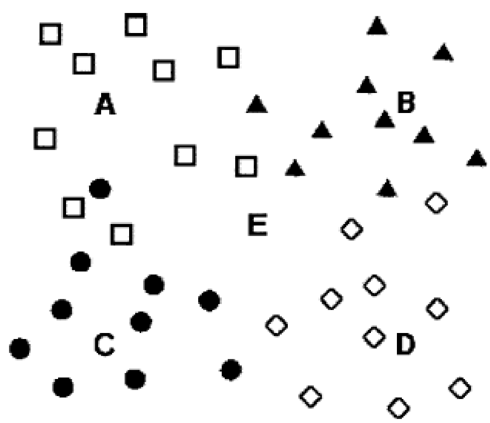
Generalno, da bismo izračunali Bulmanov kredibilitet, neophodno je prvo odrediti očekivanu vrednost procesne varijanse i varijansu hipotetičkih sredina za jednu opservaciju, pa zatim uvrstiti broj raspoloživih opservacija n u formulu. Ako se ocenjuju broj zahteva i neto premija, onda je broj n u jedinici za izloženost riziku, a ako se ocenjuje iznos šteta, onda je n izraženo u broju zahteva za odštetu (n izraženo u mernoj jedinici imenioca kvantiteta koji ocenjujemo[58]).

■**Primer 2.17** *Primer za gađanje mete*

Stefan Filbrik (Stephen Philbrick)[59] je predstavio odličan primer sa gađanjem meta, kojim je ilustrovao ideje Bulmanovog kredibiliteta.

Pretpostavimo:

- postoje četiri strelca i svaki gađa svoju metu, koje su međusobno različite.
- pogotci svakog strelca su distribuirani oko njihovih meta, označenih slovima A, B, C, i D, gde je očekivana vrednost jednaka lokaciji njihovih meta.
- svaki hitac može biti ispaljen od strane bilo kog strelca sa jednakom verovatnoćom (ispaljuju isti broj hitaca), pa svakoj meti dajemo istu težinu.



Slika 2.3

Ako su mete raspoređene kao na Slici 2.3, hitci svakog strelca će se grupisati oko njihovih meta. Hitci različitih strelaca su označeni različitim simbolima, kao na primer, hitac strelca B je označen sa trouglom itd.

Sredina populacije, t.j. srednja vrednost sve četiri mete A, B, C, i D, označena je sa E. U slučaju da ne znamo koji strelac je ispalio hitac, ocenili bismo mesto pogotka sa E, odnosno, a priori ocena μ je E .

Slučajnim procesom biramo jednog od strelaca i njegov identitet nam nije poznat. On će ispaliti hitac ka svojoj meti sa određene razdaljine. Naš zadatak je da, na osnovu lokacije ispaljenog hitca, napravimo najbolju ocenu lokacije sledećeg pogotka za istog strelca. Koristeći Bulmanov kredibilitet upoređujemo dve ocene lokacije. Jedna ocena bi bila između srednje

[58] Neto premija = ukupna šteta/ukupna izloženost= (učestalost zahteva)(iznos štete). Učestalost zahteva= broj zahteva/izloženost. Iznos štete= ukupna šteta/broj zahteva.

[59] Vidi [11]

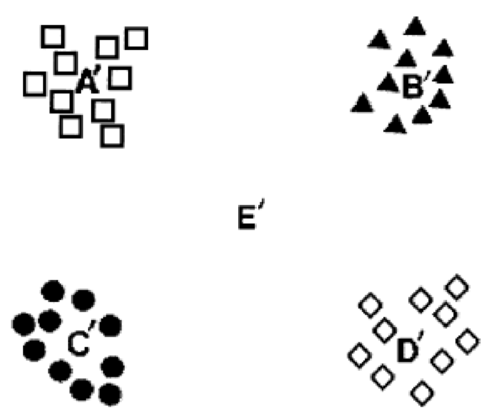
vrednosti opservacija i srednje vrednosti meta, a priori sredina E. Što je veći kredibilitet pripisan opservaciji, ocena je bliže opservaciji, i suprotno, manji kredibilitet pripisan podacima znači da će ocena biti bliža E. Kada bismo striktno pratili proceduru Bejzove analize, sledeći korak bi bio obračun nove korigovane verovatnoće da je hitac ispalio strelac A, B, C ili D, a potom bismo računali novo E na osnovu revidiranih težina. Međutim, u ovom primeru, nije nam namera da se bavimo eksplicitnim kalkulacijama, nego da intuitivno opravdamo osnov za kalkulaciju koja bi bila odrađena.

Pretpostavimo da strelci nisu savršeni, ne pogode uvek svoju metu. Prosečno rasipanje pogodaka oko meta svakog strelca je očekivana vrednost procesne varijanse υ. Ako su strelci dobri (precizani), υ je malo i hitci su usko rasuti oko meta, s druge strane, za lošeg strelca υ je veliko i hitci nisu usko rasuti oko mete. Što je strelac bolji, hitac sadrži više informacija. Što je strelac lošiji, opservacija o lokaciji pogotka sadrži više slučajnog šuma. Stoga, opservacijama dobrog strelca dajemo više težine (uz sve ostale uslove nepromenjene), u odnosu na opservacije lošeg strelca. Dakle, za boljeg strelca, veći je kredibilitet:

Strelac	**Grupisanje hitaca**	**Očekivana vrednost procesne varijanse**	**Sadržaj informacija**	**Kredibilitet pripisan opservaciji**
Dobar	Usko	Malo	Visok	Veći
Loš	Široko	Veliko	Nizak	Manji

Tabela 2.3

Kada je strelac bolji, slučajne fluktuacije u hitcima su manje prisutne i uzoračka sredina je relativno bolja ocena u odnosu na a priori sredinu. Razmotrimo sada Sliku 2.4, gde pretpostavljamo da su svi strelci bolji u odnosu na slučaj predstavljen Slikom 2.3. Ovde ilustrujemo da je za manju očekivanu vrednost procesne varijanse veći kredibilitet, t.j. pripisujemo veći kredibilitet opservacijama. Sada je mnogo lakše reći koji strelac je najverovatnije ispalio određeni hitac, samo na osnovu lokacije pogotka.



Slika 2.4

Ako su mete bliže raspoređene (uz sve ostale uslove nepromenjene), manja je varijacija u hipotetičkim sredinama, i a priori sredina postaje relativno bolja ocena u odnosu na uzoračku sredinu. Stoga, što su mete udaljenije jedna od druge, lakše je razlikovati hitce različitih strelaca i daje se veći kredibilitet opservacijama. Ako svaka meta predstavlja hipotetičku sredinu hitaca jednog strelca, onda se rasipanje meta može kvantifikovati kao varijansa hipotetičkih sredina a.

Mete	Varijansa hipotetičkih sredina	Sadržaj informacija	Kredibilitet pripisan opservacijama
Bliže	Velika	Nizak	Manji
Udaljenije	Mala	Visok	Veći

Tabela 2.4

Pogledajmo Sliku 2.5. gde je ilustrovana ideja da, što su mete udaljenije jedna od druge, to je veći kredibilitet pripisan opservacijama, odnosno, za veće a, veći je i kredibilitet. Uporedimo sada Sliku 2.5 i Sliku 2.3. Vidimo da je na Slici 2.5 lakše odrediti koji strelac je ispalio hitac samo na osnovu lokacije pogotka, drugim rečima, kredibilitet pojedinačne opservacije se povećao.

Slika 2.5

Do sada smo razmatrali rezultate slučaja kada je ispaljen samo jedan hitac. Ako se broj ispaljenih hitaca poveća, kredibilitet pripisan srednjoj vrednosti većeg broja opservacija pogodaka je veći u odnosu na jednu opservaciju. Veći broj posmatranih pogodaka za nepoznatog strelca, daje više informacija i stoga možemo pripisati veći kredibilitet srednjoj vrednosti opservacija.

Sve do sada pomenute karakteristike su obuhvaćene formulom za Bulmanov kredibilitet:

$$Z = \frac{n}{n+k} = \frac{na}{na+v}$$

gde kada υ raste, Z opada i kada a ili n raste, Z takođe raste.

Karakteristika primera sa gađanjem mete	**Matematička kvantifikacija**	**Bulmanov kredibilitet**
Bolji strelac	manje υ	Veći
Udaljenije mete	veće a	Veći
Više hitaca	veće n	Veći

Tabela 2.5

Ukupna varijansa posmatranih pogodaka jednaka je zbiru očekivane vrednosti procesne varijanse υ i varijanse hipotetičkih sredina a. Nesavršenost strelaca je kvantifikovana sa υ, a sa pretpostavkom da strelci nisu savršeni kažemo da je υ pozitivno. Čak i da se sve mete nalaze na istom mestu, postojala bi varijansa u posmatranim rezultatima. Drugo, varijansa zbog rasutosti meta je kvantifikovana sa a, i kada su mete rasute, a je pozitivno. Čak i da su strelci savršeno precizni, i dalje bi postojala varijansa u posmatranim rezultatima kada strelac gađa različite mete.□

Sada obratimo pažnju i na druge primere.

■Primer 2.18

Prisetimo se Primera 2.13. Odredi Bulmanovu ocenu za $E(X_3|0,1)$.

Rešenje:

Iz Primera 2.13 imamo:

$$\mu(D) = \mathrm{E}(X_j|\mathrm{D}) = 0.4, \quad \mu(L) = \mathrm{E}(X_j|\mathrm{L}) = 0.7,$$
$$\pi(D) = 0.75, \qquad \pi(L) = 0.25,$$

te je očekivana vrednost i varijansa za hipotetičke sredine

$$\mu = \sum_\theta \mu(\theta)\pi(\theta) = 0.4(0.75) + 0.7(0.25) = 0.475,$$
$$a = \sum_\theta \mu(\theta)^2\pi(\theta) - \mu^2 = 0.16(0.75) + 0.49(0.25) - 0.475^2 = 0.016875.$$

Očekivanu vrednost procesne varijanse dobijamo sa:

$$\upsilon(D) = \mathrm{Var}(X_j|\mathrm{D}) = 0^2(0.7) + 1^2(0.2) + 2^2(0.1) - 0.4^2 = 0.44,$$
$$\upsilon(L) = \mathrm{Var}(X_j|\mathrm{L}) = 0^2(0.5) + 1^2(0.3) + 2^2(0.2) - 0.7^2 = 0.61,$$
$$\upsilon = \sum_\theta \upsilon(\theta)\pi(\theta) = 0.44(0.75) + 0.61(0.25) = 0.4825.$$

Tada iz (2.32)

$$k = \frac{\upsilon}{a} = \frac{0.4825}{0.016875} = 28.5926$$

a iz (2.31)

$$Z = \frac{2}{2 + 28.5926} = 0.0654.$$

Dakle, sledeća očekivana vrednost je 0.0654(0.5) + 0.9346(0.475) = 0.4766, što je najbolja linearna aproksimacija Bejzove premije (dobijene u Primeru 2.13). □

■Primer 2.19

Koristeći postavku zadatka iz Primera 2.15., odredi Bulmanovu ocenu.

Rešenje:

Za ovaj model,

$$\mu(\Theta) = \Theta^{-1}, \qquad \mu = \mathrm{E}(\Theta^{-1}) = \frac{\beta}{\alpha - 1},$$
$$\upsilon(\Theta) = \Theta^{-2}, \qquad \upsilon = \mathrm{E}(\Theta^{-2}) = \frac{\beta^2}{(\alpha - 1)(\alpha - 2)},$$
$$a = \mathrm{Var}(\Theta^{-1}) = \frac{\beta^2}{(\alpha - 1)(\alpha - 2)} - \left(\frac{\beta}{\alpha - 1}\right)^2 = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)},$$
$$k = \frac{\upsilon}{a} = \alpha - 1,$$
$$Z = \frac{n}{n + k} = \frac{n}{n + \alpha - 1},$$
$$P_c = \frac{n}{n + \alpha - 1}\bar{X} + \frac{\alpha - 1}{n + \alpha - 1}\frac{\beta}{\alpha - 1},$$

što se poklapa sa Bejzovom ocenom.

Za ovaj primer, mogli smo primeniti i *alternativnu analizu*. Neka je jedna opservacija $S = X_1 + \cdots + X_n$. Na osnovu pretpostavki problema, S ima sredinu $n\Theta^{-1}$ i varijansu $n\Theta^{-2}$. Znamo da S ima Gama raspodelu, ali u ovom trenutku ta informacija nije neophodna, pošto Bulmanova aproksimacija zahteva samo vrednost momenata. Pogledajmo sledeću kalkulaciju,

$$\mu = \frac{n\beta}{\alpha - 1}, \upsilon = \frac{n\beta^2}{(\alpha - 1)(\alpha - 2)}, a = \frac{n^2\beta^2}{(\alpha - 1)^2(\alpha - 2)},$$
$$k = \frac{\alpha - 1}{n}, Z = \frac{1}{1 + k} = \frac{n}{n + \alpha - 1}.$$

Primetimo da je sada u formuli za Z obim uzorka jednak 1, što oslikava činjenicu da imamo samo jednu opservaciju S. Kako je $S = n\bar{X}$, Bulmanova ocena je

$$P_c = \frac{n}{n + \alpha - 1}n\bar{X} + \frac{\alpha - 1}{n + \alpha - 1}\frac{n\beta}{\alpha - 1},$$

što je prethodni odgovor uvećan n puta. To je zato što sada ocenjujemo sledeću vrednost za S umesto sledeću vrednost za X. Međutim, faktor kredibiliteta Z je isti, bez obzira da li ocenjujemo X_{n+1} ili sledeću vrednost za S. □

Mogli smo primetiti da Bulmanov kredibilitet meri relativni odnos korisnosti jednog ocenjivača, srednje vrednosti opservacija, u odnosu na drugog ocenjivača, a priori sredinu. Za specifičan slučaj kada je Z = 50%, dva ocenjivača su jednako dobri ili jednako loši. Generalno govoreći, Bulmanov kredibilitet je relativna mera vrednosti informacija sadržanih u opservacijama u odnosu na a priori sredinu.

2.4.4. Bulman–Štraubov model

Za Bulmanov model kažemo da je najjednostavniji model kredibiliteta, zato što dopušta da istorijske štete (slučajne promenljive $X_1, \cdots, X_n$ za dati rizik) osiguranika budu jednako raspodeljene za svaku posmatranu godinu. Ova pretpostavka ne dozvoljava varijacije u jedinici za izloženost riziku ili veličini štete, što se lako može narušiti u realnim situacijama. Na primer:

- Broj vozila u vlasništvu klijenta sa automobilskim osiguranjem se može menjati vremenom.
- Štete tokom prve godine pokrića polise osiguranja mogu oslikavati samo deo godine, zato što je polisa zaključena na neuobičajen datum.
- Količina radne snage za nosioca polise osiguranja o zaštiti na radu, se može menjati iz jedne godine u drugu.
- Iznos zarađenih premija jedne riziko klase može varirati iz godine u godinu.

Dakle, izloženost riziku za nastanak štete može varirati i pretpostavljamo da se može meriti sa: količinom premija osiguranja, brojem zaposlenih, platnim spiskom, brojem osiguranih automobila, brojem zahteva itd.

Da bismo rešili problem pomenutih varijacija, razmotrimo generalizaciju Bulmanovog modela. Ideja u Bulman-Štraubovom modelu (takođe se naziva i empirijski Bejzov model[60]) je da pretpostavimo da su sredine za slučajne promenljive jednake za posmatrani rizik, ali da su varijanse obrnuto proporcionalne veličini (t.j. izloženosti) rizika tokom svakog posmatranog perioda, na primer, kada je rizik dva puta veći, procesna varijansa je prepolovljena.

Pretpostavimo da su slučajne promenljive $X_1, \cdots, X_n$ nezavisne, uslovno za Θ, sa zajedničkom sredinom

$$\mu(\theta) = \mathrm{E}(X_j | \Theta = \theta)$$

i uslovnim varijansama

$$\mathrm{Var}\big(X_j \big| \Theta = \theta\big) = \frac{\upsilon(\theta)}{m_j},$$

[60] Vidi [10]

gde X_j[61] predstavlja broj zahteva za odštetu, monetarnu štetu, ili neki drugi kvantitet od interesa po jedinici izloženosti riziku, a m_j je poznata konstanta, mera za izloženost riziku. Primetimo da je m_j proporcionalno samo veličini rizika i da procesna varijansa slučajne promenljive opada sa porastom izloženosti riziku.

Na koji način treba da kombinujemo slučajne promenljive $X_1, \cdots, X_n$ povezane sa posmatranim rizikom (ili grupom rizika) da bismo ocenili hipotetičku sredinu $\mu(\theta)$? Težinska aritmetička sredina, koja koristi izloženosti m_j, će dati linearni ocenjivač za $\mu(\theta)$ sa minimalnom varijansom. Prvo definišimo ukupnu izloženost

$$m = m_1 + m_2 + \cdots + m_n$$

a zatim i težinsku aritmetičku sredinu

$$\bar{X} = \sum_{j=1}^{n} \frac{m_j}{m} X_j, \tag{2.33}$$

gde m_j može biti broj meseci u godini j tokom kojih je osiguranik bio izložen riziku, ili broj članova grupe u godini j, ili količina prihoda od premija za polisu osiguranja tokom godine j.

Sada, kao i u Bulmanovom modelu, neka su

$$\mu = \mathrm{E}[\mu(\Theta)], \qquad \upsilon = \mathrm{E}[\upsilon(\Theta)],$$

i

$$a = \mathrm{Var}[\mu(\Theta)],$$

gde je očekivana vrednost uzeta preko svih parametara rizika θ u populaciji.

Sledeće određujemo bezuslovne momente, iz (2.27) $\mathrm{E}(X_j) = \mu$, a iz (2.29) $\mathrm{Cov}(X_i, X_j) = a$, ali

$$\begin{aligned}\mathrm{Var}(X_j) &= \mathrm{E}\big[\mathrm{Var}(X_j \,|\Theta)\big] + \mathrm{Var}\big[\mathrm{E}(X_j \,|\Theta)\big] \\ &= \mathrm{E}\left[\frac{\upsilon(\Theta)}{m_j}\right] + \mathrm{Var}[\mu(\Theta)] \\ &= \frac{\upsilon}{m_j} + a\,.\end{aligned}$$

Da bismo odredili kredibilnosnu premiju (2.21), rešavamo sistem normalnih jednačina (2.18) i (2.20) i dobijamo $\tilde{\alpha}_0, \tilde{\alpha}_1, \dots, \tilde{\alpha}_n$. Sada koristeći (2.27), jednačina centriranosti (2.18) postaje

$$\mu = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j\, \mu,$$

ili kada sredimo izraz

$$\sum_{j=1}^{n} \tilde{\alpha}_j = 1 - \frac{\tilde{\alpha}_0}{\mu}. \tag{2.34}$$

Za $i = 1, \dots, n$, (2.20) postaje

[61] X_j može biti broj zahteva za odštetu po godini pokrivenosti za kuću ili može biti racio štete.

$$a = \sum_{\substack{j=1 \\ j\neq i}}^{n} \tilde{\alpha}_j\, a + \tilde{\alpha}_i \left(\frac{v}{m_i} + a\right) = \sum_{j=1}^{n} \tilde{\alpha}_j\, a + \frac{v\tilde{\alpha}_i}{m_i},$$

ili drugačije

$$\tilde{\alpha}_i = \frac{a}{v} m_i \left(1 - \sum_{j=1}^{n} \tilde{\alpha}_j\right) = \frac{a}{v}\frac{\tilde{\alpha}_0}{\mu} m_i, \qquad i = 1, \dots, n. \tag{2.35}$$

Sledeće koristimo (2.34) i (2.35), te dobijamo

$$1 - \frac{\tilde{\alpha}_0}{\mu} = \sum_{j=1}^{n} \tilde{\alpha}_j = \sum_{i=1}^{n} \tilde{\alpha}_i = \frac{a}{v}\frac{\tilde{\alpha}_0}{\mu} \sum_{i=1}^{n} m_i = \frac{a\tilde{\alpha}_0 m}{\mu v},$$

a potom

$$\tilde{\alpha}_0 = \frac{\mu}{1 + am/v} = \frac{v/a}{m + v/a}\mu,$$

i kao rezultat imamo

$$\tilde{\alpha}_j = \frac{a\tilde{\alpha}_0}{\mu v} m_j = \frac{m_j}{m + v/a}.$$

Konačno, kredibilnosna premija (2.21) iznosi

$$\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j = Z\bar{X} + (1 - Z)\mu, \tag{2.36}$$

gde je $Z = \frac{m}{m+k}$ i $\bar{X} = \sum_{j=1}^{n} \frac{m_j}{m} X_j$, sa $k = \frac{v}{a}$ iz (2.32).

Vidimo da kredibilnosna premija (2.36) još uvek ima oblik (2.7), samo što je u ovom slučaju m ukupna izloženost za osiguranika, i Bulman-Štraubov faktor kredibiliteta Z zavisi od m. Nastavimo dalje, $\bar{X}$ je težinska aritmetička sredina za X_j, sa težinama proporcionalnim m_j i obrnuto proporcionalnim varijansi za individualno X_j, odnosno, slučajna promenljiva sa manjom varijansom treba da ima veću težinu. Sa stanovišta grupne interpretacije, X_j je prosečna šteta m_j članova grupe tokom godine j, pa je $m_j X_j$ ukupna šteta grupe tokom godine j. Dalje, $\bar{X}$ je ukupna prosečna šteta po članu grupe tokom n godina, stoga je kredibilnosna premija za grupu u n + 1 godini $m_{n+1}[Z\bar{X} + (1 - Z)\mu]$ za m_{n+1} članova grupe u sledećoj godini.

Posmatrajmo jednu opservaciju $\bar{X}$, za koju je procesna varijansa

$$\mathrm{Var}(\bar{X}|\theta) = \sum_{j=1}^{n} \frac{{m_j}^2}{m^2} \frac{v(\theta)}{m_j} = \frac{v(\theta)}{m},$$

a očekivana vrednost procesne varijanse je v/m. Varijansa hipotetičkih sredina je i dalje a, stoga je $k = v/am$, a kako raspolažemo samo jednom opservacijom $\bar{X}$, faktor kredibiliteta je kao i ranije

$$Z = \frac{1}{1 + \upsilon/(am)} = \frac{m}{m + \upsilon/a}\,. \tag{2.37}$$

■Primer 2.20

Pretpostavimo da u godini j imamo N_j zahteva za odštetu od m_j polisa, $j = 1, \dots, n$, gde individualna polisa osiguranja ima Poasonovu raspodelu sa parametrom Θ, a parametar Θ ima Gama raspodelu sa parametrima α i β. Ako će u godini n + 1 biti m_{n+1} polisa, odredi Bulman–Štraubovu ocenu za broj zahteva u godini n + 1.

Rešenje:

Da bismo zadovoljili uslove ovog modela, neka $X_j = N_j/m_j$ za $j = 1, \dots, n$ i pošto N_j ima Poasonovu raspodelu sa sredinom $m_j\Theta$,

$$\mathrm{E}\big(X_j\big|\Theta = \theta\big) = \mathrm{E}\left(\frac{1}{m_j}N_j\middle|\Theta = \theta\right) = \frac{1}{m_j}\mathrm{E}\big(N_j\big|\Theta = \theta\big) = \Theta = \mu(\Theta)$$

i

$$\mathrm{Var}\big(X_j\big|\Theta = \theta\big) = \mathrm{Var}\left(\frac{1}{m_j}N_j\middle|\Theta = \theta\right) = \frac{1}{m_j}\mathrm{Var}\big(N_j\big|\Theta = \theta\big) = \frac{\Theta}{m_j} = \frac{\upsilon(\Theta)}{m_j}.$$

Sada,

$$\mu = \mathrm{E}(\Theta) = \alpha\beta, \qquad a = \mathrm{Var}(\Theta) = \alpha\beta^2, \qquad \upsilon = \mathrm{E}(\Theta) = \alpha\beta,$$
$$k = \frac{1}{\beta}, \qquad Z = \frac{m}{m + 1/\beta} = \frac{m\beta}{m\beta + 1},$$

pa je ocena za jednog osiguranika

$$P_c = \frac{m\beta}{m\beta + 1}\bar{X} + = \frac{1}{m\beta + 1}\alpha\beta,$$

gde $\bar{X} = m^{-1}\sum_{j=1}^{n} m_j X_j$

Konačno, za godinu *n* + 1, ocena je $m_{n+1}P_c$. □

Pretpostavke u Bulman–Štraubovom modelu su možda previše restriktivne da bi oslikavale realnost. Hewitt[62] je uočio da se veliki rizici ne ponašaju na isti način kao agregatni nezavisni mali rizici. Čak su oni i više varijabilni, što ćemo demonstrirati sledećim primerom.

■Primer 2.21

Neka su $X_1, \dots, X_n$ međusobno nezavisni i uslovno raspodeljeni pri uslovu Θ sa uslovnom sredinom $\mathrm{E}\big(X_j\big|\Theta\big) = \mu(\Theta)$ i uslovnom varijansom $\mathrm{Var}\big(X_j\big|\Theta\big) = \omega(\Theta) + \upsilon(\Theta)/m_j$. Pokaži da ovaj model podržava Hewitt-ovu opservaciju i odredi kredibilnosnu premiju.

Rešenje:

---

[62] Hewitt, C., Jr. , “Loss Ratio Distributions—A Model,” Proceedings of the Casualty Actuarial Society, LIV, 70–88., 1967

Posmatrajmo nezavisne rizike i i j sa izloženošću m_i i m_j i sa srednjom vrednosti Θ. Kada agregiramo, varijansa prosečne štete iznosi

$$\begin{aligned}\mathrm{Var}\left(\frac{m_i X_i + m_j X_j}{m_i + m_j}\middle|\Theta\right) &= \left(\frac{m_i}{m_i + m_j}\right)^2 \mathrm{Var}(X_j|\Theta) + \left(\frac{m_j}{m_i + m_j}\right)^2 \mathrm{Var}(X_j|\Theta) \\ &= \frac{{m_i}^2 + {m_j}^2}{(m_i + m_j)^2}\omega(\Theta) + \frac{1}{m_i + m_j}\upsilon(\Theta),\end{aligned}$$

dok pojedinačan rizik sa izloženošću $m_i + m_j$ ima varijansu $\omega(\Theta) + \upsilon(\Theta)/(m_i + m_j)$, što je veće.

Sledeće određujemo

$$\begin{aligned}\mathrm{E}(X_j) &= \mathrm{E}\big[\mathrm{E}(X_j|\Theta)\big] = \mathrm{E}[\mu(\Theta)] = \mu, \\ \mathrm{Var}(X_j) &= \mathrm{E}\big[\mathrm{Var}(X_j|\Theta)\big] + \mathrm{Var}\big[\mathrm{E}(X_j|\Theta)\big] \\ &= \mathrm{E}\left[\omega(\Theta) + \frac{\upsilon(\Theta)}{m_j}\right] + \mathrm{Var}[\mu(\Theta)] \\ &= \omega(\Theta) + \frac{\upsilon(\Theta)}{m_j} + a\,,\end{aligned}$$

i, za $i \neq j$, $\mathrm{Cov}(X_i, X_j) = a$ kao u (2.29). I dalje je jednačina centriranosti

$$\mu = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j\, \mu,$$

stoga i

$$\sum_{j=1}^{n} \tilde{\alpha}_j = 1 - \frac{\tilde{\alpha}_0}{\mu}.$$

Sada jednačina (2.20) postaje

$$\begin{aligned}a &= \sum_{j=1}^{n} \tilde{\alpha}_j\, a + \tilde{\alpha}_i\left(\omega + \frac{\upsilon}{m_i}\right) \\ &= a\left(1 - \frac{\tilde{\alpha}_0}{\mu}\right) + \tilde{\alpha}_i\left(\omega + \frac{\upsilon}{m_i}\right), \qquad i = 1, \dots, n.\end{aligned}$$

dakle,

$$\tilde{\alpha}_i = \frac{a\tilde{\alpha}_0/\mu}{\omega + \upsilon/m_i}.$$

Sumirajući sa obe strane izraza

$$\frac{a\tilde{\alpha}_0}{\mu}\sum_{j=1}^{n}\frac{m_j}{\upsilon + \omega m_j} = \sum_{j=1}^{n}\tilde{\alpha}_j = 1 - \frac{\tilde{\alpha}_0}{\mu},$$

i tada

$$\tilde{\alpha}_0 = \frac{1}{\frac{a}{\mu}\sum_{j=1}^{n}\frac{m_j}{\upsilon + \omega m_j} + \frac{1}{\mu}} = \frac{\mu}{1 + am^*},$$

gde

$$m^* = \sum_{j=1}^{n} \frac{m_j}{\upsilon + \omega m_j}.$$

Sada

$$\tilde{\alpha}_j = \frac{am_j}{\upsilon + \omega m_j} \frac{1}{1 + am^*},$$

i kredibilnosna premija je

$$\frac{\mu}{1 + am^*} + \frac{a}{1 + am^*} \sum_{j=1}^{n} \frac{m_j X_j}{\upsilon + \omega m_j}.$$

Možemo napraviti sumu da bismo definisali težinsku aritmetičku sredinu opservacija sa

$$\bar{X} = \frac{\sum_{j=1}^{n} \frac{m_j}{\upsilon + \omega m_j} X_j}{\sum_{j=1}^{n} \frac{m_j}{\upsilon + \omega m_j}} = \frac{1}{m^*} \sum_{j=1}^{n} \frac{m_j}{\upsilon + \omega m_j} X_j.$$

Ako sada postavimo

$$Z = \frac{am^*}{1 + am^*},$$

onda je kredibilnosna premija

$$Z\bar{X} + (1 - Z)\mu.$$

Primetimo da kada izloženost m_j teži beskonačnosti, faktor kredibiliteta postaje

$$Z \to \frac{an/\omega}{1 + an/\omega} < 1.$$

Dakle, bez obzira na veličinu rizika, verodostojnost podataka je uvek ograničena. □

Dalje možemo generalizovati problem, tako što ćemo pustiti da varijansa od $\mu(\Theta)$ zavisi od izloženosti. Ovo ima smisla ako verujemo da je sklonost, sa kojom dati rizik odstupa od sredine, zavisna od veličine. Na primer, veći rizik se pri sklapanju polise osiguranja razmatra sa većom pažnjom, pa su i ekstremna odstupanja od sredine manje verovatna.

■Primer 2.22

Prisetimo se Primera 2.21 Odredi formulu kredibiliteta ako je $\mathrm{Var}[\mu(\Theta)] = a + b/m$, *gde je ukupna izloženost za grupu* $m = \sum_{j=1}^{n} m_j$.

Rešenje:

Sada imamo

$$\mathrm{E}(X_j) = \mathrm{E}\big[\mathrm{E}(X_j|\Theta)\big] = \mathrm{E}[\mu(\Theta)] = \mu,$$
$$\mathrm{Var}(X_j) = \mathrm{E}\big[\mathrm{Var}(X_j|\Theta)\big] + \mathrm{Var}\big[\mathrm{E}(X_j|\Theta)\big]$$

$$= \mathrm{E}\left[\omega(\Theta) + \frac{v(\Theta)}{m_j}\right] + \mathrm{Var}[\mu(\Theta)]$$
$$= \omega(\Theta) + \frac{v(\Theta)}{m_j} + a + \frac{b}{m},$$

i, za $i \neq j$,

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= \mathrm{E}\big[\mathrm{E}(X_i X_j | \Theta)\big] - \mu^2 \\ &= \mathrm{E}[\mu(\Theta)^2] - \mu^2 \\ &= a + \frac{b}{m}.\end{aligned}$$

Primećujemo da su sve kalkulacije iz Primera 2.21 primenljive i ovde, kada se a zameni sa$a + b/m$. Sada faktor kredibiliteta iznosi

$$Z = \frac{(a + \frac{b}{m})m^*}{1 + (a + \frac{b}{m})m^*},$$

a kredibilnosna premija je

$$Z\bar{X} + (1 - Z)\mu,$$

sa $\bar{X}$ i m^* je definisano kao u Primeru 2.21. □

2.4.5. Tačan kredibilitet

Kažemo da je kredibilitet *tačan*, kada je Bejzova premija jednaka kredibilnosnoj premiji. U Primeru 2.19 je kredibilnosna premija jednaka Bejzovoj premiji. Prisetimo se sada rešenja sistema normalnih jednačina, $\tilde{\alpha}_0, \tilde{\alpha}_1, \dots, \tilde{\alpha}_n$ odakle dobijamo kredibilnosnu premiju $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j$ koja minimizuje

$$Q = \mathrm{E}\left\{\left[\mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right]^2\right\},$$

$$Q_1 = \mathrm{E}\left\{\left[\mathrm{E}(X_{n+1} | \mathbf{X}) - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right]^2\right\}$$

i

$$Q_2 = \mathrm{E}\left[\left(X_{n+1} - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right)^2\right].$$

Primetimo da možemo utvrditi da li je kredibilitet tačan bez izračunavanja kredibilnosne premije. Ako je Bejzova premija linearna funkcija od $X_1, \dots, X_n$,

$$\mathrm{E}(X_{n+1}\,|\mathbf{X}) = \alpha_0 + \sum_{j=1}^{n} \alpha_j X_j,$$

onda je jasno da u (2.22) kvantitet Q_1 ima minimalnu vrednost, jednaku nuli, kada je $\tilde{\alpha}_j = \alpha_j$ za $j = 0,1, \dots, n$. Stoga je kredibilnosna premija $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j = \alpha_0 + \sum_{j=1}^{n} \alpha_j X_j = \mathrm{E}(X_{n+1}\,|\mathbf{X})$ i kredibilitet je tačan.

Da sumiramo, Bulmanov ocenjivač $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j$ je *„najbolja linearna“ aproksimacija Bejzove premije* za funkciju srednjekvadratne greške Q_1.

Prisetimo se linearne regresije,

$$Y_i = \alpha_0 + \sum_{j=1}^{n} \alpha_j X_j + \epsilon_i,$$

gde

$$\epsilon_i \sim N(0, \sigma^2) \text{ for } i = 1, \dots, m.$$

$\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_n$ su izabrani tako da minimizuju

$$Q = \mathrm{E}\left[\sum_{i=1}^{m} {\epsilon_i}^2\right] = \mathrm{E}\left[\sum_{i=1}^{m} (Y_i - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j)^2\right].$$

Dakle, linija regresije $\hat{Y}_i = \hat{\alpha}_0 + \sum_{j=1}^{n} \hat{\alpha}_j X_j$ odgovara kredibilnosnoj premiji $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j$ i opservacija Y_i odgovara Bejzovoj premiji $\mathrm{E}[X_{n+1}|\mathbf{X} = \mathbf{x}_i]$, $i = 1, \dots, m$.

Tačan kredibilitet se generalno javlja u situaciji kada je Bulmanov (i Bulman–Štraubov) model u vezi sa članovima linearno eksponencijalne familije raspodela ili sa njihovim konjugovanim priornim raspodelama.

Neka su $X_j|\Theta = \theta$ međusobno nezavise slučajne promenljive iz linearno eksponencijalne familije raspodela sa raspodelom verovatnoća za $j = 1, \dots, n+1$,

$$f_{X_j|\Theta}(x_j|\theta) = \frac{p(x_j)e^{-\theta x_j}}{q(\theta)},$$

i Θ ima raspodelu

$$\pi(\theta) = \frac{[q(\theta)]^{-k} e^{-\mu k \theta}}{c(\mu, k)}, \qquad \theta_0 < \theta < \theta_1, \tag{2.38}$$

gde $-\infty \le \theta_0 < \theta_1 \le \infty$ i neka su $\pi(\theta_0) = \pi(\theta_1) = 0$. Za sada, μ i k posmatramo kao obične parametre od $\pi(\theta)$.

Iz (1.9) u odeljku 1.5. znamo da je

$$\mu(\theta) = \mathrm{E}(X_j|\Theta = \theta) = -\frac{q'(\theta)}{q(\theta)}.$$

Želimo da nađemo $\mathrm{E}[\mu(\Theta)]$. Iz jednačine (2.38)

$$\ln \pi(\theta) = -k \ln q(\theta) - \mu k \theta - \ln c(\mu, k)$$

što, kada integralimo po θ, daje

$$\frac{\pi'(\theta)}{\pi(\theta)} = -\frac{kq'(\theta)}{q(\theta)} - \mu k,$$

ili drugačije zapisano

$$\pi'(\theta) = k[\mu(\theta) - \mu]\pi(\theta). \tag{2.39}$$

Sada integralimo (2.39) od θ_0 do θ_1

$$\pi(\theta_1) - \pi(\theta_0) = k\int_{\theta_0}^{\theta_1} \mu(\theta)\pi(\theta)d\theta - k\mu\int_{\theta_0}^{\theta_1} \pi(\theta)d\theta,$$

što implicira da $0 = k\mathrm{E}[\mu(\Theta)] - k\mu$, ili ekvivalentno,

$$\mathrm{E}[\mu(\Theta)] = \mu, \tag{2.40}$$

čime demonstriramo da izbor za simbol μ, u (2.38) nije slučajan.

Sada razmatramo posteriornu raspodelu $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$ za Bulmanov metod. To je proporcionalno sa

$$\left[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\right]\pi(\theta),$$

dakle proporcionalno je i sa

$$\begin{aligned}\left[\prod_{j=1}^{n} \frac{e^{-\theta x_j}}{q(\theta)}\right][q(\theta)]^{-k}e^{-\mu k\theta} \\ &= [q(\theta)]^{-(n+k)}e^{-\theta(\mu k + n\bar{x})} \\ &= [q(\theta)]^{-k_*}e^{-\mu_* k_* \theta},\end{aligned} \tag{2.41}$$

gde

$$k_* = n + k$$
$$\mu_* = \frac{\mu k + n\bar{x}}{n+k} = \frac{n}{n+k}\bar{x} + \frac{k}{n+k}\mu.$$

Vidimo da je (2.41) proporcionalno sa gustinom raspodele oblika (2.38) gde su μ i k zamenjeni sa μ_* i k_* respektivno. Dakle,

$$\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{[q(\theta)]^{-k_*}\, e^{-\mu_*\, k_*\, \theta}}{c(\mu_*, k_*)}, \qquad \theta_0 < \theta < \theta_1.$$

Uzmimo u obzir (2.16) i razvijajući na isti način kao i za(2.40), odredimo Bejzovu premiju

$$\begin{aligned}\mathrm{E}(X_{n+1}\,|\mathbf{X} = \mathbf{x}) &= \int_{\theta_0}^{\theta_1} \mu(\theta)\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta \\ &= \mu_* \\ &= Z\bar{x} + (1 - Z)\mu,\end{aligned}$$

gde $Z = n/(n+k)$. Primetimo da je premija oblika (2.7), a kako je i linearna funkcija od x_j-ova, kredibilitet mora biti tačan, t.j. kredibilnosna premija je

$$\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j = Z\bar{X} + (1-Z)\mu = \mathrm{E}(X_{n+1} \,|\mathbf{x}\,).$$

Kako su $X_j|\Theta$ jednako raspodeljeni za $j = 1, \dots, n$, Bulmanov model je primenljiv, stoga je i (2.30) primenljivo, odnosno k mora da zadovoljava (2.32). Podsetimo se da iz (1.10) u odeljku 1.5. imamo

$$v(\theta) = \mathrm{Var}\big(X_j\big|\Theta = \theta\,\big) = -\mu'(\theta).$$

Sada diferenciramo (2.39) da bismo dobili

$$\begin{aligned}\pi''(\theta) &= k\mu'(\theta)\pi(\theta) + k^2[\mu(\theta)-\mu]^2\pi(\theta)\\ &= kv(\theta)\pi(\theta) + k^2[\mu(\theta)-\mu]^2\pi(\theta),\end{aligned}$$

i integraljenje po θ , od θ_0 do θ_1, daje

$$\begin{aligned}\pi'(\theta_1) - \pi'(\theta_0) &= -kE[v(\Theta)] + k^2\mathrm{E}\{[\mu(\Theta)-\mu]^2\}\\ &= -kv + k^2 a,\end{aligned}$$

gde $\mu(\Theta)$ ima sredinu μ i $\mathrm{E}\{[\mu(\Theta)-\mu]^2\} = \mathrm{Var}[\mu(\Theta)] = a$. Ako je $\pi'(\theta_1) = \pi'(\theta_0) = 0$, onda je implicirano da $k = v/a$ i (2.32) je zadovoljeno.

2.5. Napomene i praktični problemi

U ovom odeljku ćemo pokriti razne teme bitne za primenu teorije kredibiliteta.

Da li je greška, koja nastaje kada koristimo kredibilnosnu premiju umesto Bejzove premije, vredna brige? Videli smo da restrikcije postavljene za linearno rešenje i nisu tako loše, s obzirom da često dobijamo tačan kredibilite. Štaviše, pristup linearnog kredibiliteta podrazumeva samo pretpostavku ili ocenu prva dva momenta, dok Bejzov pristup zahteva da raspodela slučajne promenljive bude potpuno definisana. Stoga, ova neparametarska karakteristika čini linearan pristup snažnijim, što može nadomestiti manjak preciznosti.

Može se zaključiti da, inferiorno ponašanje kredibilnosnog ocenjivača, u poređenju sa Bejzovim ocenjivačem, nastaje zbog debelih repova obe raspodele. Aktuari američke asocijacije[63] su predložili 1990. godine, da se rep raspodele može stanjiti(olakšati) tako što će se u radu, umesto originalnih podataka, koristiti logaritam vrednosti podatka. Ideja je da se koristi

[63] Casualty Actuarial Society, *Foundations of Casualty Actuarial Science*, Arlington, VA: Casualty Actuarial Society, 1990

logaritam podataka, i na osnovu takvih podataka, koristeći linearni kredibilitet, odredi ocena srednje vrednosti raspodele logaritama, pa se rezultat (konačna ocena) eksponira. Konačno, pristrasnost, koju ova procedura unosi, se može ispraviti množenjem svih ocena sa konstantom, takvom da se uzoračka sredina ocene poklapa sa uzoračkom sredinom originalnih podataka.

Teorija kredibiliteta pretpostavlja da je proces nastanka štete slučajan i da su događaji nastanka štete međusobno nezavisni, što ne mora uvek biti tačno. Na primer, vremenske nepogode mogu izazvati lavinu zahteva za odštetu, poledica na ulicama može izazvati više sudara za samo nekoliko dana, ili, tokom kratkog vremenskog perioda, velike hladnoće mogu izazvati veliki broj požara usled korišćenja nebezbednih metoda zagrevanja kuća. Zadatak svakog aktuara je da pokuša da izdvoji ove specijalne događaje, ali ih nije uvek moguće identifikovati i napraviti odgovarajuće izmene.

Kod razmatranja statičnog modela, zbog „homogenosti u vremenu“, oceniti distribuciju sledeće opservacije postaje isto što i oceniti bilo koju buduću opservaciju, bez obzira koliko daleko se ona nalazi u budućnosti. U Poglavlju 7 u [4] možete razmotriti model Kalmanovog filtera, gde rastojanje u budućnosti postaje relevantno pri ocenjivanju. Glavna karakteristika ovog modela je da dozvoljava uvođenje vremenski zavisnih parametara i ne zahteva integraciju pri određivanju posteriorne raspodele. S druge strane, najveća mana modela je što pomenute rezultate postiže uvođenjem restrikcije na posmatranje normalne raspodele.

Rasuđivanje često igra veliku ulogu u odabiru parametara kredibiliteta. Kredibilnosni parametri se često biraju tako da oslikavaju aktuarovu želju da balansira između adekvatne reakcije i stabilnosti. Što se veća težina da istorijskim podacima, to tekući podaci imaju veći uticaj na ocenu, odnosno ocene više reaguju na promene u tekućim podacima. Naravno, ovo se postiže na uštrb stabilnosti ocena.
Takođe, komplement kredibiliteta treba birati tako da bude relativno stabilan i relevantan za kvantitet koji se ocenjuje.[64]

Generalno, relativno male razlike u parametrima za kredibilnosnu formulu nemaju veliki uticaj na ocene. U praksi, nije neophodno precizno odrediti najbolje parametre za kredibilnosnu formulu[65], već je obično dovoljno oceniti k u okviru dve vrednosti. Kako aktuari često određuju k na osnovu ličnog rasuđivanja, dobra ocena je zapravo najbolja ocena koja se može dobiti.

Najvažnija razlika između modela klasičnog kredibiliteta i Bulmanovog kredibiliteta jeste da Bulmanov kredibilitet nikada ne dostiže vrednost $Z = 1$, koja predstavlja asimptotu krive. Oba modela su efektivna u popravljanju stabilnosti i preciznosti ocena i obe formule daju približno iste kredibilnosne težine, ako je standard za potpuni kredibilitet teorije klasičnog kredibiliteta n_0, 7 do 8 puta veći od Bulmanovog kredibilnosnog parametra k.[66]

---

[64] Boor, “The Complement of Credibility,” Proceedings of the Casualty Actuarial Society LXXXIII, 1996.
[65] Mahler’s “An Actuarial Note on Credibility Parameters,” Proceedings of the Casualty Actuarial Society LXXIII, 1986.
[66] Vidi [2]. Ovde pretpostavljamo da je formula Bulmanovog kredibiliteta izražena u odnosu na broj podnetih zahteva za odštetu. Ukoliko nije, potrebno je napraviti konverziju.

Aktuari odlučuju koji će model primeniti na osnovu podataka kojima raspolažu i na osnovu cilja koji žele da postignu. Ako je cilj da se generiše što preciznija premija polise osiguranja, gde se za meru preciznosti koriste najmanji kvadrati, onda je Bulmanov kredibilitet najbolji izbor. Ako su ocene za v i a poznate ali se teško izračunavaju, onda se može koristiti klasičan kredibilitet. Vidi [2]. Ako aktuar imam razumnu ocenu priorne raspodele, onda izbor spada na Bejzovu analizu, ali ovaj metod je najteži za primenu i izaziva najviše poteškoća kada se rezultati metoda žele objasniti osobama koje nisu aktuari.

Sa čvrstom osnovom za određivanje kredibilnosne premije, jedina prepreka koja ostaje je da se numerički ocene kvantiteti a i v u Bulmanovoj formulaciji, ili da se definiše priorna raspodela u Bejzovoj formulaciji. Ovim pitanjem ćemo se pozabaviti u narednom poglavlju.

Poglavlje 3

3. Empirijsko Bejzovo ocenjivanje parametara

Do sada smo bili u prilici da odredimo numeričke vrednosti posmatranih kvantiteta, zato što smo pretpostavili da su raspodele $f_{X_j|\Theta}(x_j|\theta)$ i $\pi(\theta)$ poznate. Iako je to zgodno pretpostaviti zarad ilustracije metodologije, u praksi često nemamo predznanje o tim raspodelama. Praktična upotreba teorije zahteva da se nepoznati parametri ocene na osnovu podataka, kako bi model što bolje oslikavao stvarnost.

Generalno, nepoznati parametri su oni koji su u vezi sa strukturnom raspodelom $\pi(\theta)$, zbog čega se zovu *strukturni parametri*. Terminologija koju koristimo prati Bejzov okvir predstavljen u prethodnom poglavlju. Striktno govoreći, u kontekstu Bejzove metodologije pretpostavlja se da su svi parametri poznati i nema potrebe za ocenjivanjem. Ponekad ovaj potpuni Bejzov pristup nije zadovoljavajući (npr. kada imamo malo ili uopšte nemamo priorne informacije, kao što je slučaj u novoj liniji osiguranja) i moramo koristiti raspoložive informacije da bismo ocenili strukturne (priorne) parametre. Ovakav pristup se zove *empirijsko Bejzovo ocenjivanje parametara*. Razmotrićemo tri slučaja kroz neparametarsko, semiparametarsko i parametarsko ocenjivanje.

Prvi slučaj, kada su $f_{X_j|\Theta}(x_j|\theta)$ i $\pi(\theta)$ neodređeni (kao na primer u Bulmanovom ili Buman-Štraubovom modelu gde je potrebno znati samo prva dva momenta raspodele), zove se *neparametarski* slučaj. Ako se pretpostavi da je $f_{X_j|\Theta}(x_j|\theta)$ parametarskog oblika (npr. Poasonova, normalna ili neka druga raspodela) ali ne i $\pi(\theta)$, onda se za ovakav problem kaže da je *semiparametarske* prirode. Treći, ujedno i tehnički najkomplikovaniji slučaj, zove se *parametarski* slučaj, kada se za obe vrednosti $f_{X_j|\Theta}(x_j|\theta)$ i $\pi(\theta)$ pretpostavi da su parametarskog oblika.

Odluka o tome da li da se koristi paramearski oblik ili ne, zavisi delom od date situacije a delom od procene i znanja osobe koja se bavi analizom. Na primer, analiza bazirana na broju zahteva za odštetu može uključiti pretpostavku da je $f_{X_j|\Theta}(x_j|\theta)$ Poasonovog oblika, dok odabir za parametarski oblik od $\pi(\theta)$ ne mora biti razuman.

Sve parametarske pretpostavke koje imamo, treba da budu reflektovane (što je više moguće) u parametarskom ocenjivanju. Na primer, u slučaju Poasonove raspodele, pošto je sredina jednaka varijansi, ista ocena se treba koristiti za oba kvantiteta. Neparametarski ocenjivači normalno ne moraju biti efikasniji od ocenjivača za odabrani parametarski model, pod pretpostavkom da je odabrani model odgovarajući. Ovo može biti relevantno pri odluci o tome da li izabrati parametarski model. Konačno, činjenica da neparametarski modeli imaju tu prednost da su odgovarajući za veliki proj različitih situacija, pomaže nam da eliminišemo dodatni teret parametarske pretpostavke.

U ovom poglavlju pretpostavljamo da za svakog osiguranika $r \geq 1$ imamo opservacije o štetama $\boldsymbol{X}_i = (X_{i1}, \cdots, X_{in_i})^T$ za $i = 1, \ldots, r$ po jedinici izloženosti. Pretpostavimo da su slučajni vektori $\{\boldsymbol{X}_i, i = 1, \ldots, r\}$ statistički nezavisni (iskustvo različitih osiguranika je nezavisno). (Nepoznati)Parametar rizika za i-tog osguranika je θ_i, $i = 1, \ldots, r$, i još pretpostavljamo da su $\theta_1, \ldots, \theta_r$ vrednosti međusobno nezavisnih i slučajno raspodeljenih slučajnih promenljivih Θ_i sa strukturnom raspodelom $\pi(\theta_i)$. Neka su za fiksno i, (uslovne) slučajne promenljive $X_{ij}|\theta_i$ međusobno nezavisne sa raspodelama $f_{X_{ij}|\Theta}(x_{ij}|\theta_i)$, $j = 1, \ldots, n_i$.

Dva uobičajena slučaja daju podatke ovakvog oblika. Jedan je određivanje klasifikacije premija, a drugi je ocenjivanje iskustva. U oba slučaja, i određuje klasu ili grupu, a j određuje individualnog člana. Drugi slučaj je kao i prvi, kada i i dalje određuje klasu ili grupu, ali sada j određuje godinu a opservacija je prosečna šteta za tu godinu. Bez obzira na potencijalnu postavku slučaja, sa r označavamo broj entiteta (osiguranika).

Moguće je da znamo vektor izloženosti riziku $\boldsymbol{m}_i = (m_{i1}, \cdots, m_{in_i})^T$ za osiguranika i, gde je $i = 1, \ldots, r$. Ukoliko nam taj vektor nije poznat (i ukoliko je odgovarajuće), možemo postaviti da je $m_{ij} = 1$ za svako i i j. Neka je

$$m_i = \sum_{j=1}^{n_i} m_{ij}\,, i = 1, \ldots, r$$

ukupna istorijska izloženost za osiguranika i, i neka je

$$\bar{X}_i = \frac{1}{m_i} \sum_{j=1}^{n_i} m_{ij} X_{ij}\,, i = 1, \ldots, r$$

istorijska prosečna šteta. Dalje, neka je ukupna izloženost

$$m = \sum_{i=1}^{r} m_i = \sum_{i=1}^{r} \sum_{j=1}^{n_i} m_{ij}$$

a ukupna prosečna istorijska šteta je

$$\bar{X} = \frac{1}{m} \sum_{i=1}^{r} m_i \bar{X}_i = \frac{1}{m} \sum_{i=1}^{r} \sum_{j=1}^{n_i} m_{ij} X_{ij}. \tag{3.1}$$

Parametri koje treba da ocenimo zavise od pretpostavki u vezi sa raspodelama $f_{X_{ij}|\Theta}(x_{ij}|\theta_i)$ i $\pi(\theta)$.

Za Bulman-Štraubovu formulaciju posmatramo dodatne kvantitete. Hipotetička sredina (ne zavisi od j) je

$$\mathrm{E}\left(X_{ij}\middle|\Theta_i = \theta_i\right) = \mu(\theta_i)$$

a procesna varijansa je

$$\mathrm{Var}\left(X_{ij}\middle|\Theta_i = \theta_i\right) = \frac{\upsilon(\theta_i)}{m_{ij}}.$$

Strukturni parametri su

$$\mu = \mathrm{E}[\mu(\Theta_i)], \qquad \upsilon = \mathrm{E}[\upsilon(\Theta_i)]$$

i

$$a = \mathrm{Var}[\mu(\Theta_i)].$$

Potrebno je oceniti μ, υ, i a (kada su nepoznati) na osnovu opservacija. Kredibilnosna premija za štetu sledeće godine (po jedinici izloženosti) za osiguranika *i* iznosi

$$Z_i\bar{X}_i + (1 - Z_i)\mu, i = 1, \dots, r, \tag{3.2}$$

gde je

$$Z_i = \frac{m_i}{m_i + k}, \qquad k = \frac{\upsilon}{a}.$$

Ako su ocenjivači μ, υ, i a označeni sa $\hat{\mu}$, $\hat{\upsilon}$, i $\hat{a}$ respektivno, onda se kredibilnosna premija može zameniti ocenjivačem

$$\hat{Z}_i\bar{X}_i + \left(1 - \hat{Z}_i\right)\hat{\mu}, \tag{3.3}$$

gde je

$$\hat{Z}_i = \frac{m_i}{m_i + \hat{k}}, \qquad \hat{k} = \frac{\hat{\upsilon}}{\hat{a}}.$$

Primetimo da, čak i kada su $\hat{\upsilon}$ i $\hat{a}$ centrirani ocenjivači za υ i a isto se ne može reći za $\hat{k}$ *i* $\hat{Z}_i$. Konačno, kredibilnosna premija za pokriće svih jedinica izloženosti m_{in_i+1} za osiguranika *i* u sledećoj godini bi bila (3.3) pomnožena sa m_{in_i+1}.

3.1. Neparametarsko ocenjivanje

U ovom odeljku razmatramo centrirane ocene za μ, υ, i a. Da bismo ilustrovali ideju, počećemo sa jednostavnim primerom Bulmanovog tipa.

■Primer 3.1

Neka je $n_i = n > 1$ za sve i i $m_{ij} = 1$ za svako i i j. Odnosno, za svakog osiguranika i, imamo vektor sa štetama

$$\boldsymbol{X}_i = (X_{i1}, \cdots, X_{in})^T, i = 1, \dots, r.$$

Dalje, uslovno za $\Theta_i = \theta_i$, X_{ij} ima sredinu

$$\mu(\theta_i) = E\left(X_{ij}\middle|\Theta_i = \theta_i\right)$$

i varijansu

$$\upsilon(\theta_i) = Var\left(X_{ij}\middle|\Theta_i = \theta_i\right),$$

i $X_{i1}, \cdots, X_{in}$ su (uslovno) nezavisni. Takođe, istorijske štete različitih osiguranika su nezavisne, tako da ako je $i \neq s$, onda su X_{ij} i X_{st} nezavisni. U tom slučaju,

$$\bar{X}_i = n^{-1} \sum_{j=1}^{n} X_{ij} \ i \ \bar{X} = r^{-1} \sum_{i=1}^{r} \bar{X}_i = (rn)^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n} X_{ij}.$$

Odredi centrirane ocenjivače za Bulmanove kvantitete.

Rešenje:

Centrirani ocenjivač za μ je

$$\hat{\mu} = \bar{X}$$

zato što je

$$\mathrm{E}(\hat{\mu}) = (rn)^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n} \mathrm{E}(X_{ij}) = (rn)^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n} \mathrm{E}[\mathrm{E}(X_{ij} | \Theta_i)]$$
$$= (rn)^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n} \mathrm{E}[\mu(\Theta_i)] = (rn)^{-1} \sum_{i=1}^{r} \sum_{j=1}^{n} \mu = \mu.$$

Za ocenu v i a, koristimo sledeći rezultat. Neka su $Y_1, \cdots, Y_k$ međusobno nezavisne (ali ne moraju biti jednako raspodeljene) slučajne promenljive sa jednakim sredinama i varijansama $\mu = \mathrm{E}[Y_i]$ i $\sigma^2 = \mathrm{Var}(Y_j)$. Neka je $\bar{Y} = k^{-1} \sum_{j=1}^{k} Y_j$. Tada je

$$\mathrm{E}(\bar{Y}) = k^{-1} \sum_{j=1}^{k} Y_j = \mu,$$
$$\mathrm{Var}(\bar{Y}) = k^{-2} \sum_{j=1}^{k} \mathrm{Var}(Y_j) = \frac{\sigma^2}{k}.$$

Sledeće razmotrimo statistiku $\sum_{j=1}^{k} (Y_j - \bar{Y})^2$. Drugačije se može zapisati

$$\begin{aligned}
\sum_{j=1}^{k} (Y_j - \bar{Y})^2 &= \sum_{j=1}^{k} [(Y_j - \mu) + (\mu - \bar{Y})]^2 \\
&= \sum_{j=1}^{k} [(Y_j - \mu)^2 + 2(Y_j - \mu)(\mu - \bar{Y}) + (\mu - \bar{Y})^2] \\
&= \sum_{j=1}^{k} (Y_j - \mu)^2 + 2(\mu - \bar{Y}) \sum_{j=1}^{k} (Y_j - \mu) + \sum_{j=1}^{k} (\bar{Y} - \mu)^2 \\
&= \sum_{j=1}^{k} (Y_j - \mu)^2 + 2(\mu - \bar{Y})(k\bar{Y} - k\mu) + k(\bar{Y} - \mu)^2,
\end{aligned}$$

što pojednostavljuje

$$\sum_{j=1}^{k}(Y_j-\bar{Y})^2=\sum_{j=1}^{k}\left(Y_j-\mu\right)^2-k(\bar{Y}-\mu)^2. \tag{3.4}$$

Kada uzmemo očekivanje sa obe strane, dobijamo

$$\begin{aligned}\mathrm{E}\left[\sum_{j=1}^{k}\left(Y_j-\bar{Y}\right)^2\right]&=\sum_{j=1}^{k}\mathrm{E}\left[\left(Y_j-\mu\right)^2\right]-k\mathrm{E}[(\bar{Y}-\mu)^2]\\&=\sum_{j=1}^{k}\mathrm{Var}(Y_j)-k\mathrm{Var}(\bar{Y})\\&=k\sigma^2-k\left(\frac{\sigma^2}{k}\right)=(k-1)\sigma^2.\end{aligned}$$

Dakle,

$$\mathrm{E}\left[\frac{1}{k-1}\sum_{j=1}^{k}\left(Y_j-\bar{Y}\right)^2\right]=\sigma^2, \tag{3.5}$$

Stoga je $\sum_{j=1}^{k}\left(Y_j-\bar{Y}\right)^2/(k-1)$ centrirani ocenjivač za varijansu od Y_j.

Da bismo ocenili υ, razmotrimo

$$\hat{\upsilon}_i=\frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij}-\bar{X}_i\right)^2, \tag{3.6}$$

Prisetimo se da su za svako i slučajne promenljive $X_{i1},\cdots,X_{in}$ nezavisne, uslovno za $\Theta_i=\theta_i$. Stoga, $\hat{\upsilon}_i$ je centrirana ocena za $Var\left(X_{ij}\middle|\Theta_i=\theta_i\right)=\upsilon(\theta_i)$. Bezuslovno,

$$\mathrm{E}(\hat{\upsilon}_i)=\mathrm{E}[\mathrm{E}(\hat{\upsilon}_i|\Theta_i)]=\mathrm{E}[\upsilon(\Theta_i)]=\upsilon,$$

i $\hat{\upsilon}_i$ je centrirano za υ. Stoga, centrirana ocenjivač za υ je

$$\hat{\upsilon}=r^{-1}\sum_{i=1}^{r}\hat{\upsilon}_i=\frac{1}{r(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}\left(X_{ij}-\bar{X}_i\right)^2. \tag{3.7}$$

Sada se vratimo ocenjivanju parametra a. Počinjemo sa

$$\mathrm{E}(\bar{X}_i|\Theta_i=\theta_i)=n^{-1}\sum_{j=1}^{n}\mathrm{E}\left(X_{ij}\middle|\Theta_i=\theta_i\right)=n^{-1}\sum_{j=1}^{n}\mu(\theta_i)=\mu(\theta_i).$$

Dakle

$$\mathrm{E}(\bar{X}_i)=\mathrm{E}[\mathrm{E}(\bar{X}_i|\Theta_i)]=\mathrm{E}[\mu(\Theta_i)]=\mu$$

i

$$\begin{aligned}\mathrm{Var}(\bar{X}_i)&=\mathrm{Var}[\mathrm{E}(\bar{X}_i|\Theta_i)]+\mathrm{E}[\mathrm{Var}(\bar{X}_i|\Theta_i)]\\&=\mathrm{Var}[\mu(\Theta_i)]+\mathrm{E}\left[\frac{\upsilon(\Theta_i)}{n}\right]=a+\frac{\upsilon}{n}.\end{aligned}$$

Dakle, $\bar{X}_1, \dots, \bar{X}_r$ su nezavisne sa zajedničkom sredinom μ i varijansom $a + \frac{v}{n}$. Njihova uzoračka sredina je $\bar{X} = r^{-1} \sum_{i=1}^{r} \bar{X}_i$. Posledično, centrirani ocenjivač za $a + \frac{v}{n}$ je $(r-1)^{-1} \sum_{i=1}^{r} (\bar{X}_i - \bar{X})^2$. Pošto već imamo centrirani iocenjivač za v, centrirani ocenjivač za a je

$$\begin{aligned} \hat{a} &= \frac{1}{r-1} \sum_{i=1}^{r} (\bar{X}_i - \bar{X})^2 - \frac{\hat{v}}{n} \\ &= \frac{1}{r-1} \sum_{i=1}^{r} (\bar{X}_i - \bar{X})^2 - \frac{1}{rn(n-1)} \sum_{i=1}^{r} \sum_{j=1}^{n} \left(X_{ij} - \bar{X}_i\right)^2. \qquad (3.8) \ \square \end{aligned}$$

Ovi ocenjivači mogu slično izgledati. Razmotrimo sada jednofaktorsku analizu varijanse, u kojoj svaki osiguranik predstavlja jedno delovanje. Ocenjivač za v (3.7) je unutrašnja srednjekvadratna greška. Prvi sabirak u ocenjivaču za a (3.8) je među (takođe se naziva delovanje) srednjekvadratna greška podeljena sa n. Hipoteza da sva delovanja imaju iste srednje vrednosti je prihvatljiva kada među srednjekvadratna greška mala u odnosu na unutrašnju srednjekvadratnu grešku, odnosno kada je $\hat{a}$ malo u odnosu na $\hat{v}$. Ovakav odnos implicira da će $\hat{Z}$ biti blizu nule i da će se mali kredibilitet pripisati vrednostima $\bar{X}_i$. Ovako i treba da bude kada su osiguranici esencijalno jednaki.

Zbog razlike u (3.8) moguće je da $\hat{a}$ bude negativno. Ukoliko se to desi, postavljamo da je $\hat{a}$ $=\hat{Z} = 0$. Ovaj slučaj je ekvivalentan slučaju kada je F test statistika u analizi varijanse manja od 1, slučaj koji uvek vodi ka prihvatanju hipoteza o jednakim sredinama.

■Primer 3.2

Prisetimo se Primera 3.1 Pretpostavimo da kao numeričku ilustraciju imamo $r = 2$ osiguranika sa $n = 3$ godina iskustva za svakog od njih. Neka su štete $\boldsymbol{x}_1 = (3,5,7)^T$ i $\boldsymbol{x}_2 = (6,12,9)^T$. Odredi Bulmanove kredibilnosne premije za svakog osiguranika.

Rešenje:

Imamo

$$\bar{X}_1 = \frac{1}{3}(3+5+7) = 5, \qquad \bar{X}_2 = \frac{1}{3}(6+12+9) = 9,$$

pa je $\bar{X} = \frac{1}{2}(5+9) = 7$. Tada je $\hat{\mu} = 7$. Sledeće imamo

$$\hat{v}_1 = \frac{1}{2}[(3-5)^2 + (5-5)^2 + (7-5)^2] = 4,$$

$$\hat{v}_2 = \frac{1}{2}[(6-9)^2 + (12-9)^2 + (9-9)^2] = 9,$$

pa je $\hat{v} = \frac{1}{2}(4+9) = \frac{13}{2}$. Tada je

$$\hat{a} = [(5-7)^2 + (9-7)^2] - \frac{1}{3}\hat{v} = \frac{35}{6}.$$

Sledeće, $\hat{k} = \frac{\hat{v}}{\hat{a}} = \frac{39}{35}$ i ocenjeni faktor kredibiliteta je $\hat{Z} = \frac{3}{3+\hat{k}} = \frac{35}{48}$. Ocenjene kredibilnosne premije su

$$\hat{Z}\bar{X}_1 + (1-\hat{Z})\hat{\mu} = \left(\frac{35}{48}\right)(5) + \left(\frac{13}{48}\right)(7) = \frac{133}{24},$$
$$\hat{Z}\bar{X}_2 + (1-\hat{Z})\hat{\mu} = \left(\frac{35}{48}\right)(9) + \left(\frac{13}{48}\right)(7) = \frac{203}{24},$$

za osiguranik 1 i 2, respektivno. □

Vratimo se sada generalnijoj Bulman-Štraubovoj postavci opisanoj ranije u ovom odeljku. Imamo $\mathrm{E}(X_{ij}) = \mathrm{E}[\mathrm{E}(X_{ij}|\Theta_i)] = \mathrm{E}[\mu(\Theta_i)] = \mu$. Dakle,

$$\mathrm{E}(\bar{X}_i|\Theta_i) = \sum_{j=1}^{n_i} \frac{m_{ij}}{m_i}\mathrm{E}(X_{ij}|\Theta_i) = \sum_{j=1}^{n_i} \frac{m_{ij}}{m_i}\mu(\Theta_i) = \mu(\Theta_i),$$

implicira da

$$\mathrm{E}(\bar{X}_i) = \mathrm{E}[\mathrm{E}(\bar{X}_i|\Theta_i)] = \mathrm{E}[\mu(\Theta_i)] = \mu$$

Konačno

$$\mathrm{E}(\bar{X}) = \frac{1}{m}\sum_{i=1}^{r} m_i \mathrm{E}(\bar{X}_i) = \frac{1}{m}\sum_{i=1}^{r} m_i\mu = \mu$$

Pa je očigledno centrirani ocenjivač za μ

$$\hat{\mu} = \bar{X}. \tag{3.9}$$

Da bismo ocenili v i a u Bulman-Štraubovom okviru, potrebna nam je statistika koja je više generalizovana u odnosu na (3.5). Vidimo rešenje primera koji sledi.

■Primer 3.3

Neka su $X_1, \cdots, X_n$ međusobno nezavisne sa jednakim sredinama $\mu = E(X_j)$ i varijansom $Var(X_j) = \beta + \frac{\alpha}{m_j}, \alpha, \beta > 0$ i svako $m_j \geq 1$. Pretpostavljamo da su vrednosti m_j poznate. Neka je $m = \sum_{j=1}^{n} m_j$ i razmotrimo ocenjivače

$$\bar{X} = \frac{1}{m}\sum_{j=1}^{n} m_j X_j \;\; i \;\; \hat{\mu}_1 = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Pokaži da su oba ocenjivača centrirana za μ i uporedi njihove srednjekvadratne greške. Takođe izračunaj očekivanu vrednost sume kvadrata što se može iskoristitu za ocenu α i β.

Rešenje:

Prvo razmotrimo $\bar{X}$:

$$\mathrm{E}(\bar{X}) = \frac{1}{m}\sum_{j=1}^{n} m_j\,\mathrm{E}(X_j) = \frac{1}{m}\sum_{j=1}^{n} m_j\,\mu = \mu,$$
$$\mathrm{Var}(\bar{X}) = \frac{1}{m^2}\sum_{j=1}^{n} {m_j}^2\,\mathrm{Var}(X_j)$$

$$= \frac{1}{m^2}\sum_{j=1}^{n} m_j{}^2\left(\beta + \frac{\alpha}{m_j}\right)$$
$$= \frac{\alpha}{m} + \frac{\beta}{m^2}\sum_{j=1}^{n} m_j{}^2.$$

Za ocenjivača $\hat{\mu}_1$ se lako pokazuje da je centriran. Takođe imamo

$$\mathrm{Var}(\hat{\mu}_1) = \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}(X_j)$$
$$= \frac{1}{n^2}\sum_{j=1}^{n}\left(\beta + \frac{\alpha}{m_j}\right)$$
$$= \frac{\beta}{n} + \frac{\alpha}{n^2}\sum_{j=1}^{n}\frac{1}{m_j}.$$

Sada razmotrimo odnos ove dve varijanse (pošto su oba ocenjivača centrirani, njihove srednjekvadratne greške su jednake njihovim varijansama). Razlika je

$$\mathrm{Var}(\bar{X}) - \mathrm{Var}(\hat{\mu}_1) = \alpha\left(\frac{1}{m} - \frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{m_j}\right) + \beta\left(\frac{1}{m^2}\sum_{j=1}^{n} m_j{}^2 - \frac{1}{n}\right).$$

Koeficijent β je nenegativan. Da bismo to videli, primetimo

$$\frac{1}{n}\sum_{j=1}^{n} m_j{}^2 \geq \left(\frac{1}{n}\sum_{j=1}^{n} m_j\right)^2 = \frac{m^2}{n^2}$$

(leva strana nejednakosti je kao uzorački drugi momenat, a desna strana nejednakosti je kao kvadrat uzoračke sredine) i pomnožimo obe strane sa nm^{-2}. Da bismo pokazali da je α nepozitivno, primetimo

$$\frac{n}{\sum_{j=1}^{n}\frac{1}{m_j}} \leq \frac{1}{n}\sum_{j=1}^{n} m_j = \frac{m}{n}$$

(harmonijska sredina je uvek manja ili jednaka aritmetičkoj sredini), kada pomnožimo obe strane nejednakosti sa n, i onda uzmemo recipročne vrednosti obeju strana. Dakle, sa odgovarajućim izborom za α i β, razlika varijansi može biti pozitivna ili negativna.

S osvrtom na sumu kvadrata, primetimo

$$\sum_{j=1}^{n} m_j(X_j - \bar{X})^2 = \sum_{j=1}^{n} m_j[X_j - \mu + \mu - \bar{X})]^2$$
$$= \sum_{j=1}^{n} m_j\left(X_j - \mu\right)^2 + 2\sum_{j=1}^{n} m_j\left(X_j - \mu\right)(\mu - \bar{X}) + \sum_{j=1}^{n} m_j(\mu - \bar{X})^2$$

$$= \sum_{j=1}^{n} m_j\left(X_j-\mu\right)^2 + 2(\mu-\bar{X})\sum_{j=1}^{n} m_j\left(X_j-\mu\right) + m(\mu-\bar{X})^2$$
$$= \sum_{j=1}^{n} m_j\left(X_j-\mu\right)^2 + 2(\mu-\bar{X})m(\bar{X}-\mu) + m(\mu-\bar{X})^2$$
$$= \sum_{j=1}^{n} m_j\left(X_j-\mu\right)^2 - m(\bar{X}-\mu)^2. \tag{3.10}$$

Kad uzmemo očekivanje, dobijamo

$$\mathrm{E}\left[\sum_{j=1}^{n} m_j\left(X_j-\bar{X}\right)^2\right] = \sum_{j=1}^{n} m_j \mathrm{E}\left[\left(X_j-\mu\right)^2\right] - m\mathrm{E}[(\bar{X}-\mu)^2]$$
$$= \sum_{j=1}^{n} m_j \mathrm{Var}\left(X_j\right) - m\mathrm{Var}(\bar{X})$$
$$= \sum_{j=1}^{n} m_j\left(\beta+\frac{\alpha}{m_j}\right) - \frac{\beta}{m}\sum_{j=1}^{n} {m_j}^2 - \alpha$$

dakle

$$\mathrm{E}\left[\sum_{j=1}^{n} m_j\left(X_j-\bar{X}\right)^2\right] = \beta\left(m-\frac{1}{m}\sum_{j=1}^{n} {m_j}^2\right) + \alpha(n-1). \tag{3.11}$$

Jednačina (3.11) daje centrirani ocenjivač u generalnijim situacijama nego u (3.5), koji dobijamo sa izborom $\alpha = 0$ i $m_j = 1$ za $j = 1, 2, \dots, n$, što implicira da je $m = n$. Takođe, za $\beta = 0$, (3.11) nam dozvoljava da odredimo ocenjivača za α kada je svako X_j aritmetička sredina m_j nezavisnih opservacija sa sredinama μ i varijansama α. U svakom slučaju, m_j (stoga i m) su poznati.

Sada se vratimo na problem ocenjivanja vrednosti υ u Bulman-Štraubovom okviru. Očigledno,
$\mathrm{E}\left(X_{ij}|\Theta_i\right) = \mu(\Theta_i)$ i $\mathrm{Var}\left(X_{ij}|\Theta_i\right) = \upsilon(\Theta_i)/m_{ij}$ za $j = 1, \dots, n_i$. Razmotrimo

$$\hat{\upsilon}_i = \frac{\sum_{j=1}^{n_i} m_{ij}\left(X_{ij}-\bar{X}_i\right)^2}{n_i-1}, i = 1, \dots, r. \tag{3.12}$$

Uslovno za Θ_i , koristeći (3.11) sa $\beta = 0$ i $\alpha = \upsilon(\Theta_i)$, tada je $\mathrm{E}(\hat{\upsilon}_i|\Theta_i) = \upsilon(\Theta_i)$, što implicira da je bezuslovno

$$\mathrm{E}(\hat{\upsilon}_i) = \mathrm{E}[\mathrm{E}(\hat{\upsilon}_i|\Theta_i)] = \mathrm{E}[\upsilon(\Theta_i)] = \upsilon,$$

pa je $\hat{\upsilon}_i$ centrirano za υ za $i = 1, \dots, r$. Drugi centrirani ocenjivač za υ je onda težinska sredina $\hat{\upsilon} = \sum_{i=1}^{r} \omega_i \hat{\upsilon}_i$, gde je $\sum_{i=1}^{r} \omega_i = 1$. Ako težine izaberemo proporcionalno sa to $n_i - 1$, težine uz X_{ij} su m_{ij}. Odnosno sa $\omega_i = \frac{n_i - 1}{\sum_{i=1}^{r} n_i - 1}$, dobijamo centriranog ocenjivača za υ , naime,

$$\hat{\upsilon} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{n_i} m_{ij} \left(X_{ij} - \bar{X}_i\right)^2}{\sum_{i=1}^{r} (n_i - 1)}. \tag{3.13}$$

Sada se vratimo ocenjivanju a. Prisetimo se da su za fiksno i, slučajne promenljive $X_{i1}, \cdots, X_{in_i}$ međusobno nezavisne, uz uslov Θ_i. Dakle,

$$\begin{aligned} \operatorname{Var}(\bar{X}_i | \Theta_i) &= \sum_{j=1}^{n_i} \left(\frac{m_{ij}}{m_i}\right)^2 \operatorname{Var}\left(X_{ij} \middle| \Theta_i\right) = \sum_{j=1}^{n_i} \left(\frac{m_{ij}}{m_i}\right)^2 \frac{\upsilon(\Theta_i)}{m_{ij}} \\ &= \frac{\upsilon(\Theta_i)}{{m_i}^2} \sum_{j=1}^{n_i} m_{ij} = \frac{\upsilon(\Theta_i)}{m_i}. \end{aligned}$$

Tada je bezuslovno,

$$\begin{aligned} \operatorname{Var}(\bar{X}_i) &= \operatorname{Var}[\mathrm{E}(\bar{X}_i | \Theta_i)] + \mathrm{E}[\operatorname{Var}(\bar{X}_i | \Theta_i)] \\ &= \operatorname{Var}[\mu(\Theta_i)] + \mathrm{E}\left[\frac{\upsilon(\Theta_i)}{m_i}\right] = a + \frac{\upsilon}{m_i}. \end{aligned} \tag{3.14}$$

Da sumiramo, , $\bar{X}_1, \dots, \bar{X}_r$ su nezavisni sa jednakim sredinama μ i varijansama $\operatorname{Var}(\bar{X}_i) = a + \frac{\upsilon}{m_i}$. Dalje, $\bar{X} = \frac{1}{m} \sum_{i=1}^{r} m_i \bar{X}_i$. Sada (3.11) možemo opet iskoristiti sa $\beta = a$ i $\alpha = \upsilon$ da bismo dobili

$$\mathrm{E}\left[\sum_{i=1}^{r} m_i (X_i - \bar{X})^2\right] = a\left(m - m^{-1} \sum_{i=1}^{r} {m_i}^2\right) + \upsilon(r-1).$$

Centrirani ocenjivač za a se može dobiti kada umesto υ uvrstimo vrednost njegovog ocenjivača $\hat{\upsilon}$ i rešimo po a. Odnosno, centrirani ocenjivač za a je

$$\hat{a} = \left(m - m^{-1} \sum_{i=1}^{r} {m_i}^2\right)^{-1} \left[\sum_{i=1}^{r} m_i (X_i - \bar{X})^2 - \hat{\upsilon}(r-1)\right] \tag{3.15}$$

sa $\hat{\upsilon}$ datim u (3.13).

Jednačine (3.9), (3.13) i (3.15) daju centrirane ocenjivače za vrednosti μ, υ, i a, respektivno. Oni su neparametarski i ne zahtevaju pretpostavke u vezi sa raspodelom slučajnih promenljivih. Sigurno, ovo nisu jedini (centrirani) ocenjivači koji mogu da se odrede, i moguće je da $\hat{a} < 0$. U tom slučaju, a je vrlo verovatno blizu nule, i ima smisla da se postavi $\hat{Z} = 0$. Dalje, Bulmanovi ocenjivači iz Primera 3.1 se dobijaju sa $m_{ij} = 1$ i $n_i = n$. Konačno, ovi ocenjivači su esencijalno ocenjivači maksimalne verodostojnosti u slučaju gde su $X_{ij} | \Theta_i$ i Θ_{i_i} normalno raspodeljeni, stoga ocenjivači imaju dobre statističke osobine.

Pri korišćenju razvijenih formula, javlja se jedan problem. U prošlosti, štete i-tog osiguranika su sakupljene uz izloženost od m_i. Ukupna istorijska šteta (Total Loss) za sve osiguranike je TL=$\sum_{i=1}^{r} m_i\bar{X}_i$. Da smo naplatili kredibilnosnu premiju kao što je prethodno dato, ukupna premija (Total Premium) bi bila

$$\begin{aligned}\mathrm{TP} &= \sum_{i=1}^{r} m_i[\hat{Z}_i\bar{X}_i + (1-\hat{Z}_i)\hat{\mu}] \\ &= \sum_{i=1}^{r} m_i(1-\hat{Z}_i)(\hat{\mu}-\bar{X}_i) + \sum_{i=1}^{r} m_i\bar{X}_i \\ &= \sum_{i=1}^{r} m_i\frac{\hat{k}}{m_i+\hat{k}}(\hat{\mu}-\bar{X}_i) + \sum_{i=1}^{r} m_i\bar{X}_i.\end{aligned}$$

Često je poželjno da *TL* bude jednako *TP*, zato što bilo koje povećanje premije, da bi dobilo odobrenje regulatora, mora da se bazira na ukupnom iznosu zahteva za odštetu iz prošlog perioda. Iako prilagođeni kredibilitet ima smisla, što praktičnog, što teorijskog, takođe je dobra ideja da se total održi nepromenjen. Dakle, treba da bude

$$0 = \sum_{i=1}^{r} m_i\frac{\hat{k}}{m_i+\hat{k}}(\hat{\mu}-\bar{X}_i)$$

ili

$$\hat{\mu}\sum_{i=1}^{r}\hat{Z}_i = \sum_{i=1}^{r}\hat{Z}_i\bar{X}_i$$

ili

$$\hat{\mu} = \frac{\sum_{i=1}^{r}\hat{Z}_i\bar{X}_i}{\sum_{i=1}^{r}\hat{Z}_i}. \tag{3.16}$$

Odnosno, umesto da koristimo (3.9) da bismo izračunali $\hat{\mu}$, koristimo kredibilnosnu težinsku sredinu individualnih uzoračkih sredina. Oba metoda proizvode centrirani ocenjivač (za date $\hat{Z}_i$), ali ovaj drugi ima prednost što očuvava vrednost ukupnih zahteva za odštetu. Treba primetiti kada koristimo (3.15), vrednost $\bar{X}$ iz (3.1) i dalje treba koristiti, što se može zaključiti iz argumenata sa najmanjim kvadratima. Konačno, iz Primera 3.3 i imajući u vidu oblik varijanse $\mathrm{Var}(\bar{X}_j)$ iz (3.14), težine u (3.16) daju najmanju bezuslovnu varijansu za $\hat{\mu}$.

■Primer 3.4

U Tabeli 3.1 su dati istorijski podaci za dve grupe osiguranika. Odredi ocenjene kredibilnosne premije koje treba naplatiti svakoj grupi u četvrtoj godini.

	Grupa	Godina 1	Godina 2	Godina 3	Godina 4
Ukupna šteta	1	-	10,000	13,000	-
Broj osiguranika		-	50	60	75

u grupi					
Ukupna šteta	2	18,000	21,000	17,000	-
Broj osiguranika u grupi		100	110	105	90

Table 3.1 Podaci za primer 3.4

Rešenje:

Prvo moramo da odredimo u svakoj grupi prosečan broj zahteva po osiguraniku za svaku prethodnu godinu. Imamo $n_1 = 2$ godine iskustva za grupu 1 i $n_2 = 3$ za grupu 2. Nije bitno iz koje godine su opservacije za grupu 1, pa za notaciju biramo

$$m_{11} = 50 \text{ i } X_{11} = \frac{10{,}000}{50} = 200.$$

Slično,

$$m_{12} = 60 \text{ i } X_{12} = \frac{13{,}000}{60} = 216.67.$$

Tada je

$$m_1 = m_{11} + m_{12} = 50 + 60 = 110,$$

$$\bar{X}_1 = \frac{10{,}000 + 13{,}000}{110} = 209.09.$$

Za grupu 2,

$$m_{21} = 100 \text{ i } X_{21} = \frac{18{,}000}{100} = 180,$$
$$m_{22} = 110 \text{ i } X_{21} = \frac{21{,}000}{110} = 190.91,$$
$$m_{23} = 105 \text{ i } X_{23} = \frac{17{,}000}{105} = 161.90.$$

Tada je

$$m_2 = m_{21} + m_{22} + m_{23} = 100 + 110 + 105 = 315,$$

$$\bar{X}_2 = \frac{18{,}000 + 21{,}000 + 17{,}000}{315} = 177.78.$$

Sada, $m_1 + m_2 = 110 + 315 = 425$. Ukupna sredina je

$$\hat{\mu} = \bar{X} = \frac{10{,}000 + 13{,}000 + 18{,}000 + 21{,}000 + 17{,}000}{425} = 185.88.$$

Alternativnu ocenu (3.16) za μ ćemo izračunati kasnije. Sada,

$$\hat{v} = \frac{\begin{array}{c}50(200 - 209.09)^2 + 60(216.67 - 209.09)^2 + 100(180 - 177.78)^2 \\ +110(190.91 - 177.78)^2 + 105(161.90 - 177.78)^2\end{array}}{(2-1) + (3-1)}$$

$$= 17{,}837.87,$$

pa je

$$\hat{a} = \frac{110(209.09 - 185.88)^2 + 315(177.78 - 185.88)^2 - (17{,}837.87)(1)}{425 - (110^2 + 315^2)/425}$$
$$= 380.76.$$

Tada je $\hat{k} = \frac{\hat{v}}{\hat{a}} = 46.85$. Ocenjeni faktori kredibiliteta za ove dve grupe su

$$\hat{Z}_1 = \frac{110}{110 + 46.85} = 0.70, \qquad \hat{Z}_2 = \frac{315}{315 + 46.85} = 0.87.$$

Po osiguraniku, ocenjena kredibilnosna premija za grupu 1 je

$$\hat{Z}_1\bar{X}_1 + (1 - \hat{Z}_1)\hat{\mu} = (0.70)(209.09) + (0.30)(185.88) = 202.13,$$

pa je ukupna ocenjena kredibilnosna premija za celu grupu

$$75(202.13) = 15{,}159.75.$$

Za grupu 2

$$\hat{Z}_2\bar{X}_2 + (1 - \hat{Z}_2)\hat{\mu} = (0.87)(177.78) + (0.13)(185.88) = 178.83,$$

a je ukupna ocenjena kredibilnosna premija za celu grupu

$$90(178.83) = 16{,}094.70.$$

Za alternativnog ocenjivača bismo koristili

$$\hat{\mu} = \frac{0.70(209.09) + 0.87(177.78)}{0.70 + 0.87} = 191.74.$$

Kredibilnosne premije su

$$(0.70)(209.09) + (0.30)(191.74) = 203.89,$$
$$(0.87)(177.78) + (0.13)(191.74) = 179.59.$$

Ukupna istorijska kredibilnosna premija je 110(203.89)+315(179.59)=78,998.75. Osim greške zaokruživanja, ovaj total se poklapa sa ukupnim štetama 79,000. □

U nastavku analize pretpostavljamo da su parametri μ, v i a nepoznati i potrebno ih je oceniti, što ne mora uvek biti slučaj. Takođe, neka je $n_i > 1$ i $r > 1$. Ako je $n_i = 1$, tako da postoji samo jedna jedinica izloženosti riziku za iskustvo osiguranika i, onda je teško dobiti informacije o procesnoj varijansi $v(\Theta_i)$, stoga i v. Slično, ako je $r = 1$, postoji samo jedan osiguranik i teško je dobiti informaciju o varijansi hipotetičkih sredina a. U ovim situacijama su potrebne jače pretpostavke, kao što je znanje o jednom ili više parametara (npr. neto premija ili manualna premija μ) ili parametarska pretpostavka koja implicira funkcionalnu zavisnost između parametara.

Da bismo ilustrovali ove ideje, pretpostavimo da je npr. manualna premija μ poznata, ali je potrebno oceniti a i v. U ovom slučaju (3.13) se može koristiti za ocenu v pošto je centrirana, bez obzira na to da li je μ poznata ili ne. Slično, (3.15) je i dalje centrirani ocenjivač za a. Međutim, ako je μ poznato, alternativni centrirani ocenjivač za a je

$$\tilde{a} = \sum_{i=1}^{r} \frac{m_i}{m} (\bar{X}_i - \mu)^2 - \frac{r}{m} \hat{v},$$

gde je $\hat{v}$ dobijeno iz (3.13). Da bismo potvrdili centriranost, primetimo

$$\begin{aligned} \mathrm{E}(\tilde{a}) &= \sum_{i=1}^{r} \frac{m_i}{m} \mathrm{E}[(\bar{X}_i - \mu)^2] - \frac{r}{m} \mathrm{E}(\hat{v}) \\ &= \sum_{i=1}^{r} \frac{m_i}{m} \mathrm{Var}(\bar{X}_i) - \frac{r}{m} v \\ &= \sum_{i=1}^{r} \frac{m_i}{m} \left(a + \frac{v}{m_i}\right) - \frac{r}{m} v = a. \end{aligned}$$

Ako postoje opservacije samo za jednog osiguranika, neophodno je koristiti ovakav pristup. Očigledno (3.12) daje ocenjivača za v samo na osnovu istorijskih podataka od osiguranika *i*, i centrirani ocenjivač za *a* samo na osnovu istorijskih podataka od osiguranika *i* je

$$\tilde{a}_i = (\bar{X}_i - \mu)^2 - \frac{\hat{v}_i}{m_i} = (\bar{X}_i - \mu)^2 - \frac{\sum_{j=1}^{n_i} m_{ij} \left(X_{ij} - \bar{X}_i\right)^2}{m_i (n_i - 1)},$$

što je centrirano zato što je $\mathrm{E}[(\bar{X}_i - \mu)^2] = \mathrm{Var}(\bar{X}_i) = a + \frac{v}{m_i}$ i $\mathrm{E}(\hat{v}_i) = v$.

■Primer 3.5

Za grupu osiguranika, istorijski podaci su dati u Tabeli 3.2. Ako je manualna premija po osobi 500 na godišnjem nivou, oceni ukupnu kredibilnosnu premiju za treću godinu.

	Godina 1	**Godina 2**	**Godina 3**
Ukupna šteta	60,000	70,000	-
Broj osiguranika u grupi	125	150	200

Table 3.2 Podaci za Primer 3.5

Rešenje:

Imamo (radi lakše notacije pretpostavićemo da je ova grupa osiguranik i, $m_{i1} = 125$, $X_{i1} = 60{,}000/125 = 480$, $m_{i2} = 150$, $X_{i2} = \frac{70{,}000}{150} = 466.67$, $m_i = m_{i1} + m_{i2} = 275$, i $\bar{X}_i = (60{,}000 + 70{,}000)/275 = 472.73$. Tada je

$$\hat{v}_i = \frac{125(480 - 472.73)^2 + 150(466.67 - 472.73)^2}{2 - 1} = 12{,}115.15,$$

i, sa $\mu = 500$. $\tilde{a}_i = (472.73 - 500)^2 - (12{,}115.15/275) = 699.60$. Sada ocenimo k sa $\hat{v}_i/\tilde{a}_i = 17.32$. Ocenjeni faktor kredibiliteta je $m_i/(m_i + \hat{v}_i/\tilde{a}_i) = 275/(275 + 17.32) = 0.94$.

Ocenjena kredibilnosna premija po osobi je onda 0.94(472.73) + 0.06(500) = 474.37, a ocenjena ukupna kredibilnosna premija za godinu 3 je 200(474.37) = 94.874. □

Treba napomenuti da se ocenjivanje parametara a i υ, na osnovu podataka samo jednog osiguranika (kao što je slučaj u Primeru 3.5) , ne savetuje, osim ako ne postoji alternativa zato što ocenjivači $\hat{\upsilon}_i$ i $\tilde{a}_i$ imaju visoku varijabilnost. Naročito, mi efektivno ocenjujemo a na osnovu jedne opservacije ($\bar{X}_i$). U ovakvim situacijama se snažno sugeriše da se pokuša sa prikupljanjem više informacija.

3.2. Poluparametarsko ocenjivanje

U nekim situacijama može biti razumno pretpostaviti parametarski oblik za uslovnu raspodelu $f_{X_{ij}|\Theta}(x_{ij}|\theta_i)$. Posmatrana situacija može sugerisati da je takva pretpostavka razumna ili priorne informacije mogu implicirati njenu adekvatnost.

Na primer, kada posmatramo brojeve zahteva za odštetu, može biti razumno pretpostaviti da broj zahteva $m_{ij}X_{ij}$ za osiguranika i u godini j ima Poasonovu raspodelu sa sredinom $m_{ij}\Theta_i$ za dato $\Theta_i = \theta_i$. Dakle $\mathrm{E}(m_{ij}X_{ij}|\Theta_i) = \mathrm{Var}(m_{ij}X_{ij}|\Theta_i) = m_{ij}\Theta_i$, što implicira da je $\mu(\Theta_i) = \upsilon(\Theta_i) = \Theta_i$ pa je u ovom slučaju $\mu = \upsilon$. Umesto da koristimo (3.13) za ocenu vrednosti υ, mogli bismo koristiti $\hat{\mu} = \bar{X}$ za ocenu vrednosti υ .

■Primer 3.6

U prošloj godini, distribucija osiguranika sa polisom auto osiguranja je data u Tabeli 3.3 sa brojem zahteva za odštetu. Za svakog osiguranika, na osnovu istorijskog iskustva odredi kredibilnosnu ocenu za broj zahteva sledeće godine, uz pretpostavku da broj zahteva za svakog osiguranika ima Poasonovu raspodelu.

Broj zahteva	Broj osiguranika
0	1,563
1	271
2	32
3	7
4	2
Ukupno	**1,875**

Table 3.3 Podaci za Primer 3.6

Rešenje:

Pretpostavimo da imamo $r = 1{,}875$ osiguranika, $n_i = 1$ po jednu godinu iskustva za svakog, i izloženost $m_{ij} = 1$. Za osiguranika i (gde $i = 1, \dots, 1{,}875$), neka $X_{ij}|\Theta_i = \theta_i$ ima Poasonovu raspodelu sa sredinom θ_i tako da je $\mu(\Theta_i) = \upsilon(\Theta_i) = \Theta_i$ i $\mu = \upsilon$. Kao u Primeru 3.1,

$$\bar{X} = \frac{1}{1{,}875}\left(\sum_{i=1}^{1{,}875} X_{i1}\right)$$
$$= \frac{0(1{,}563) + 1(271) + 2(32) + 3(7) + 4(2)}{1{,}875} = 0.194.$$

Sada je

$$\begin{aligned}\mathrm{Var}(X_{i1}) &= \mathrm{Var}[\mathrm{E}(X_{i1}|\Theta_i)] + \mathrm{E}[\mathrm{Var}(X_{i1}|\Theta_i)] \\ &= \mathrm{Var}[\mu(\Theta_i)] + \mathrm{E}[\upsilon(\Theta_i)] = a + \upsilon = a + \mu.\end{aligned}$$

Dakle, centrirana ocena za $a + \upsilon$ je uzoračka varijansa

$$\frac{\sum_{i=1}^{1{,}875}(X_{i1} - \bar{X})^2}{1{,}874} = \frac{\begin{array}{c}1{,}563(0-0.194)^2 + 271(1-0.194)^2 \\ +32(2-0.194)^2 + 7(3-0.194)^2 + 2(4-0.194)^2\end{array}}{1{,}874}$$
$$= 0.226.$$

Posledično, $\hat{a} = 0.226 - 0.194 = 0.032$ i $\hat{k} = 0.194/0.032 = 6.06$ i faktor kredibiliteta Z je $1/(1 + 6.06) = 0.14$. Ocenjena kredibilnosna premija za broj zahteva po osiguraniku je $(0.14)X_{i1} + (0.86)(0.194)$, gde je X_{i1} jednako 0, 1, 2, 3, ili 4, u zavisnosti od osiguranika. □

Primetimo da je u ovom slučaju $\mu = \upsilon$, tako da je potrebna samo jedna godina iskustva po osiguraniku.

■Primer 3.7

Neka se posmatra verovatnoća da individua u grupi podnese zahtev za odštetu (npr. u pitanju je grupno životno osiguranje), i pretpostavimo da verovatnoća varira od jednog do drugog osiguranika. Tada $m_{ij}X_{ij}$ može da predstavlja broj m_{ij} individua u godini j za osiguranika i koji je podneo zahtev. Razvij model kredibiliteta za ovu situaciju.

Rešenje:

Ako je verovatnoća nastanka štete Θ_i za osiguranika i, tada je razuman model za opisivanje ovog efekta $m_{ij}X_{ij}$ sa binomnom raspodelom sa parametrima m_{ij} i θ_i za dato $\Theta_i = \theta_i$. Tada je

$$\mathrm{E}\left(m_{ij}X_{ij}\middle|\Theta_i\right) = m_{ij}\Theta_i \quad \text{i} \quad \mathrm{Var}\left(m_{ij}X_{ij}\middle|\Theta_i\right) = m_{ij}\Theta_i(1-\Theta_i),$$

pa je $\mu(\Theta_i) = \Theta_i$ sa $\upsilon(\Theta_i) = \Theta_i(1 - \Theta_i)$. Dakle

$$\mu = \mathrm{E}[\Theta_i], \qquad \upsilon = \mu - \mathrm{E}[(\Theta_i)^2],$$

i

$$a = \mathrm{Var}[\Theta_i] = \mathrm{E}[(\Theta_i)^2] - \mu^2 = \mu - \upsilon - \mu^2.$$ □

U ovom primeru postoji funkcionalna veza između parametara μ, υ, i a koja sledi iz parametarske pretpostavke, što često fasilituje ocenjivanje parametara.

3.3. Parametarsko ocenjivanje

Ako imamo potpuno parametarske pretpostavke u vezi sa raspodelama $f_{X_{ij}|\Theta}\left(x_{ij}\middle|\theta_i\right)$ i $\pi(\theta_i)$ za $i=1,\dots,r$ i $j=1,\dots,n_i$, tada nam je na raspolaganju ceo set tehnika ocenjivanja parametara, pored već razmatranih neparametarskih metoda. Sada ćemo razmatrati ocenu maksimalne verodostojnosti. Za svakog osiguranika i, zajednička raspodela za $\boldsymbol{X}_i=(X_{i1},\cdots,X_{in_i})^T$, za uslovno Θ_i, za $i=1,\dots,r$, data je sa

$$f_{\mathbf{X}_i}(\mathbf{x}_i)=\int\left[\prod_{j=1}^{n_i}f_{X_{ij}|\Theta}\left(x_{ij}\middle|\theta_i\right)\right]\pi(\theta_i)d\theta_i. \tag{3.17}$$

Funkcija verodostojnosti je data sa

$$L=\prod_{i=1}^{r}f_{\mathbf{X}_i}(\mathbf{x}_i). \tag{3.18}$$

Ocenjivači maksimalne verodostojnosti za parametre se tada biraju tako da maksimizuju L ili ekvivalentno ln$L$.

■Primer 3.8

Neka je $n_i=n$ za $i=1,\dots,r$ i $m_{ij}=1$. Neka $X_{ij}|\Theta_i$ ima Poasonovu raspodelu sa sredinom Θ_i, odnosno

$$f_{X_{ij}|\Theta}\left(x_{ij}\middle|\theta_i\right)=\frac{{\theta_i}^{x_{ij}}e^{-\theta_i}}{x_{ij}!},\qquad x_{ij}=0,1,\dots,$$

i neka Θ_i ima eksponencijalnu raspodelu sa sredinom μ,

$$\pi(\theta_i)=\frac{1}{\mu}e^{-\theta_i/\mu},\qquad \theta_i>0.$$

Odredi ocenjivač maksimalne verodostojnosti za μ.

Rešenje:

Jednačina (3.17) postaje

$$\begin{aligned}f_{\mathbf{X}_i}(\mathbf{x}_i)&=\int_0^\infty\left(\prod_{j=1}^{n}\frac{{\theta_i}^{x_{ij}}e^{-\theta_i}}{x_{ij}!}\right)\frac{1}{\mu}e^{-\frac{\theta_i}{\mu}}d\theta_i\\&=\left(\prod_{j=1}^{n}x_{ij}!\right)^{-1}\frac{1}{\mu}\int_0^\infty{\theta_i}^{\sum_{j=1}^{n}x_{ij}}e^{-\theta_i\left(n+\frac{1}{\mu}\right)}d\theta_i\\&=C(\mathbf{x}_i)\frac{1}{\mu}\left(n+\frac{1}{\mu}\right)^{-\sum_{j=1}^{n}x_{ij}-1}\int_0^\infty\frac{\beta(\beta\theta_i)^{\alpha-1}e^{-\beta\theta_i}}{\Gamma(\alpha)}d\theta_i,\end{aligned}$$

gde se $C(\mathbf{x}_i)$ može izraziti u kombinatorijskom obliku kao

$$C(\mathbf{x}_i)=\binom{\sum_{j=1}^{n}x_{ij}}{x_{i1}\ x_{i2}\cdots x_{in}};$$

$$\beta=n+\frac{1}{\mu},$$

i

$$\alpha = \sum_{j=1}^{n} x_{ij} + 1.$$

Prepoznajemo da je integral sa gama gustinom sa parametrima α i $1/\beta$, stoga je jednak 1, pa je

$$f(\mathbf{x}_i) = C(\mathbf{x}_i)\frac{1}{\mu}\left(n + \frac{1}{\mu}\right)^{-\sum_{j=1}^{n} x_{ij} - 1}.$$

Ako to uvrstimo u (3.18) dobijamo

$$L(\mu) \propto \mu^{-r}\left(n + \frac{1}{\mu}\right)^{-\sum_{i=1}^{r}\sum_{j=1}^{n} x_{ij} - r}.$$

Dakle,

$$l(\mu) = \ln L(\mu) = -r\ln\mu - \left(r + \sum_{i=1}^{r}\sum_{j=1}^{n} x_{ij}\right)\ln\left(n + \frac{1}{\mu}\right) + c,$$

gde je c konstanta koja ne zavisi od μ. Diferenciranjem dobijamo

$$l'(\mu) = -\frac{r}{\mu} - \frac{r + \sum_{i=1}^{r}\sum_{j=1}^{n} x_{ij}}{n + \frac{1}{\mu}}\left(-\frac{1}{\mu^2}\right).$$

Ocenjivač maksimalne verodostojnosti $\hat{\mu}$ za μ se dobija kada je $l'(\hat{\mu}\,) = 0$, odakle se dobija

$$\frac{r}{\hat{\mu}} = \frac{r + \sum_{i=1}^{r}\sum_{j=1}^{n} x_{ij}}{\hat{\mu}(\hat{\mu}n + 1)}$$

pa je

$$\hat{\mu}n + 1 = 1 + \frac{1}{r}\sum_{i=1}^{r}\sum_{j=1}^{n} x_{ij}$$

ili

$$\hat{\mu} = \frac{1}{nr}\sum_{i=1}^{r}\sum_{j=1}^{n} x_{ij}.$$

Ali ovo je isto kao kod neparametarske ocene dobijene u Primeru 3.1. To se desilo zato što imamo da je $\mu(\theta_i) = \theta_i$ na osnovu Poasonove pretpostavke i $\mathrm{E}[\mu(\Theta_i)] = \mathrm{E}[\Theta_i]$, koje je isto μ koje smo koristili u eksponencijalnoj raspodeli $\pi(\theta_i)$.

Dalje, takođe je i $\upsilon(\theta_i) = \theta_i$ (na osnovu Poasonove raspodele), pa je $\upsilon = \mathrm{E}[\upsilon(\Theta_i)] = \mu$. Takođe, $a = \mathrm{Var}[\mu(\Theta_i)] = \mathrm{Var}[\Theta_i] = \mu^2$ na osnovu eksponencijalne pretpostavke za $\pi(\theta_i)$. Stoga su za υ i a ,ocenjivači maksimalne verodostojnosti $\hat{\mu}$ i $\hat{\mu}^2$, usled invarijantnosti ocena maksimalne verodostojnosti u odnosu na transformacije parametara. Slično, ocenjivači maksimalne verodostojnosti za $k = \frac{\upsilon}{a}$, faktor kredibiliteta Z, i kredibilnosnu premiju $Z\bar{X}_i + (1 - Z)\mu$ su $\hat{k} = \hat{\mu}^{-1} = \bar{X}^{-1}$, $\hat{Z} = n/(n + \hat{\mu}^{-1})$, i $\hat{Z}\bar{X}_i + (1 - \hat{Z})\hat{\mu}$, respektivno. Još napominjemo da je kredibilitet tačan za ovaj model, tako da je Bejzova premija jednaka kredibilnosnoj premiji. □

■Primer 3.9

Neka je $n_i = n$ za sve i i $m_{ij} = 1$. Pretpostavimo da je $X_{ij}|\Theta_i \sim N(\Theta_i, \upsilon)$,

$$f_{X_{ij}|\Theta}\left(x_{ij}|\theta_i\right) = (2\pi\upsilon)^{-1/2}\exp\left[-\frac{1}{2\upsilon}\left(x_{ij}-\theta_i\right)^2\right], -\infty < x_{ij} < \infty,$$

i $\Theta_i \sim N(\mu, a)$, *tako da je*

$$\pi(\theta_i) = (2\pi a)^{-1/2}\exp\left[-\frac{1}{2a}(\theta_i-\mu)^2\right], -\infty < \theta_i < \infty.$$

Odredi ocenjivače maksimalne verodostojnosti za parametre.

Rešenje:

Imamo $\mu(\theta_i) = \theta_i$ i $\upsilon(\theta_i) = \theta_i$. Dakle, $\mu = \mathrm{E}[\mu(\Theta_i)]$, $\upsilon = \mathrm{E}[\upsilon(\Theta_i)]$ i $a = \mathrm{Var}[\mu(\Theta_i)]$, su konzistentni sa prethodnom upotrebom μ, υ i a. Sada treba da odredimo ocenjivače maksimalne verodostojnosti za parametre μ, υ i a. Posmatrajmo sada $\bar{X}_i = n^{-1}\sum_{j=1}^{n} X_{ij}$. Uslovno za Θ_i, X_{ij} su međusobno nezavisne slučajne promenljive sa raspodelom $N(\Theta_i, \upsilon)$, što implicira da je $\bar{X}_i|\Theta_i \sim N\left(\Theta_i, \frac{\upsilon}{n}\right)$. Pošto je $\Theta_i \sim N(\mu, a)$, može se pokazati da je bezuslovno $\bar{X}_i \sim N\left(\mu, a + \frac{\upsilon}{n}\right)$. Stoga je gustina za $\bar{X}_i$, sa $\omega = a + \frac{\upsilon}{n}$,

$$f(\bar{x}_i) = (2\pi\omega)^{-1/2}\exp\left[-\frac{1}{2\omega}(\bar{x}_i-\mu)^2\right], \qquad -\infty < \bar{x}_i < \infty.$$

S druge strane, za uslovljeno sa Θ_i, imamo

$$f(\bar{x}_i) = \int_{-\infty}^{\infty}(2\pi\upsilon)^{-1/2}\exp\left[-\frac{n}{2\upsilon}\left(x_{ij}-\theta_i\right)^2\right](2\pi a)^{-1/2}\exp\left[-\frac{1}{2a}(\theta_i-\mu)^2\right]d\theta_i,$$

Ignorišući izraze koji ne sadrže μ, υ ili a, $f(\bar{x}_i)$ je proporcionalno sa

$$f(\bar{x}_i) = \int_{-\infty}^{\infty}\left\{\prod_{j=1}^{n}(2\pi\upsilon)^{-1/2}\exp\left[-\frac{1}{2\upsilon}\left(x_{ij}-\theta_i\right)^2\right]\right\}(2\pi a)^{-1/2}\exp\left[-\frac{1}{2a}(\theta_i-\mu)^2\right]d\theta_i,$$

što je proporcionalno sa

$$\upsilon^{-n/2}a^{-1/2}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\upsilon}\sum_{j=1}^{n}\left(x_{ij}-\theta_i\right)^2-\frac{1}{2a}(\theta_i-\mu)^2\right]d\theta_i.$$

Koristeći izraz iz (3.4) drugačije zapisujemo

$$\sum_{j=1}^{n}\left(x_{ij}-\theta_i\right)^2 = \sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2 + n(\bar{x}_i-\theta_i)^2,$$

što znači da je $f(\mathbf{x}_i)$ proporcionalno sa

$$\upsilon^{-n/2}a^{-1/2}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2\upsilon}\left[\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2+n(\bar{x}_i-\theta_i)^2\right]-\frac{1}{2a}(\theta_i-\mu)^2\right\}d\theta_i,$$

a to je proporcionalno sa

$$\upsilon^{-(n-1)/2}\exp\left[-\frac{1}{2\upsilon}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2\right]f(\bar{x}_i)$$

koristeći drugi izraz dat gore za gustinu $f(\bar{x}_i)$ od $\bar{X}_i$. Tada (3.18) daje

$$L \propto \upsilon^{-\frac{r(n-1)}{2}} \exp\left[-\frac{1}{2\upsilon}\sum_{i=1}^{r}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2\right]\prod_{i=1}^{r} f(\bar{x}_i).$$

Sada ćemo predstaviti invarijantnost ocenjivača maksimalne verodostojnosti za transformacije parametara i koristićemo μ, υ i $\omega = a + \frac{\upsilon}{n}$ umesto μ, υ i a. Ovo znači da je

$$L \propto L_1(\upsilon)L_2(\mu,\omega),$$

gde je

$$L_1(\upsilon) = \upsilon^{-\frac{r(n-1)}{2}} \exp\left[-\frac{1}{2\upsilon}\sum_{i=1}^{r}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2\right]$$

a

$$L_2(\mu,\omega) = \prod_{i=1}^{r} f(\bar{x}_i) = \prod_{i=1}^{r}\left\{(2\pi\omega)^{-1/2}\exp\left[-\frac{1}{2\omega}(\bar{x}_i-\mu)^2\right]\right\}.$$

Ocenjivač maksimalne verodostojnosti $\hat{\upsilon}$ za υ se dobija maksimizacijom $L_1(\upsilon)$, a ocenjivač maksimalne verodostojnosti $(\hat{\mu},\hat{\omega})$za (μ,ω) se dobija maksimizacijom $L_2(\mu,\omega)$. Uzimajući logaritme, dobijamo

$$l_1(\upsilon) = -\frac{r(n-1)}{2}\ln\upsilon - \frac{1}{2\upsilon}\sum_{i=1}^{r}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2,$$

$$l'_1(\upsilon) = -\frac{r(n-1)}{2\upsilon} + \frac{1}{2\upsilon^2}\sum_{i=1}^{r}\sum_{j=1}^{n}\left(x_{ij}-\bar{x}_i\right)^2,$$

i sa $l'(\hat{\mu}) = 0$ dobijamo

$$\hat{\upsilon} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n}\left(X_{ij}-\bar{X}_i\right)^2}{r(n-1)}.$$

Pošto je $L_2(\mu,\omega)$ uobičajena normalna verodostojnost, ocenjivači maksimalne verodostojnosti su jednostavno empirijska sredina i varijansa. Odnosno,

$$\hat{\mu} = r^{-1}\sum_{i=1}^{r}\bar{X}_i = (rn)^{-1}\sum_{i=1}^{r}\sum_{j=1}^{n}X_{ij} = \bar{X}$$

i

$$\hat{\omega} = \frac{1}{r}\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2.$$

Ali $a = \omega - \frac{\upsilon}{n}$, pa je ocena maksimalne verodostojnosti za a

$$\hat{a} = \frac{1}{r}\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2 - \frac{\sum_{i=1}^{r}\sum_{j=1}^{n}\left(X_{ij}-\bar{X}_i\right)^2}{rn(n-1)}.$$

Treba napomenuti da su dobijeni ocenjivači maksimalne verodostojnosti $\hat{\mu}$ i $\hat{\upsilon}$ su jednaki sa neparametarskim centriranim ocenjivačima u Bulmanovom modelu u Primeru 3.1. Ocenjivač maksimalne verodostojnosti $\hat{a}$ je skoro isti kao neparametarski centrirani ocenjivač, jedina razlika je u deliocu, *r* umesto *r*-1 u prvom izrazu. □

Zaključak

> „...poslovanje se i dalje suočava sa velikim brojem problema, od kojih će mnoge, znamo, aktuari konačno morati da reše. Pustimo ga -aktuara neživotnog osiguranja o kome sam do sada pričao- da nastavi da se rva sa ovim problemima, znajući dobro da ima ogromnu prednost u posedovanju naučnog uma i naučnih metoda; sa ovima on će, sa svojim zaslugama, biti pozvan da igra veću i najodgovorniju ulogu u poslovanju neživotnog osiguranja.“.

Francis S. Perrymana[67]

Po mišljenju velikog broja statističara, postoje dva međusobno isključujuća pristupa analizi podataka. „Klasična“ ili „frekventistička“ teorija, koja se sastoji od intervala poverenja i testiranja hipoteza, se najviše koristi i najrasprostranjenija je u tipičnim statističkim tekstovima. S druge strane, „Bejzova“ statistika, analiza zasnovana na Bejzovoj teoremi, privukla je malu grupu strastvenih pristalica.

Prema Bejzovoj statistici pojam verovatnoće se tumači na subjektivan način, gde verovatnoća predstavlja stepen uverenja koji se koriguje u trenutku kada informacije ili podaci postanu raspoloživi. Bejzov pristup pretpostavlja da su samo trenutno posmatrani podaci relevantni za analizu i da je raspodela populacije ta koja je promenljiva. Ovaj pristup odlikuje upotreba priornih informacija i izražava verodostojnost u vidu stepena verovanja(uverenja) da će se dati događaj realizovati. Priorne informacije postoje u izobilju i potiču iz prethodnih studija, objavljenih radova, intuicije istraživača, mišljenja nezavisnih eksperata, iz pogodnosti, srodnosti, neodređenosti itd.

Do početka devedesetih godina prošlog veka, Bejzovi metodi nisu bili toliko vidljivi zbog poteškoća u računanju vrednosti višestrukih integrala. Međutim, primena ovih metoda je svakako bila veoma rasprostranjena, npr. u oblastima koje se bave kliničkim ispitivanjem metodologija, rezultata, ekonomskom evaluacijom troškova, donošenjem odluka u kliničkoj praksi, lekovima, meta-analizom, sistemom prismotre, ekranizacijom i mapiranjem bolesti i metodama preživljavanja.[68] Treba još pomenuti i primenu u analizi neuronskih mreža (uz upotrebu brain-imaging tehnologije) i njihovim vezama sa načinom ponašanja i formiranjem ličnosti.

Bejzova paradigma se već dugi niz godina smatra najpogodnijom za primenu u analizi različitih modela iz oblasti aktuarstva, osiguranja i upravljanja rizicima. Po ugledu na aktuarska društva, koja su tokom vremena koristila Bejzov i Bulmanov kredibilitet za optimizaciju

[67] Presidential Address-CAS XXV, p. 291 (1939)
[68] Diskusija Rozenberga (Marjorie A. Rosenberg), profesora sa Univerziteta u Viskonsinu u SAD (University of Wisconsin). Vidi [14].

preciznog određivanja premija, i finansijski analitičari su počeli da koriste ovakav kredibilitet kako bi optimizovali svoje procene beta koeficijenata i kretanja akcija.[69]

Ono što se najviše zamera Bejzovom pristupu, jeste suštinski i njegova najveća prednost. Glavni aspekt Bejzove analize koji se najčešće kritikuje je odabir priorne raspodele, sa prigovorom da lična priroda ove raspodele uklanja auru objektivnosti, koja se očekuje oko naučnog istraživanja.

Teorija kredibiliteta je jedan od temelja aktuarske nauke, bez kog moderno osiguranje ne bi postojalo u obliku u kom ga danas poznajemo. Teorija kredibiliteta obezbeđuje set kvantitativnih alata, koji omogućuju osiguravaču da se izbori sa slučajnošću podataka, koji se koriste za predviđanje budućih događaja ili šteta. Drugim rečima teorija omogućuje osiguravaču da, na osnovu istorijskih podataka, koriguje buduću premiju za određeni rizik ili riziko grupu.

Predstavljena je *teorija kredibiliteta ograničene fluktuacije*, koja se drugačije naziva i *klasična teorija kredibiliteta*. Ova teorija obezbeđuje mehanizam za pripisivanje potpunog ili parcijalnog kredibiliteta iskustvu osiguranika (istorijskim podacima), i posledično pokušava da ograniči efekat koji će slučajne fluktuacije u opservacijama imati na ocene.

Drugi predstavljeni pristup se zove *teorija kredibiliteta visoke preciznosti.* Ovaj pristup je formalizovao Bulman u svom klasičnom radu iz 1967[70], gde je predstavio statistički okvir u kome je teorija kredibiliteta doživela svoj razvoj i procvat. Bulmanov pristup se drugačije naziva *kredibilitet najmanjih kvadrata*, zato što je cilj pristupa da se minimizuje srednjekvadratna greška između ocene i stvarne vrednosti parametra koji se ocenjuje.

U praksi često nemamo predznanje u vezi sa raspodelama za posmatrane kvantitete. Praktična upotreba izložene teorije zahteva da se nepoznati parametri ocene na osnovu podataka, kako bi model što bolje oslikavao stvarnost. Zbog toga smo razmatrali *empirijsko Bejzovo ocenjivanje parametara* kroz neparametarsko, semiparametarsko i parametarsko ocenjivanje.

Literatura, koja se bavi temom Bejzove statistike i njene primene, kako u teoriji kredibiliteta tako i u drugim oblastima, veoma je opširna, prevashodno na engleskom jeziku. Nemoguće je sve relevantne informacije staviti u okvir jedne knjige, a pogotovo na veoma detaljan način predstaviti teoriju u okviru master rada. U ovom radu je cilj bio da se ukratko predstavi osnovna Bejzova paradigma, uz osvrt na široku primenu Bejzove analize i na razlike u odnosu na frekventističku analizu. Takođe, cilj je bio da se kroz primere objasni teorija kredibiliteta, kako bi čitalac imao razumevanje o stvarnoj prirodi kredibiliteta. Nadam se da smo u tome uspeli i da je predstavljena materija čitaocu razumljiva i korisna.

[69] Vidi [13].
[70] Bühlmann, H. (1967), "*Experience Rating and Credibility*," ASTIN Bulletin, 4, 199–207.

BIBLIOGRAFIJA

[1] Klugman S., Panjer H., Willmont G., “Loss Models – from data to decisions“ - 2nd ed., Wiley series, 2004

[2] Mahler H.C. and Dean C.G., “Foundations of Casualty Actuarial Science”- “Chapter 8th, Credibility” by Casualty Actuarial Society , 4th Ed., 2001

[3] Dean C.G., “Topics in credibility theory”, 2005

[4] Klugman S. A., „Bayesian statistics in actuarial science with emphasis on credibility“, Kluwer academic publisher, 1992

[5] Longley-Cook L.H., “An Introduction to Credibility Theory”, Proceeding of the Casualty Actuarial Society, XLIX, 194–221., 1962

[6] Don Behan, Credibility Theory, Southeastern Actuarial Conference, 18th of June 2009

[7] Lozanov-Crvenković Z., „Statistika“, Faculty of Sciences, University of Novi Sad, 2012

[8] Rajter-Ćirić D., „Verovatnoća“, Faculty of Sciences, University of Novi Sad, 2009

[9] Klugman Stuart, „Sample size selection for multiple samples- A brief introduction to credibility theory and an example featuring race-based insurance premiums“, Drake University, for presentation at NYU – March 4, 2004

[10] Behan F. Donald, “Statistical Credibility Theory”, Presented to the Southeastern Actuarial Conference, June 18, 2009

[11] Philbrick Stephen, “An Examination of Credibility Concepts,” PCAS, 1981.

[12] Online course “Bayesian Statistics” provided by Duke University in North Carolina, USA, for Coursera https://www.coursera.org/learn/bayesian

[13] Papić Blagojević N. i Stefanović N.,Uloga i znača j teori je kredibiliteta u aktuarskoj praksi, UDK: 368.91(085.7), BIBLID: 0352-3713 (2014); 31, (4–6): 42–53, pregledni naučni rad

[14] Makov, U. (2001). Principal applications of Bayesian methods in actuarial science: A perspective, North American Actuarial Journal, p. 53.

[15] Casella George, Bayesians and Frequentists - Models, Assumptions, and Inference, ACCP 37th Annual Meeting, Philadelphia, 2008

[16] Jorge López Puga, Martin Krzywinski & Naomi Altman, Nature Methods 12, 377–378 (2015); published online 29 April 2015; corrected after print 24 September 2015.

[17] International Society for Bayesian Analysis https://bayesian.org/

[18] Casualty Actuarial Society http://www.casact.org/

Biografija

Bojana Bobar je rođena 15.08.1990. u Novom Sadu. Osnovnu školu „Kosta Trifković“ završila je 2005. godine u Novom Sadu kao nosilac Vukove diplome. Uporedo je 2005. godine završila osnovnu muzičku školu „Josip Slavenski“ u Novom Sadu, u klasi Tatjane Čerović. Iste godine upisala je opšti smer Gimnazije „Svetozar Marković“, koju je završila 2009. godine kao odličan đak. Po završetku gimnazije upisala je osnovne akademske studije na Prirodno-matematičkom fakultetu u Novom Sadu, smer primenjena matematika, modul matematika finansija, koje je završila u septembru 2012. godine. sa prosečnom ocenom 9.71. Potom je upisala master akademske studije na Prirodno-matematičkom fakultetu u Novom Sadu, smer primenjena matematika, modul matematika finansija. Položila je sve ispite predviđene nastavnim planom i programom master studija 2014. godine.

Stipendista je fonda za mlade talente Republike Srbije i na završnoj godini osnovnih i na završnoj godini master akademskih studija, i bila je stipendista Fondacije „Privrednik“. Tokom studija bila je aktivan član marketing tima na departmanu na kom je studirala. Zimski semestar školske 2013/2014. godine provela je na Univerzitetu u Gentu, kao nosilac stipendije Evropske komisije, programa Erasmus Mundus Action 2, Basileus Project IV. Koautor je pobedničkog projekta na studentskom takmičenju „Dokaži se 100%“ Customer Satisfaction University Award – Banca Intesa Beograd, nakon čega je obavljala stručnu praksu u sedištu Banca Intesa u Beogradu. U januaru 2015. godine izabrana je za projekat “Mladi lideri” kompanije Delta Holding u Beogradu, u okviru kojeg je radila 6 meseci u kompaniji Danubius d.o.o. Kao stipendista nemačke vlade obavljala je šestomesečnu praksu 2015. godine u sedištu osiguravajuće kompanije Allianz u Minhenu. Potom je 2016. godine, u želji za daljim usavršavanjem, obavljala i četvoromesečnu praksu u aktuarskom departmanu sedišta osiguravajuće kompanije Allianz Global Corporate & Specialty u Minhenu.

Novi Sad, oktobar, 2016. Bojana Bobar