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

Kristina Popadić

Analiza preživljavanja sa primenama u zdravstvenom osiguranju

- master rad -

Mentor:
prof. dr Dora Seleši

Novi Sad, 2015.

Sadržaj

Predgovor

Analiza preživljavanja predstavlja jedan od naprednih metoda istraživanja koji se koristi u mnogim različitim naukama. Naziv ovog metoda oslikava njegove početke u medicinskim i demografskim proučavanjima mortaliteta. Nakon predstavljanja proporcionalnog hazardnog modela od strane Koksa 1972. godine, empirijska analiza podataka o istoriji događaja je postala popularnija i doživela još širu primenu.

Analiza preživljavanja je deo statistike koji se bavi analizom vremena do pojave nekog događaja. Predmet posmatranja je slučajna promenljiva koja ima pozitivne vrednosti, kao što su na primer u medicini vreme do smrti, dužina boravka u bolnici, vreme do oporavka prilikom primene neke terapije, ili u ekonomiji trajanje štrajka, nezaposlenosti i slično. U osiguranju ima značajnu primenu jer se koristi kao bitna alatka u proračunima premija za različite polise. Na osnovu velikih baza podataka statističari su u mogućnosti da utvrde učestalost smrtnosti u zavisnosti od godina i zdravstvenog stanja osiguranika. Kada se uračunaju moguće isplate po osnovu polise, može se doći do iznosa premije. Cilj ovog rada je upoznavanje sa osnovnim pojmovima analize preživljavanja i pokazivanje primera primene u teoriji zdravstvenog osiguranja.

Rad započinjemo definisanjem osnovnim pojmovima u analizi preživljavanja, kao što su na primer funkcija preživljavanja i hazardna stopa. Zatim u drugom poglavlju opisujemo Koksov poluparametarski proporcionalni model koji omogućava uključivanje cenzurisanih podataka u proceni, što je veoma bitno jer oni imaju veliki udeo u bazama podataka koje se koriste u analizi preživljavanja. Ovaj model je značajan i zbog mogućnosti da se uključe vremenski zavisne promenjlive. Dalje ćemo pokazati kako se metodom maksimalne verodostojnosti ocenjuju koeficijenti Koksovog modela. U nastavku se upoznajemo sa stohastičkim procesima i definišemo Markov stohastički proces. Upotrebom Markovog procesa moguće je opisati modele sa više stanja koji se u praksi analize preživljavanja često sreću. Na kraju trećeg poglavlja su definisani intenziteti prelaza i pokazana je veza između intenziteta prelaza i verovatnoća prelaza preko Kolmogorovljevih diferencijalnih jednačina.

Četvrti deo rada je posvećen upoznavanju sa elementima iz aktuarstva. Definisane su aktuarske sadašnje vrednosti novčanih tokova koji se javljaju u osiguranju, a zatim je objašnjen pojam rezervi i princip ekvivalencije koji se koristi za računanje premije osiguranja. U osiguranju postoji mnogo modela sa više stanja pa smo na kraju ovog poglavlja pokazali i kako se računaju aktuarske sadašnje vrednosti u Markovim modelima sa više stanja.

U poslednjem poglavlju ćemo prikazati jednu od primena prethodno opisanih metoda analize preživljavanja u zdravstvenom osiguranju. U osiguranju za dugoročnu negu visina isplate beneficije od strane osiguravača zavisi od tipa nege koja se pruža osiguraniku, s tim da u toku trajanja osiguranja može doći do promena tipa nege pa je kao takav model idealan primer za primenu Markovih modela sa više stanja. Nakon predstavljanja ovog tipa osiguranja preko Markovog modela, pokazaćemo kako se intenziteti prelaza mogu izračunati pomoću Koksovog modela. U završetku rada ćemo objasniti kako se preko ocenjenih intenziteta prelaza, upotrebom aktuarskih formula dobijenih u četvrtom poglavlju mogu izračunati premije osiguranja za dugoročnu negu.

1. **Analiza preživljavanja**

Analiza podataka o vremenu preživljavanja (tačnije vremenu do nekog događaja, istoriji događaja ili podataka o trajanju) igra bitnu ulogu u medicini, biologiji, demografiji, ekonomiji, aktuarstvu i drugim naukama. Analiza preživljavanja se razlikuje od ostalih polja statistike zbog pojave cenzurisanih podataka. U procesu prikupljanja, neki podaci sadrže samo parcijalne informacije o pojavi koja se posmatra i zbog toga ovakve vrste nepotpunih opservacija povlače potrebu za specijalnim vrstama statističkih metoda. U prvom poglavlju ovog rada ćemo navesti osnovne karakteristike analize preživljavanja i objasniti pojam cenzurisanja. U nastavku poglavlja ćemo definisati funkcije kojima se opisuje vreme preživljavanja kao neprekidna slučajna promenljiva i pokazati ekvivalenciju definisanih funkcija. U praksi se vreme preživljavanja često predstavlja diskretnim modelom pa ćemo na kraju poglavlja izvesti i osnovne formule koje se u tom slučaju koriste.

1. 2. Osnovni pojmovi u analizi preživljavanja

Analiza preživljavanja predstavlja kolekciju statističkih procedura koje se koriste u izučavanju i proračunu vremena do pojave događaja od interesa. Predmet posmatranja analize preživljavanja je specijalna vrsta slučajne promenljive koja je definisana kao vreme propasti, odnosno vreme preživljavanja subjekta za koji se zna da je postojao na početku perioda posmatranja. Drugim rečima, pod vremenom preživljavanja za neki subjekat ovde podrazumevamo godine, mesece, nedelje ili dane koji proteknu od trenutka kad se subjekat počeo posmatrati do trenutka kad se događaj od interesa ostvari. U medicini se za događaj najčešće uzima smrt, pojava bolesti ili oporavak pacijenta pod nekom terapijom, ali događaj se može definisati i kao na primer kraj štrajka, dobijanje posla, bankrot preduzeća i slično.

Mi ćemo u ovom radu pretpostaviti da se posmatra realizacija samo jednog događaja, iako postoje analize u kojima postoji više događaja od interesa koji se ostvaruju nad jednim subjektom. U tom slučaju razlikujemo problem sa rekurentnim događajima i problem paralelnih rizika. Neki primeri rekurentnih događaja su: višestruki povraćaji iz remisije od leukemije, pacijent koji se leči od srčane bolesti više puta dobija infarkt i slično. Problem paralelnih rizika se javlja kada propast jednog subjekta može biti posledica bar dva uzroka, kao npr. ako se za propast smatra smrt neke osobe, ona može nastati usled bolesti, nesrećnog slučaja, kao samoubistvo ili ubistvo.

Prikupljanje podataka u analizi preživljavanja se obavlja u određenom vremenskom periodu, a prema načinu prikupljanja podataka studije mogu biti:

1) Unakrsne studije
2) Longitudinalne studije

Unakrsne studije najčešće koriste aktuari i demografi jer oni rade sa velikim uzorcima. U ovakvim studijama se prvo definiše studijska grupa, odnosno daje se opis subjekata čije će se vreme preživljavanja posmatrati, npr. neki primeri studijske grupe su populacija grada, pripadnici nacije, osiguranici u osiguravajućoj kompaniji, korisnici penzionog plana, itd. Zatim se bira i fiksira period posmatranja. Na početku ovog perioda postoji mnogo subjekata koji su već u studijskoj grupi i oni se tada i počinju posmatrati. Međutim, u studijsku grupu mogu ući subjekti i u toku perioda posmatranja dok neki drugi subjekti mogu i izaći pre kraja perioda a da se propast kod njih ne ostvari. Ovakvi ulasci i izlasci u toku trajanja studija se nazivaju migracije. Velika prednost unakrsnih studije jeste da se mogu posmatrati vremena preživljavanja koja su poduža.

U medicinskim istraživanjima su češći manji uzorci, pa se u tom slučaju primenjuje longitudinalni oblik studije. Dok se u unakrsnim studijama prvo fiksira period pa se iz tog perioda posmatraju subjekti koji odgovaraju studijskoj grupi, u longitudinalnim studijama se unapred sastavlja studijska grupa koja se zatim kontinuirano prati. Ovakav tip studija se može koristiti samo kada je vreme do propasti relativno kratko.

Longitudinalne studije mogu biti dizajnirane tako da omogućavaju potpune podatke ili kao studije sa nepotpunim podacima. Kada se unapred odabrana grupa subjekata od interesa, koja se u praksi naziva kohorta, posmatra sve dok se očekivani događaj ne ostvari na svakom subjektu iz grupe onda dobijamo potpune podatke. Ovako dizajnirane studije međutim, mogu biti veoma dugačke i skupe ako kohortu čine zdrave osobe pa se zato u takvim slučajevima longitudinalne studije nekada okončavaju i pre nego što se zabeleži trenutak propasti za sve subjekte. Postoje dva načina na koja se ovakvo okončavanje vrši, prvi način je na osnovu unapred definisanog datuma a drugi način je na osnovu unapred definisanog broja ostvarenih očekivanih događaja. U oba slučaja, pojaviće se subjekti kod kojih se očekivani događaj nije ostvario za vreme studije, odnosno za njih se tada ne zna da li će se nakon okončavanja studija događaj ostvariti i za koje vreme ili se eventualno neće ni ostvariti. Takvi slučajevi predstavljaju nepotpune podatke koje nazivamo i cenzurisanim podacima.

Cenzurisani podaci su veoma česti u analizi preživljavanja i predstavljaju značajan analitički problem. Cenzurisani podaci su oni podaci koji u sebi sadrže neke informacije o vremenu preživljavanja subjekta ali ne i egzaktno vreme preživljavanja. Zbog toga se u tom slučaju vreme koje protekne od početka posmatranja datog subjekta do trenutka cenzurisanja naziva cenzurisano vreme preživljavanja a ukoliko je vreme propasti za nekog subjekta poznato onda ćemo vreme koje protekne od početka posmatranja tog subjekta do njegove propasti nazivati egzaktno vreme preživljavanja.

Cenzurisanje se osim u longitudinalnim studijama sa nepotpunim podacima može pojaviti:

1) Ako se u toku studije neka osoba “izgubi”, odnosno ukoliko se praćenje nekog subjekta iz nepoznatog razloga ne nastavi
2) Usled povlačenja nekog subjekta iz studije zbog nekog događaja koji nije od interesa, na primer kada je događaj od interesa zaposlenje neke osobe a u toku posmatranja ta osoba umre

Da li je vreme preživljavanja cenzurisano ili nije pokazuje slučajna promenljiva koju ćemo označavati grčkim slovom δ. Ukoliko se očekivani događaj za neki subjekat ostvario onda će za taj subjekat važiti $\delta = 1$ a ako je vreme preživljavanja tog subjekta cenzurisano tada je $\delta = 0$.

1.2. Raspodela vremena preživljavanja

Vreme preživljavanja nekog subjekta se obeležava sa T, a sa malim slovom t obeležavamo realizovanu vrednost slučajne promenljive T. Na primer, ako T predstavlja broj meseci koje pacijent provede na hemotarapiji pre smrti i ako nas zanima da li je pacijent preživeo najmanje 5 meseci na hemoterapiji onda se u stvari pitamo da li je T veće od $t = 5$. Pošto slučajna promenljiva T predstavlja vreme, ona može uzimati samo nenegativne vrednosti, odnosno važi da $T \geq 0$. Kao i svaka slučajna promenjliva i vreme preživljavanja može biti neprekidna, diskretna ili mešovita slučajna promenljiva.

Pretpostavimo da je T apsolutno neprekidna slučajna promenljiva i da je $F(t)$ njena funkcija raspodele. U praksi se u zavisnosti od cilja analize, odnosno radi iskazivanja različitih zaključaka obrade podataka, raspodela vremena preživljavanja opisuje preko sledeće tri funkcije:

1. Funkcija preživljavanja
2. Gustina verovatnoće vremena preživljavanja
3. Hazardna funkcija

Ove tri funkcije su matematički ekvivalentne, odnosno ako nam je raspodela opisana jednom od ovih funkcija, na jedinstven način možemo odrediti druge dve funkcije. Sada ćemo opisati svaku od ovih funkcija i pokazati njihovu ekvivalenciju.

Funkcija preživljavanja

Funkcija preživljavanja, koju ćemo obeležavati sa $S(t)$, predstavlja verovatnoću da je vreme preživljavanja subjekta duže od t, tj.:

$$S(t) = P\{vreme\ preživljavanja\ je\ duže\ od\ t\} = P\{T > t\} \tag{1.2.1}$$

Kako po definiciji za funkciju raspodele važi $F(t) = P\{T < t\}$, dobijamo da je:

$$S(t) = 1 - P\{vreme\ preživljavanja\ nije\ duže\ od\ t\} = 1 - F(t) \tag{1.2.2}$$

Osobine funkcije $S(t)$:

1) $S(t)$ je monotono nerastuća funkcija – verovatnoća da će osoba živeti duže od t godina ne raste povećanjem godina t
2) $S(0) = 1$ – verovatnoća da će na početku perioda posmatranja osoba biti živa je jednaka 1
3) $S(\infty) = 0$ – verovatnoća da će osoba živeti beskonačno godina je 0

Funkcija $S(t)$ se često naziva i kumulativna stopa preživljavanja. Kako bi opisao tok funkcije preživljavanja, Berkson (1942) je uveo grafičku prezentaciju funkcije preživljavanja. Grafik funkcije $S(t)$ se naziva kriva preživljavanja.

Gustina verovatnoće (funkcija intenziteta verovatnoće) vremena preživljavanja

Kao i svaka neprekidna slučajna promenljiva, vreme preživljavanja T ima svoju gustinu verovatnoće. Gustina verovatnoće vremena preživljavanja, u oznaci $f(t)$, predstavlja graničnu vrednost verovatnoće da se događaj smrti realizuje u veoma kratkom intervalu od t do $t + \Delta t$ po jedinici priraštaja Δt, odnosno verovatnoću umiranja u veoma kratkom intervalu po jedinici vremena. Matematički ovu definiciju zapisujemo na sledeći način:

$$f(t) = \frac{\lim\limits_{\Delta t \to 0} P[osoba\ umire\ u\ intervalu\ (t,\ t + \Delta t)]}{\Delta t} \tag{1.2.3}$$

Osobine funkcije $f(t)$:

1) $f(t)$ je nenegativna funkcija
2) Površina između grafika funkcije $f(t)$ i x – ose je 1.

Funkcija $f(t)$ predstavlja bezuslovnu stopu preživljavanja, što znači da ona prikazuje verovatnoću smrtnosti u trenutku t koristeći samo informaciju da je osoba živa u trenutku $t = 0$.

Hazardna funkcija (hazardna stopa)

Hazardna funkcija $h(t)$ vremena preživljavanja T predstavlja uslovnu stopu preživljavanja.

Hazardna funkcija ili hazardna stopa predstavlja verovatnoću realizacije smrti u veoma kratkom vremenskom intervalu pod uslovom da je posmatrana osoba živa na početku tog intervala, odnosno definišemo je kao graničnu vrednost u obliku:

$$h(t) = \frac{\lim\limits_{\Delta t \to 0} P[osoba\ umire\ u\ (t,\ t\ +\ \Delta t)\ pod\ uslovom\ da\ je\ živa\ u\ trenutku\ t]}{\Delta t} \tag{1.2.4}$$

Hazardna stopa je poznata i kao trenutna (nasuprot intervalnoj) stopa preživljavanja, sila mortaliteta i kao uslovna stopa mortaliteta. Takođe je možemo posmatrati i kao stopu preživljavanja u odnosu na određene godine ako t u formuli 1.2.4 predstavlja godine.

Kumulativna hazardna funkcija se definiše kao:

$$H(t) = \int_0^t h(x)dx \tag{1.2.5}$$

Kasnije ćemo pokazati da važi sledeća veza između kumulativne hazardne funkcije i funkcije preživljavanja:

$$H(t) = -\ln S(t) \tag{1.2.6}$$

Iz jednakosti 1.2.6 i osobina funkcije $S(t)$ možemo zaključiti da je:

1) $H(t) = 0$, jer je $S(t) = 1$
2) $H(\infty) = \infty$, jer je $S(\infty) = 1$
3) Između 0 i ∞ vrednost funkcije $H(t)$ može biti bilo koji realan broj

Ekvivalentnost funkcija $S(t)$, $f(t)$ i $h(t)$

Kao što smo napomenuli, tri funkcije koje smo definisali su ekvivalentne što ćemo sada i pokazati.

U teoriji verovatnoće važi da je bezuslovna mera jednaka proizvodu uslovne mere i verovatnoće realizacije uslova. Ako ovaj odnos primenimo u našem slučaju dobijamo da je bezuslovna stopa preživljavanja $f(t)$ jednaka proizvodu uslovne stope preživljavanja $h(t)$ i verovatnoće uslova da je posmatrana osoba živa u trenutku t. Ovaj uslov se može zapisati kao verovatnoća $P\{T \geq t\}$, što je po definiciji jednako sa $S(t)$. Odnosno, dobijamo sledeću formulu za hazardnu stopu $h(t)$:

$$h(t) = \frac{f(t)}{S(t)} \tag{1.2.7}$$

Pošto je funkcija intenziteta verovatnoće $f(t)$, tj. gustina verovatnoće izvod funkcije raspodele verovatnoće $F(t)$ dobijamo da je:

$$f(t) = \frac{d}{dt} F(t) = \frac{d}{dt}(1 - S(t)) = -S'(t) \tag{1.2.8}$$

Odnosno vidimo da ako znamo funkciju preživljavanja možemo lako dobiti funkciju intenziteta verovatnoće a kombinacijom formula 1.2.7 i 1.2.8 dobijamo i sledeću formulu za hazardnu funkciju:

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt}\ln S(t) \tag{1.2.9}$$

Integraljenjem obe strane jednakosti 1.2.9 na intervalu $(0, t)$ i korišćenjem da je $S(t) = 1$ dobijamo:

$$-\int_0^t h(x)dx = \ln S(t) \tag{1.2.10}$$

Odavde vidimo da važi formula 1.2.6 za kumulativnu hazardnu stopu $H(t)$. Dalje, iz prethodne jednakosti možemo izraziti $S(t)$ u zavisnosti od $h(t)$:

$$S(t) = exp\left[-\int_0^t h(x)dx\right] \tag{1.2.11}$$

A zatim na osnovu 1.7 i 1.11 dobijamo i funkciju intenziteta izraženu preko hazardne funkcije:

$$f(t) = h(t)exp\left[-\int_0^t h(x)dx\right] \tag{1.2.12}$$

Ukoliko nam je poznata funkcija intenziteta, odnosno gustina verovatnoće $f(t)$ možemo odrediti njoj odgovarajuću funkciju raspodele $F(t)$ čijim uvrštavanjem u definiciju funkcije $S(t)$ dobijamo:

$$S(t) = 1 - \int_0^t f(x)dx \tag{1.2.13}$$

A iz osobina funkcije $f(t)$ znamo da je njen integral nad celom oblašću definisanosti jednak 1 pa je:

$$S(t) = \int_t^\infty f(x)dx \tag{1.2.14}$$

Zatim kada uvrstimo $S(t)$ u ovakvom obliku u 1.2.9 dobićemo i formulu za hazardnu funkciju. Dakle pokazali smo da na osnovu bilo koje od funkcija koje definišu raspodelu od T možemo dobiti preostale dve funkcije, odnosno funkcija preživljavanja $S(t)$, funkcija intenziteta verovatnoće $f(t)$ i hazardna funkcija $h(t)$ su ekvivalentne.

1.3. Vreme preživljavanja kao diskretna slučajna promenljiva

Kada je T diskretna slučajna promenljiva sa vrednostima $t_0 < t_1 < t_2 < \cdots$ onda je funkcija preživljavanja nerastuća, sa leve strane neprekidna stepenasta funkcija. Ako verovatnoću da je t_k vreme do realizacije propasti označimo sa $p_k = P(T = t_k)$, onda je verovatnoća da je vreme preživljavanja duže od t_k, odnosno vrednost funkcije preživljavanja u $T = t_k$:

$$S(t_k) = P(T > t_k) = p_{k+1} + p_{k+2} + \cdots = \sum_{i:\, t_i > t_k} p_i$$

Primetimo, verovatnoća da je t_k vreme koje protekne do realizacije propasti je jednaka razlici verovatnoće da je vreme preživljavanja duže od t_{k-1} i verovatnoće da je vreme preživljavanja duže od t_k, tj.:

$$p_k = S(t_{k-1}) - S(t_k)$$

Hazardnu stopu u diskretnom slučaju možemo definisati kao uslovnu verovatnoću u sledećem obliku:

$$h(t_k) = P(vreme\ do\ realizacije\ propasti\ je\ t_k | vreme\ preživljavanja\ je\ najmanje\ t_k)$$

Odnosno dobijamo sledeći odnos hazardne stope i funkcije preživljavanja:

$$h(t_k) = P(T = t_k | T \geq t_k) = \frac{P(T = t_k)}{P(T \geq t_k)} = \frac{p_k}{S(t_{k-1})} = 1 - \frac{S(t_k)}{S(t_{k-1})}$$

Na osnovu prethodne jednakosti možemo izvesti rekurentnu vezu:

$$\begin{aligned} S(t_k) &= S(t_{k-1}) \cdot \big(1 - h(t_k)\big) = S(t_{k-2}) \cdot \big(1 - h(t_{k-1})\big) \cdot \big(1 - h(t_k)\big) \\ &= S(t_0) \cdot \big(1 - h(t_1)\big) \cdot \big(1 - h(t_2)\big) \cdot \ldots \cdot \big(1 - h(t_{k-1})\big) \cdot \big(1 - h(t_k)\big) \end{aligned}$$

Pošto u studiju ulaze samo subjekti kod kojih na početku perioda posmatranja događaj od interesa nije ostvaren, odnosno važi da je $P(T > t_0) = S(t_0) = 1$, dobijamo:

$$S(t_k) = \prod_{1 \leq i \leq k} \big(1 - h(t_i)\big) \qquad (1.3.1)$$

2. Ocena funkcije preživljavanja, Koksov model

Parametarski metodi određivanja funkcije preživljavanja su moćni ako je poznata raspodela vremena preživljavanja, ali u praksi je ona najčešće nepoznata. U neparametarskim metodima nije neophodno poznavati raspodelu vremena preživljavanja, a najčešće korišćen model je Koksov proporcionalni hazardni model. Hazardna funkcija u ovom modelu može biti bilo kakvog oblika, ali je pretpostavka da su hazardne stope za bilo koje dve različite osobe proporcionalne i da je njihov odnos nezavisan od vremena. Takođe je bitno napomenuti da se pri oceni ovog modela koristi metoda slična metodi maksimalne verodostojnosti a to je metoda maksimalne parcijalne verodostojnosti. Statistički zaključak koji se izvodi na osnovu ove dve metode je sličan. Ovaj deo rada je posvećen upoznavanju Koksovog modela i metode maksimalne parcijalne verodostojnosti. U prvom delu poglavlja je predstavljena forma Koksovog proporcionalnog modela i način tumačenja ocenjenog modela. Pristup oceni ovog modela je različit u slučaju kada u skupu prikupljenih podataka nema ponovljenih vremena preživljavanja i u slučaju kada ih ima, što je u nastavku poglavlja i objašnjeno na primerima. Poseban značaj primene metode parcijalne verodostojnosti se vidi na kraju ovog poglavlja gde je prikazan način obrade podataka u slučaju kada se vreme preživljavanja predstavlja modelom sa vremenski zavisnim promenljivama.

2.1. Forma Koksovog modela

Osnovna karakteristika Koksovog proporcionalnog hazardnog modela je da su hazardne stope dve različite osobe proporcionalne, tj ako pretpostavimo da na vreme preživljavanja utiče p faktora, odnosno imamo p nezavisnih promenljivih u modelu $X_1, X_2, \dots, X_p$ i ako su vrednosti tih promenljivih za neku osobu $\boldsymbol{X_1} = \left(X_{11}, X_{21}, \dots, X_{p1}\right)$ a za neku drugu osobu $\boldsymbol{X_2} = \left(X_{12},\ X_{22}, \dots, X_{p2}\right)$, onda važi:

$$\frac{h(t \mid \boldsymbol{X_1})}{h(t \mid \boldsymbol{X_2})} = c$$

Gde je c konstanta koja ne zavisi od vremena. Drugim rečima, odnos rizika od umiranja između dve osobe je fiksan u odnosu na vreme, tj. ne menja se vremenenom bez obzira koliko dugo te osobe živele.

Na osnovu ove osobine, hazardnu stopu za osobu kojoj odgovaraju vrednosti promenljivih $\boldsymbol{X} = (X_1,\ X_2, \dots, X_p)$ možemo zapisati u obliku proizvoda osnovne hazardne stope $h_0(t)$ koja zavisi samo od vremena i funkcije $g(\boldsymbol{X})$ koja zavisi samo od vrednosti promenljivih u modelu, tj. možemo posebno posmatrati uticaj vremena i uticaj faktora na hazardnu stopu.

$$h(t\,|\boldsymbol{X}) = h_0(t) \cdot g(\boldsymbol{X})$$

Kad u ovom obliku napišemo hazardnu funkciju, lako možemo potvrditi da je hazardni odnos dve različite osobe fiksan u odnosu na vreme:

$$\frac{h(t \mid \boldsymbol{X_1})}{h(t\,|\boldsymbol{X_2})} = \frac{h_0(t) \cdot g(\boldsymbol{X_1})}{h_0(t) \cdot g(\boldsymbol{X_2})} = \frac{g(\boldsymbol{X_1})}{g(\boldsymbol{X_2})}$$

Osobinu da odnos hazardnih stopa dve osobe zavisi samo od razlika između vrednosti promenljivih koje odgovaraju tim osobama ali ne i od vremena poseduju ekponencijalni regresioni model i Weibullov regresioni model. Koks je u svom modelu hazardne stope pretpostavio da je $g(\boldsymbol{X})$ ekponencijalna funkcija promenljivih, odnosno ako je $\boldsymbol{b} = (b_1, b_2, \dots, b_p)$ vektor koeficijenata koji stoje uz promenljive $X_1, X_2, \dots, X_p$, onda je Koksov oblik hazardne funkcije:

$$h(t\,|\boldsymbol{X}) = h_0(t) \cdot exp\{\boldsymbol{b} \cdot \boldsymbol{X}\} = h_0(t) \cdot exp\{b_1 \cdot X_1 + b_2 \cdot X_2 + \cdots + b_p \cdot X_p\}$$

Ako u prethodnoj formuli za vrednosti promenljivih uzmemo $\boldsymbol{X} = (0, 0, \dots, 0)$ dobićemo

$$h(t\,|0, 0, \dots, 0) = h_0(t)$$

što znači da je vrednost $h_0(t)$ jednaka vrednosti hazardne funkcije kada su sve promenljive zanemarene. Zbog toga se $h_0(t)$ naziva osnovna hazardna stopa.

Koeficijenti $b_1, b_2, \dots, b_p$ iz Koksovog modela se mogu oceniti koristeći razne metode ocene na osnovu prikupljenih podataka, a zatim se pomoću ocenjenih koeficijenata mogu iskazati veličine efekata odgovarajućih promenljivih na hazardnu stopu. Ako u odnosu hazardnih stopa dve osobe, sa vrednostima promenljivih $\boldsymbol{X_1} = (X_{11}, X_{21}, \dots, X_{p1})$ i $\boldsymbol{X_2} = (X_{12}, X_{22}, \dots, X_{p2})$, za vrednosti svih promenljivih osim vrednosti promenljive X_1 stavimo nule, koeficijent b_1 će prikazivati kako promena vrednosti promenljive X_1 utiče na odnos hazardnih stopa.

$$\frac{h(t\,|X_{11}, 0, \dots, 0)}{h(t\,|X_{12}, 0, \dots, 0)} = \frac{h_0(t) \cdot exp\{b_1 \cdot X_{11} + b_2 \cdot 0 + \cdots + b_p \cdot 0\}}{h_0(t) \cdot exp\{b_1 \cdot X_{12} + b_2 \cdot 0 + \cdots + b_p \cdot 0\}} = exp\{b_1 \cdot (X_{11} - X_{12})\}$$

Kada zanemarimo promenljive $X_2, \dots, X_p$, prema vrednostima ocenjenih koeficijenta b_1 možemo utvrditi tri različita efekta promenljive X_1:

1) Ako je $\widetilde{b_1} = 0$ onda promenjliva X_1 ne utiče na hazardnu stopu, odnosno promena vrednosti promenljive X_1 ne utiče na vreme preživljavanja neke osobe, tj.:

$$X_{11} \neq X_{12} \wedge \widetilde{b_1} = 0 \implies \frac{h(t\,|X_{11}, 0, \dots, 0)}{h(t\,|X_{12}, 0, \dots, 0)} = exp\{0 \cdot (X_{11} - X_{12})\} = 1$$

2) Ako je $\widetilde{b_1} < 0$ onda pad vrednosti promenljive X_1 uzrokuje povećanje vrednosti hazardne stope, odnosno tada osoba 1 koja ima veću vrednost promenljive X_1 nego osoba 2 ima i duže predviđeno vreme preživljavanja, tj.:

$$X_{11} > X_{12} \wedge \widetilde{b_1} < 0 \Rightarrow \frac{h(t\,|X_{11}, 0, \dots, 0)}{h(t\,|X_{12}, 0, \dots, 0)} = exp\{\widetilde{b_1} \cdot (X_{11} - X_{12})\} < 1$$

3) Ako je $\widetilde{b_1} > 0$ onda pad vrednosti promenljive X_1 uzrokuje smanjenje vrednosti hazardne stope, odnosno tada osoba 1 koja ima veću vrednost promenljive X_1 nego osoba 2 ima i kraće predviđeno vreme preživljavanja, tj.:

$$X_{11} > X_{12} \wedge \widetilde{b_1} > 0 \Rightarrow \frac{h(t\,|X_{11}, 0, \dots, 0)}{h(t\,|X_{12}, 0, \dots, 0)} = exp\{\widetilde{b_1} \cdot (X_{11} - X_{12})\} > 1$$

Naš cilj je da odredimo promenljive u modelu koje imaju značajan uticaj na vreme preživljavanja. Tačnije, mi želimo da pronađemo podskup promenjivih koje značajnije utiču na hazardnu stopu a samim tim i na dužinu preživljavanja neke osobe. Kao što smo rekli, koje promenljive su značajne ćemo otkriti preko odgovarajućih koeficijenata, odnosno ako je $\widetilde{b_i} = 0$ promenljiva X_i nije značajna a ako je $\widetilde{b_i} \neq 0$ onda je značajna. Kako bi ocenio koeficijente $b_1, b_2, \dots, b_p$, Koks je koristio funkciju parcijalne verodostojnosti koja se zasniva na uslovnoj verovatnoći propasti pod pretpostavkom da u uzorku nema pacijenata sa jednakim vremenima preživljavanja. S obzirom da se u praksi često sreću ponovljena vremena preživljavanja, Koksova funkcije parcijalne verodostojnosti je kasnije modifikovana tako da se može primeniti i kad je pomenuta pretpostavka narušena.

2.2. Funkcija parcijalne verodostojnosti u slučaju kada nema ponovljenih vremena preživljavanja

Metodom maksimalne verodostojnosti se dobijaju ocene parametara pri kojima je verovatnoća realizacije dobijenog uzorka najveća. Ona se dakle koristi za ocenu parametarskih modela a funkcija verodostojnosti se konstruiše na osnovu raspodela zavisnih promenljivih u tom modelu. U Koksovom modelu nema pretpostavki o raspodeli zavisne promenljive odnosno o raspodeli vremena preživljavanja pa se zato ne može koristiti tipičan način konstrukcije funkcije verodostojnosti. Za ocenu parametara Koksovog modela koristi se Koksova verodostojnost koja se zasniva na poretku ostvarenih događaja a umesto gustine raspodele koriste se hazardne funkcije. Pre nego što pojasnimo formu funkcije Koksove verodostojnosti uvešćemo definiciju rizičnog skupa.

Pretpostavimo da postoji k različitih necenzurisanih vremena preživljavanja od n pacijenata, što znači da ima $n - k$ cenzurisanih vremena preživljavanja. Nek su vremena poređana u rastući niz $t_1 < t_2 < \cdots < t_n$ i neka su $\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_p}$ odgovarajući vektori nezavisnih promenljivih. Za svako egzaktno vreme preživljavanja t_i definasaćemo rizični skup u oznaci $R(t_i)$ koji predstavlja skup indeksa koji odgovaraju vremenima preživljavanja koja nisu kraća od t_i:

$$R(t_i) = \{j \in \{1,2,..n\} \big| t_j \geq t_i\}$$

Formiranje Koksove verodostojnosti ćemo prikazati pomoću narednog primera.

Primer: Neka se uzorak sastoji od 6 osoba sa različitim vremenima preživljavanja, pri čemu su treće i peto cenzurisana vremena. U tabeli su vremena preživljavanja poređana u rastućem redosledu, od osobe koja je prva umrla do osobe koja je poslednja umrla. Radi jednostavnosti zapisa, nezavisne promenljive nismo specifikovali za ovaj primer, ali smo pretpostavili da ima p nezavisnih promenljivih, tj. $\boldsymbol{X_i} = (X_{1i}, X_{2i}, \dots, X_{pi})$

Osoba	**Vreme preživljavanja**	**Cenzurisano**	X
$\boldsymbol{a_1}$	20	Ne	$\boldsymbol{X_1}$
$\boldsymbol{a_2}$	21	Ne	$\boldsymbol{X_2}$
$\boldsymbol{a_3}$	26	Da	$\boldsymbol{X_3}$
$\boldsymbol{a_4}$	28	Ne	$\boldsymbol{X_4}$
$\boldsymbol{a_5}$	29	Da	$\boldsymbol{X_5}$
$\boldsymbol{a_6}$	31	Ne	$\boldsymbol{X_6}$

Tabela 2.2.1: Podaci bez ponovljenih vremena preživljavanja

Šta je zajedničko za ove podatke?

- Vremena trajanja se mogu poređati
- U početnom trenutku t_0 su svi u riziku od umiranja
- Nakon prvog događaja umiranja, skup onih u riziku se smanjuje za 1 element
- Skup pacijenata u riziku se sukcesivno smanjuje kako se događaji umiranja ostvaruju

Koksova verodostojnost je jednaka verovatnoći da je da je osoba sa skupom vrednosti $\boldsymbol{X_1}$ umrla prva, a zatim da je umrla osoba sa skupom vrednosti $\boldsymbol{X_2}$, pa osoba kojoj odgovara $\boldsymbol{X_4}$ i da je poslednja umrla osoba sa $\boldsymbol{X_6}$. Verovatnoća da je baš osoba sa skupom vrednosti $\boldsymbol{X_i}$ umrla i - ta odnosno da joj odgovara vreme preživljavanja t_i je jednaka:

$$\frac{P(osoba\ kojoj\ odgovara\ \boldsymbol{X_i}\ je\ umrla\ u\ t_i)}{P(neka\ osoba\ koja\ je\ živa\ do\ vremena\ \ t_i\ tj.\ u\ riziku\ je\ od\ umiranja, je\ umrla\ u\ t_i)}$$

Pošto osobe umiru nezavisno jedna od druge, verovatnoća iz imenioca je jednaka zbiru verovatnoća umiranja za sve osobe koje su u riziku u vremenu t_i

$$\frac{P(osoba\ kojoj\ odgovara\ \boldsymbol{X_i}\ je\ umrla\ u\ t_i)}{\sum_{j\in R(t_i)} P(osoba\ kojoj\ odgovara\ \boldsymbol{X_j}\ je\ umrla\ u\ t_i)}$$

Sada verovatnoće umiranja u vremenu t_i zamenjujemo verovatnoćama umiranja u intervalu $[t_i, t_i + \Delta)$

$$\frac{P(osoba\ kojoj\ odgovara\ \boldsymbol{X_i}\ je\ umrla\ u\ [t_i, t_i + \Delta))}{\sum_{j\in R(t_i)} P(osoba\ kojoj\ odgovara\ \boldsymbol{X_j}\ je\ umrla\ u\ [t_i, t_i + \Delta))}$$

Kada odredimo graničnu vrednost prethodnog izraza kada $\Delta \rightarrow 0$ dobijamo hazardne stope umesto početnih verovatnoća

$$\frac{h(t_i|\boldsymbol{X_i})}{\sum_{j\in R(t_i)} h(t_i|\boldsymbol{X_j})} = \frac{h_0(t_i)exp\{\boldsymbol{bX_i}\}}{\sum_{j\in R(t_i)} h_0(t_i)\, exp\{\boldsymbol{bX_j}\}}$$

Nakon skraćivanja osnovne hazardne stope, dobijamo da je doprinos verodostojnosti uzorka od strane elementa uzorka kome pripada t_i, odnosno verodostojnost tog elementa uzorka:

$$\frac{exp\{\boldsymbol{bX_i}\}}{\sum_{j\in R(t_i)} exp\{\boldsymbol{bX_j}\}}$$

Kada pomnožimo verodostojnost za sve elemente uzorka kod kojih vreme preživljavanja nije cenzurisano dobijamo verovatnoću da je ostvaren baš poredak iz uzorka odnosno dobijamo Koksovu funkciju verodostojnosti. Ako uvedemo oznaku:

$$g_i := exp\{\boldsymbol{bX_i}\}$$

za naš primer dobijamo:

$$\frac{g_1}{g_1 + g_2 + g_3 + g_4 + g_5 + g_6} \cdot \frac{g_2}{g_2 + g_3 + g_4 + g_5 + g_6} \cdot \frac{g_4}{g_4 + g_5 + g_6} \cdot \frac{g_6}{g_6}$$

Kao što vidimo, za konstrukciju funkcije verodostojnosti su korišćena samo egzaktna vremena preživljavanja pa se zbog toga Koksova verodostojnost naziva parcijalna verodostojnost. Metod parcijalne verodostojnosti ne koristi eksplicitno cenzurisane podatke, ali ih ni ne zanemaruje. Tj. osobe kojima pripadaju cenzurisana vremena preživljavanja koja nisu duža od t_i se nalaze u rizičnom skupu $R(t_i)$ i na taj način doprinose informacije za verodostojnost uzorka.

Ako uvedemo funkciju $\delta(t_i)$ koja će imati vrednost 1 ako je t_i egzaktno vreme preživljavanja a vrednost 0 ako je t_i cenzurisano vreme preživljavanja, onda funkciju parcijalne verodostojnosti možemo zapisati u sledećem obliku:

$$L_P = \prod_{i=1}^{n} \left(\frac{exp\{\boldsymbol{bX_i}\}}{\sum_{j \in R(t_i)} exp\{\boldsymbol{bX_j}\}} \right)^{\delta(t_i)}$$

Nakon konstrukcije funkcije parcijalne verodostojnosti, naredni koraci ocene metodom maksimalne verodostojnosti su isti kao i u slučaju standardne funkcije verodostojnosti. Pre određivanja maksimuma funkcije parcijalne verodostojnosti, radi jednostavnosti koristi se logaritam od L_P:

$$ln(L_P) = \sum_{i=1}^{n} \delta(t_i) \cdot \left(\boldsymbol{bX_i} - ln\left(\sum\nolimits_{j \in R(t_i)} exp\{\boldsymbol{bX_j}\} \right) \right)$$

Nepoznati parametri ovog Koksovog proporcionalnog modela su $b_1, b_2, \dots, b_p$, što znači da ćemo maksimum funkcije $ln(L_P)$ dobiti tako što ćemo parcijalne izvode po b_l, $l = 1,2, \dots p$ izjednačiti sa nulom i rešiti dobijeni sistem od p jednačina sa p nepoznatih.

$$\frac{\partial ln(L_P)}{\partial b_l} = \sum_{i=1}^{n} \delta(t_i) \cdot \left(X_{li} - \frac{\sum_{j \in R(t_i)} exp\{\boldsymbol{bX_j}\} \cdot X_{lj}}{\sum_{j \in R(t_i)} exp\{\boldsymbol{bX_j}\}} \right)$$

Jednačine koje se dobijaju su nelinearne pa se ovaj sistem rešava nekim iterativnim postupkom.

2.3. Funkcija parcijalne verodostojnosti u slučaju kada ima ponovljenih vremena preživljavanja

U praksi se nekada može desiti da se u uzorku nađe više osoba sa istim vremenima preživljavanja. U tom slučaju se ne može koristiti prethodno opisan oblik funkcije parcijalne verodostojnosti, pa ćemo sada opisati postupak ocene kada u uzorku ima ponovljenih vremena preživljavanja.

Pretpostavimo da imamo n vremena preživljavanja u uzorku i da je među njima k necenzurisanih različitih vremena koja su poređena u rastućem redosledu $t_{(1)} < t_{(2)} < \dots < t_{(k)}$. Sa m_i ćemo označiti broj pacijenata čije je egzaktno vreme preživljavanja $t_{(i)}$, dakle $m_i = 1$ ako postoji samo jedna osoba kojoj je egzaktno vreme preživljavanja $t_{(i)}$, a $m_i > 1$ ako postoji više osoba sa egzaktnim vremenom preživljavanja $t_{(i)}$. $R(t_{(i)})$ i dalje predstavlja rizični skup za $t_{(i)}$, a sa r_i ćemo označiti broj elemenata u $R(t_{(i)})$, tj.

$$r_i := |R(t_{(i)})|.$$

Iz svakog skupa $R(t_{(i)})$ možemo na slučajan način izvući skup od m_i osoba koji ćemo obeležiti sa u_j. Ukupan broj različitih skupova u_j sa m_i osoba iz rizičnog skupa $R(t_{(i)})$ pri čemu je $r_i = |R(t_{(i)})|$ ćemo označiti sa C_{r_i,m_i}, a dobija se pomoću sledeće formule:

$$C_{r_i,m_i} := \binom{|R(t_{(i)})|}{|u_j|} = \binom{r_i}{m_i} = \frac{r_i!}{m_i! \cdot (r_i - m_i)!}$$

Kako se u praksi određuju vrednosti oznaka koje smo uveli ćemo pojasniti sledećim primerom.

Primer:

Osoba	**Vreme preživljavanja**	**Cenzurisano**	**X**
$\boldsymbol{a_1}$	20	Ne	$\boldsymbol{X_1}$
$\boldsymbol{a_2}$	21	Da	$\boldsymbol{X_2}$
$\boldsymbol{a_3}$	21	Da	$\boldsymbol{X_3}$
$\boldsymbol{a_4}$	26	Ne	$\boldsymbol{X_4}$
$\boldsymbol{a_5}$	26	Ne	$\boldsymbol{X_5}$
$\boldsymbol{a_6}$	26	Ne	$\boldsymbol{X_6}$
$\boldsymbol{a_7}$	28	Ne	$\boldsymbol{X_7}$
$\boldsymbol{a_8}$	29	Da	$\boldsymbol{X_8}$
$\boldsymbol{a_9}$	31	Ne	$\boldsymbol{X_9}$
$\boldsymbol{a_{10}}$	31	Ne	$\boldsymbol{X_{10}}$

Tabela 2.3.1: Podaci sa ponovljenim vremenima preživljavanja

Posmatrajmo uzorak veličine $n = 10$ sa $k = 4$, tj. različita egzaktna vremena preživljavanja su $t_{(1)} = 20$, $t_{(2)} = 26$, $t_{(3)} = 28$ i $t_{(4)} = 31$. Broj osoba sa tim vremenima je redom $m_1 = 1$, $m_2 = 3$, $m_3 = 1$ i $m_4 = 2$. Odredimo sada C_{r_i,m_i} za vreme preživljavanja $t_{(2)} = 26$. Odgovarajući rizični skup je:

$$R(t_{(2)}) = \{osobe\ čije\ je\ vreme\ preživljavanja\ nije\ kraće\ od\ t_{(2)} = 26\}$$

Odnosno u ovom skupu će se naći osobe a_4, a_5, a_6, a_7, a_8, a_9 i a_{10}, pa je broj elemenata ovog skupa je $r_2 = 7$. Dakle dobijamo da je za $t_{(2)} = 26$ sa $r_2 = 7$ i $m_2 = 3$:

$$C_{r_2,m_2} = \frac{r_2!}{m_2! \cdot (r_2 - m_2)!} = \frac{7!}{3! \cdot (7-3)!} = 35$$

Znači postoji ukupno 35 mogućih načina da iz skupa $R(t_{(2)})$ izvučemo $m_2 = 3$ osobe, odnosno postoji 35 različitih skupova u_j za $R(t_{(2)})$. Jedan od tih skupova je na primer: $u_j = \{osobe\ sa\ vremenima\ preživljavanja\ 26, 26\ i\ 29\} = \{a_4, a_5, a_8\}$.

Vratimo se na problem konstrukcije funkcije parcijalne verodostojnosti. Ako pretpostavimo da je t neprekidno, za m_i osoba sa vremenom preživljavanja $t_{(i)}$ logično je pretpostaviti da ta vremena nisu identična nego su posledica nepreciznog merenja. Da su precizna merenja moguća, "tačna" vremena preživljavanja bi mogla biti poređana u rastući niz i mogli bismo primeniti metod za slučaj u kome nema ponovljenih vremena. S obzirom da takva preciznost nije moguća i da nam je egzaktni poredak tih vremena nepoznat, moramo razmotriti sve moguće poretke. Za svako $t_{(i)}$ m_i odgovarajućih jednakih vremena se mogu poređati na $m_i!$ različitih načina. Za svaki od tih načina ćemo zatim moći odrediti funkciju parcijalne verodostojnosti na isti način koji smo koristili kad nismo imali ponovljena vremena. Pošto je ovakav način konstrukcije besmislen kad je m_i veliko, uglavnom se koriste aproksimacije funkcije parcijalne verodostojnosti. Kad je svako m_i malo u odnosu na odgovarajuće r_i, najčešće korišćeni metodi su:

1) Metod Breslowa:

$$L_B = \prod_{i=1}^{n} \left(\frac{exp\left\{\boldsymbol{b z_{u_j^*}}\right\}}{\left[\sum_{j\in R(t_i)} exp\{\boldsymbol{bX_j}\}\right]^{m_i}} \right)^{\delta(t_i)}$$

2) Metod Efrona:

$$L_E = \prod_{i=1}^{n} \left(\frac{exp\left\{\boldsymbol{b z_{u_j^*}}\right\}}{\prod_{j=1}^{m_i}\left[\sum_{j\in R(t_i)} exp\{\boldsymbol{bX_j}\} - \left((j-1)/m_i\right)\cdot \sum_{j\in u_i^*} exp\{\boldsymbol{bX_j}\}\right]} \right)^{\delta(t_i)}$$

U prethodnim formulama je za svako egzaktno vreme preživljavanja $t_{(i)}$ definisana vrednost:

$$\boldsymbol{z_{u_j^*}} := \sum_{i\in u_j^*} \boldsymbol{X_i} = \left(\sum_{i\in u_j^*} X_{1i}, \sum_{i\in u_j^*} X_{2i}, \dots, \sum_{i\in u_j^*} X_{pi} \right) = \left(z_{1u_j^*}, z_{2u_j^*}, \dots, z_{pu_j^*} \right)$$

pri čemu je u_i^* skup koji se sastoji od m_i osoba čija su vremena preživljavanja $t_{(i)}$.

U slučaju kada nema ponovljenih vremena preživljavanja, L_B i L_E se svode na običnu funkciju parcijalne verodostojnosti.

2.4. Funkcija parcijalne verodostojnosti u slučaju vremenski zavisnih promenljivih

Glavna karakteristika Koksovog proporcionalnog modela jeste da se hazardna stopa može rastaviti na osnovnu hazardnu stopu koja zavisi samo od vremena i ekponencijalnu funkciju nezavisnih promenljivih a kao posledica nastaje činjenica da su hazardne stope dve osobe proporcionalne i ne zavise od vremena. Ova karakteristika postaje narušena kada se u modelu pojave nezavisne promenljive koje su zavisne od vremena. U praksi se često sreću modeli u kojima postoje i promenljive koje zavise od vremena i one koje ne zavise.

Postoje dve vrste vremenski zavisnih promenljivih:

1) Promenljive koje se višekratno mere u različitim vremenskim trenucima do nastanka očekivanog događaja ili do kraja studije (mi ćemo u ovom delu rada pretpostaviti da je zavisnost od vremena ove vrste)
2) Promenljive čije se vremenske promene mogu opisati poznatom matematičkom funkcijom i promenljive koje imaju različite vrednosti usled dejstva terapije, starosti ili promene medicinskog stanja.

U slučaju vremenski zavisnih promenljivih se i dalje može koristiti Koksov model ali se tada naziva prošireni Koksov model. Hazardna stopa će i dalje imati isti oblik osim što će neke nezavisne promenljive biti funkcije vremena.

Pretpostavimo da vremena preživljavanja zavise od p nezavisnih promenljivih od kojih su prvih p_1 nezavisne od vremena a preostalih $p - p_1$ zavise od vremena. Tj. za vektor nezavisnih promenljivih $\boldsymbol{X}(t) = (X_1, \dots, X_{p_1}, X_{p_1+1}(t), \dots, X_p(t))$ hazardna stopa je oblika:

$$h(t \mid \boldsymbol{X}(t)) = h_0(t) \cdot exp\left\{b_1 \cdot X_1 + \cdots + b_{p_1} X_{p_1} + b_{p_1+1} \cdot X_{p_1+1}(t) + \cdots + b_p \cdot X_p(t)\right\}$$

a funkcija parcijalne verodostojnosti pod pretpostavkom da nema ponovljenih vremena preživljavanja je:

$$L_P = \prod_{i=1}^{n} \left(\frac{exp\left\{b_1 \cdot X_1 + \cdots + b_{p_1} X_{p_1} + b_{p_1+1} \cdot X_{p_1+1}(t) + \cdots + b_p \cdot X_p(t)\right\}}{\sum_{j \in R(t_i)} exp\left\{b_1 \cdot X_1 + \cdots + b_{p_1} X_{p_1} + b_{p_1+1} \cdot X_{p_1+1}(t) + \cdots + b_p \cdot X_p(t)\right\}} \right)^{\delta(t_i)}$$

Ocena modela metodom maksimalne verodostojnosti se ne može direktno primeniti kao u slučaju vremenski nezavisnih promenljivih, nego se u ovom slučaju prvo vrši specijalna transformacija prikupljenih podataka kako bi se "izbacila" vremenska zavisnost. Sada ćemo na primeru pokazati kako se konstruiše funkcija parcijalne verodostojnosti u slučaju da u modelu ima nezavisnih promenljivih koje zavise od vremena.

Primer: Cilj studije je da proceni da li profili biomarkera mogu da se upotrebe za procenu rizika i rano otkrivanje raka bešike kod grupe radnika koji su na svom radnom mestu izloženi uticaju benzidina usled čega imaju rizik od raka bešike. Na pregledu obavljenom na početku studija radnici nisu imali rak bešike. U narednih 7 godina su vršeni ponovni pregledi u svrhu provere postojanja raka bešike, s tim da broj i učestalost pregleda nisu bili isti za svakog pacijenta. Vreme preživljavanja (izraženo u mesecima) u ovoj studiji je vreme koje je proteklo od početka studije do pregleda na kom je ustanovljeno da pacijent ima rak, a ukoliko pacijent ni na poslednjem pregledu u toku studija nije imao rak onda je za njega vreme preživljavanja vreme od početka studije do poslednjeg pregleda i ono se smatra cenzurisanim vremenom. U modelu postoji jedna vremenski nezavisna promenljiva koja predstavlja nivo izloženosti benzidinu, u oznaci LEX (nivo izloženosti je utvrđen prema radnoj poziciji u fabrici), i tri vremenski zavisne promenljive a to su godine pacijenta i biomarkeri $M1$ i $M2$ (ako je biomarker pozitivan onda je $M1 = 1$, tj. $M2 = 1$ a ako je negativan onda je $M1 = 0$, tj. $M2 = 0$). Vrednost biomarkera i status raka bešike ($status = 1$ ako pacijent ima rak, $status = 0$ ako nema) su mereni na početku i na svakom ponovnom pregledu. U tabeli su prikazani vrednosti svih promenljivih pri prvom merenju, a zatim vremena ponovnih merenja sa tadašnjim vrednostima biomerkara i statusom raka.

Pacijent	Godine	Lex	M1	M2	status
$\boldsymbol{a_1}$	58,97	15	0	0	0
	Drugo merenje nakon 30,82 meseci		0	0	0
	Treće merenje nakon 36,40 meseci		1	0	0
	Četvrto merenje nakon 41,26 meseci		1	0	0
	Peto merenje nakon 48,33 meseci		0	0	0
	Šesto merenje nakon 54,83 meseci		0	0	0
	Sedmo merenje nakon 60,71 meseci		1	0	0
	Osmo merenje nakon 66,17 meseci		0	0	0
	Deveto merenje nakon 73,07 meseci		1	0	0
$\boldsymbol{a_2}$	47,82	36	1	0	0
	Drugo merenje nakon 42,94 meseci		0	0	0
	Treće merenje nakon 67,06 meseci		0	0	0
$\boldsymbol{a_3}$	58,64	180	0	0	0
	Drugo merenje nakon 24,61 meseci		1	0	1
$\boldsymbol{a_4}$	62,26	192	0	1	0
	Drugo merenje nakon 31,47 meseci		1	0	0
	Treće merenje nakon 36,17 meseci		0	0	1

Tabela 2.4.1: Model sa vremenski zavisnim promenljivama

U slučaju vremenski zavisnih promenljivih, podaci se razlažu na vremenske intervale u kojima će vrednosti vremenski zavisnih promenljivih biti fiksne. Vremenski intervali su oblika $(start, stop]$, gde se za $stop$ vrednost uzima vreme u kom se promena desila a ako se u tom vremenu posmatrani događaj (pojava raka) nije ostvario onda je $cenz = 0$ a u suprotnom je $cenz = 1$. Način razlaganja podataka je prikazan u narednoj tabeli.

Pacijent	Start	Stop	Početne god.	Lex	M10	M20	God.	M1	M2	Status
a_1	0,00	30,82	58,97	15	0	0	61,54	0	0	0
	30,82	41,26	58,97	15	0	0	62,01	1	0	0
	41,26	54,83	58,97	15	0	0	63,00	0	0	0
	54,83	60,71	58,97	15	0	0	64,03	1	0	0
	60,71	66,17	58,97	15	0	0	64,49	0	0	0
	66,17	73,07	58,97	15	0	0	65,06	1	0	0
a_2	0,00	67,06	47,82	36	1	0	53,40	0	0	0
a_3	0,00	24,61	58,64	180	0	0	60,70	1	0	1
a_4	0,00	31,47	62,26	192	0	1	64,88	1	0	0
	31,47	36,17	62,26	192	0	1	65,27	0	0	1

Tabela 2.4.2: Model sa vremenski zavisnim promenljivama – sređeni podaci

Kao što vidimo, pacijenti kod kojih se vrednosti promenljivih nisu menjale tokom vremena posmatranja se pojavljuju samo jednom u tabeli, dok se pacijenti kod kojih su se vrednosti menjale pojavljuju više puta. Broj redova tabele koji odgovara nekom pacijentu je jednak broju ukupnih promena svih vremenski zavisnih promenljivih. Na primer, pacijent a_3 je tokom vremena posmatranja imao fiksne vrednosti biomarkera pa se u tabeli nalazi samo jednom. Za razliku od njega, pacijent a_4 je na početnom merenju imao vrednosti biomarkera $M1 = 0$ i $M2 = 1$, a zatim su mu na drugom merenju ustanovljene nove vrednosti oba biomerka $M1 = 1$ a $M2 = 0$, a do trećeg merenje promenila mu se opet vrednost biomarkera $M1$ pa je na poslednjem merenju iznosila $M1 = 0$. Dakle, kod pacijenta a_4 su na naredna dva merenja zabeležene promene vrednosti vremenski zavisnih promenljivih, pa će se podaci o tom pacijentu razložiti na dva reda nove tabele. Primetimo da iako su bile uočene tri promene, one su uočene u dva različita vremenska trenutka pa se razlaganje vrši na dva reda. Takođe je bitno napomenuti da ukoliko u modelu postoje i vremenski nezavisne promenljive, njihove vrednosti se prepisuju u svaki red.

S obzirom da se u tabeli podaci o jednom pacijentu iz uzorka mogu pojaviti više puta, može se pretpostaviti da će ti skupovi podataka biti korelisani, pa se u tom slučaju ne bi mogli primeniti standardni metodi ocene parametara. Korelisanost, međutim, nije prisutna nakon razlaganja početnih podataka u ovakvom tipu analize a problem može nastati u slučaju kada se kao predmet analize javlja više događaja nad jednim subjektom. Kao što znamo, parcijalna verodostojnost je jednaka proizvodu verodostojnosti svih egzaktnih vremena preživljavanja iz uzorka. Pri konskrukciji verodostojnosti svakog vremena preživljavanja koristi se najviše jedan red podataka od nekog pacijenta, onaj koji se odnosi na vremenski interval u kom je dati pacijent u riziku. Ovo je razlog zašto se u ovakvim studijama nezavisnost korištenog uzorka ne narušava.

3. Modeli sa više stanja u analizi preživljavanja

Predmet analize preživljavanja je vreme preživljavanja nekog pacijenta. Iskustvo nekog pacijenta se može predstaviti modelom sa dva stanja u kojima pacijent može biti, a to su stanje „živ“ i stanje „mrtav“, i tada se posmatra vreme za koje će pacijent preći iz prvog u drugo stanje. Ovakav model predstavlja najjednostavniji model u analizi preživljavanja, dok se njegovim uopštavanjem mogu dobiti mnogo komplikovaniji modeli sa više od dva stanja u kojima pacijenti mogu menjati stanja više puta ili čak vraćati se u neko stanje u kom su prethodno već bili. U longitudinalnim medicinskim studijama, pacijenti se posmatraju u nekom vremenskom intervalu a podaci, odnosno vrednosti promenljivih, se prikupljaju u određenim vremenskim trenucima. U slučaju kada se više različitih događaja mogu ostvariti od strane jednog pacijenta, najčešće se studije opisuju modelom sa više stanja.

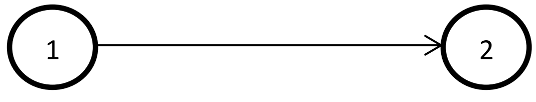
3.1. Modeli sa više stanja i neki primeri

Model sa više stanja je model stohastičkog procesa koji dozvoljava da predmet posmatranja prolazi kroz konačan broj stanja. Promena stanja se naziva prelaz ili događaj. Stanja mogu biti:

- Prelazna – nakon izlaska iz prelaznog stanja, moguć je ponovni ulazak u isto stanje
- Striktno prelazna – nakon izlaska iz striktno prelaznog stanja, ponovni ulazak je onemogućen
- Apsorbujuća – ukoliko se predmet posmatranja nađe u stanju koje je apsorbujuće iz njega ne može izaći

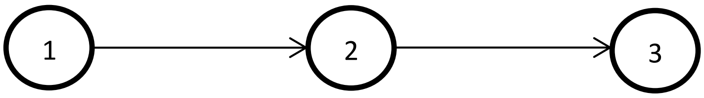
Kompleksnost modela zavisi od broja stanja kao i od njihove vrste, odnosno zavisi od mogućih prelaza. Sada ćemo dati neke primere osnovnih modela sa više stanja.

1) Model mortaliteta je najjednostavniji model sa stanjima “živ” i “mrtav” i jednim mogućim prelazom, prelazom iz stanja “živ” u stanje “mrtav”. U ovom modelu je stanje “živ” striktno prelazno a stanje “mrtav” je apsorbujuće.



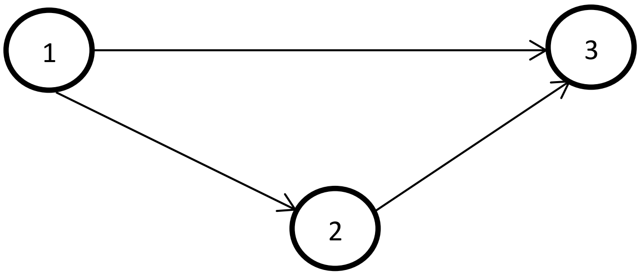
Slika 3.1.1: Model mortaliteta

2) Progresivni model sa k stanja je model koji ima $k-1$ striktno prelaznih stanja i jedno apsorbujuće stanje. Entitet prelazi redom kroz sva stanja, odnosno iz stanja i prelazi u stanje $i+1$, $i=1,2,\ldots,k-1$. Na slici je primer progresivnog modela sa 3 stanja.



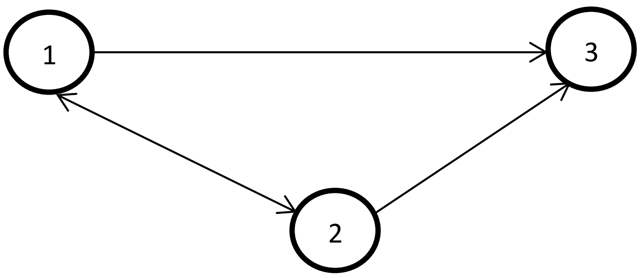
Slika 3.1.2: Progresivni model sa 3 stanja

3) Model bolest – smrt je model sa dva striktno prelazna stanja “zdrav” i “bolestan” i jednim apsorbujućim stanjem “smrt”. Mogući prelazi su opisani na slici 4, gde vidimo da ako se pacijent jednom razboli on ne može ozdraviti a umreti može i ako je zdrav i ako je bolestan.



Slika 3.1.3: Model bolest – smrt

4) Dvosmerni model bolest – smrt ima ista stanja kao prethodni model s tim da je u ovom modelu omogućen prelaz iz stanja “bolestan” u stanje “zdrav”.



Slika 3.1.4: Dvosmerni model bolest – smrt

3.2. Stohastički procesi i lanci Markova

Proces sa više stanja je stohastički proces $Z(t), \mathrm{t} \in \mathrm{T}$ sa konačnim skupom stanja $\mathcal{S} = \{1, 2, \ldots, N\}$, gde je T skup indeksa vremena. Ako je $\mathrm{T} = [0, \tau]$, $\tau > 0$, $Z(t)$ je neprekidni stohastički proces a ako je $\mathrm{T} = \{0, 1, 2, \ldots\}$, onda je $Z(t)$ diskretni stohastički proces. Vrednost procesa u vremenu t predstavlja stanje zauzeto u tom trenutku.

Evolucija procesa tokom vremena, tj. istorija H_{t-} predstavlja σ-algebru koju čine opservacije procesa nad intervalom $[0, t)$ ili $\{0, 1, 2, \ldots, \mathrm{t} - 1\}$, odnosno istorija je skup stanja u kojima je proces bio pre trenutka t.

Proces sa više stanja je u potpunosti određen preko verovatnoća prelaza između stanja i i j. Verovatnoća prelaza, u oznaci $P_{ij}(t, u)$, je verovatnoća da se proces nalazi u stanju j u trenutku u pod uslovom da se u trenutku t nalazio u stanju i i da je H_{t-} njegova istorija tj:

$$P_{ij}(t, u) := P(Z(u) = j \,|Z(t) = i, H_{t-}), \quad i, j \in \mathcal{S}, t, u \in T, t \leq u$$

Stohastički procesi kod kojih verovatnoća da se ostvari neko stanje u budućnosti zavisi samo od sadašnjeg stanja bez obzira na prošlost procesa, odnosno uslovno od sadašnjosti, budućnost i prošlost su nezavisni se nazivaju lanci Markova.

Dakle, stohastički proces $Z(t), \mathrm{t} \in \mathrm{T}$, je lanac Markova ako za svako t i u, $\mathrm{t} < u$, iz skupa T i za svaka dva proizvoljna stanja $i, j \in \mathcal{S}$ važi svojstvo Markova, tj. važi:

$$P(Z(u) = j \,|Z(t) = i, H_{t-}) = P(Z(u) = j \,|Z(t) = i)$$

Kod lanaca Markova verovatnoća prelaza ne zavisi od istorije H_{t-}, odnosno definiše se na sledeći način:

$$P_{ij}(t, u) := P(Z(u) = j \,|Z(t) = i)$$

Bitno je napomenuti da po ovoj definiciji u slučaju verovatnoće $P_{ii}(t, u)$ proces ne mora ostati u stanju i u vremenskom intervalu (t, u), tj. bitno je da se na početku i na kraju intervala nalazi u tom stanju a u međuvremenu proces može da prođe i kroz neka druga stanja iz skupa $\mathcal{S}$. Nasuprot ovim verovatnoćama definišu se verovatnoće zadržavanja stanja, u kojima proces ostaje u stanju i tokom vremenskog intervala $[t, u]$.

Verovatnoće zadržavanja stanja se obeležavaju sa $P_{\underline{ii}}(t, u)$ a definišu se na sledeći način:

$$P_{\underline{ii}}(t, u) = P(Z(z) = i, \forall z \in [t, u] | Z(t) = i), t < u$$

Jednu osobinu verovatnoća prelaza su primetili Chapman i Kolmogorov, a to je da je verovatnoća da se u trenutku u proces nađe u stanju j ako se u trenutku t, $t < u$ nalazio u stanju i jednaka sumi prolazaka kroz sva moguća međustanja k u nekom trenutku w iz intervala $[t, u]$. Ova osobina je opisana jednačinama Chapman-Kolmogorova:

$$P_{ij}(t,u) = \sum_{k \in \mathcal{S}} P_{ik}(t,w) \cdot P_{kj}(w,u) \qquad t \leq w \leq u \quad i,j \in \mathcal{S}$$

Preko sličnih jednačina se mogu izračunati bezuslovne verovatnoće $P(Z(u) = j)$ pomoću početne verovatnoće $a_i := P(Z(0) = i)$ i odgovarajućih uslovnih verovatnoća, tj.:

$$P(Z(u) = j) = \sum_{i \in \mathcal{S}} a_i \cdot P_{ij}(0,u)$$

Verovatnoće zadržavanja stanja takođe zadovoljavaju jednačine Chapman-Kolmogorova ali se one razlikuju od jednačina za verovatnoće prelaza, odnosno za verovatnoću $P_{\underline{ii}}(t,u)$ i proizvoljno $w \in [t,u]$ važi:

$$P_{\underline{ii}}(t,u) = P_{\underline{ii}}(t,w) \cdot P_{\underline{ii}}(w,u)$$

3.3. Intenziteti prelaza i diferencijalne jednačine Kolmogorova za neprekidne lance Markova

Intenziteti prelaza se definišu za svako $i \neq j$ i $t \geq 0$ na sledeći način:

$$\mu_{ij}(t) := \lim_{dt \to 0} \frac{P_{ij}(t, t+dt)}{dt} = \lim_{dt \to 0} \frac{P(Z(t+dt) = j | Z(t) = i)}{dt}$$

pri čemu pretpostavljamo da granične vrednosti iz definicije postoje i da su intenziteti integrabilni na kompaktnim intervalima.

Intenzitet prelaza možemo protumačiti kao verovatnoću prelaza iz jednog neko drugo stanje u veoma kratkom intervalu $[t, t+dt)$.

Zatim, možemo definisati za svako $i \in \mathcal{S}$ i $t \geq 0$ intenzitet zadržavanja stanja:

$$\mu_{ii}(t) := -\sum_{j \neq i} \mu_{ij}(t)$$

Sumiranjem svih intenziteta prelaza koji se odnose na izlazak iz nekog stanja i dobijamo intenzitet napuštanja stanja i:

$$\mu_i(t) := \sum_{j \neq i} \mu_{ij}(t)$$

Intenzitet $\mu_i(t)$ možemo predstaviti i kao uslovnu verovatnoću napuštanja stanja i u kratkom vremenskom interval $[t\,,t{+}dt)$ i prelaska u bilo koje drugo stanje, pri uslovu da se stohastički proces nalazi u stanju i u trenutku t. I u tom slučaju, koristeći definiciju intenziteta prelaza, intenzitet napuštanja stanja i možemo zapisati u vidu sledeće formule:

$$\mu_i(t) := \sum_{j \neq i} \lim_{dt \to 0} \frac{P_{ij}(t, t+dt)}{dt}$$

$$= \lim_{dt \to 0} \sum_{j \neq i} \frac{P_{ij}(t, t+dt)}{dt}$$

$$= \lim_{dt \to 0} \frac{1 - P_{ij}(t, t+dt)}{dt}$$

Veza između verovatnoća prelaza i intenziteta prelaza u modelu Markova sa više stanja se može opisati pomoću diferencijalnih jednačina Kolmogorova. Postoje backward i forward diferencijalne jednačine Kolmogorova. S obzirom da se dolazi do istih rezultata pomoću obe vrste jednačina, mi ćemo govoriti samo o forward diferencijalnim jednačinama.

Forward diferencijalne jednačine Kolmogorova su definisane za sva stanja i, j i za sve vremenske trenutke $t, u, 0 \leq t \leq u$ sa graničnim uslovom $P_{ij}(t,t) = \delta_{ij}$ na sledeći način:

$$\frac{d}{dt} P_{ij}(t,u) = \sum_{k \neq j} P_{ik}(t,u) \cdot \mu_{kj}(u) - P_{ij}(t,u) \cdot \mu_j(u)$$

$$= \sum_{k \neq j} P_{ik}(t,u) \cdot \mu_{kj}(u) - P_{ij}(t,u) \cdot \sum_{k \neq j} \mu_{jk}(u)$$

$$= \sum_{k \neq j} P_{ik}(t,u) \cdot \mu_{kj}(u) - P_{ij}(t,u) \cdot \mu_{jk}(u)$$

Interpretacija ove definicije koju su dali Haberman i Pitacco (1999) glasi: Verovatnoće prelaza započinju stanjem i u trenutku t. Leva strana forward diferencijalne jednačine Kolmogorova predstavlja promenu verovatnoće ulaska u stanje j u odnosu na kratak vremenski interval $[t, t+dt)$, dok desna strana predstavlja razliku verovatnoće ulaska u stanje j iz proizvoljnog početnog stanja $k, k \neq j$ i verovatnoće napuštanja stanja j u kratkom vremenskom intervalu $[t, t+dt)$.

Sada ćemo na primeru pokazati kako se dobijaju rešenja jednačina Kolmogorova. Koristićemo model sa slike 3 iz prvog dela ovog poglavlja. U ovom modelu je skup stanja $\mathcal{S} = \{1, 2, 3\}$ a prelazi $2 \to 1$, $3 \to 2$ i $3 \to 1$ su nemogući pa su odgovarajući intenziteti prelaza i verovatnoće prelaza jednake nuli.

Koristeći poslednji oblik formule za jednačine Kolmogorova i na osnovu činjenice da važi $\mu_{21}(u) = \mu_{31}(u) = 0$ i $P_{21}(t,u) = P_{31}(t,u) = 0$ za prelaze iz stanja 1 u ostala stanja skupa $\mathcal{S}$, dobijamo:

$$\begin{aligned}\frac{d}{dt}P_{11}(t,u) &= P_{12}(t,u)\cdot\mu_{21}(u) + P_{13}(t,u)\cdot\mu_{31}(u) - P_{11}(t,u)\cdot(\mu_{12}(u)+\mu_{13}(u))\\ &= -P_{11}(t,u)\cdot(\mu_{12}(u)+\mu_{13}(u))\end{aligned}$$

$$\begin{aligned}\frac{d}{dt}P_{12}(t,u) &= P_{11}(t,u)\cdot\mu_{12}(u) + P_{13}(t,u)\cdot\mu_{32}(u) - P_{12}(t,u)\cdot(\mu_{21}(u)+\mu_{23}(u))\\ &= P_{11}(t,u)\cdot\mu_{12}(u) - P_{12}(t,u)\cdot\mu_{23}(u)\end{aligned}$$

$$\begin{aligned}\frac{d}{dt}P_{13}(t,u) &= P_{11}(t,u)\cdot\mu_{13}(u) + P_{12}(t,u)\cdot\mu_{23}(u) - P_{13}(t,u)\cdot(\mu_{31}(u)+\mu_{32}(u))\\ &= P_{11}(t,u)\cdot\mu_{13}(u) + P_{12}(t,u)\cdot\mu_{23}(u)\end{aligned}$$

Analogno dobijamo prelaze iz stanja 2 i 3:

$$\frac{d}{dt}P_{21}(t,u) = 0$$

$$\frac{d}{dt}P_{22}(t,u) = -P_{22}(t,u)\cdot\mu_{23}(u)$$

$$\frac{d}{dt}P_{23}(t,u) = P_{22}(t,u)\cdot\mu_{23}(u)$$

$$\frac{d}{dt}P_{31}(t,u) = 0$$

$$\frac{d}{dt}P_{32}(t,u) = 0$$

$$\frac{d}{dt}P_{33}(t,u) = 0$$

Granični uslovi su za ovaj primer:

$$P_{11}(t,t) = P_{22}(t,t) = P_{33}(t,t) = 1,$$

$$P_{ij}(t,t) = 0, \qquad \forall i,j \in \{1,2,3\}, i \neq j$$

Princip rešavanja jednačina koje sadrže samo $P_{11}(t,u)$ odnosno $P_{22}(t,u)$ je isti pa ćemo samo pokazati za $P_{11}(t,u)$.

$$\frac{d}{dt}P_{11}(t,u) = -P_{11}(t,u) \cdot (\mu_{12}(u) + \mu_{13}(u))$$

$$\frac{1}{P_{11}(t,u)} \cdot \frac{d}{dt}P_{11}(t,u) = -(\mu_{12}(u) + \mu_{13}(u))$$

$$\frac{d}{dt}lnP_{11}(t,u) = -(\mu_{12}(u) + \mu_{13}(u))$$

$$\int_t^u \frac{d}{dz}lnP_{11}(t,z)dz = -\int_t^u (\mu_{12}(z) + \mu_{13}(z))dz$$

$$lnP_{11}(t,u) - lnP_{11}(t,t) = -\int_t^u (\mu_{12}(z) + \mu_{13}(z))dz$$

Kad iskoristimo granični uslov $P_{11}(t,t) = 1$ dobijamo rešenje:

$$P_{11}(t,u) = exp\left\{-\int_t^u (\mu_{12}(z) + \mu_{13}(z))dz\right\}$$

Analognim postupkom se dobija:

$$P_{22}(t,u) = exp\left\{-\int_t^u \mu_{23}(z)dz\right\}$$

Na osnovu osobine matrice prelaza $\sum_{j\in\mathcal{S}} P_{ij}(t,u) = 1 \quad \forall i \in \mathcal{S}$, dobijamo:

$$P_{13}(t,u) = 1 - P_{11}(t,u) - P_{12}(t,u)$$

$$P_{23}(t,u) = 1 - P_{22}(t,u)$$

$$P_{33}(t,u) = 1$$

Ostaje da još izvedemo rešenje za $P_{12}(t,u)$. Diferencijalna jednačina koje odgovara ovoj verovatnoći prelaza je nehomogena linearna diferencijalna jednačina prvog reda za koju znamo:

Ako je diferencijalna jednačina oblika:

$$\frac{dy}{dx} + p(x) \cdot y = g(x)$$

Onda je rešenje dobijeno metodom varijacije konstanti:

$$y(x,c) = c \cdot e^{-r(x)} + e^{-r(x)} \cdot \int^x g(s) \cdot e^{r(s)}\, ds$$

Gde je $r(x) = \int^{x} p(s)ds$.

U našem slučaju je jednačina oblika:

$$\frac{d}{dt}P_{12}(t,u) + \mu_{23}(u) \cdot\ P_{12}(t,u) = P_{11}(t,u) \cdot \mu_{12}(u)$$

Pa se metodom varijacije konstanti dobija:

$$\begin{aligned}
P_{12}(t,u) = c \cdot exp\left\{-\int_t^u \mu_{23}(z)dz\right\} + exp\left\{-\int_t^u \mu_{23}(z)dz\right\} \\
\cdot\left(\int_t^u P_{11}(t,z) \cdot \mu_{12}(z) \cdot exp\left\{\int_t^z \mu_{23}(s)ds\right\} dz\right) \\
= c \cdot exp\left\{-\int_t^u \mu_{23}(z)dz\right\} \\
+ \int_t^u P_{11}(t,z) \cdot \mu_{12}(z) \cdot exp\left\{-\int_t^u \mu_{23}(z)dz\right\} \cdot exp\left\{-\int_z^t \mu_{23}(s)ds\right\} dz \\
= c \cdot exp\left\{-\int_t^u \mu_{23}(z)dz\right\} \\
+ \int_t^u P_{11}(t,z) \cdot \mu_{12}(z) \cdot exp\left\{-\int_z^u \mu_{23}(s)ds\right\} dz = c \cdot exp\left\{-\int_t^u \mu_{23}(z)dz\right\} \\
+ \int_t^u P_{11}(t,z) \cdot \mu_{12}(z) \cdot P_{22}(z,u)dz
\end{aligned}$$

Sada kada uvrstimo granični uslov $P_{12}(t,t) = 0$ dobijamo da je $c = 0$, pa je rešenje:

$$P_{12}(t,u) = \int_t^u P_{11}(t,z) \cdot \mu_{12}(z) \cdot P_{22}(z,u)dz$$

4. Računanje premije osiguranja u stohastičkom modelu

Osnovni predmet izučavanja u aktuarstvu je rizik. Rizik se najčešće definiše kao mogućnost da se u budućnosti ostvari događaj koji za posledicu ima finansijski gubitak. Osiguranje je jedan oblik upravljanja rizikom, prvenstveno usmeren na smanjenje finansijskih gubitaka. Osiguranik, tj. osoba koja želi da se zaštiti od rizika, potpisuje ugovor koji se naziva polisa osiguranja sa kompanijom koja se bavi osiguranjem, tj. sa osiguravajućom organizacijom. Polisom osiguranja se osiguravajuća organizacija obavezuje da će u slučaju da se ostvari događaj koji povlači finansijski gubitak, isplatiti novčane beneficije kojima će osiguranik pokriti taj gubitak. Zauzvrat osiguranik plaća ugovorenu novčanu sumu koja se naziva premija. Na taj način osiguranje predstavlja prenos rizika sa osiguranika na osiguravajuću organizaciju, uz plaćanje premije osiguranja.
Osiguranje je moguće praktično za sve vrste nepovoljnih događaja za koje su vreme i mesto događaja neizvesni, učestalost predvidiva i za koje je gubitak znatan ali ne i katastrofalan, kako bi osiguravajuća društva i mogla i imala interes da organizuju osiguranje. Danas postoje brojni tipovi osiguranja a najčešći su osiguranje automobila, osiguranje za slučaj smrti, zdravstveno osiguranje, osiguranje imovine, putničko osiguranje, penzijsko osiguranje itd.
Cilj svake osiguravajuće organizacije je da ukupno isplaćene beneficije za jednu vrstu osiguranja pokrije ukupno naplaćenim premijama za to osiguranje, s tim da se ne može unapred znati koliki će biti ukupan iznos isplaćenih beneficija. Iz tog razloga je uglavnom u praksi zastupljen stohastički pristup izračunavanju visine premije koji ćemo predstaviti u ovom poglavlju. Prvo ćemo definisati aktuarske sadašnje vrednosti, a zatim ćemo uvesti pojam rezervi i pokazati kako se određuje iznos premije na osnovu principa ekvivalencije.
U zavisnosti od vrste osiguranja, beneficije se mogu isplaćivati u vidu periodičnih anuiteta ili jednokratno. Periodični anuiteti su npr. prisutni u penzionom osiguranju u kome osiguranik prima ugovorene iznose anuiteta periodično od trenutka kada dostigne ugovorom definisanu starost do trenutka smrti. Jedan primer osiguranja sa jednokratnom ispatom beneficije predstavlja osiguranje za slučaj smrti gde se isplata beneficije izvršava nakon smrti osiguranika u korist nekih trećih lica, najčešće članova porodice. Postoje i osiguranja u kojima se kombinuju načini isplate i različiti iznosi beneficija u zavisnosti od stanja u kom se nalazi osiguranik a takve vrste osiguranja se modeliraju pomoću Markovih lanaca o kojima ćemo govoriti na kraju ovog poglavlja.

4.1. Aktuarske sadašnje vrednosti beneficija i premije osiguranja

Neka je T diskretna slučajna promenljiva koja predstavlja vreme završetka trajanja polise osiguranja, pri čemu je vreme trajanja polise od trenutka kupovine polise do trenutka kada se osigurani događaj ostvari. Iako nije neophodno, u nekim slučajevima je skup mogućih vrednosti slučajne promenljive T ograničen sa gornje strane. Ukoliko postoji gornje ograničenje, tada vreme završetka trajanja polise odgovara minimumu vrednosti gornjeg ograničenja i dužine vremena do ostvarenja osiguranog događaja. Mi ćemo u narednim definicijama pretpostaviti da postoji gornja granica koju ćemo obeležavati sa τ, što znači da će skup mogućih vrednosti slučajte promenljive T biti oblika $\{0, 1, \dots, \tau\}$.

Novčani tokovi koji se javljaju u toku trajanja polise osiguranja su:

1) $c(k) -$ jednokratna isplata beneficije u trenutku k u slučaju da se osigurani događaj ostvari u vremenskom trenutku k,
2) $b(k) -$ anuitet koji se isplaćuje u trenutku k ukoliko se do tad osigurani događaj nije još ostvario
3) $\pi(k)$ – premija koja se plaća u trenutku k ukoliko se do tad osigurani događaj nije još ostvario

Sada ako bismo odredili sadašnje vrednosti ovih novčanih tokova, ne bismo direktno dobili realne brojeve već slučajne promenljive. Za razliku od determinističkih modela gde se tačno zna dužina novčanih tokova, u stohastičkom modelu je ona neizvesna i zavisi od vremena ostvarenja osiguranog događaja, pa se u ovom slučaju moraju računati aktuarske sadašnje vrednosti. Aktuarske sadašnje vrednosti predstavljaju očekivane sadašnje vrednosti novčanih tokova.

Ako pretpostavimo da je diskontni faktor v fiksan, a da se beneficije i premije plaćaju na godišnjem nivou, dobijamo sledeće aktuarske sadašnje vrednosti prethodno definisanih novčanih tokova:

1) Aktuarska sadašnja vrednost vektora jednokratne beneficije $\boldsymbol{c}$, pri čemu je $c(k)$ beneficija isplaćena kada je $T = k$
$$A_\tau(\boldsymbol{c}, v) := \sum_{k=0}^{\tau} P\{T = k\} \cdot v^k \cdot c(k) = \sum_{k=0}^{\tau} p_k \cdot v^k \cdot c(k)$$
2) Aktuarska sadašnja vrednost vektora anuiteta $\boldsymbol{b}$, pri čemu je $b(k)$ anuitet isplaćen ako je $T > k$
$$A_\tau(\boldsymbol{b}, v) := \sum_{k=0}^{\tau} P\{T > k\} \cdot v^k \cdot b(k) = \sum_{k=0}^{\tau} S(k) \cdot v^k \cdot b(k)$$

3) Aktuarska sadašnja vrednost vektora premija $\boldsymbol{\pi}$, pri čemu je $\pi(k)$ premija plaćena ako je $T > k$

$$A_\tau(\boldsymbol{\pi}, v) := \sum_{k=0}^{\tau} P\{T > k\} \cdot v^k \cdot \pi(k) = \sum_{k=0}^{\tau} S(k) \cdot v^k \cdot \pi(k)$$

4.2. Aktuarske rezerve i princip ekvivalencije

U slučaju determinističkog modela za neku polisu osiguranja i proizvoljno $t \in \{0, 1, \dots, \tau\}$, aktuarske rezerve u trenutku t se definišu kao suma novca koja je potrebna osiguravajućoj kompaniji u trenutku t kako bi obezbedila ispunjavanje svojih budućih obaveza iz ugovora. Mi ćemo u ovom poglavlju rada predstaviti stohastički pristup rezervama u diskretnom modelu.

Fiksirajmo proizvoljno $r \in \{0, 1, \dots, \tau - 1\}$ i pretpostavimo da se do tog trenutka osiguravajući događaj nije ostvario, onda možemo definisati slučajnu promenljivu ${}_rL$ koja će predstavljati sadašnju vrednost neto novčanog toka u trenutku r. Neto novčani tok je razlika između svih budućih beneficija koje će osiguravajuća organizacija isplatiti i svih premija koje će biti naplaćene, a njegova vrednost zavisi od vremena ostvarenja osiguranog događaja. Za slučajnu promenjlivu T, ${}_rL$ možemo zapisati u sledećem obliku:

$${}_rL = b(r) + b_{r+1} \cdot v + \dots + b(T-1) \cdot v^{T-1} + c(T) \cdot v^T - (\pi(r) + \dots + \pi(T-1) \cdot v^{T-1})$$

Slučajna promenljiva ${}_rL$ se često naziva i potencijalni gubitak u trenutku r. Bitno je napomenuti i da ova slučajna promenljiva takođe može imati i negativne realizovane vrednosti i u tom slučaju ta realizovana vrednost predstavlja dobitak osiguravajuće kompanije.

Sada uvodimo definiciju rezervi u trenutku r za stohastički model na sledeći način:

$${}_rV = E({}_rL)$$

Dakle, aktuarske rezerve u trenutku r predstavljaju očekivanu vrednost potencijalnog gubitka u trenutku r.

Slučajna promenljiva ${}_0L$, koja se često obeležava samo sa L, je jednaka razlici sadašnje vrednosti svih budućih beneficija i sadašnje vrednosti svih budućih premija u toku trajanja polise. Po principu ekvivalencije se za premije uzimaju one vrednosti za koje je

$$E(L) = 0$$

Odnosno po principu ekvivalencije aktuarske rezerve moraju biti jednake nuli u trenutku kada polisa osiguranja počinje da važi.

4.3. Markovi procesi u osiguranju

U prethodnom poglavlju smo govorili o osiguranju u kome postoji jedan osigurani događaj koji drugim rečima predstavlja prelazak osiguranika iz jednog stanja u drugo stanje, npr. ako je osigurani događaj smrt osiguranika onda je prvo stanje živ a drugo stanje preminuo. Međutim, postoje mnogo kompleksnije polise osiguranja i u tom slučaju se za modeliranje kretanja osiguranika kroz stanja koriste stohastički procesi. Mi ćemo u ovom poglavlju predstaviti način računja aktuarskih vrednosti i premije kada stohastički procesi imaju svojstvo Markova. Definiciju i osnovne osobine Markovih procesa smo dali u glavi 4.

Jedan način uopštavanja je polisa u kojoj se isplaćuju različiti iznosi beneficija za različite uzroke smrti. Ako postoji m uzroka smrti, tada posmatramo Markov lanac sa $m+1$ stanja u kome za osiguranika u stanju 0 znači da nije preminuo, dok stanje $j,\ j = 1, 2, \dots, m$ predstavlja smrt usled uzroka j. Još kompleksniji model nastaje omogućavanjem većeg broja prelaza među stanjima. Na primer, ako se beneficije isplaćuju za razvoj neke izlečive bolesti, može se desiti da se osiguranik razboli i ozdravi više puta u toku trajanja polise osiguranja.

Pretpostavimo da se razvoj stanja osiguraka može predstaviti diskretnim procesom Markova sa konačnim skupom stanja $\mathcal{S} = \{0, 1, \dots, N\}$ pri čemu je stanje 0 početno stanje u kome osiguranik ne prima beneficije ali može da plaća premiju. U tom slučaju se Markov lanac za polisu osiguranja sa rokom prestanka pokrića τ definiše na sledeći način:

$$Z:\ T \times \mathcal{S} \to \mathcal{S},\ \ \mathcal{S} = \{0, 1, \dots, N\}, T = \{0, 1, \dots, \tau\}$$

Pri čemu $Z(t) = i$ znači da se osiguranik u trenutku t nalazi u stanju i.

Kao što smo rekli, za razliku od prethodnog poglavlja gde su se pojavljivala samo dva stanja, sada imamo više stanja a samim tim i više različitih novčanih tokova, pa zbog toga uvodimo nove oznake:

1) $c_{ij}(t) -$ iznos beneficije koju će osiguranik primiti za prelazak iz stanja i u stanje j u trenutku t, $i, j \in \{0, 1, \dots, N\}$, $i < j$
2) $b_i(t) -$ vrednost anuiteta u trenutku t koji osiguranik prima dok je u stanju i, $i \in \{1, \dots, N\}$
3) $\pi_i(k)$ – vrednost premije koju osiguranik plaća u trenutku k ukoliko se u tom trenutku nalazi u stanju i, $i \in \{0, 1, \dots, N\}$

Princip računanja aktuarskih sadašnjih vrednosti je sličan, s tim da ćemo prvo definisati verovatnoće prelaza. Verovatnoća prelaza predstavlja verovatnoću da se osiguranik u trenutku t nalazi u stanju j, ako se u trenutku s nalazio u stanju i, i za nju ćemo uvesti oznaku:

$$P_{ij}(s,t) = P\{Z(t) = j | Z(s) = i\}$$

Pretpostavimo da se u trenutku $t = 0$ osiguranik nalazi u početnom stanju 0 u kome ne prima beneficije, onda dobijamo da su aktuarske sadašnje vrednosti u zavisnosti od stanja date pomoću sledećih formula:

1) Aktuarska sadašnja vrednost vektora beneficije $\boldsymbol{c_{ij}}$ u trenutku $t = 0$

$$A_\tau(\boldsymbol{c_{ij}}, v) := \sum_{t=1}^{\tau} P_{0i}(0, t-1) \cdot P_{ij}(t-1, t) \cdot v^t \cdot c_{ij}(t)$$

2) Aktuarska sadašnja vrednost vektora beneficije $\boldsymbol{b_i}$ u trenutku $t = 0$

$$A_\tau(\boldsymbol{b_i}, v) := \sum_{t=1}^{\tau} P\{Z(t) = i\} \cdot v^t \cdot b_i(t)$$

3) Aktuarska sadašnja vrednost vektora premije $\boldsymbol{\pi_i}$ u trenutku $t = 0$

$$A_\tau(\boldsymbol{\pi_i}, v) := \sum_{t=1}^{\tau} P\{Z(t) = i\} \cdot v^t \cdot \pi_i(t)$$

Nakon računanja aktuarskih sadašnjih vrednosti beneficija i premije za svako stanje posebno, neophodno je izračunati i aktuarske sadašnje vrednosti ukupnih beneficija i ukupnih premija, tj. računamo sledeće sume:

1) Aktuarska sadašnja vrednost ukupnih beneficija $\boldsymbol{B}$ u trenutku $t = 0$

$$A_\tau(\boldsymbol{B}, v) := \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} A_\tau(\boldsymbol{c_{ij}}, v) + \sum_{i=1}^{N} A_\tau(\boldsymbol{b_i}, v)$$

2) Aktuarska sadašnja vrednost ukupnih premija $\boldsymbol{\Pi}$ u trenutku $t = 0$

$$A_\tau(\boldsymbol{\Pi}, v) := \sum_{i=0}^{N} A_\tau(\boldsymbol{\pi_i}, v)$$

Podsetimo se, po principu ekvivalencije aktuarska sadašnja vrednost rezervi mora biti jednaka nuli na početku važenja polise što se drugim rečima znači da aktuarska sadašnja vrednost ukupnih beneficija i aktuarska sadašnja vrednost ukupnih premija moraju biti jednake. Dakle po principu ekvivalencije, premija se dobija rešavanjem sledeće jednačine:

$$A_\tau(\mathbf{B}, v) = A_\tau(\boldsymbol{\Pi}, v)$$

5. Primena

Danas postoji mnogo vrsta različitih osiguranja. Kako bismo napravili model koji prikazuje primenu teorije koja je razrađena u prethodnim poglavljima, mi ćemo u ovom radu predstaviti osiguranje za dugoročnu negu u Nemačkoj. U Nemačkoj je 1995. godine uvedeno obavezno osiguranje za dugoročnu negu. Prilikom zaključivanja ovakvog tipa osiguranja, osiguravač sklapa polisu sa osiguranikom bez prethodnih saznanja o zdravstvenom stanju osiguranika. Beneficije od osiguranja se isplaćuju ukoliko se osiguranik nađe u zdravstvenom stanju koje zahteva negu medicinskog osoblja u vidu kućnih poseta ili prelazak u dom za zdravstvenu negu. Visina beneficije zavisi od obima i oblika neophodne nege. Naš cilj je procena uticaja teškoće bolesti, oblika neophodne nege (kod kuće ili u domu), godina i pola osiguranika na trajanje zahteva za negu. U ove svrhe ćemo koristiti Koksov poluparametarski proporcionalni hazardni model. Ovaj model omogućava uključivanje cenzurisanih podataka u procenu što je veoma bitno jer oni imaju veliki udeo u dostupnim bazama podataka. Takođe je pogodan zbog mogućnosti upotrebe vremenski zavisnih promenljivih kao što su u našem slučaju teškoća bolesti, oblik neophodne nege i godine osiguranika. Pomoću Koksovog modela takođe možemo oceniti i intenzitete prelaza među nivoima. Svrha ovog poglavlja je prikazivanje kako se koristeći ocenjenih intenziteta prelaza i Markovog modela sa više stanja može doći do formule za računanje premije za plan pomenutog tipa osiguranja.

Dugoročna nega podrazumeva pružanje neophodne nege i podrške osiguranicima koji na osnovu svog stanja imaju pravo na nju. Nega se može pružati kod kuće i u specijalizovanom domu za zdravstvenu negu, a prema dužini vremena neophodnog za pružanje nege na dnevnom nivou razlikujemo tri nivoa:

1) Prvi nivo nege: Obolelom osiguraniku je potrebno najmanje 90 minuta dnevno za obavljanje svakodnevnih aktivnosti (odlazak u krevet, kupanje, ishrana...)
2) Drugi nivo nege: Obolelom osiguraniku je potrebno najmanje 180 minuta dnevno za obavljanje svakodnevnih aktivnosti
3) Treći nivo nege: Obolelom osiguraniku je potrebno najmanje 300 minuta dnevno za obavljanje svakodnevnih aktivnosti

5.1. Koksov model

Za svakog pacijenta postoje podaci o datumu rođenja, polu pacijenta, datumu kada je pacijent zatražio negu i o obliku i nivou nege koju je zatražio. U zavisnosti od daljeg razvoja stanja pacijenta na raspolaganju mogu biti podaci o promeni nivoa, odnosno oblika nege kao i datum te promene, datum smrti i eventualno datum prestanka pružanja nege zbog ozdravljenja ili poboljšanja zdravstvenog stanja. Primer raspoložive baze podataka neke osiguravajuće organizacije je dat u sledećoj tabeli:

<table>
<tr><th>Pacijent</th><th>Datum rođenja</th><th>Pol</th><th>Datum smrti</th><th>Početna nega</th></tr>
<tr><td>1</td><td>12/20/1963</td><td>Ženski</td><td>-</td><td>1. nivo kod kuće</td></tr>
<tr><td rowspan="2">2</td><td>1/10/1954</td><td>Muški</td><td>-</td><td>2. nivo u domu</td></tr>
<tr><td colspan="4">- Datuma 4/16/1997 pacijent 2 je ozdravio</td></tr>
<tr><td>3</td><td>3/3/1965</td><td>Muški</td><td>12/31/1996</td><td>3. nivo u domu</td></tr>
<tr><td rowspan="4">4</td><td>1/2/1975</td><td>Ženski</td><td>12/4/1998</td><td>1. nivo kod kuće</td></tr>
<tr><td colspan="4">- Datuma 8/30/1996 pacijent 4 je prešao na 1. nivo nege u domu</td></tr>
<tr><td colspan="4">- Datuma 7/17/1997 pacijent 4 je prešao na 2. nivo nege u domu</td></tr>
<tr><td colspan="4">- Datuma 9/1/1998 pacijent 4 je prešao na 3. nivo nege u domu</td></tr>
<tr><td rowspan="4">5</td><td>3/30/1967</td><td>Muški</td><td>-</td><td>1. nivo kod kuće</td></tr>
<tr><td colspan="4">- Datuma 12/1/1996 pacijent 5 je prešao na 2. nivo nege kod kuće</td></tr>
<tr><td colspan="4">- Datuma 12/1/1997 pacijent 5 je prešao na 3. nivo nege kod kuće</td></tr>
<tr><td colspan="4">- Datuma 2/1/1998 pacijent 5 je prešao na 3. nivo nege u domu</td></tr>
<tr><td>6</td><td>4/1/1940</td><td>Muški</td><td>-</td><td>3. nivo u domu</td></tr>
<tr><td rowspan="2">7</td><td>4/5/1939</td><td>Muški</td><td>7/5/1998</td><td>2. nivo u domu</td></tr>
<tr><td colspan="4">- Datuma 2/15/1998 pacijent 7 je prešao na 3. nivo nege u domu</td></tr>
<tr><td rowspan="2">8</td><td>12/5/1950</td><td>Ženski</td><td>1/12/1997</td><td>1. nivo u domu</td></tr>
<tr><td colspan="4">- Datuma 1/31/1996 pacijent 8 je prešao na 3. nivo nege u domu</td></tr>
<tr><td>9</td><td>4/2/1980</td><td>Ženski</td><td>-</td><td>1. nivo u domu</td></tr>
<tr><td>10</td><td>7/4/1971</td><td>Ženski</td><td>10/3/1996</td><td>2. nivo u domu</td></tr>
</table>

Tabela 5.1.1: Podaci o korisnicima osiguranja za dugoročnu negu

Ovakav skup podataka odgovara skupu podataka koji se dobija u longitudinalnim studijama sa unapred definisanim datumom završetka posmatranja. Vreme preživljavanja u ovom primeru predstavlja vreme trajanja nege pacijenta koje ćemo izražavati u godinama, dakle definišemo:

$T :=$ broj godina tokom kojih je pacijentu neophodna nega

Pacijenti koji su na kraju perioda posmatranja ostali u nekom nivou i obliku nege predstavljaju cenzurisana posmatranja. S obzirom da je potpuni oporavak pacijenta, odnosno prestanak potrebe za negom zbog poboljšanja stanja veoma redak mi ćemo pretpostaviti da je nemoguć a takve podatke ćemo smatrati cenzurisanim podacima.

Za analizu ovih podataka i procenu trajanja nege pacijenta možemo koristiti Koksov model sa sledećim vremenski zavisnim promenljivama:

- X_{pol} – pol pacijenta ($X_{pol} = 1$ za ženski pol a $X_{pol} = 0$ za muški pol)
- $X_{god}(t)$ – starost pacijenta u trenutku promene stanja
- $X_{n1}(t)$ – 1. nivo nege u trenutku t ($X_{n1}(t) = 1$ ako se pacijentu pruža 1. nivo nege a u suprotnom je $X_{n1}(t) = 0$)
- $X_{n2}(t)$ – 2. nivo nege u trenutku t ($X_{n2}(t) = 1$ ako se pacijentu pruža 2. nivo nege a u suprotnom je $X_{n2}(t) = 0$)
- $X_{dom}(t)$ – promenljiva koja pokazuje mesto nege ($X_{dom}(t) = 1$ ako je pacijent smešten u domu za negu, a $X_{dom}(t) = 0$ ako mu se nega pruža kod kuće

Primetimo da ukoliko su $X_{n1}(t) = X_{n2}(t) = 0$, pacijent je u 3. nivou nege. Na osnovu ovako definisanih promenljivih dobijamo Koksov model sledećeg oblika:

$$h(t\,|\boldsymbol{X}) = h_0(t) \cdot exp\{b_1 \cdot X_{n1}(t) + b_2 \cdot X_{n2}(t) + b_3 \cdot X_{dom}(t) + b_4 \cdot X_{god}(t) + b_5 \cdot X_{pol}\}$$

Kao što vidimo, ovaj model sadrži vremenski zavisne promenljive, pa se pre obrade podataka, tabela mora transformisati na način opisan u poglavlju 2.4. U tabeli 5.1.2 su prikazani podaci u obliku pogodnom za dalji rad.

Pacijent	**Start**	**Stop**	X_{pol}	X_{god}	X_{n1}	X_{n2}	X_{dom}	**Status**
1	0,00	3,75	1	31,30	1	0	0	0
2	0,00	2,04	0	41,25	0	1	1	0
3	0.00	1.75	0	30.10	0	0	1	1
4	0.00	1.41	1	20.26	1	0	0	0
	1.41	2.29	1	21.67	1	0	1	0
	2.29	3.42	1	22.55	0	1	1	0
	3.42	3.68	1	23.68	0	0	1	1
5	0.00	1.67	0	28.02	1	0	0	0
	1.67	2.67	0	29.69	0	1	0	0
	2.67	2.84	0	30.69	0	0	0	0
	2.84	3.75	0	30.86	0	0	1	0
6	0.00	3.75	0	55.04	0	0	1	0
7	0.00	2.88	0	56.03	0	1	1	0
	2.88	3.26	0	58.90	0	0	1	1
8	0.00	0.83	1	44.35	1	0	1	0
	0.83	1.79	1	45.18	0	0	1	1
9	0.00	3.75	1	15.01	1	0	1	0
10	0.00	1.51	1	23.76	0	1	1	1

Tabela 5.1.2: Podaci o osiguranicima razloženi na vremenske intervale u kojima su vrednosti svih promenljivih fiksne

Nakon ocene koeficijenata Koksovog modela i osnovne hazardne stope, mogu se upoređivati hazardne stope pacijenata koji imaju različite vrednosti nekih promenljivih i na taj način izvoditi određeni zaključci. Na primer, ako uporedimo dva pacijenta, ženu i muškarca starosti 50 godina koji se nakon 6 meseci od dana zahtevanja nege nalaze u domu, pri čemu im je neophodan 2. stepen nege dobijamo:

$$h(0{,}5|\boldsymbol{X}_{\boldsymbol{žena}}) = \widetilde{h_0}(0{,}5) \cdot exp\{\widetilde{b_2} + \widetilde{b_3} + \widetilde{b_4} \cdot 50 + \widetilde{b_5}\}$$

$$h(0{,}5|\boldsymbol{X}_{\boldsymbol{muškarac}}) = \widetilde{h_0}(0{,}5) \cdot exp\{\widetilde{b_2} + \widetilde{b_3} + \widetilde{b_4} \cdot 50\}$$

Odnosno dobijamo da je odnos njihovih hazardnih stopa:

$$\frac{h(0{,}5|\boldsymbol{X}_{\boldsymbol{žena}})}{h(0{,}5|\boldsymbol{X}_{\boldsymbol{muškarac}})} = exp\{\widetilde{b_5}\}$$

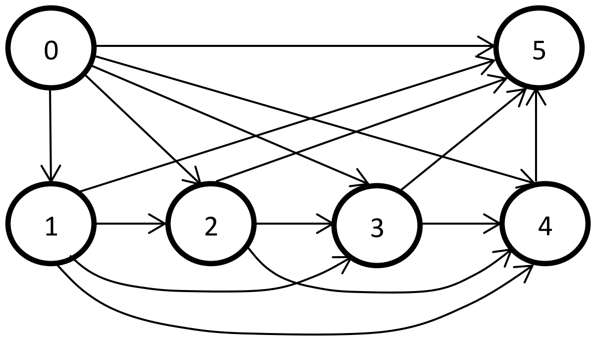
A zatim na osnovu ocenjene vrednosti koeficijenta b_5 možemo zaključiti da li pol pacijenta utiče na dužinu njegovog zahteva za negu. Način izvođenja zaključka u zavisnosti od vrednosti koeficijenta je opisan u poglavlju 2.1. ovog rada.

5.2. Markov model

Promene stanja pacijenta, odnosno prelazi se mogu predstaviti pomoću Markovog modela. Zbog načina isplaćivanja beneficija koji ćemo objasniti u narednom poglavlju, u ovom radu ćemo koristiti 6 mogućih stanja osiguranika. Dakle, skup stanja $\mathcal{S} = \{0, 1, 2, 3, 4, 5\}$ ćemo definisati na sledeći način:

1) $Stanje\ 0$:= osiguranik je zdrav
2) $Stanje\ 1$:= osiguraniku je potreban 1. nivo nege kod kuće
3) $Stanje\ 2$:= osiguraniku je potreban 2. nivo nege kod kuće
4) $Stanje\ 3$:= osiguraniku je potreban 3. nivo nege kod kuće
5) $Stanje\ 4$:= osiguraniku je potrebna nege u domu
6) $Stanje\ 5$:= osiguranik je preminuo

Primetimo da stanje 4 obuhvata sva tri nivoa nege a redosled po kom su definasana stanja $1, 2, 3$ i 4 odgovara obimu neophodne nege i količini sredstava da se ona pruži u datom stanju, odnosno dati redosled stanja odražava rastući poredak težine stanja pacijenta. Pošto se u podacima retko sreću poboljšanja stanja, mi ćemo pretpostaviti da su ona nemoguća i da je moguć prelaz samo u teže stanje, pri čemu je stanje 5 apsorbujuće stanje. Ovako definisani prelazi se predstavljaju sledećim Markovim modelom:



Slika 5.2.1: Markov model osiguranja za dugoročnu negu

Sada nas zanimaju ocene intenziteta prelaza $\mu_{ij},\ i,j \in \mathcal{S}$. Možemo primetiti da hazardne stope po definiciji odgovaraju intenzitetima prelaza iz stanja "$živ$" u stanje "$preminuo$", što nam omogućava da i intenzitete prelaza ocenimo pomoću Koksovog modela. Prelazi $1 \to 5$, $2 \to 5$, $3 \to 5$ i $4 \to 5$ se mogu direktno dobiti iz ocenjenog Koksovog modela iz prethodnog poglavlja fiksiranjem vrednosti promenljivih $X_{n1}(t),\ X_{n1}(t)$ i $X_{dom}(t)$ na sledeći način:

$$X_{n1}(t) = 1, X_{n2}(t) = X_{dom}(t) = 0 \text{ za } \mu_{15}$$

$$X_{n2}(t) = 1, X_{n1}(t) = X_{dom}(t) = 0 \text{ za } \mu_{25}$$

$$X_{n1}(t) = X_{n2}(t) = X_{dom}(t) = 0 \text{ za } \mu_{35}$$

$$X_{dom}(t) = 1, X_{n1}(t) = X_{n2}(t) = 0 \text{ za } \mu_{45}$$

Intenziteti prelaza iz stanja 0 u ostala stanja se ne mogu odrediti na osnovu podataka iz internih izvora kompanije, pa se za μ_{05} koriste javno dostupne godišnje stope mortaliteta a za prelaze iz stanja 0 u stanja $1, 2, 3$ i 4 se mogu upotrebiti godišnje stope realizacije zahteva za negu iz nekih eksternih izvora.

Ostali intenziteti prelaza se ocenjuju konstrukcijom novog Koksovog modela, pri čemu se dobijaju i nove vrednosti koeficijenata b_i. Prilikom konstrukcije Koksovog modela se uzimaju promenljive $X_{god}(t)$ i $X_{pol}(t)$, tj. modeli će biti oblika:

$$\mu_{ij}(t\,|\boldsymbol{X}) = \mu_{ij0}(t) \cdot exp\{b_1 \cdot X_{god}(t) + b_2 \cdot X_{pol}\}$$

Pre ocenjivanja koeficijenata, neophodno je podatke “prilagoditi” za svaki intenzitet posebno. Na primer, ukoliko ocenjujemo intenzitet prelaza μ_{12} vreme preživljavanja T predstavlja vreme koje protekne od trenutka kada je pacijentu pružena nega prvog nivoa kod kuće do trenutka kad je pacijent prešao u drugi nivo kod kuće. Dakle, neophodno je formirati podskup pacijenata koji su se tokom perioda posmatranja našli u stanju 1 u bilo kom trenutku, pri čemu je njihovo zdravstveno stanje do tog trenutka nebitno, odnosno smatra se da su oni ušli u studiju u trenutku ulaska u stanje 1. U ovom slučaju svi pacijenti koji tokom perioda posmatranja nisu ušli u stanje 2 predstavljaju cenzurisane podatke. U narednoj tabeli smo prikazali kako ovaj postupak izgleda na našem primeru prilikom ocene intenziteta prelaza μ_{12}.

Pacijent	Start	Stop	X_{pol}	X_{god}	X_{n1}	X_{n2}	X_{dom}	Status
1	0,00	3,75	1	31,30	1	0	0	0
4	0.00	1.41	1	20.26	1	0	0	0
5	0.00	1.67	0	28.02	1	0	0	1

Tabela 5.2.1: Podaci o pacijentima koji su se nalazili u stanju 1 tokom vremena posmatranja

5.3. Premija

Pretpostavimo da osiguravajuća organizacija nudi osiguranje za dugoročnu negu u kome se premije plaćaju na godišnjem nivou u fiksnom iznosu. S druge strane, osiguravajuća organizacija se obavezuje da osiguraniku u zavisnosti od nivoa nege isplaćuje različite iznose beneficije. Unapred se ugovara ukupan godišnji iznos sume koji će biti isplaćivan ukoliko je pacijent primljen u dom za negu, a za niže nivoe nege se isplaćuje deo te ugovorene sume po sledećem pravilu:

1) Za prvi nivo nege kod kuće isplaćuje se 25% od ugovorene sume
2) Za drugi nivo nege kod kuće isplaćuje se 50% od ugovorene sume
3) Za treći nivo nege kod kuće isplaćuje se 75% od ugovorene sume

Novčani tokovi koji se pojavljuju u ovom obliku osiguranja su:

- π – fiksna godišnja premija koju plaća osiguranik
- b_i – ukupna suma svih beneficija koju osiguranik dobija na godišnjem nivou dok je u stanju i, $i = 1, 2, 3, 4$

Kako bi se mogao primeniti način računanja aktuarksih sadašnjih vrednosti novčanih tokova iz poglavlja 4.3., neophodno je prvobitno utvrditi vreme završetka polise osiguranja. Za vreme završetka polise se uzima aktuarska starost pacijenta pri smrti koju ćemo obeležavati sa w. Dakle, w predstavlja starost pacijenta, izraženu u godinama, za koju važi da je $P(pacijent\ živi\ duže\ od\ w\ godina) = 0$

Sada možemo precizno definisati vreme trajanja polise. To je vreme koje protekne od trenutka zaključivanja ugovora do smrti osiguranika, tj. ako sa θ obeležimo period trajanja polise onda je $\theta = [0, w - x)$. Kako bi se primenio princip ekvivalencije neophodno je utvrditi aktuarske vrednosti svih novčanih tokova u trenutku izdavanja polise osiguranja. Sa $A_{w-x}(\mathbf{\Pi}, v)$ ćemo označiti aktuarsku vrednost svih premija plaćenih u toku trajanja polise a sa $A_{w-x}(\mathbf{B}, v)$ aktuarsku vrednost ukupno isplaćenih beneficija za negu na nivou i u toku trajanja polise. Ako pretpostavimo da nam je data fiksna godišnja diskontna stopa v, dobijamo sledeće formule za aktuarske sadašnje vrednosti novčanih tokova na početku perioda trajanja polise:

$$A_\tau(\mathbf{\Pi}, v) = \sum_{t=0}^{w-x-1} P\{osiguranik\ je\ zdrav\ u\ periodu\ [0,t]\} \cdot v^t \cdot \pi$$

$$A_\tau(\mathbf{B}, v) = \sum_{t=1}^{w-x-1} P\{osiguranik\ je\ \ zdrav\ u\ \ s=0, a\ u\ s=t\ je\ u\ stanju\ i\} \cdot v^t \cdot b_i$$

Nakon određivanja aktuarskih vrednosti, primenom principa ekvivalencije dobija se iznos godišnje premije π:

$$\pi = \frac{B_{0,b_1}(0) + B_{0,b_2}(0) + B_{0,b_3}(0) + B_{0,b_4}(0)}{\sum_{t=0}^{w-x-1} P\{osiguranik\ je\ zdrav\ u\ periodu\ [0,t]\} \cdot v^t}$$

Postupak računanja verovatnoća koje se nalaze u prethodnim formulama ćemo objasniti u narednom poglavlju.

5.4. Računanje diskretnih verovatnoća prelaza

Kao što smo rekli, stanje osiguranika predstavlja Markov proces sa skupom stanja $\mathcal{S} = \{0,1,2,3,4,5\}$ i verovatnoćama prelaza $P_{ij}(s,t)$.

U ovom poglavlju ćemo opisati jedan način na koji se mogu izračunavati verovatnoće:

1) $P\{osiguranik\ je\ zdrav\ u\ periodu\ [0,t]\} = P_{\underline{00}}(0,t) = P\{Z(t) = 0|Z(0) = 0\}$
2) $P\{osiguranik\ je\ \ zdrav\ u\ \ s = 0, a\ u\ trenutku\ s = t\ je\ u\ stanju\ i\} = P_{0i}(0,t) = P\{Z(t) = i|Z(0) = 0\}$

Pošto se u životnom osiguranju premije računaju na osnovu godišnjih stopa prelaza, neophodno je prvo odrediti godišnje verovatnoće prelaza. Godišnja verovatnoća prelaza je verovatnoća da se osiguranik u narednoj godini nalazi u stanju j, ako se u godini t nalazio u stanju i, a označavaćemo je sa:

$p_{ij}(t) = P\{Z(t+1) = j|Z(t) = i\}$

Na osnovu formule iz 1.3.1 iz prvog poglavlja se može pokazati da se godišnje verovatnoće prelaza mogu dobiti na osnovu intenziteta prelaza pomoću sledeće formule:

$$p_{ij}(t) = 1 - \prod_{t \le t_k < t+1} \left(1 - \mu_{ij}(t_k)\right) = \sum_{t \le t_k < t+1} \mu_{ij}(t_k) \cdot \prod_{t \le t_s < t_k} \left(1 - \mu_{ij}(t_s)\right)$$

Primetimo da stope prelaza $\mu_{ij}(t_k)$ zavise od starosti i pola, što znači da se verovatnoće računaju posebno za oba pola i za različite godine starosti, odnosno dobijaju se premije u zavisnosti od pola i starosti osiguranika.

Sada nakon dobijanja godišnjih verovatnoća prelaza izvešćemo formule za dobijanja verovatnoća prelaza $P_{ij}(s,t)$ koje su nam potrebne za računanje premije. Pošto smo za naš Markov model pretpostavili da je moguć samo prelazak u gora stanja, za verovatnoće zadržavanja stanja $P_{\underline{ii}}(t,t+1)$, $i \in \{0,1,2,3,4\}$ važi sledeće:

$$P_{\underline{ii}}(t,t+1) = 1 - \left(P_{ii+1}(t,t+1) + \cdots + P_{i5}(t,t+1)\right) = 1 - \sum_{i<j\le 5} P_{ij}(t,t+1)$$

Dalje, kako po definiciji važi jednakost $P_{ii}(t,t+1) = p_{ii}(t)$, sledi da je:

$$P_{\underline{ii}}(t,t+1) = 1 - \sum_{i<j\le 5} p_{ij}(t)$$

A zatim koristeći jednačine Chapman-Kolmogorova i rekurzivne veze možemo izvesti i za proizvoljne godine:

$$P_{\underline{ii}}(s,t) = \prod_{k=0}^{t-s-1} P_{\underline{ii}}(s+k, s+k+1) = \prod_{k=0}^{t-s-1} \left(1 - \sum_{i<j\le 5} p_{ij}(s+k)\right)$$

Na osnovu osnovnih osobina verovatnoće se izvode i formule za verovatnoće prelaza u gora stanja, odnosno sve verovatnoće koje su potrebne za računanje premije za naš model se mogu izračunati pomoću sledećih formula:

$$P_{\underline{00}}(0,t) = \prod_{k=0}^{t-1} \left(1 - p_{01}(k) - p_{02}(k) - p_{03}(k) - p_{04}(k) - p_{05}(k)\right)$$

$$P_{\underline{11}}(0,t) = \prod_{k=0}^{t-1} \left(1 - p_{12}(k) - p_{13}(k) - p_{14}(k) - p_{15}(k)\right)$$

$$P_{\underline{22}}(0,t) = \prod_{k=0}^{t-1} \left(1 - p_{23}(k) - p_{24}(k) - p_{25}(k)\right)$$

$$P_{\underline{33}}(0,t) = \prod_{k=0}^{t-1} \left(1 - p_{34}(k) - p_{35}(k)\right)$$

$$P_{\underline{44}}(0,t) = \prod_{k=0}^{t-1}\big(1 - p_{45}(k)\big)$$

$$P_{01}(0,t) = \sum_{k=0}^{t-1} P_{\underline{00}}(0,k) \cdot p_{01}(k) \cdot P_{\underline{11}}(k+1,t)$$

$$P_{02}(0,t) = \sum_{k=0}^{t-1} \Big(P_{\underline{00}}(0,k) \cdot p_{02}(k) + P_{01}(0,k) \cdot p_{12}(k)\Big) \cdot P_{\underline{22}}(k+1,t)$$

$$P_{03}(0,t) = \sum_{k=0}^{t-1} \Big(P_{\underline{00}}(0,k) \cdot p_{03}(k) + P_{01}(0,k) \cdot p_{13}(k) + P_{02}(0,k) \cdot p_{23}(k)\Big) \cdot P_{\underline{33}}(k+1,t)$$

$$P_{04}(0,t) = \sum_{k=0}^{t-1} \Big(P_{\underline{00}}(0,k) \cdot p_{04}(k) + P_{01}(0,k) \cdot p_{14}(k) + P_{02}(0,k) \cdot p_{24}(k) + P_{03}(0,k) \cdot p_{34}(k)\Big) \cdot P_{\underline{44}}(k+1,t)$$

$$P_{05}(0,t) = 1 - P_{\underline{00}}(0,t) - P_{01}(0,t) - P_{02}(0,t) - P_{03}(0,t) - P_{04}(0,t)$$

Zaključak

Postoje mnogi drugi načini za formiranje premije u kojima se uzima u obzir stohastička priroda gubitka. Osim principa ekvivalencije postoje i princip standardne devijacije, princip varijanse i drugi. Može se pronaći obimna literatura koja se bavi upoređivanjem različitih principa i pronalaskom najboljih formula za računanje premije.

Mi smo u ovom radu koristili Markov proces kako bismo predstavili više stanja u modelu osiguranja. Kao što smo rekli, za primer smo uzeli osiguranje za dugoročnu negu u Nemačkoj, međutim uslovi iz polise mogu biti drugačije ugovoreni pa se tada i Markov model mora izmeniti u vidu promene broja stanja ili tipa prelaza. Takođe je bitno napomenuti da postoje i drugi načini ocena intenziteta i verovatnoća prelaza iz Markovog modela. Koksov model koji smo mi koristili je dosta komplikovan pristup zbog potrebe za transformacijom podataka, ali postoje statistički paketi koji olakšavaju računanje hazardnih stopa a najpopularniji je program R.

Literatura

[1] C. CZADO, F. RUDOLPH, *Application of Survival Analysis Methods in Long Term Care Insurance*, Insurance: Mathematics and Economics 31, 395-413, 2002.

[2] T. KONSTANTOPOULOS, *Notes on Survival Models*, School of Mathematical and Computer Sciences, Heriot-Watt University, 2006.

[3] T. M. THERNEAU, P. M. GRAMBSCH, *Modeling Survival Data: Extending the Cox Model*, Springer – Verlag, New York, 2000.

[4] E. T. LEE, J. WENYU WANG, *Statistical Methods for Survival Data Analysis*, *Third Edition*, J.Wiley & Sons, Inc., Publication, New Jersey, 2003.

[5] S. D. PROMISLOW, *Fundamentals of Actuarial Mathematics*, J.Wiley & Sons, Ltd, England, 2006.

[6] D. G. KLEINBAUM, M. KLEIN, *Survival Analysis, A Self – Learning Text, Second Edition*, Springer Science+Business Media, Inc., New York, 1996

[7] F. HELMS, *Estimating LTC Premiums using GEEs for Pseudo – Values*, Technical University Munich, 2003

[8] A. WIENKE, *Frailty Models in Survival Analysis*, Chapman and Hall/CRC Biostatistics Series, 2010.

[9] T. THERNEAU, C. CROWSON, *Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model*, Mayo Clinic, 2015.

[10] T. M. POWELL, M. E. BAGNELL, *Your "Survival" Guide to Using Time – Dependent Covariates*, SAS Global Forum, page 168, 2012.

[11] L. THOMAS, E. M. REYES, *Tutorial: Survival Estimation for Cox Regression Models with Time – Varying Coefficients Using SAS and R*, Journal of Statistical Software, Volume 61, October 2014.

[12] B. BAJIĆ, M. KRESOJA, *Analiza preživljavanja*, seminarski rad, Prirodno – matematički fakultet, Univerzitet u Novom Sadu, 2011.

[13] D. R. COX, *Partial likelihood*, Biometrika 62, 269-276, 1975.

[14] J. P. KLEIN, *Survival Distributions and Their Characteristics*, Encyclopedia of Biostatistics, 2005.

[15] F. HELMS, C. CZADO, S. GSCHLÖßL, *Calculation of LTC Premiums based on direct estimates of transition probabilities*, ASTIN Bulletin, 35, 455-469, 2005.

[16] L. MEIRA-MACHADO, J. DE UNA/ALVAREZ, C. CADARSO-SUAREZ, P. K. ANDERSEN, *Multi – state models for the analysis of time – to – event data*, Statistical methods in medical research 18(2), 195-222, 2009.

[17] www.wikipedia.org

[18] sada2013.sciencesconf.org/16748/document

[19] www.public.iastate.edu/~kkoehler/stat565/coxph.4page.pdf

Biografija

Kristina Popadić je rođena 28.05.1990. u Subotici. Završila je Osnovnu školu „Miloš Crnjanski“ kao nosilac Vukove diplome, a zatim upisala i završila kao đak generacije ekonomsku srednju školu „Bosa Milićević“ u Subotici. 2009. godine je upisala osnovne akademske studije Primenjene matematike na Prirodno – matematičkom fakultetu u Novom Sadu, modul Matematika finansija. Nakon zavšenih osnovnih studija, 2012. godine upisavši master studije istog usmerenja, nastavila je svoje studiranje na Prirodno – matematičkom fakultetu. Položila je sve ispite predviđene nastavnim planom i programom uz prosečnu ocenu 9.71 i time stekla uslov za odbranu master rada.

Novi Sad, 2015. Kristina Popadić

UNIVERZITET U NOVOM SADU
PRIRODNO – MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR

Identifikacioni broj:
IBR

Tip dokumentacije: Monografska dokumentacija
TD

Vrsta rada: Master rad
VR

Autor: Kristina Popadić
AU

Mentor: prof. dr Dora Seleši
MN

Naslov rada: Analiza preživljavanja sa primenama u zdravstvenom osiguranju
NR

Jezik publikacije: srpski (latinica)
JP

Jezik izvoda: s / e
JI

Zemlja publikovanja: Srbija
ZP

Uže geografsko područje: Vojvodina
UGP

Godina: 2015.
GO

Izdavač: Autorski reprint
IZ

Mesto i adresa: Novi Sad, Departman za matematiku I informatiku, Prirodno – matematički fakultet, Univerzitet u Novom Sadu, Trg Dositeja Obradovića 4
MA

Fizički opis rada: (5, 56, 0, 7, 5, 0, 0) – (broj poglavlja, strana, literalnih citata, tabela, slika, grafika, priloga)
FO

Naučna oblast: Matematika
NO

Naučna disciplina: Aktuarska matematika
ND

Predmetna odrednica / Ključne reči: Funkcija preživljavanja, Hazardna funkcija, Koksov model, Lanac Markova, jednačine Chapman-Kolmogorova, Intenziteti prelaza, Aktuarske sadašnje vrednosti, Aktuarske rezerve, Princip ekvivalencije
PO, UDK

Čuva se: Biblioteka Departmana za matematiku i informatiku Prirodno – matematičkog fakulteta Univerziteta u Novom Sadu
ČU

Važna napomena:
VN

Izvod: Cilj ovog rada je predstavljanje matematičkih modela koji se mogu koristiti za računanje premija u osiguranju. U praksi postoje različiti principi računanja premije, a u ovom radu je korišćen princip ekvivalencije po kome vrednosti ukupno naplaćenih premija i ukupno isplaćenih beneficija moraju biti jednake u trenutku izdavanja polise osiguranja. Kako je u tom trenutku iznos ukupno isplaćenih beneficija nepoznat, zadatak aktuara je da predviđanjem trenutka ostvarenja osiguranog događaja odrede njihov očekivani iznos. U tu svrhu se vreme do ostvarenja osiguranog događaja predstavlja kao vreme preživljavanja i koriste se modeli iz analize preživljavanja. U ovom radu su nakon uvodnog predstavljanja analize preživljavanja opisani Koksov model i Markov model, a zatim je prikazan način računanja aktuarskih sadašnjih vrednosti novčanih tokova u osiguranju i izvedena je formula za premiju osiguranja za dugoročnu negu u Nemačkoj.
IZ

Datum prihvatanja teme od strane NN veća: 2.3.2015
DP

Datum odbrane: Septembar 2015.
DO

Članovi komisije:
KO

Predsednik: **dr Danijela Rajter Ćirić**, redovni profesor,
Prirodno – matematički fakultet,
Univerzitet u Novom Sadu,

Mentor: **dr Dora Seleši**, vanredni profesor,
Prirodno – matematički fakultet,
Univerzitet u Novom Sadu,

Član: **dr Sanja Rapajić**, vanredni profesor,
Prirodno – matematički fakultet,
Univerzitet u Novom Sadu.

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCE
KEY WORDS DOKUMENTATION

Accession number:
ANO

Identification number:
INO

Document type: Monograph type
DT

Type of record: Master’s thesis
CC

Author: Kristina Popadić
AU

Mentor: Dora Seleši, Ph.D.
MN

Title: Survival analysis with applications in health care insurance
TI

Language of text: Serbian
LT

Language of abstract: English
LA

Country of publication: Serbia
CP

Locality of publication: Vojvodina
LP

Publication year: 2015.
PY

Publisher: Author’s reprint
PU

Publ. place: Novi Sad, Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad, Trg Dositeja Obradovića 4
PP

Physical description: (5, 56, 0, 7, 5, 0, 0) – (number of chapters, pages, references, tables, pictures, graphs, additional lists)
PD

Scientific field: Mathematics
SF

Scientific discipline: Actuaial Mathematics
SD

Subject / Key words: Survival function, Hazard function, Cox model, Markov chains, Chapman-Kolmogorov equations, Transition intensities, Actuarial present values, Actuarial reserves, Equivalence principle
SKW, UC

Holding data: The Library of Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad
HD

Note:
N

Abstract: The purpose of this paper is to present mathematical models which can be used for insurance premiums calculation. In practice, there are different principles that can be used for calculating the premium, but in this paper we presented equivalence principle. According to this principle, the total value of premiums and the total value of benefits should be equal at the time insurance policy is issued. Since in that moment, the exact amount of paid out benefits is still unknown, actuaries have to predict when insured event will happen and then calculate the expected value of total benefits. For those purposes, the time until the insured event happens is considered as a survival time and survival analysis models are used. In this paper, after survival analysis introduction we described Cox model and Makrov model. Moreover, we presented how actuarial present values for insurance cash flows are calculated and we derived the formula for calculating premium for long term care insurance in Germany.
AB

Accepted by the Scientific Board on: 2.3.2015
ASB

Defended: September 2015.
DE

Thesis defend board:
DB

President: **Danijela Rajter Ćirić**, **Ph.D.** Full Professor,
Faculty of Science,
University of Novi Sad,

Supervisor: **dr Dora Seleši**, **Ph.D.** Associate Professor,
Faculty of Science,
University of Novi Sad,

Member: **dr Sanja Rapajić**, **Ph.D.** Associate Professor,
Faculty of Science,
University of Novi Sad.